

Parte 1 - Leitura documentária para indexação:
abordagens interdisciplinares

Mariângela Spotti Lopes Fujita
Roberta Caroline Vesu Alves
Carlos Cândido de Almeida
(orgs.)

SciELO Books / SciELO Livros / SciELO Libros

FUJITA, M. S. L., ALVES, R. C. V., and ALMEIDA, C. C. Leitura documentária para indexação: abordagens interdisciplinares. In.: FUJITA, M. S. L., ALVES, R. C. V., and ALMEIDA, C. C., eds. *Modelos de leitura documentária para indexação: abordagens teóricas interdisciplinares e aplicações em diferentes tipos de documentos* [online]. Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2020, pp. 68-240. ISBN: 978-65-8654-607-1. Available from: <http://books.scielo.org/id/96v3r>. <https://doi.org/10.36311/2020.978-65-86546-07-1>.



All the contents of this work, except where otherwise noted, is licensed under a [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Todo o conteúdo deste trabalho, exceto quando houver ressalva, é publicado sob a licença [Creative Commons Atribuição 4.0](https://creativecommons.org/licenses/by/4.0/).

Todo el contenido de esta obra, excepto donde se indique lo contrario, está bajo licencia de la licencia [Creative Commons Reconocimiento 4.0](https://creativecommons.org/licenses/by/4.0/).

PARTE 1
LEITURA DOCUMENTÁRIA PARA
INDEXAÇÃO: ABORDAGENS
INTERDISCIPLINARES

3

ANÁLISE DOCUMENTAL DE TEXTOS NARRATIVOS FICCIONAIS: APORTES TEÓRICOS DE *ABOUTNESS* PARA IDENTIFICAÇÃO DE TEMAS

João Batista Ernesto de MORAES

Larissa de Mello LIMA

Mariana da Silva CAPRIOLI

RESUMO: Este trabalho tem o objetivo de avaliar o Percurso Gerativo de Sentido enquanto metodologia para os textos narrativos de ficção, para tanto caracteriza-se a questão do *aboutness* em relação à Análise Documental mostrando formas de interlocução entre as mesmas. Explica-se que no processo de análise documental focam-se nos pormenores dos processos; a análise, pois esta representa a decomposição do todo para a extração dos elementos abstratos constitutivos do texto, de maneira semelhante aos procedimentos do Percurso Gerativo de Sentido. No que tange especificamente no terreno da identificação de conteúdo dos textos narrativos de ficção, como subsídio, principalmente, às atividades de indexação, surge a proposta de se analisar de que maneira a questão do *aboutness* se articula em textos dessa natureza. Para tanto, apresenta-se a proposta de se utilizar os parâmetros do Percurso Gerativo de Sentido, como elemento instrumental que pode auxiliar nesse processo.

PALAVRAS-CHAVE: Percurso Gerativo de Sentido. Textos narrativos de ficção. Análise Documental. *Aboutness*.

ABSTRACT: This work has the objective of evaluating the Gerative Direction of Sense as a methodology for the fictional narrative texts, for both characterizing the issue of aboutness in relation to the Documentary Analysis showing forms of interlocution between them. It is explained that in the process of documentary analysis they focus on the details of the processes; to be analyzed, since this represents the decomposition of the whole for the extraction of the abstract constitutive elements of the text, in a similar way to the procedures of the Gerativo Direction of Direction. In what specifically refers to the field of content identification of narrative fiction texts, as mainly to indexing activities, the idea arises of analyzing how the question of aboutness is articulated in texts of this nature. In order to do so, we present the proposal of using the parameters of the Generative Path of Sense, as an instrumental element that can help in this process.

KEYWORDS: Gerativo Direction of Direction. Narrative texts of fiction. Documental analysis. Aboutness.

INTRODUÇÃO

Este capítulo surgiu com o propósito de se caracterizar a Análise Documental⁴ e, principalmente, como a questão do *aboutness* se insere no processo de AD, tendo em vista que a intenção primeira deste estudo é o de se avaliar o Percurso Gerativo de Sentido enquanto metodologia de análise de textos narrativos de ficção.

Assim, num primeiro momento, observar-se-á a inserção da AD na Ciência da Informação para em seguida, observar-se seus elementos constitutivos, as perspectivas epistemológicas, bem com as suas relações interdisciplinares, especialmente com a Linguística.

Por fim, trata-se da questão do *aboutness*, seu uso na literatura nacional, e alguns conceitos cunhados por vários autores, até centrar-se na concepção apresentada por Beghtol, uma vez que esta última vem ao encontro das propostas deste estudo, ou seja, de se propor ferramentas de análise dos textos narrativos de ficção.

2 CIÊNCIA DA INFORMAÇÃO E ANÁLISE DOCUMENTAL

Parece ser consensual que a denominação Ciência da Informação surgiu como resultado dos trabalhos realizados no quadro das conferências do *Georgia Institute of Technology* (abreviadamente ‘*Georgia Tech*’), realizadas

⁴ Doravante denominada AD.

em 1961 e 1962 (ROBREDO; BRÄSCHER 2010, p. 24). Como fruto destas conferências, é comumente destacado o trabalho publicado por Shera e Cleveland (1967), no qual os autores formulam uma das primeiras definições de Ciência da Informação, na seguinte conformidade:

Ciência da Informação é a que investiga as propriedades e comportamento da informação, as forças que regem o fluxo da informação e os meios de processamento da informação para um máximo de acessibilidade e uso. O processo inclui a origem, disseminação, coleta, organização, armazenamento, recuperação, interpretação e uso da informação. (SHERA; CLEVELAND, 1967, p. 265, tradução nossa).

Ainda na década de 1960, Borko (1968) elabora a seguinte definição de Ciência da Informação, segundo o autor, baseado em artigo de Robert S. Taylor⁵:

Information science is a discipline that investigates the properties and behavior of information, the forces governing the flow of information, and the means of processing information for optimum accessibility and usability. It is concerned with that body of knowledge relating to the origination, collection, organization, storage, retrieval, interpretation, transmission, transformation, and utilization of information. (BORKO, 1968, p. 3).

Pode-se observar claramente uma afinidade nos discursos de Borko e de Shera e Cleveland, ambas focando a informação e seus processos internos e externos. Porém, o primeiro, em seu artigo, além da definição, também faz uma caracterização do que seria a Ciência da Informação: “It has both a pure science component, which inquires into the subject without regard to its application, and an applied science component, which develops services and products.” (BORKO, 1968, p. 3).

Desta forma, Borko (1968) destaca que a Ciência da Informação tem tanto uma dimensão teórica quanto aplicada. Para Guimarães (2008, p. 39):

⁵ TAYLOR, R.S. Professional aspects of Information Science and Technology. In: CUADRA, C.A. (ed.). *Annual Review of Information Science and Technology*. New York: John Wiley & Sons, 1966. v. 1.

Reconhece-se uma relação orgânica entre o saber da Ciência da Informação, caracterizador de uma base conceitual específica, e os fazeres que se materializam nas distintas atividades profissionais da área de informação: os primeiros, aqueles norteando teórica e metodologicamente os segundos (e oferecendo elementos para explicar seus fenômenos) e, em sentido oposto, os fazeres profissionais oferecendo espaço de experimentação e de efetivação social dos saberes científicos da área. Em outras palavras, pode-se dizer que a Ciência da Informação nutre teórica e metodologicamente, os fazeres arquivístico e bibliotecário enquanto a Arquivologia e a Biblioteconomia atuam como campos de aplicação e de validação das teorias e metodologias oriundas da Ciência da Informação.

Desta forma, pode-se pensar a AD como um elemento que se insere no trânsito entre o fazer e o teórico, pois faz parte da prática biblioteconômica, e arquivístico possivelmente, ao mesmo tempo em que, enquanto metodologia de análise, apresenta aspectos de um processo de análise embasado em aspectos teóricos, principalmente a partir de aspectos teóricos da Linguística⁶, dentro do que se denomina Tratamento Temático da Informação.

Assim, considerando que a Ciência da Informação se articula primordialmente no trinômio *produção, organização e uso* da informação, (GUIMARÃES, 2008, p. 40) tem-se a AD como um recurso para a organização, já que esta apresenta um conjunto de procedimentos metodológicos voltados à definição do conteúdo temático de documentos de modo a permitir a recuperação, o acesso e o uso da informação neles contida. Conforme observa Guimarães (2009, p. 105):

Observa-se que um conhecimento registrado (informação socializada) requer que sobre ele se desenvolvam processos específicos de organização que propiciem sua passagem da dimensão pública para a dimensão privada, ou seja, a construção do conhecimento individual (...) que, por sua vez gerará uma nova informação registrada para ser socializada, caracterizando um verdadeiro helicóide informacional.

⁶ A este respeito veja-se, dentre outros: SMIT (1978); CINTRA (1983).

Deve-se destacar que o autor instaura um novo paradigma, pois até então se falava em círculo informacional, uma imagem inadequada uma vez que o final do círculo fica no seu começo, diferente do helicóide, cujo início e fim tendem ao infinito.

Como destacam Guimarães, Moraes e Guarido (2007, p. 94) a AD insere-se como uma das vertentes de um universo epistemológico do TTI, ao lado da *indexação* (de tradição inglesa, com ênfase nos instrumentos de tratamento temático) e da *catalogação de assunto* (de tradição norte-americana, cuja ênfase recai sobre o catálogo enquanto produto desse tratamento).

Em sentido mais estrito, e considerando que a AD abriga, em seu bojo, e paradoxalmente, como já destacou Guimarães (2003, p. 112), etapas de análise e de síntese, pode-se dizer que os maiores esforços devem centrar-se nos procedimentos metodológicos que visam a *desconstruir* ou *decompor* o conteúdo temático do documento para que se possa, assim, posteriormente, reconstruí-lo por meio de representações documentais, como o resumo ou o índice.

A concepção de AD provém, principalmente, dos estudos de Jean-Claude Gardin, a partir de suas experiências nas décadas de 60 e de 70. Para Gardin (1969) na AD é essencial a linguagem documentária para converter o conteúdo dos textos em informação e, para tanto, foi proposto um modelo de linguagem, constituída por um conjunto de termos (o léxico), por relações entre as unidades lexicais, determinadas *a priori* (o eixo paradigmático) e por uma sintaxe que articula os encadeamentos entre os termos da linguagem, em face de um documento específico (o eixo sintagmático (GARDIN, 1973). Nota-se claramente nesta definição a influência de uma das dicotomias saussurianas (Sintagma X Paradigma⁷), o que evidencia, também, que a base das teorias gardinianas sobre AD repousa em teorias linguísticas.

A partir de uma análise seletiva da literatura internacional acerca das definições de AD, tem-se⁸:

⁷ Por um lado, no discurso, os termos estabelecem entre si, em virtude de seu encadeamento, relações baseadas no caráter linear da língua. [...] Tais combinações, que se apóiam na extensão, podem ser chamadas de sintagmas. [...] Por outro lado, fora do discurso, as palavras que oferecem algo de comum se associam na memória e assim se formam grupos dentro dos quais imperam relações muito diversas. [...] Chamá-la-emos relações associativas. (SAUSSURE, 1999, p. 142-143).

⁸ Na Espanha, a AD é denominada como análise documental de conteúdo.

Quadro 1 - Definições de Análise Documental

AUTOR(ES)	DEFINIÇÕES
RUIZ PEREZ ⁹	Conjunto de operações necessárias para a extração da informação contida nas fontes primárias de modo a prepará-la para sua posterior recuperação e utilização.
CHAUMIER ¹⁰	Operação ou conjunto de operações visando a representar o conteúdo de um documento sob uma forma distinta de seu estado original, com o fim de facilitar a consulta ou a posterior localização
GARDIN et alli ¹¹	Toda operação ou grupo de operações que buscam a representação de um documento sob uma forma distinta da original , seja por tradução, resumo ou indexação, de modo a facilitar a recuperação por especialistas interessados
GARCÍA GUTIÉRREZ ¹²	Técnica documental que permite, mediante uma operação intelectual objetiva, a identificação e transformação dos documentos em produtos que facilitem a consulta aos originais, em áreas de controle documental, e com o objetivo último de servir à comunidade científica
PINTO; GALVEZ ¹³	Processo duplo de identificação e representação do texto / documento

Fonte: Elaborado pelos autores a partir de Guimarães, Moraes, Guarido (2007, p. 98).

Analisando-se as distintas definições de análise documental apresentadas pela literatura internacional, observam-se que os aspectos de *identificação*, *extração* e *representação* da informação predominam nas definições, havendo mesmo autores que consideram os dois aspectos, identificação e representação.

Em suma, pode-se dizer que a área de análise documental de conteúdo pode ser definida como um conjunto de procedimentos de

⁹ RUIZ PEREZ, 1992, p. 51.

¹⁰ CHAUMIER, 1988, p. 17.

¹¹ GARDIN *et al.* 1981, p. 29.

¹² García Gutierrez Apud RUIZ PEREZ, 1992, p. 59.

¹³ PINTO; GALVEZ, 1999, p. 31.

natureza analítico/sintética, que envolve os processos de análise do conteúdo temático dos documentos, e sua síntese, por meio da condensação ou da representação em linguagens documentárias.

Dessa concepção alguns elementos merecem destaque:

Quadro 2 - Análise documental de Conteúdo

PROCEDIMENTOS	CARACTERÍSTICAS
Processos	o conteúdo da área se dá por meio de uma sequência lógica de procedimentos
Análise	a decomposição de um todo em seus elementos constitutivos, buscando um sentido informativo
Conteúdo temático	o conjunto de elementos documentais que refletem a dimensão informativa (a função original) do documento
Representação	o conteúdo temático passa a ser expresso de maneira padronizada conforme parâmetros previamente estabelecidos

Fonte: Guimarães; Moraes; Guarido (2007, p. 94).

Para fins deste estudo, focam-se as questões dos *processos*, pois estes envolvem uma operação mental de abstração; a *análise*, pois esta representa a decomposição do todo para a extração dos elementos abstratos constitutivos do texto, de maneira semelhante aos procedimentos do Percurso Gerativo de Sentido que têm por objetivo chegar ao *conteúdo temático* dos textos narrativos de ficção, para que se possa fazer a *representação* dos conteúdos a partir dos conteúdos extraídos.

Por fim, deve-se ressaltar que a ênfase da AD quanto à análise do documento científico como o paradigma documental para o tratamento de conteúdo se encontra amplamente difundida na tradição da Biblioteconomia, da Documentação e da Ciência da Informação enquanto fundamentação de seus procedimentos técnicos. Neste sentido, não apenas a tradição francesa de AD, mas também os estudos de classificação e de indexação têm seu foco no texto científico, visualizando-o como paradigma de suporte de informação.

Contudo, García-Marco et al. (2010) chamam a atenção para outros aspectos que começam a se fazer presentes.

The revolution of today is about multimedia information, and a great deal of such information does not suit into the frame of traditional scientific reports and articles. They are much nearer to fiction and narrative documents. In fact, narrative-based communications are gaining considerable room in disciplines like education—with multimedia and video as increasing important tools in face-to-face, blended and virtual education—, psychology and even in fields like business and management in general, where the “story-telling” movement is gaining importance in marketing, corporative culture management and political communication. (GARCÍA-MARCO et al., 2010, p. 395).

Com relação à afirmação que os estudos tradicionais sempre enfocaram os documentos científicos, entretanto, Eriksson (2005) faz a seguinte ressalva:

Although it often may seem so, the classification and indexing of imaginative literature and fiction in particular is not an issue that has only been discussed in the last few decades. Presumably the first serious article on the subject appeared in 1898, when the noteworthy British librarian Ernest Baker wrote about the classification of fiction in the *Library World* (Baker, 1898), and in the beginning of the 20th century a prolonged discussion concerning the same issue took place in the American Library Association (ALA) after John Thomson had initiated a major fiction classification project at the Wagner Institute Branch of the Free Library of Philadelphia. (ERIKSSON, 2005, p. 1).

A rigor, nos últimos trinta anos, a questão dos textos literários parece ter voltado à tona, pois podem ser encontrados os estudos de Pejtersen (1978, 1979, 1983, 1984, 1998), Beghtol (1986, 1992, 1994, 1995, 1997), dentre outros. Porém, deve-se observar que tais estudos concentram seu foco nas questões de indexação, classificação ou recuperação da informação, ou seja, as preocupações recaem antes sobre a questão dos

produtos documentais, do que sobre a questão da análise do documento, tal como preconizado pela AD.

Deve-se acrescentar, ainda, que tais produtos documentais, em sua maioria, focam a questão do gênero (romance, poesia, drama), da nacionalidade (Romance Inglês, Romance Brasileiro), ou mesmo rotulado sob o genérico “ficção”.

Como exemplo, vejam-se os critérios propostos por Veiga e Martha (1986) para a classificação de obras literárias:

Na classificação de uma obra literária, no sentido estrito, devemos considerar os seguintes elementos:

- Língua em que foi escrita a obra
- Gênero ou forma
- Tema ou assunto
- Nacionalidade do escritor
- Local e data da publicação
- Teoria, ciência e técnica literária (VEIGA; MARTHA, 1986, p. 13)

Porém, no desenvolvimento do trabalho, as autoras desenvolvem a questão do *tema ou assunto* em conjunto com a questão *Gênero ou forma*, servindo-se para tal de uma citação do trabalho de Afrânio Coutinho, na seguinte conformidade:

Um gênero é, no princípio, uma tendência do espírito humano. A tendência do espírito humano tornou-se um gênero, e um gênero cria suas leis; mas no fundo ele fica sempre uma tendência do espírito. Em muitos casos, a dificuldade é grande em introduzir distinções entre uns e outros gêneros, de modo a torná-los nitidamente classificáveis sob uma qualquer categoria. (COUTINHO, 1978, p. 46).

Como sugestão de categorias de análise, García-Marco (2006, p. 395) sugere:

With the multimedia revolution, a growing percentage of content is leaving the relatively unambiguous path of scientific and journalistic genres and approaching the semantic jungles of creative literature. This kind of material requires the complex kind of content analysis

that is characteristic of artistic disciplines, that can be summarized in the simplified model that was proposed by the Art historian Panofsky (1955) —also applicable analogically to the analysis of fiction—: description of common subjects; identification of the specific names and exemplars of these common subjects; and interpretation of the abstracts subjects that are being represented through the former..

A classificação de Panofsky¹⁴, referida por García-Marco, é utilizada principalmente para a indexação de imagens, como pode ser visto em Layne (2002), e conta com as seguintes categorias:

Tabela 1 - Classificação de Panofsky

OBJETO DA INTERPRETAÇÃO	ATO DA INTERPRETAÇÃO
I. <i>Tema primário ou natural</i> - (A) fatural, (B) expressional – constituindo o mundo dos motivos artísticos.	<i>Descrição pré-iconográfica</i> (e análise pseudoformal)
II. <i>Tema secundário ou convencional,</i> constituindo o mundo das <i>imagens,</i> <i>estórias e alegorias.</i>	<i>Análise iconográfica.</i>
III. <i>Significado intrínseco ou conteúdo,</i> constituindo o mundo dos <i>valores</i> <i>“simbólicos”.</i>	<i>Interpretação iconológica.</i>

Fonte: Panofsky (2002).

E é exatamente o foco no documento e em seu processo de análise que interessa particularmente a este estudo, uma vez que se pretende privilegiar o estudo do próprio documento literário como forma de se obter o conteúdo temático dos documentos.

Além disso, deve-se observar que o documento que é objeto da AD, notadamente em bibliotecas, não é apenas o documento científico, pois em algumas dessas unidades de informação, tais como Bibliotecas Públicas e Bibliotecas Escolares, o texto narrativo é objeto de especial

¹⁴ A princípio, a classificação de Panofsky não será utilizada neste estudo.

atenção, seja no desenvolvimento de coleções, seja na organização das mesmas para atender ao usuário. Além destas bibliotecas, deve-se ter em mente, também, as Bibliotecas Universitárias, notadamente da área de Letras, as quais abrigam um grande acervo de textos narrativos que, por sua vez, necessita ser precisa e objetivamente tratado como suporte ao ensino e à pesquisa em seus cursos de Graduação e Pós-Graduação.

Porém, no momento em que a “matéria prima” da AD consiste em textos narrativos de ficção, a complexidade aumenta consideravelmente, pois o modelo metodológico de identificação de conceitos elaborado para textos científicos, pautado na leitura documental de partes canônicas do texto, tais como título, subtítulo, resumo, etc. (CHAUMIER, 1988), não se aplica a textos narrativos, pois a estrutura destes textos assume outras formas de evidenciar seu conteúdo.

2.1 ABOUTNESS

Em artigo publicado no de 2006, Moraes e Guimarães (2006) davam os primeiros passos no sentido de estabelecer ligações entre o *aboutness/meaning* e Percurso Gerativo de Sentido. Segundo os autores, o texto partia do resgate teórico das concepções de *aboutness* como suporte à análise documental em documentos científicos e, em seguida, buscava-se caracterizar o texto narrativo em sua estrutura de conteúdo, privilegiando-se os percursos temático e figurativo.

Com base em tais elementos, foi desenvolvida uma análise comparativa das dimensões de *aboutness* e *meaning* previstas por Beghtol, com os percursos temático e figurativo abordados por Greimas, de modo a estabelecer o cotejo entre os traços distintivos de ambas as teorias. Por fim, chegou-se à conclusão que um ponto em comum unia estas duas concepções teóricas: a delimitação específica do conteúdo do documento (MORAES; GUIMARÃES, 2006, p. 74).

Desta forma, pretende-se discutir neste seção a questão do *aboutness*, principiando pela discussão do uso do termo em Português, passando pelo surgimento e estudos sobre o conceito, e finalizando com algumas considerações que envolvem a proposta deste estudo, ou seja, a de se propor uma metodologia para a análise documental de textos narrativos

de ficção, principalmente levando em conta as conclusões de Moraes e Guimarães, (2006).

2.2 CONSIDERAÇÕES SOBRE O USO DO TERMO ABOUTNESS EM LÍNGUA PORTUGUESA

Nas publicações em Língua Portuguesa sobre *aboutness*, uma questão precede às discussões sobre o conceito: a exata tradução do termo. Analisando-se a literatura publicada no Brasil, parece não haver um consenso sobre a utilização do termo.

Desta forma, podem ser encontradas as seguintes traduções para o Português: “atênência” (ALVARENGA, 2001); “tematicidade” (MEDEIROS, 1986; FUJITA, 2003, 2004; MORAES; GUIMARÃES, 2006). Pode-se até mesmo encontrar o termo sendo usado como um sinônimo para “assunto”, embora se deva ressaltar que este último termo também apresenta divergências teóricas e conceituais na área de Ciência da Informação.

Em uma das primeiras publicações em Língua Portuguesa na qual aparece o termo *aboutness*, Baranow (1983) faz a seguinte consideração:

[...] o primeiro passo na indexação é determinar do que trata um determinado documento. É o que poderíamos chamar de ‘concernência’ (do verbo **concernir**, adj. **concernente**), termo mais adequado ao vernáculo do que o esdrúxulo anglicismo ‘sobrecidade’ (**aboutness**). (BARANOW, 1983, p. 25).

Em estudo sobre questões terminológicas em Ciência da Informação, Medeiros (1986, p.140) recomenda a tradução por *tematicidade*, observando ser esta a tradução mais indicada, pois, apesar de ser um neologismo, “[...] foi criado observando-se os padrões gramaticais da língua portuguesa (derivação do adjetivo **temático** com o sufixo — (d) ade)”. A autora descarta o uso do termo *sobrecidade*, pois

[...] constitui-se num neologismo por tradução de morfemas erroneamente combinados em vernáculo, uma vez que, de acordo com as regras gramaticais de derivação, as preposições em português (ex. sobre) não são passíveis de derivação lexical (ingl. *aboutness*). (MEDEIROS, 1986, p. 140).

Dias e Naves (2010) optam pelo termo *atinência*, a mesma opção adotada por Briquet de Lemos ao traduzir para o Português a obra *Indexação e Resumos: teoria e prática*, de Lancaster (2004). Na página 13 da referida obra, o tradutor faz a seguinte observação: *O autor emprega os termos ingleses about e aboutness. O primeiro traduzimos por 'trata de' e o segundo por 'atinência'. Outros traduzem aboutness por 'tematicidade', 'temática', 'acerca-de', 'ser acerca-de', 'ser sobre-algo', etc.*

2.2.3 QUESTÕES SOBRE O ABOUTNESS

Ao abordar a questão das perspectivas de utilização em bibliotecas universitárias do Percorso Gerativo de Sentido, Antonio (2008, p. 119) faz a seguinte sugestão:

A criação de uma base beta que insira obras de ficção indexadas não apenas pelos grandes assuntos obtidos no Bibliodata, mas dos assuntos que tratam realmente da obra. Assim, poderá realizar uma avaliação da revocação e precisão na recuperação da informação dessas obras, bem sua utilização.

A pesquisadora chegou a tal conclusão ao observar que os bibliotecários da Rede Unesp de Bibliotecas não dispõem de ferramental específico para a Análise Documental de Conteúdo de obras narrativas de ficção. Desta forma, uma base já existente não apenas facilitaria o trabalho de Análise Documental, mas também diminuiria em muito a possibilidade de certas inadequações no processo de representação de obras de ficção.

A rigor, a questão das incompletudes ou desvios na representação não é privilégio apenas das obras de ficção, pois como ressalta Castañon Moreno (1992, p. 5, grifo nosso), entre os fatores de erro, omissão ou inconsistência na análise e posterior representação temática de um documento está o fato de que “[...] *o bibliotecário se preocupar mais em ajustar-se a um sistema de classificação ou de indexação do que em desenvolver uma análise documental que lhe permita compreender o tema exposto no texto de uma obra.*”.

Esta preocupação também pode ser observada em Beghtol (1986, 1992, 1994, 1995, 1997), quando trata especificamente das obras de ficção.

Tratando-se de textos científicos, ou textos de ficção, o problema fundamental reside na definição – e consequentemente – na identificação do *tema* do documento, como forma de evitar eventuais incompletudes ou desvios caso o analista venha a desenvolver seu processo interpretativo sem a adoção de determinados parâmetros pré-definidos (MORAES; GUIMARÃES, 2006, p. 74).

Para Brown (1976), o tema do documento é a forma intelectual do conhecimento expresso no documento que, segundo o autor, engloba os seguintes itens:

- Forma de conhecimento (o que é)
- Conceitos temáticos (sobre o que é)
- Formas de apresentação (como é).

Essa discussão remete à questão do *aboutness*, que já há algumas décadas vem sendo objeto de estudo e de preocupação para os teóricos da análise documental de conteúdo (MORAES; GUIMARÃES, 2006, p. 74).

Cumprе ressaltar que a questão do *aboutness* foi pela primeira vez discutida, na área de Ciência da Informação, por John Hutchins, em 1977 (tal como ressaltado por Caffo, 1988, p. 24), ao afirmar que o problema crucial da área reside em identificar “*what documents are about*”. Na concepção de Hutchins (1977, p. 17) a primeira pergunta a ser feita é “*what is meant by the topic of a document in the context of an information system?*”

Para o autor, uma resposta superficial seria que o *tema* é o *assunto* descrito no índice ou vocabulário relacionado ao documento e admite, no entanto, que raramente há uma verdadeira relação entre a entrada no vocabulário controlado e o que de fato o documento trata, e dá como exemplo os sistemas de indexação pós-coordenados, cuja entrada pode representar somente uma parte do conteúdo do documento, ou seja, a descrição do assunto é apenas uma forma de expressão de parte do tema de um documento.

Uma segunda questão abordada por Hutchins (1977, p. 17) é “What do we mean by the content of a document?” e afirma que para respondermos a isto devemos estar certos quanto à distinção entre o ‘*sense*’

de uma expressão linguística e a ‘*reference*’ com o objeto, indivíduo que esta expressão representa.

Em outras palavras, pode-se afirmar que uma expressão só pode fazer sentido e ser encarada como conteúdo relevante, se se toma como ponto de partida o relacionamento existente entre um termo e os outros termos que compõem o texto, sendo que esse sistema de relacionamento é o que de fato pode determinar o seu sentido e a sua relevância.

Não se pode afirmar que Hutchins tenha bebido, ou não, nas fontes da linguística, mas o fato é que tal constatação em muito lembra o conceito de *valor* do signo linguístico, ou seja, o conteúdo de um signo só é verdadeiramente determinado pelo concurso do que existe fora dela (SAUSSURE, 1999, p. 134). Há ainda, certa conexão com o conceito de isotopia – o “parentesco” dos traços distintivos (GREIMAS; COURTÉS, 2008, p. 400), preconizado por Greimas.

Outro autor que trabalhou com a questão do aboutness foi MacCafferty (1977), citado por Begthol (1986, p. 86). Para o autor, *o mesmo documento pode ter diferentes significados para o mesmo leitor em diferentes tempos, mas o documento, por si só, é imutável, possui um aboutness fundamental* (BEGTHOL, 1986, p. 85).

Maron (1977), partindo da constatação sobre o ponto principal de um sistema de recuperação de informações, observa: “Since *about* is at the heart of indexing, how are we to formulate any proper theory of indexing if we cannot explicate precisely the key concept of *about*?” (MARON, 1977, p. 38).

Desta forma, ao investigar possíveis comportamentos de análise e busca de documentos, Maron (1977) simula algumas possíveis situações cotidianas de troca de informação entre indivíduos, buscando exemplificar a dificuldade de exprimir a compreensão sobre um documento. Para o autor, a leitura de uma obra gera uma experiência interior sobre o que ela trata, mas não necessariamente permite que se esteja apto a relatá-la com precisão, ou instruir alguém a como fazê-lo.

Assim, a formulação de um conceito de *aboutness* é buscado pelo autor, através da proposição de três tipos de *about*: *S-about* (*subjective about*); *O-about* (*objective about*); *R-about* (*retrieval about*).

Estes conceitos de *about* propostos por Maron (1977) podem ser sintetizados no quadro 01:

Quadro 3 - *About* segundo Maron

CONCEITO	DEFINIÇÃO
<i>S-about</i> (<i>subjective about</i>)	- o <i>about</i> interior, gerada da experiência psicológica de análise, leitura de uma obra - é pouco passível de ser expressa ou descrita pelo indivíduo:
<i>O-about</i> (<i>objective about</i>)	- refere-se ao conceito que pode ser observado externamente em uma situação de recuperação de informação - refere-se ao atual ou potencial comportamento de busca ou pesquisa de documentos
<i>R-about</i> (<i>retrieval about</i>)	- pode ser observado e mensurado - refere-se à probabilidade de satisfação dos usuários com um determinado documento a partir de uma determinada busca

Fonte: baseado em Maron (1977, p. 40).

De qualquer forma, mesmo contribuindo com suas propostas de conceituação, Maron (1977) reconhece a dificuldade dessas definições:

Let us attempt to clarify further what *about* means. First of all, we are suggesting that there are several different *about* concepts. Perhaps we should say that *about* can be interpreted from several different points of view and that each different viewpoint gives rise to a different interpretation of the meaning of *about*. (MARON, 1977, p. 40).

Mesmo com a complexidade das definições, torna-se interessante destacar que o autor leva em conta não apenas o que o documento em si pode representar, mas também leva em consideração o contexto de busca, bem como o contexto de recuperação, ou seja, contextos diferentes podem fazer com que o documento tenha significações também diferentes.

Outro autor que abordou a questão do *aboutness* foi Ingwersen (1992). Para o autor, o *aboutness* pode ser observado de quatro pontos de vista: do autor, do indexador, do usuário, e da solicitação.

O quadro 02 sintetiza os pontos de vista do autor:

Quadro 4 - Tipos de aboutness segundo Ingwersen

TIPO DE ABOUTNESS	CONCEITO
<i>aboutness</i> do autor (<i>author aboutness</i>)	determinado pelo autor através da sua linguagem natural, direta
<i>aboutness</i> do indexador (<i>indexer aboutness</i>)	implica na transformação da linguagem do autor em uma linguagem documental aceita ou criada pelo indexador ou sistema do qual faz parte
<i>aboutness</i> do usuário (<i>user aboutness</i>)	formulado a partir da necessidade do usuário por novos conhecimentos
<i>aboutness</i> da solicitação (<i>request aboutness</i>)	o <i>aboutness</i> da requisição conforme formulada pelo usuário.

Fonte: Baseado em Ingwersen (1992, p. 50).

Apesar de interessante, o conceito de *aboutness* exposto por Ingwersen merece algumas considerações. A rigor, se considerarmos o *aboutness* do autor diferente do *aboutness* do indexador, poder-se-ia considerar que se tratam de duas coisas diferentes, seja porque o segundo não representou corretamente o primeiro, seja porque o segundo vai ser uma entidade diferente do primeiro, sem o estabelecimento de uma relação. Ademais, se considerarmos os dois outros tipos de *aboutness* – do usuário, da solicitação, o mesmo comentário pode ser válido, ou seja, se não há uma relação explícita entre as entidades e suas representações, pode-se afirmar que há uma incompreensão sobre o primeiro, ou algum tipo de desvio ou equívoco na representação. Assim, para efeitos deste estudo, a classificação de Ingwersen (1992) é apenas citada, sem que haja uma concordância quanto a seus pressupostos.

Por outro lado, Caffo (1988) apresenta outra visão acerca do *aboutness*, por considerar a relação entre “[...] o discurso do documento acerca de um conceito e o conceito teoricamente definido como uma parte do conhecimento geral.” (CAFFO, 1988, p. 23). A autora destaca ainda que a questão do *aboutness*, especificamente no âmbito das atividades de indexação, dedica-se não à análise conceitual do complexo teórico do

conhecimento, mas a partir da dimensão conceitual de um conhecimento materializado, no caso o documento, tal como preconizado por Michael Buckland (1991), dentre outros autores.

A discussão do *aboutness* assume uma dupla vertente que, como ressalta Beghtol (1986), encontra reflexo nas concepções de Faithorne¹⁵ quanto a um *Extensional aboutness*, ou assunto inerente ao documento, em contraposição a um *Intensional aboutness*, ou razão ou objetivo pelo qual o documento é adquirido ou consultado; de Boyce¹⁶, ao contrapor a *Topicality* à *Informativeness*, e de Van Dijk¹⁷, ao referir-se a *relevância normal* e *relevância diferencial*.

Beghtol (1986) expõe o seu ponto de vista nos seguintes termos:

Whatever terms are chosen, a distinction between ‘aboutness’ and ‘meaning’, as the terms used here, seems justifiable on the assumption that a document has an intrinsic subject, an ‘aboutness’, that is at least to some extent independent of the temporary usage to which an individual might put one or more of its meanings. (BEGHTOL, 1986, p. 85).

Sintetizando esses dois níveis de conteúdo informacional, Beghtol (1986, p. 85) refere-se a um *aboutness* propriamente dito, e a um *meaning*. Assim, tem-se, segundo a autora, o *aboutness* como algo intrínseco ao documento, de natureza relativamente permanente, integrando a essência do mesmo, ao passo que o *meaning* é mutável em função do local, do momento histórico, do interesse do usuário, etc.

Nesse sentido, Moraes e Guimarães (2006, p. 75) destacam que

[...] os hábitos classificatórios da tradição bibliotecária, por vezes demasiadamente ligados às formas temáticas pelas quais um dado documento será interessante ao usuário, acabam por deixar de lado a abordagem da temática intrínseca ao documento, atuando por comparação, o que encontra reflexos no alerta de Faithorne no sentido de que a preocupação reside antes na decisão se dois

¹⁵ FAITHORNE, R. A. Content analysis, specification and control. *ARIST*, Easter Timur, n. 4, p. 73-109, 1969.

¹⁶ BOYCE, B. Beyond topicality: a two-stage view of relevance and the retrieval process. *Information Processing and Management*, London, n. 18, p. 105-109, 1982

¹⁷ VAN DIJK, T. A. Relevance assignment in discourse comprehension. *Discourse Process*, Philadelphia, n. 2, p. 113-126, 1979.

documentos possuem o mesmo assunto do que, efetivamente, qual é o assunto do documento.

Sintetizando a questão, Beghtol (1986) destaca que o assunto do documento pode ser considerado como resultado de sua tematicidade permanente, ao que se contrapõe ao eventual caráter contextual do(s) significado(s).

Como consequência, a análise do *aboutness* consiste no processo pelo qual o classificador ou indexador discerne os aspectos relativos a um tópico daqueles que refletem a temática geral do documento e busca elementos da linguística textual de modo a transformar a estrutura superficial do documento em uma estrutura lógica proposicional profunda, na seguinte conformidade:

- a) Transforms the surface structure of the document into its deep propositional logical structure;
- b) Transforms the surface structure of the classification system into its deep propositional logical structure;
- c) Joins these two deep propositional structures using the actual system of the classification schedules and instructions as an artificially-constructed virtual system for expressing the natural language actual system of primary documents;
- d) Transforms the resulting single propositional structure of classification system applied to document' back into the surface structure of the classification schedules, generates the apposite notation. (BEGTHOL, 1986, p. 102).

Em suma, tem-se um conteúdo intrínseco ao documento (*aboutness*) paralelamente a outros que advém de questões externas, como a necessidade informacional do usuário (*meanings*).

Porém, há que se chamar a atenção para o fato de se observar, *no âmbito dos estudos de tematicidade, que os mesmo têm, via de regra, voltado sua atenção para o texto científico, no mais das vezes considerado como paradigma para estudos de análise documental* (MORAES; GUIMARÃES, 2008, p. 34).

No entanto, um outro gênero documental merece atenção no que tange aos estudos sobre aboutness, ou seja, o texto narrativo de ficção, principalmente se considerarmos que este constitui grande parte do acervo em bibliotecas públicas e escolares, conforme já afirmado anteriormente.

Adentrando especificamente no terreno da identificação de conteúdo dos textos narrativos de ficção, como subsídio, principalmente, às atividades de indexação, surge a proposta de se analisar de que maneira a questão do *aboutness* se articula em textos dessa natureza. Para tanto, apresenta-se a proposta de se utilizar os parâmetros do Percurso Gerativo de Sentido, em seus três patamares, como elemento instrumental que pode auxiliar nesse processo.

REFERÊNCIAS

ALVARENGA, L. A Teoria do Conceito Revisitada em Conexão com Ontologias e metadados no Contexto das Bibliotecas Tradicionais e Digitais. *DataGramZero: Revista de Ciência da Informação*, Brasília, v. 2, n. 6, dez. 2001.

ANTONIO, D. M. *Análise do percurso temático e do percurso figurativo para identificação de temas em contos*. 2008. Dissertação (Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2008.

BARANOW, U. G. Perspectivas na contribuição da Lingüística e de áreas afins à Ciência da Informação. *Ciência da Informação*, Brasília, v. 12, n. 1, p. 23-35, 1983.

BEGHTOL, C. Bibliographic classification theory and text linguistics: aboutness analysis, intertextuality and the cognitive act of classifying documents. *Journal of Documentation*, London, v. 42, n. 2, p. 84-113, jun. 1986.

BEGHTOL, C. Domain analysis, literary warrant, and consensus: the case of fiction studies. *Journal of the American Society for Information Science*, New York, v. 46, n. 1, p. 30-44, 1995.

BEGHTOL, C. Stories: applications of narrative discourse analysis to issues in information storage and retrieval. *Knowledge Organization*, Baden-Baden, v. 24, n. 2, p. 64-71, 1997.

BEGHTOL, C. *The classification of fiction: the development of a system based on theoretical principles*. Metuchen: Scarecrow Press, 1994.

BEGHTOL, C. Toward a theory of fiction analysis for information storage and retrieval. In: WILLIAMSON, N. J.; HUDON, M. *Classification research for knowledge representation and organization*. Amsterdam: Elsevier, 1992. p. 32-48.

- BORKO, H. Information Science – what is it? *American Documentation*, Hoboken, v. 19, n. 1, p.3-5, 1968.
- BROWN, A. G. *Introduction to subject indexing*. London: Clive Bingley; Hamden (Conn.): Linnet, 1976.
- BUCKLAND, M. Information as a thing. *Journal of the American Society of Information Science*, Journal of the American Society of Information Science, v. 42, n. 5, p. 351-360, Jun. 1991.
- CAFFO, R. *Analise e indicizzazione dei documenti: l'accesso per soggetto all'informazione*. Milano: Bibliografica, 1988.
- CASTAÑON MORENO, B. Análisis temático documental. *Investigación bibliotecológica*, México, v. 6, n. 12, p. 4-10, ene./jun. 1992.
- CHAUMIER, J. Indexação: conceito, etapas e instrumentos. *Revista Brasileira de Biblioteconomia e Documentação*, São Paulo, v. 21, n. 1/2, p. 63-79, jan./jun. 1988.
- CINTRA, A.M.M. Estratégias de leitura em documentação. In: SMIT, J.W. (org.). *Análise documentária: a análise da síntese*. Brasília: IBICT, p. 28-35, 1983.
- COUTINHO, A. *Notas de teoria literária*. Rio de Janeiro: Civilização Brasileira, 1978.
- DIAS, E.W.; NAVES, M.M.L. *Análise de assunto: teoria e prática*. São Paulo: Ergon, 2010. p. 262-268.
- ERIKSSON, R. The classification and indexing of imaginative literature. In: ASIS&T SIG/CR CLASSIFICATION RESEARCH WORKSHOP, 16., 2005. *Proceedings [...]*. Charlotte, Oct. 29, 2005.
- FUJITA, M. S. L. A identificação de conceitos no processo de análise de assunto para indexação. *Revista Digital de Biblioteconomia e Ciência da Informação*, Campinas, v. 1, n. 1, p. 60-90, jul./dez. 2003.
- FUJITA, M. S. L. A Leitura Documentária na perspectiva de suas variáveis: leitor-texto-contexto. *DataGramaZero: Revista de Ciência da Informação*, Rio de Janeiro, v. 5, n. 4, ago. 2004.
- GARCÍA-MARCO, F. J. Understanding the categories and dynamics of multimedia information: a model for analysing multimedia information. In: INTERNATIONAL ISKO CONFERENCE, 9., 2006, Vienna. *Proceedings [...]*. Frankfurt/Main, Indeks Verlag, 2006, p. 395-404.
- GARCIA-MARCO, F. J.; MORAES, J. B. E.; GARCIA-MARCO, L. F.; GUIMARÃES, J. A. C. Knowledge organization on fiction and narrative documents: a challenge in the age of multimedia revolutions. In: GNOLLI, C.; MAZZOCCHI, F. (org.). *Paradigms and conceptual systems in knowledge organization*. Würzburg: Ergon, 2010. p. 262-268.
- GARDIN, J. C. Document analysis and linguistics. *Journal of Documentation*, London, v. 29, n. 2, p. 137-168, 1973.

- GARDIN, J. C. *et al. La logique du plausible: essais d'épistemologie pratique*. Paris: Maison des Sciences de l'Homme, 1981.
- GARDIN, J. C. Semantic analysis procedures in the sciences of man. *Social Science Information*, London, v. 8, n. 1, p. 17-42, 1969.
- GREIMAS, A. J.; COURTÉS, J. *Dicionário de semiótica*. São Paulo: Contexto, 2008.
- GUIMARÃES, J. A. C. A análise documental no âmbito do tratamento da informação: elementos históricos e conceituais. In: RODRIGUES, G. M.; LOPES, I. L. (org.). *Organização e representação do conhecimento na perspectiva da Ciência da Informação*. Brasília: Thesaurus, 2003. p. 100-118.
- GUIMARÃES, J. A. C. A dimensão teórica do tratamento temático da informação e suas interlocuções com o universo científico da International Society for Knowledge Organization (ISKO). *Revista Ibero-Americana de Ciência da Informação*, Brasília, v. 1, n. 1, p. 77-99, 2008.
- GUIMARÃES, J. A. C. Abordagens teóricas de tratamento temático da informação (TTI): catalogação de assunto, indexação e análise documental. In: GARCÍA MARCO, F. J. *Avances y perspectivas en sistemas de información y documentación*. Zaragoza: Ibersid, 2009, p. 105-117.
- GUIMARÃES, J. A. C.; MORAES, J. B. E.; GUARIDO, M. D. M. Análisis documental de contenido de textos narrativos: bases epistemológicas y perspectivas metodológicas. In: GARCÍA MARCO, F. J. (org.). *Avances y perspectivas en sistemas de información y documentación en entorno digital*. Zaragoza: Pressas Universitarias de Zaragoza, 2007. p. 93-100.
- HUTCHINS, J.W. On the problem of 'aboutness' in document analysis. *Journal of Informatics*, East Anglia, v. 1, n. 1, p. 17-35, 1977.
- INGWERSEN, Peter. *Information Retrieval Interaction*. London: Taylor Graham, 1992.
- LANCASTER, F.W. *Indexação e resumos*. Brasília: Briquet de Lemos, 2004.
- LAYNE, S. S. Subject access to art images. In: BACA, M. (ed.). *Introduction to art image access: issues, tools, standard, strategies*. Los Angeles: Getty Research Institute, 2002. p. 1-19.
- MACCAFERTY, M. 'About aboutness', Co-ordinate Indexing Group colloquium held at Aslib, April 18, 1977. *Journal of Informatics*, London, v. 3, n. 1, p. 121-122, 1977.
- MARON, M. E. On Indexing, retrieval and the meaning of about. *Journal of the American Society for Information Science*, New York, n. 28, n. 1, p. 38-43, 1977.
- MEDEIROS, M.B.B. Terminologia brasileira em ciência da informação: uma análise. *Revista Ciência da Informação*, Brasília, v. 15, n. 2, p. 135-142, jul./dez. 1986.
- MORAES, J. B. E.; GUIMARÃES, J. A. C. Análisis documental de contenido de textos literarios narrativos: en busca del diálogo entre las concepciones de aboutness/meaning y de recorrido temático/recorrido figurativo. *Scire*, Zaragoza, v. 12, p. 71-84, 2006.

MORAES, J. B. E.; GUIMARÃES, J.A.C. Análise documental de conteúdo de textos literários narrativos: em busca de um diálogo entre as concepções de aboutness/meaning e percurso temático/percurso figurativo *In*: GASPAR, N.R.; ROMÃO, L.M.S. *Discurso e Texto: multiplicidade de sentidos na Ciência da Informação*. São Carlos: EDUFSCar, 2008. p. 35-45.

PANOFSKY, E. *Significado nas artes visuais*. São Paulo: Perspectiva, 2002.

PEJTERSEN, A.M. Fiction and library classification. *Scandinavian Public Library Quarterly*, Stockholm, n.1, p. 5-12, 1978.

PEJTERSEN, A.M. The meaning of 'about' in fiction indexing and retrieval. *Aslib Proceedings*, Bingley, v. 31, n. 5, p. 251-257, May 1979.

PEJTERSEN, A.M.; ALBRECHTSEN, H.; SANDELIN, R.; LUNDGREN, L.; VALTONEN, R. The Scandinavian book house: indexing methods and OPAC development for subject access to Scandinavian fiction literature. *Advances in classification research*, Medford, v. 6, p. 99-113, 1998.

PEJTERSEN, A.M.; AUSTIN, J. Fiction retrieval: experimental design and evaluation of a search system based on users' value criteria: part 1. *Journal of Documentation*, Bingley, v. 39, n. 4, 230-246, 1983.

PEJTERSEN, A.M.; AUSTIN, J. Fiction retrieval: experimental design and evaluation of a search system based on users' value criteria: part 2. *Journal of Documentation*, Bingley, v. 40, n. 1, 25-35, 1984.

PINTO, M.; GALVEZ, C. *Análisis documental de contenido*. Madrid: Síntesis, 1999.

RUIZ PÉREZ, R. *El análisis documental: bases terminológicas, conceptualización y estructura operativa*. Granada: Ed. Universidad de Granada, 1992.

ROBREDO, Jaime; BRÄSCHER, Marisa (org.). *Passeios pelo bosque da informação: estudos sobre representação e organização da informação e do conhecimento*. Brasília: IBICT, 2010. Disponível em: https://repositorio.unb.br/bitstream/10482/7950/6/CAPITULO_MetriasInformacaoHistoria.pdf. Acesso em: 15 mai. 2019.

SAUSSURE, F. *Curso de linguística geral*. São Paulo: Cultrix, 1999.

SHERA, J. H.; CLEVELAND, D. B. History and foundations of information science. *Annual Review of Information Science and Technology*, Washington, v. 2, p. 249-275, 1967.

SMIT, J. W. Documentação e Linguística: inter-relação e campos de pesquisa. *Revista Brasileira de Biblioteconomia e Documentação*, São Paulo, v. 11, n. 1, p. 33-42, 1978.

VEIGA, E. A.; MARTHA, M. O. B. *Classificação de obras literárias*. Porto Alegre: Sagra, 1986.

4

TIPOS E ESTRUTURAS DE TEXTOS NA ÓTICA DA ORGANIZAÇÃO E RECEPÇÃO DE INFORMAÇÕES

Nair Yumiko KOBASHI

RESUMO: Apresentação de metodologia de elaboração de informações documentárias (resumo e indexação) fundamentada nas abordagens sobre tipos de textos e discursos, e sua interpretação, desenvolvidos na área das Ciências da Linguagem. Adota-se a perspectiva de tipologização (descritivo, narrativo expositivo e dissertativo) proposta pela Linguística textual. Os tipos de textos são, em seguida, descritos e esquematizados em constituintes. Tal esquematização permite delinear uma metodologia de análise de textos tendo como horizonte as ações de informação próprias da Ciência da Informação. Sob essa ótica, propõem-se formas de condensar textos e representá-los, cujo parâmetro mais importante é a interpretação orientada pela intencionalidade, fator básico para garantir equivalência de conteúdo entre texto de partida e resumos e índices.

PALAVRAS-CHAVE: Análise documentária. Tipos de textos. Informação documentária. Elaboração de resumos. Indexação.

ABSTRACT: Presentation of a methodology to elaborate documentary information (abstracting and indexing) based on the approaches on types of texts and discourses, and their interpretation, developed in the area of Linguistics and Semiotics. It was adopted the classification of texts (descriptive, expository and narrative narrative) proposed by Textual linguistics. The types of texts are then described and schematized into constituents. Such a schematization allows to delineate a methodology of analysis of texts having as

horizon the actions of information proper to Information Science. From this perspective, we propose ways to condense and represent texts, whose most important parameter is intentionality-oriented interpretation. This is a basic factor to guarantee equivalence between text and abstracts and indexes.

KEYWORDS: Documentary analysis. Types of texts. Documentary information. Abstracting. Indexing.

1 INTRODUÇÃO

A organização de informações, na perspectiva da Ciência da Informação, tem como tarefa central propor métodos e técnicas para elaborar informações documentárias, aqui definidas como representações condensadas de textos. A função principal da informação documentária é filtrar informação para promover a circulação e o acesso aos documentos produzidos nas várias esferas das atividades sociais. Tais representações, armazenadas em sistemas de informação, são consideradas fundamentais para responder às necessidades de informação, para se conhecer o que é publicado, tomar decisões, aprofundar saberes. Deve-se acrescentar que os conteúdos desses dispositivos vêm sendo crescentemente usados para avaliar produção científica e propor políticas de pesquisa.

Os textos, que se individualizam pelos assuntos abordados e pela forma de apresentá-los, são a matéria prima da elaboração de informações documentárias. No âmbito das ações de informação realizadas em instituições de memória (bibliotecas, museus, arquivos), os textos são representados de modo específico. De um lado, são identificados os dados que contornam o texto: o título, o subtítulo, o lugar de publicação, a editora, a data. São dados paratextuais (ou pré-textuais) inseparáveis do texto (MAINGUENEAU, 2006, p. 105). De outro, representa-se o conteúdo mesmo dos textos, ou seja, sobre o que e como o autor (individual ou coletivo) abordou um determinado tema.

Neste capítulo focalizamos, de forma especial, o conceito de texto, sua tipologia, a compreensão e interpretação desses objetos e as operações para condensá-los e representá-los documentariamente (elaboração de resumos e indexação). Não será, portanto, discutida a representação de dados pré-textuais.

A representação de conteúdos requer o auxílio de teorias e métodos que permitam transformar um objeto (o texto) em outro texto por meio de operações interpretativas específicas. Nessa perspectiva, os textos são analisados para estabelecer distinção entre informação principal e acessória; em seguida, as informações selecionadas são combinadas e convertidas em resumos ou expressos por meio de conceitos (indexação). Um fato que deve ser destacado, as representações documentárias são úteis se permitirem filtrar informações para oferecer respostas adequadas às demandas dos usuários.

Elaborar resumos e indexar são, muitas vezes, consideradas operações técnicas que não necessitam de procedimentos metódicos. No entanto, deve-se ter presente que essas operações são atos comunicacionais realizados no interior de instituições que têm missões específicas. Nessa medida, são também específicos os tipos de usuários e informações a serem tratadas.

Decorre do que foi dito que os textos de partida são desestruturados para hierarquizar as informações neles contidos. Os fragmentos selecionados são reorganizados sob a forma de novos textos condensados. Estes últimos são utilizados em buscas por informação em bases de dados. Inicialmente, a busca é feita por meio de termos, operação que resulta em listas de registros. Os resumos são filtros adicionais que permitem escolher, dos conjuntos obtidos, os documentos que poderão ser efetivamente úteis.

Diversas teorias e métodos podem subsidiar as operações acima apresentadas. Neste capítulo julgamos serem centrais: o conceito de texto, sua classificação, interpretação e condensação, aspectos desenvolvidos nos itens abaixo.

2 TEORIAS SOBRE A CLASSIFICAÇÃO DE TEXTOS

Apresentamos neste item, de forma resumida, aspectos das teorias e conceitos sobre a classificação de textos elaborados no campo da Semiótica e da Linguística. Estes conceitos encontram-se sistematizados em perspectiva pragmática, qual seja a sua pertinência para sustentar a elaboração de informações documentárias.

2.1 CONCEITO DE TEXTO

Neste capítulo, consideramos texto e discurso como termos sinônimos. As correntes que analisam textos ou discursos compartilham a ideia de que estes são objetos que apresentam um sentido global; não são, portanto, inscrições que se expressam pela mera justaposição de palavras ou de frases. Nessa medida, os textos definem-se por sua “autonomia e fechamento” (DUCROT; TODOROV, 1972, p. 375), características garantidas pela progressão, pela coesão e coerência (DUCROT; TODOROV, 1972). Essas propriedades referem-se às variadas formas de repetição de ideias: reiteração de pontos de vista, exemplificação, definições de termos, entre outros. São essas características que permitem condensar textos. Dito de outro modo: as reiterações, exemplos, definições, podem ser suprimidos sem comprometer a informação substantiva veiculada pelo texto.

Adam (1989, p. 20), por sua vez, caracteriza as propriedades do texto de forma didática: “A textualidade pode ser definida como um equilíbrio delicado entre continuidade-repetição, de um lado, e progressão de informação, de outro.” Essas propriedades referem-se às variadas formas de reiteração de pontos de vista, exemplificação, definições de termos, entre outros. São essas características que permitem condensar textos. Dito de outro modo: as reiterações, exemplos, definições, podem ser suprimidos sem comprometer a informação substantiva veiculada pelo texto.

Os objetos textuais apresentam, portanto, estruturas convencionais, socialmente construídas, aptas a comunicar ideias, experiências, histórias, descrever objetos, expressar sentimentos, convencer.

Embora os textos possam ser descritos de diferentes maneiras (oral, escrito, audiovisual), neste capítulo trataremos apenas de textos verbais escritos não-ficcionais, também chamados de textos pragmáticos (técnicos e científicos). Não trataremos, portanto, de textos literários. Justificamos, a seguir, a distinção entre texto literário e não-literário e as razões para limitar este capítulo à abordagem de textos não-literários. Semelhante distinção tem, como veremos a seguir, uma função operacional.

A função estética é o aspecto divisor de águas entre texto literário e não-literário. Outro aspecto importante é o estatuto da referência (CORTINA, 2000, p. 100). Como exposto na citação abaixo:

A demonstração da importância da função estética para o texto literário e da utilitária para o não-literário pode ser constatada ainda na afirmação de Valéry, que diz que, ao se resumir um texto não-literário, apreende-se o que ele tem de essencial, mas, ao resumir um texto literário, perde-se exatamente aquilo que é essencial para ele. (CORTINA, p. 2000, p. 102).

Por outro lado,

a distinção entre o literário e o não-literário pode ser pensada a partir da oposição entre o verossímil e o verídico. Uma vez que o verossímil corresponde à tentativa de reconstrução do referente do discurso, a utilização desse termo situa-se, por consequência, num contexto social, caracterizado por uma certa atitude com relação à linguagem e de sua relação com a realidade extralinguística. Nesse sentido, em um dado contexto cultural, a verossimilhança é característica dos discursos figurativos, enquanto a busca da verdade é um procedimento particular dos discursos abstratos, portanto temáticos. (GREIMAS, 1983¹⁸ apud CORTINA, 2000, p. 103-104).

As diferenças conceituais acima assinaladas são operacionais no contexto da organização da informação, questão que será retomada de forma específica no item 3 (Leitura e interpretação de textos: os problemas da recepção).

2.2 TIPOS DE TEXTOS

Admite-se que os estudos sobre os textos tiveram início no século XX, a partir de Saussure (1973), que criou a Linguística como campo de conhecimento. Nesta ciência, são estabelecidos conceitos fundantes, como a oposição entre língua e fala, ou seja, entre sistema linguístico e uso do sistema. Desta oposição derivam diversas abordagens de classificação de textos, dentre as quais a Semiótica, a Análise do discurso e a Linguística textual, que se desdobram em várias vertentes. Embora

¹⁸ GREIMAS, A.J. Le contrat de véridiction *In*: GREIMAS, A. J. *Du Sens II: essais sémiotiques*. Paris: Seuil, 1983, p. 103-113.

existam diferenças entre elas, serão destacados os traços, muitas vezes comuns, que concorrem para fundamentar as operações de elaboração de informações documentárias.

Segundo a perspectiva Semiótica, o discurso é o lugar da subjetividade, um signo constituído de um plano de expressão e de um plano de conteúdo. A Análise do discurso, na concepção de Barros (1988, p. 3) deve considerar:

Três questões básicas para a concepção do discurso e sua análise:
a) a relação do discurso com a enunciação e as condições de recepção; b) o discurso como o lugar, ao mesmo tempo do social e do individual; c) a articulação entre narrativa e discurso, isto é, o discurso constituído sobre estruturas narrativas que o sustentam.

Outros autores destacam a enunciação, fato que permite classificar os textos em didáticos, polêmicos e científicos. Há ainda uma vertente que procura observar as condições históricas de produção de forma a classificá-los em três categorias: discursos lúdicos, polêmicos e autoritários. Por fim, o texto pode ser analisado do ponto de vista da organização global, ou estrutura, o que permite classificá-los em narrativos, descritivos, explicativos e dissertativos (CORTINA, 2000).

Deve-se chamar a atenção para o aspecto dominante em cada tipo de texto. Os textos não são puros, isto é, uma narrativa pode conter aspectos polêmicos, uma descrição pode apresentar enunciados narrativos e os textos dissertativos podem apresentar segmentos narrativos como recurso argumentativo.

2.3 CLASSIFICAÇÃO DE TEXTOS

Como vimos acima, a classificação de textos baseia-se em diferentes teorias e métodos. São, grosso modo, ora tipificados pela estrutura interna (descritivo, narrativo, dissertativo), pela finalidade ou pelas condições de produção, ou pelos aspectos sintáticos e semânticos (técnico, científico, didático, jornalístico, jurídico, político, etc.).

Os primeiros estudos sistemáticos sobre os “gêneros” textuais e as “partes do discurso” são atribuídos a Aristóteles. Suas ideias,

apresentadas no âmbito da Retórica, fundamentam, ainda hoje, as modernas teorias discursivas relacionadas à persuasão (CITELLI, 1989; OKASABE, 1979).

A Retórica, para Aristóteles, não se reduz a um conjunto de normas ou à oratória, mas ao esforço de “descobrir especulativamente em qualquer dado, o persuasivo” (ARISTÓTELES apud OKASABE, 1979, p. 142). É nessa perspectiva que o estagirita afirma: “Não há senão duas partes no discurso, pois é necessário dizer qual é o tema e demonstrá-lo. [...] Dessas duas partes, uma é a proposição; a outra, a confirmação. Como se tratasse de um lado, do problema e, do outro, da demonstração.” (ARISTÓTELES apud OKASABE, 1979, p. 156).

Os textos argumentativos não escapam a essa caracterização. Contudo, a esse esquema genérico - proposição/demonstração - podem-se agregar outros constituintes, de modo a caracterizar mais amplamente as variantes de um tipo.

O texto técnico-científico é, via de regra, de natureza argumentativa, elaborado com o intuito de expor metodicamente os resultados da observação de um fenômeno. Para Gardin,

[...] a expressão concreta de um raciocínio científico é o texto científico, no qual o autor expõe as operações do espírito que o conduziram da observação de certos fatos empíricos ao enunciado de proposições denominadas de forma diversa: teses, hipóteses, interpretações, comentários, conclusões, explicações. (GARDIN, 1987, p. 4).

É no interior da atividade de conhecer, portanto, que se define o texto científico canônico: uma unidade de comunicação do saber dotada de certos elementos estruturais. No âmbito das pesquisas acadêmicas esses elementos estruturais são o Tema, o Problema, a Hipótese, a Metodologia, os Resultados e a Conclusão (ASTI-VERA, 1979; BUNGE, 1973).

Quadro 1 - Estrutura do texto científico canônico

TEMA	OBJETO DE PESQUISA
PROBLEMA	INDAGAÇÃO
HIPÓTESE	CONJETURA
METODOLOGIA	OBSERVAÇÃO
RESULTADO	INTERPRETAÇÃO
CONCLUSÃO	COMENTÁRIO FINAL

Fonte: Kobashi (1994, 2008).

O Quadro 1, acima, apresenta os constituintes básicos das dissertações, teses acadêmicas, e artigos científicos publicados em periódicos e anais de eventos científicos.

Um aspecto intrínseco aos textos que visam defender ideias e opiniões é a argumentação. Podem ser identificados nestes os seguintes constituintes: **Tese** (apresentação de um ponto de vista); **Argumentos** (evidências que sustentam o ponto de vista); **Conclusão**, (confirmação do da tese apresentada).

Quadro 2 - Estrutura do texto argumentativo

TESE	PONTO DE VISTA
ARGUMENTOS	PROVAS
CONCLUSÃO	CONFIRMAÇÃO

Fonte: Kobashi (1994, 2008).

Os artigos de opinião, sejam eles publicados em jornais diários, revistas semanais de informação, ou mesmo em revistas científicas, costumam apresentar a configuração acima.

O texto denominado expositivo (FÁVERO; KOCH, 1988) é usualmente elaborado para expor algum problema (social, político, material) que necessita de intervenção. Pode ser esquematizado da seguinte forma: Problema (aspecto problemático da realidade; Causas e

consequências do problema (identificação de fatores que concorrem para a existência do problema); Solução (alternativas de intervenção)

Quadro 3 – Estrutura do texto expositivo

PROBLEMA	QUESTÃO PROBLEMÁTICA IDENTIFICADA
CAUSAS	RAZÕES QUE CAUSAM O PROBLEMA
SOLUÇÃO	RESPOSTAS AO PROBLEMA

Fonte: Kobashi (1994, 2008).

A estrutura acima é frequentemente utilizada para a elaboração de diagnósticos e relatórios técnicos para criação ou aprimoramento de serviços.

As narrativas (relatos de experiências, textos literários como os romances) podem ser olhadas com base nas categorias propostas por Lasswell (1971), constituído de 6 categorias, como segue:

Quadro 4 - Estrutura do texto narrativo

QUEM	ACTANTES/ATORES
O QUÊ	ACONTECIMENTO
QUANDO	ASPECTO TEMPORAL
ONDE	ASPECTO ESPACIAL
COMO	MODO
PORQUÊ	RAZÃO

Fonte: Kobashi (1994, 2008).

A estrutura narrativa é apropriada para relatar fatos da vida em sociedade, como as reportagens jornalísticas, para a criação de textos ficcionais, para relatar experiências pessoais ou institucionais, em prontuários de pacientes.

O texto descritivo apresenta os traços ou características de objetos, sejam eles abstratos ou concretos, expressos em duas categorias:

Quadro 5 - Estrutura do texto descritivo

REFERENTE	OBJETO DO MUNDO
CARACTERÍSTICAS	PREDICADOS/ATRIBUTOS

Fonte: Kobashi (1994, 2008).

A estrutura acima (Quadro 5), é utilizada, em geral, em manuais de equipamentos, como também para caracterizar objetos do mundo real, emoções, sentimentos.

O trecho abaixo resume os tipos de textos e respectivas ocorrências:

Assim, o tipo narrativo ocorre em romances, contos, novelas, reportagens, noticiários, depoimentos, relatórios etc; o descritivo, na caracterização de personagens e do espaço em narrativas, guias turísticos, verbetes de enciclopédias, resenhas de jogos, relatos de experiências ou pesquisas, reportagens etc; o expositivo, em manuais didáticos, científicos, obras de divulgação etc; o argumentativo *stricto sensu*, em textos publicitários, propagandistas, peças judiciais, matérias opinativas etc; o injuntivo, em manuais de instruções, receitas culinárias, bulas de remédios etc; o preditivo, em horóscopos, profecias, boletins meteorológicos e previsões em geral. (CORTINA, 2000, p. 92).

A tipificação de textos apresentada neste item não tem a pretensão de inventariar exaustivamente todas as variantes textuais. Os tipos apresentados fundamentam a possibilidade de utilizar as estruturas textuais como paradigmas para interpretá-los e selecionar dados considerados pertinentes para os fins da documentação.

3 LEITURA E INTERPRETAÇÃO DE TEXTOS: OS PROBLEMAS DA RECEPÇÃO

A leitura, a interpretação e a recepção de textos são fenômenos estudados sob diferentes perspectivas filosóficas.

Umberto Eco, na obra “Leitor do texto literário” (1983) problematiza a leitura e a compreensão de textos com base nas relações dialéticas entre uso e interpretação (ECO, 1983, p. 62). A discussão do

autor tem como foco as possibilidades de interpretação, expressas na dicotomia leitura correta e leitura incorreta. Desenvolve estas ideias com base na tricotomia: leitura como busca da intenção do autor (*intentio auctoris*), da intenção da obra (*intentio operis*) e da intenção do leitor (*intentio lectoris*) (ECO, 1993, 2000, p. 6).

Como expõe Eco (1979), os debates sobre essas oposições, que ocorreram ao longo da história, e certamente permanecem ainda hoje, podem ser assim sintetizadas:

- a) deve-se buscar no texto aquilo que o autor queria dizer;
- b) deve-se buscar no texto aquilo que ele diz, independentemente das intenções do autor.

Só com a aceitação da segunda ponta da oposição é que se poderia, em seguida, articular a oposição entre

b1) é preciso buscar no texto aquilo que ele diz relativamente à sua própria coerência contextual e à situação dos sistemas de significação em que se respalda;

b2) é preciso buscar no texto aquilo que o destinatário aí encontra relativamente a seus próprios sistemas de significação e/ou relativamente a seus próprios desejos, pulsões, arbítrios. (ECO, 2000, p. 7).

Para Cortina (2000, p. 35), a opção **a**, acima, remete à ideia de que “interpretar um texto corresponde a tentar descobrir exatamente aquilo que seu autor pretendeu dizer.” Essa perspectiva (*intentio operis*) foi substituída, nos anos 1960-1970, pela ideia de que “[...] ler um texto significa buscar a intenção da obra, pouco importando se o autor quis dizer isso ou aquilo em seu texto. O autor é destronado e, em seu lugar, instala-se o texto [...]” (CORTINA, 2000, p. 36-37). Nesta concepção, a estrutura global do texto, isto é, os constituintes organizados segundo uma dada intenção comunicativa seria o suporte adequado para interpretar a obra. Portanto, a interpretação realizada segundo a intenção do leitor (*intentio lectoris*), subordina-se ao uso que será feito do texto.

Pode-se concluir do que foi dito que há várias possibilidades de ler e interpretar textos, porém, “[...] dizer que qualquer interpretação

de um texto é verdadeira, significa dizer que ele não tem nenhuma importância.” (CORTINA, 2000, p. 39). Assim, concorda-se com a seguinte posição do autor:

Defender que um texto tem mais de uma possibilidade de leituras não significa dizer que qualquer leitura que se produza a partir dele seja verdadeira. Uma leitura só é válida quando consegue se sustentar no próprio texto, isto é, quando é coerente com o que foi enunciado de forma explícita ou implícita. (CORTINA, 2000, p. 40).

Apresentados os aspectos teóricos da classificação de textos e as diferentes perspectivas de leitura, passamos, a seguir, à discussão da leitura e interpretação de textos no âmbito pragmático da Organização da Informação, tendo presente que:

As formas de leitura dependem, portanto, dos interesses daqueles que se põem a ler. Uma observação que se pode fazer é que o processo de interpretação do sujeito leitor será melhor toda vez que este tiver um determinado objetivo ao ler um texto. A leitura cumpre um duplo papel: informar e propiciar prazer. Dependendo, portanto, da intenção do leitor, este deverá optar por textos distintos. (CORTINA, 2000, p. 55).

4 ANÁLISE DOCUMENTÁRIA E INTERPRETAÇÃO, CONDENSAÇÃO E REPRESENTAÇÃO DE TEXTOS

Antes de discutir a interpretação na perspectiva da Análise documentária, deve-se esclarecer que partimos do pressuposto de que a leitura correta de um texto depende do objetivo com que se lê. Afirmamos, também na Introdução, que a área da organização de informações tem como tarefa central propor métodos e técnicas para elaborar informações documentárias. Para realizar semelhante tarefa, admitimos que as operações intelectuais complexas, como produzir resumos e indexar, podem ser sistematizadas em métodos e regras a partir das teorias linguísticas apresentadas nos itens 2 (Teorias sobre a caracterização e classificação de textos) e 3 (Leitura e interpretação de textos: os problemas da recepção).

4.1 CONDENSAÇÃO E REPRESENTAÇÃO DE TEXTOS

Segundo Stierle (1987, p. 88), do ponto de vista da Estética da recepção, a leitura do texto tem início com o “horizonte de expectativas”, do leitor, cuja continuidade pode ser caracterizada como uma “hermenêutica de perguntas e respostas” feitas ao texto a fim de condensá-los e representá-los.

As representações documentárias típicas podem ser olhadas como “traduções” condensadas do conteúdo informacional de textos. A tradução, como afirma Paul Ricoeur, “não implica apenas um trabalho teórico ou prático, mas também um problema ético. Levar “o leitor ao autor, levar o autor ao leitor” (RICOEUR, 2011, p. 49). Dito de outra maneira, a produção de informações documentárias (indexação e resumos) é um ato intencional de mediação.

As diretrizes e textos didáticos caracterizam a indexação e a elaboração de resumos, como produto de três operações intelectuais: a) análise, compreensão e interpretação do texto; b) seleção das informações principais e c) sua representação dentro de padrões estabelecidos por normas de documentação.

A indexação, por sua vez, é justaposição de palavras ou sintagmas presentes em uma Linguagem documentária (KOBASHI, 1994).

Tanto a elaboração de resumos quanto a indexação requerem a compreensão global de um texto, como vimos no item 2 (Teorias sobre a caracterização e classificação de textos). O grau de redução a que se deve submetê-los para obter um ou outro tipo de representação documentária sugere que, no caso da indexação, pode-se operar com as noções de tema¹⁹ e de estrutura temática; no caso do resumo, a noção de estrutura ou superestrutura textual poderá nortear a operação de seleção de informações.

A Linguagem Documentária, por ser um instrumento constituído por léxico reduzido e sintaxe precária, não permite a produção de mensagens com a mesma riqueza informacional dos resumos. Desse modo, para a indexação, interessa analisar o texto para determinar a sua macroestrutura²⁰, representada por palavras ou expressões, ditos descritores.

¹⁹ “O tema (ou tópico) de um ato de enunciação é aquilo sobre o qual o locutor fala, é o objeto do discurso [...]” (DUCROT; TODOROV, 1972, p. 345).

²⁰ A macroestrutura é definida como uma rede de proposições associadas hierarquicamente (VAN DIJK; KINTSCH, 1983).

4.2 TEMA E ESTRUTURA TEMÁTICA: CONCEITOS OPERACIONAIS

Como apontam as teorias sobre os textos, a compreensão), requer estratégias de leitura e interpretação orientadas para um fim (KINTSCH; VAN DIJK, 1978). No caso da Análise documentária, a leitura é uma atividade comprometida com intenções precisas de comunicação: elaborar produtos que permitam recuperar informação, ou seja, levar o texto ao leitor.

Os produtos documentários destinam-se a: auxiliar a tomada de decisões sobre a leitura ou não do documento original; dar ao leitor uma quantidade suficiente de informações, a fim de que o mesmo, em certas circunstâncias, não tenha necessidade de ler integralmente o original. Nessa perspectiva, a leitura documentária é, fundamentalmente, uma operação de hierarquização e seleção de informações. A seleção das informações é factível, apenas, com base na distinção entre informação essencial e informação acessória.

Na leitura documentária, a compreensão global de um texto requer a identificação do tema de um texto. Os temas são as ideias principais discutidas ou tratadas em um documento. Definem-se, portanto, como o foco principal de um texto e não os seus aspectos periféricos. O “tema”, portanto, é o elemento em torno do qual se estrutura a mensagem, é o seu núcleo informativo (KOBASHI, 1994, 2008).

A identificação do tema resulta de um processo de condensação semântica. García Gutiérrez e Lucas (1987, p. 47-49) propõem a utilização das perguntas conceptuais de Lasswell (Who, What, Whem, Where, Why) como método de abordagem analítica do texto de atualidade. Para Lasswell (1971) um ato de comunicação, como a informação jornalística (reportagem), pode ser expressa pelas respostas fornecidas a 5 perguntas, a saber: Quem, Diz o quê, Em que canal, Para quem, Com que efeito.

Sendo a Indexação uma operação que implica: a) análise e compreensão do texto; b) seleção das informações principais e c) representação das informações na linguagem utilizada no sistema, as duas primeiras operações (análise e seleção de informações), poderão ser realizadas por meio dos mecanismos de resposta às categorias da estrutura temática.

4.3 REPRESENTAÇÃO POR INDEXAÇÃO

A indexação tem como **objetivo** caracterizar o conteúdo de um documento por meio de uma linguagem documentária; suas **operações** básicas são: Identificar o tema do texto; elaborar um enunciado temático e representá-lo. Para a identificação do tema, lê-se o texto, tendo como parâmetros as categorias da estrutura temática, exposto no item 4.2 acima. Deve-se observar, ainda, que na indexação há uma operação final de conversão do enunciado temático para a Linguagem Documentária adotada pelo sistema.

A identificação do tema, ou macroestrutura de um texto (VAN DIJK; KINTSCH, 1983), é o elemento principal a ser identificado. É em torno dele que se organiza o texto.

No caso de um texto dissertativo acadêmico (dissertação ou tese), tal como esquematizado no Quadro 1, enuncia-se a questão de pesquisa como um problema que merece ser investigado. Após oferecer uma hipótese de solução e observar metodicamente as possibilidades de responder à questão de pesquisa, apresenta, na conclusão, a possível resposta para o problema. Nessa medida, o tema, via de regra, pode ser identificado tanto no segmento em que se apresenta o problema de pesquisa, quanto nas conclusões.

Quadro 6 - Estrutura do texto canônico dissertativo

SUPERESTRUTURA	INDEXAÇÃO
TEMA	
PROBLEMA	
HIPÓTESE	
METODOLOGIA	
RESULTADOS	
CONCLUSÕES	

Fonte: elaboração própria.

No Quadro 7 apresentamos a superestrutura de textos que pretendem defender um ponto de vista, contrapondo-o a outros. A tese

é o ponto de partida que orienta a organização do texto. Portanto, na indexação, os constituintes a serem analisados são a Tese e as conclusões.

Quadro 7 - Estrutura do texto argumentativo (polêmico)

SUPERESTRUTURA	INDEXAÇÃO
TESE	
ARGUMENTOS	
CONCLUSÕES	

Fonte: elaboração própria (2019).

Em face de um texto com a estrutura esquematizada abaixo, que na Linguística do texto é denominado expositivo, a leitura deve se concentrar no Problema e na Solução do Problema.

Quadro 8 - Estrutura do texto expositivo

SUPERESTRUTURA	INDEXAÇÃO
PROBLEMA	
CAUSAS	
SOLUÇÃO	

Fonte: elaboração própria.

O Quadro 9 apresenta a estrutura do texto descritivo. Este tipo de texto tem como núcleo temático um objeto concreto ou abstrato, sobre o qual são atribuídas características. Na indexação, o enunciado temático será elaborado com base no objeto do mundo.

Quadro 9 - Estrutura do texto descritivo

SUPERESTRUTURA	INDEXAÇÃO
OBJETO DO MUNDO	
PREDICADOS	

Fonte: elaboração própria.

O Quadro 10 esquematiza o texto narrativo. Este se apresenta como um relato que envolve um sujeito (actante) que realiza alguma ação (o que), contextualizado no espaço e no tempo. O tema, portanto, surge da identificação dos quatro constituintes destacados. O como e o porquê nem sempre estão presentes em textos narrativos técnico-científicos.

Quadro 10 - Estrutura do texto narrativo

SUPERESTRUTURA	INDEXAÇÃO
QUEM	
O QUE	
QUANDO	
ONDE	
COMO	
PORQUÊ	

Fonte: elaboração própria.

Na indexação dos tipos de textos apresentados acima, após a identificação do tema, elabora-se um enunciado temático que será convertido (traduzido) em descritores que sinalizam o assunto do documento. O uso das linguagens documentárias para representar conteúdos informacionais é uma operação de inclusão dos textos em classes. Assim, instaura-se uma nova ordem de sentido porque os textos são generalizados pelo apagamento dos traços que individualizam os textos.

4.4 REPRESENTAÇÃO POR RESUMOS

O **resumo** é um novo texto que representa o original de forma **condensada**, enquanto a **indexação** representa o assunto do documento por meio da justaposição de termos (descritores) de uma Linguagem Documentária. É útil relembrar aqui o conceito de Análise documentária. Segundo Jean-Claude Gardin, o criador da Análise documentária, esta é um “[...] conjunto de procedimentos utilizados para exprimir o conteúdo dos documentos científicos sob formas destinadas a facilitar a sua localização ou consulta.” (GARDIN, 1974, p. 48-49).

O resumo é um novo texto que supõe coesão e coerência, tanto quanto o texto de partida. É, portanto, uma representação mantém relação de semelhança e de contiguidade com o texto de partida.

Tal como na operação de indexação, a elaboração de resumos supõe a seleção de informações textuais, realizada com base na sua hierarquização. Segundo Van Dijk e Kintsch (1983), a superestrutura é um elemento fundamental para a compreensão dos textos porque: a) ele tem caráter convencional, sendo conhecido e reconhecido por uma comunidade linguística; b) a superestrutura configura-se como um esquema abstrato que estabelece a ordem global de um texto e se compõe de uma série de categorias, cujas possibilidades de combinação se baseiam em regras convencionais. O domínio das superestruturas permite um processamento *top-down* (dedutivo) coerente porque a leitura é feita com base em hipóteses e não pela construção do sentido através de processamento *bottom-up* (indutivo). (VAN DIJK, 1992).

Os passos da elaboração de resumos documentários podem, tal como na indexação, devem ser observados como objetivos e operações. O **objetivo** é produzir representações parafrásticas condensadas de um texto. As **operações** básicas são: identificar o tipo de texto (superestrutura) e selecionar nos constituintes as informações consideradas relevantes para a elaboração de diferentes tipos de resumos: resumo indicativo ou resumo informativo, propostas em normas de documentação. O resumo informativo supõe a condensação de cada constituinte da superestrutura e sua articulação, de modo a, em certas circunstâncias, substituir o texto de partida; o resumo indicativo, como o próprio nome sugere, é uma caracterização sumária do conteúdo. Para a elaboração de resumos, sugere-se a seleção de constituintes segundo o tipo de resumo pretendido.

Os quadros 11 a 15 apresentam a superestrutura de cada tipo de texto (segundo a classificação proposta no item 2.3 (Classificação de textos). Propõe-se que na elaboração de resumos informativos cada constituinte da superestrutura seja condensado e reestruturado em um novo texto coeso e coerente. No caso do resumo indicativo, propõe-se a inspeção e condensação dos constituintes destacados em cinza. O resumo indicativo é expresso também como um novo texto coeso e coerente.

Quadro 11 - Condensação do texto científico canônico

SUPERESTRUTURA	RESUMO INFORMATIVO	RESUMO INDICATIVO
TEMA		
PROBLEMA		
HIPÓTESE		
METODOLOGIA		
RESULTADOS		
CONCLUSÕES		

Fonte: Kobashi (1994).

Quadro 12 - Condensação do texto argumentativo (polêmico)

SUPERESTRUTURA	RESUMO INFORMATIVO	RESUMO INDICATIVO
TESE		
ARGUMENTOS		
CONCLUSÕES		

Fonte: Kobashi (1994, 2008).

Quadro 13 - estrutura do texto expositivo

SUPERESTRUTURA	RESUMO INFORMATIVO	RESUMO INDICATIVO
PROBLEMA		
CAUSAS		
SOLUÇÃO		

Fonte: Kobashi (1994).

Quadro 14 - Estrutura do texto descritivo

SUPERESTRUTURA	RESUMO INFORMATIVO	RESUMO INDICATIVO
OBJETO DO MUNDO		
PREDICADOS		

Fonte: Kobashi (1994, 2008).

Quadro 15 - Estrutura do texto narrativo

SUPERESTRUTURA	RESUMO INFORMATIVO	RESUMO INDICATIVO
QUEM		
O QUE		
QUANDO		
ONDE		
COMO		
PORQUÊ		

Fonte: Kobashi (1994, 2008).

5 CONSIDERAÇÕES FINAIS

Apresentamos neste capítulo um conjunto de teorias e métodos para subsidiar a produção de informações documentárias. Nessa perspectiva, revisitamos, inicialmente, as teorias sobre a análise e a interpretação de textos desenvolvidas no campo das Ciências da Linguagem, com destaque para a Análise do discurso e a Linguística textual. Em seguida, foram abordados os conceitos de Texto, Tipos de textos, Tema, Estrutura temática, Macroproposição semântica e Superestrutura textual. A apropriação destes conceitos, como foi discutido, fundamentam as operações de condensação e representação de conteúdos.

As teorias sobre a interpretação, inspiradas na hermenêutica (RICOEUR, 1981), foram apresentadas e discutidas para aplicação na compreensão de textos. Abordamos os problemas da recepção com base na tricotomia proposta por Umberto Eco (2000) - *Intentio auctoris, intentio opera e intentio lectoris*. A abordagem de Eco coloca em cheque a ideia de que todas as interpretações de textos são válidas ao defender a leitura literal de textos. Com base nas categorias acima, demonstra que há interpretações corretas e incorretas. De fato, como os textos são elaborados com intenções comunicativas específicas, embora eles admitam diferentes leituras, há limites a serem considerados na interpretação. A Intenção da obra e a intenção do leitor são os parâmetros que permitem distinguir a interpretação correta da incorreta.

Apresentamos, ainda, o conceito de interpretação como tradução. Compreender um texto, nesta perspectiva hermenêutica, significa lidar com o problema da fidelidade da tradução em relação ao texto de partida (RICOEUR,1981). Torna-se evidente, nesta abordagem, a tensão que acompanha o processo de manter equivalência entre o texto de partida e o texto traduzido. Esta ideia da fidelidade é fundamental na Análise documentária porque as representações documentárias devem preservar a equivalência entre o texto fonte e as versões condensadas.

Definimos, em seguida, os tipos de informações produzidas no âmbito da organização de informações, que denominamos genericamente de **Informação Documentária**. Procuramos evidenciar a importância dos tipos de textos e respectivas estruturas textuais como paradigmas importantes para selecionar informação para os fins da elaboração de resumos e indexar. Com efeito, a equivalência entre textos e representações condensadas podem desempenhar o papel de filtros para selecionar informação pertinente apenas se forem mantidas as relações de contiguidade e de semelhança com o texto-base. Dessa forma, sugerimos a adoção de parâmetros de leitura e interpretação que possam guiar metodicamente o trabalho documentário, nas suas várias etapas, desde a seleção de dados até a combinação final dos dados obtidos em um novo texto.

Afirma-se, portanto, que as atividades documentárias, em particular a indexação e a elaboração de resumos, podem ser monitoradas. Portanto, deixa-se de lado a concepção de que a análise de textos depende apenas de talento e bom senso. Assume-se que, como em todo jogo de linguagem, a produção de informações documentárias será mais efetiva se realizada com base em regras explícitas.

A questão de fundo discutida deste capítulo deriva da ideia de que o texto não se oferece à compreensão de forma direta e imediata. Não é, portanto, uma tarefa trivial. Ao contrário, requer trabalho intelectual baseado em parâmetros conceituais e metodológicos que facilitem a apreensão do texto e a identificação e seleção de informação segundo o produto a ser elaborado.

Outro aspecto a ser problematizado é a noção de superestrutura, que não deve ser compreendida de forma dogmática, já que os textos concretos não são puros. A noção de dominância permite identificar a intenção comunicativa do texto e classificá-lo no interior de tipos. Deve-

se ter em mente, portanto, que criar classificações é realizar operações cognitivas de generalização. Nesse sentido, os modelos de estruturais textuais são abstrações que não captam todos os traços dos objetos do mundo. São sempre representações simplificadas do real.

Portanto, as informações documentárias retêm apenas os traços mais gerais dos textos, o que implica a perda controlada de informações. As informações documentárias não têm valor em si. Elas representam, não substituem os textos originais.

A metodologia de análise de textos aqui proposta vem sendo continuamente testada em cursos de graduação de Biblioteconomia e em cursos de atualização profissional. Esperamos que as ideias aqui apresentadas confirmem rigor ao fazer documentário e concorram para aprimorar a elaboração de representações e, em decorrência, ampliem as condições de difundir informação contida em documentos.

REFERÊNCIAS

- ADAM, J. M. Pour une pragmatique linguistique et textuelle. In: REICHLER, C. *L'interprétation de textes*. Paris: Minuit, 1989. p. 183-219.
- ASTI-VERA. *Metodologia da pesquisa científica*. Porto Alegre: Globo, 1979.
- BARROS, D. L. P. *Teoria do discurso: fundamentos semióticos*. São Paulo: Atual, 1988.
- BUNGE, M. *La investigación científica*. Barcelona: Ariel, 1973.
- CITELLI, A. O. *Linguagem e persuasão*. São Paulo: Ática, 1989.
- CORTINA, A. *O príncipe de Maquiavel e seus leitores: uma investigação sobre o processo de leitura*. São Paulo: Ed. da UNESP, 2000.
- DUCROT, O.; TODOROV, T. *Dictionnaire encyclopédique des sciences du langage*. Paris: Seuil, 1972.
- ECO, U. *Conceito de texto*. São Paulo: T.A. Queirós, 1979.
- ECO, U. *Interpretação e superinterpretação*. São Paulo: Martins Fontes, 1993.
- ECO, U. *Leitura do texto literário: lector in fabula*. Lisboa: Presença, 1983.
- ECO, U. *Os limites da interpretação*. São Paulo: Perspectiva, 2000.
- FÁVERO, L. L.; KOCH, I. G. V. *Linguística textual: introdução*. São Paulo: Cortez, 1988.

- GARCÍA GUTIÉRREZ, A.; LUCAS, R. *Documentación automatizada de los medios informativos*. Madrid: Paraninfo, 1987.
- GARDIN, J. C. L'analyse logiciste. In: GARDIN, J. C. et al. *Systèmes experts et sciences humaines*. Paris: Eyrolles, 1987. p. 17-26.
- GARDIN, J. C. *Les analyses de discours*. Neuchâtel: Delachaux et Niestlé, 1974.
- GREIMAS, A.J. Le contrat de véridiction. In: GREIMAS, A. J. *Du Sens II: essais sémiotiques*. Paris: Seuil, 1983, p. 103-113.
- KINTSCH, W.; VAN DIJK, T. Toward a model of text comprehension and production. *Psychological Review*, London, v. 85, n. 5, p. 363-394, set. 1978.
- KOBASHI, N. Y. *A elaboração de informações documentárias: em busca de uma metodologia*. 1994. Tese (Doutorado em Ciência da Informação) - Escola de Comunicações e Arte, Universidade de São Paulo, São Paulo, 1994.
- KOBASHI, N. Y. Linguística textual e elaboração de informações documentárias: algumas reflexões. In: GASPAR, N.; ROMÃO, L. M. S (org). *Discurso e texto: multiplicidade de sentidos na Ciência da Informação*. São Carlos: EdUFSCAR, 2008. p. 47-66.
- LASSWELL, H. D. A estrutura e a função da comunicação na sociedade. In: COHN, G. *Comunicação e indústria cultural*. São Paulo: Nacional; EDUSP, 1971. p. 105-107.
- MAINGUENEAU, D. *Termos-chave da análise do discurso*. Belo Horizonte: Ed. da UFMG, 2006.
- OKASABE, H. *Argumentação e discurso político*. São Paulo: Kairós, 1979.
- RICOEUR, P. *Hermeneutics and the human sciences: essays on language, action and interpretation*. Cambridge: Cambridge University Press; Paris: Ed. de la Maison des Sciences de l'Homme, 1981.
- RICOEUR, P. *Sobre a tradução*. Belo Horizonte: Ed. da UFMG, 2011.
- STIERLE, K. ¿Qué significa "repción" en los textos de ficción? In: BURGER, et al. *Estética de la recepción*. Madrid: Arco Libros S/A, 1987. p. 87-143.
- VAN DIJK, T.; KINTSCH, W. *Strategies of discourse comprehension*. Orlando: Academic Press, 1983.
- VAN DJIK, T. *Cognição, discurso e interação*. São Paulo: Contexto, 1992.

5

ANÁLISE DOCUMENTÁRIA E SEMÂNTICA DISCURSIVA: CONTRIBUTOS PARA O TRATAMENTO TEMÁTICO DE OBJETOS INFORMACIONAIS

Raimunda Fernanda dos SANTOS

Dulce Amélia de Brito NEVES

Maria Elizabeth Baltar Carneiro de ALBUQUERQUE

RESUMO: Discute aspectos relativos aos contributos da Semântica Discursiva para a produção significativa de metadados semânticos no processo de análise documental de recursos informacionais. Destaca que os percursos temático e figurativo possuem qualidades metodológicas a serem aplicadas na leitura documental, com isso faz-se necessário que o indexador reconheça as superestruturas responsáveis pela caracterização do tipo do objeto informacional, considerando as suas especificidades. Na revisão de literatura são apontadas pesquisas na área de Organização e Representação da Informação e do Conhecimento que constata a pertinência e adequação do percurso figurativo e do percurso temático de Greimas para a análise documental de objetos informacionais. Conclui ressaltando que os contributos da Semântica Discursiva para a análise documental resultam em novo método para extrair os conteúdos de recursos informacionais, trazendo como diferencial a redução do nível de subjetividade, tornando a indexação mais próxima possível de atender os parâmetros conceituais dos documentos.

PALAVRAS-CHAVE: Análise Documentária. Semântica Discursiva. Indexação.

ABSTRACT: Discusses aspects related to the Discursive Semantics contributions to the significant production of semantic metadata in the process of documentary analysis of information resources. Emphasizes that the thematic and figurative paths have methodological qualities to be applied in documentary reading, so it is necessary that the indexer recognizes the superstructures responsible for characterizing the type of informational object, considering their specificities. Through a literature review are pointed researches in the area of Organization and Representation of Information and Knowledge that find the relevance and adequacy of the figurative and Greimas thematic path for the documentary analysis of informational objects. Concludes by emphasizing that the contributions of Discursive Semantics to documentary analysis result in a new method for extracting indicators of information resources, bringing as a differential the reduction of the subjectivity level, making the indexing closer to meet the conceptual parameters of the documents.

KEYWORD: Documentary Analysis. Discursive semantics. Indexing.

1 INTRODUÇÃO

No decurso da história, a representação está presente nas práticas sociais, na cultura popular e nos diversos contextos de construção e transfiguração da sociedade. Neste limiar, o conhecimento pode abranger o que o sujeito pode discernir logicamente e o que é possível ser observado em uma determinada realidade por meio da linguagem. Assim, as relações entre a humanidade e o conhecimento evidencia a necessidade do homem em representar e organizar o conhecimento com vistas à perpetuação, preservação, recuperação da memória social.

Partindo desse pressuposto, a representação é vista como uma ação criativa, e socialmente construída, que faz uso da linguagem e se configura como objeto de análise crítica e científica de uma realidade. A linguagem, por sua vez, é compreendida como o meio pelo qual o significado é produzido e intercambiado e se constitui como elemento fundamental para os sentidos e para a cultura (HALL, 2016).

Esse aspecto multifacetado da representação e da linguagem desafia aqueles que se propõem a estudá-las no contexto da Ciência da Informação através de pesquisas teóricas e práticas sob o viés da Organização e Representação da Informação e do Conhecimento. A representação no contexto da Ciência da Informação, além de se constituir como uma atividade pragmática, se amplia como campo de estudos teóricos e aplicados

que objetivam responder aos problemas e efeitos da informação a fim de tornar os registros informacionais acessíveis à sociedade.

A Ciência da Informação consiste em um campo interdisciplinar que se preocupa, principalmente, em tornar a informação acessível para aqueles que dela necessitam, tendo em vista que os itens informacionais, independente do seu formato²¹ (áudio, texto, imagem, vídeo, etc.), devem passar por um conjunto de tarefas de armazenamento, organização e representação para que seja possível a sua recuperação, acesso e uso em sistemas de recuperação da informação. Nesse cenário, os objetos informacionais devem passar por esse conjunto de tarefas.

A indexação, considerada por Chaumier (1988) como operação mais importante da análise documentária, visa descrever os conteúdos dos itens informacionais por meio de termos, também denominados de descritores, que se configuram como, ou se propõem a ser, pontos de acesso para fins de disponibilização das informações pertinentes às necessidades dos usuários. Desse modo, a indexação condiciona os resultados de uma estratégia de busca e a qualidade dessa atividade reflete na recuperação, no acesso e no uso da informação pela geração hodierna e futura.

Assim, todas as fases do tratamento temático da informação, como análise conceitual e tradução, fazem parte das técnicas realizadas pelos profissionais da informação, sendo influenciadas por fatores linguísticos, cognitivos, lógicos, ideológicos, contextuais e culturais, simultaneamente.

Com isso, verifica-se a imprescindibilidade do compromisso do indexador na execução da análise documentária, tendo em vista que essa operação consiste em uma atividade intelectual considerada como o limite dos aparelhos tecnológicos, uma vez que o computador, por exemplo, ainda não realiza o trabalho humano de abstrair, distinguir, compreender e assimilar o que está presente nos itens informacionais (SANTOS, 2016).

Considerando que a indexação é uma operação relacionada à semântica, faz-se necessário discutir aspectos relativos aos contributos da Semântica Discursiva para a produção significativa de metadados semânticos no processo de análise documentária de recursos informacionais.

²¹ Nesta pesquisa, o conceito de formato está relacionado à configuração física, forma dos objetos informacionais, também denominados como materiais informacionais ou itens informacionais.

2 ANÁLISE DOCUMENTÁRIA E TRATAMENTO TEMÁTICO DA INFORMAÇÃO

A análise documentária centra-se nos procedimentos voltados para a identificação e seleção de conceitos para fins de representação do conteúdo dos itens informacionais. A indexação, operação que integra a análise documentária, está metodologicamente associada a duas etapas intelectualmente distintas: a *análise de assunto* e a *tradução*.

A *análise de assunto* diz respeito à extração de conceitos que representem o conteúdo de um material informacional. Seguindo essa linha de raciocínio, essa etapa se inicia por meio da leitura técnica através da análise das principais partes ou dos elementos representativos do material informacional em questão. Nesta etapa, faz-se necessário que o indexador conheça o tipo e estrutura desse item informacional (imagem, texto, áudio, vídeo, etc.) para que a leitura seja iniciada para fins específicos. Em seguida, é efetuada a extração dos conceitos²² constituindo-se o segundo momento da análise de assunto. Essa atividade é necessária para a representação do conteúdo do material informacional em termos adequados. Desse modo, o assunto do objeto informacional é determinado com base nesses conceitos e no contexto em que o mesmo se encontra inserido. Por isso, é importante verificar a complexidade do assunto do material informacional analisado para que seja possível elencar termos que representem efetivamente o seu conteúdo.

Após a etapa de análise do assunto é realizada a *tradução* desses descritores em termos provenientes de uma linguagem de indexação, ou seja, em termos autorizados para representar o conteúdo do objeto informacional analisado. Nesse entendimento, a etapa de *tradução* consiste na representação desse material informacional em termos provenientes de instrumentos de representação da informação, denominados de linguagens de indexação ou linguagens documentárias, como listas de cabeçalhos de assunto, tesouros, ontologias e taxonomias, por exemplo.

Destarte, as linguagens documentárias são instrumentos de controle terminológico que podem ser empregados na etapa de tradução dos conceitos, visando facilitar a comunicação entre a linguagem dos usuários e a linguagem do sistema, mediante a padronização de termos para a descrição dos conteúdos dos itens informacionais. Esses instrumentos

²² Conceitos são, portanto, definidos como **unidades do conhecimento** identificadas através de enunciados verdadeiros sobre um item de referência e representados por uma forma verbal (termo ou palavra) (MEDEIROS, 1986, p. 136, grifo do autor).

são frutos de trabalhos colaborativos entre a Terminologia, Teoria da Classificação, Linguística Documentária, Ciência da Computação, Análise Documentária, Ciências Cognitivas, por exemplo.

Lancaster (2004) ressalta que esses instrumentos consistem em listas de termos ou descritores controlados, construídas para fins de tratamento e recuperação da informação objetivando controlar sinônimos²³, diferenciar homógrafos²⁴ e agregar termos cujos significados apresentem alguma relação mais estreita entre si. Diante disso, observa-se que esses instrumentos objetivam controlar as dispersões lexicais existentes nas linguagens naturais; orientar o indexador no que concerne à escolha dos termos para representar o assunto dos materiais informacionais e indicar os melhores termos para a indexação.

Motta (1987, p. 25), conceitua a linguagem documentária como:

Um sistema de vocabulário baseado em conceitos, incluindo termos preferidos (descritores), termos não preferidos (não descritores) e suas inter-relações, que se aplica a um determinado ramo do conhecimento e que se destina a controlar a terminologia utilizada para a indexação/recuperação de documentos.

Com a ampliação da quantidade de sistemas de recuperação da informação e o incremento de tecnologias aplicadas às atividades de representação, organização, recuperação, acesso e uso da informação, surgiram outros instrumentos complementares aos sistemas de classificação como as listas de cabeçalhos de assunto, taxonomias, tesauros, ontologias, mapas conceituais, por exemplo. Esses instrumentos têm sido denominados genericamente por *Knowledge Organization Systems* (KOS), Sistemas de Organização do Conhecimento (SOC), em português.

Zeng (2008) apresenta uma lista de Sistemas de Organização do Conhecimento por meio das suas tipologias, a saber:

✓ Listas de Termos

- Listas propriamente ditas (*pick lists*): conjunto de termos em ordem sequencial;

²³ Palavras que possuem significados semelhantes.

²⁴ Termos que possuem a mesma grafia, mas apresentam pronúncia e significados diferentes.

- Dicionários: listas de termos organizados alfabeticamente com as suas respectivas definições, apresentando significados variados de acordo com a aplicação desses termos;
- Glossários: listas de termos organizados alfabeticamente com as suas respectivas definições;
- Anéis de sinônimos: conjunto de descritores considerados como equivalentes para fins de representação e recuperação da informação.

✓ *Modelos semelhantes a metadados:*

- Arquivos de autoridade: listas de termos empregados para fins de controle das variações de nomes para uma autoridade;
- Diretórios: listas de nomes e informações de contato associadas;
- Gazetteers: dicionários geoespaciais de tipos e nomes de lugares.

✓ *Classificação e categorização:*

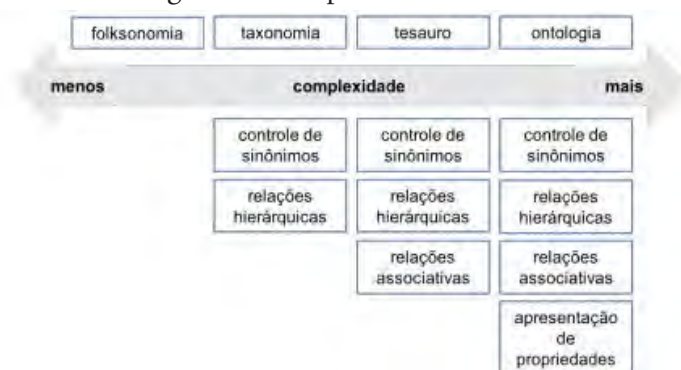
- Cabeçalhos de assunto: esquemas que apresentam um conjunto de descritores controlados para representar o assunto dos itens em uma coleção e dispõem um conjunto de diretrizes para combinar descritores em cabeçalhos compostos;
- Esquemas de categorização: esquemas de agrupamento “vagamente formados” (ZENG, 2008, p. 161);
- Taxonomias: separação de itens em grupos ordenados ou categorias, considerando características particulares;
- Esquemas de classificação: arranjos hierárquicos e facetados que apresentam notações alfabéticas ou numéricas visando representar assuntos genéricos.

✓ *Modelos de relações:*

- Tesouro: listas de termos autorizados de um domínio específico do conhecimento, constituída de base léxica composta por termos que apresentam relações hierárquicas, de equivalência e associativas entre eles.
- Redes semânticas: conjunto de descritores modelados como nós em uma rede de variados tipos de relações objetivando representar conceitos;
- Ontologias: modelos de conceitos específicos que representam relações complexas entre objetos, incluindo regras e axiomas ausentes em redes semânticas.

Considerando que as funções dos SOC diversificam conforme o grau de complexidade e o objetivo de sua aplicação, Moreira (2018) também representa em sua pesquisa a complexidade desses sistemas:

Figura 1- Complexidade dos SOC



Fonte: Moreira (2018, p. 102).

Não obstante, as vantagens das linguagens documentárias sobressaíam em relação à linguagem natural, em algumas situações a aplicação dessa última espécie de linguagem se faz apropriada mediante a aplicação dos métodos colaborativos de indexação. Lancaster (2004) afirma essa conveniência em ambientes que tratam materiais informacionais com conteúdos completos e altamente subjetivos que

viabilizam variadas interpretações - como é o caso das imagens - tendo em vista a sua capacidade de fomentar a exaustividade e especificidade da indexação e, conseqüentemente, recuperação da informação. Todavia, considera-se a importância de um tratamento linguístico mínimo dos descritores apresentados pelos usuários na atividade de representação do item informacional.

Em síntese, verifica-se que as etapas de *análise de assunto e tradução* são intelectualmente distintas, apesar de não serem diferenciadas de maneira clara nas produções científicas sobre a temática. Campos (1987) afirma que embora sejam operações distintas, elas se sucedem haja vista o caráter integrativo das mesmas, objetivando atender às necessidades de uma determinada comunidade de usuários.

Todavia, não se pode definir o assunto de um recurso informacional sem que sejam contemplados diversos aspectos, além do contexto no qual esse item informacional está inserido. Assim, corroboramos com Lucas (2000) quando a mesma revela que a leitura documentária não é resumida aos procedimentos e técnicas de tratamento dos itens informacionais, cujas etapas foram citadas anteriormente. Nas práticas de indexação devem ser consideradas as características do material informacional e os gestos de leitura desse sujeito enquanto leitor e usuário da informação, os quais são inerentes à sua formação discursiva configurada em processos de significação.

Entendendo isso, observa-se que a complexidade da indexação está associada ao fato de que a descrição do conteúdo intelectual, operação altamente subjetiva, pode ser realizada de maneira incorreta, tendo em vista a influência de fatores linguísticos, cognitivos, lógicos, ideológicos, contextuais, culturais, etc. Mesmo estando correta, essa atividade pode ser considerada insuficiente para demarcar congruentemente os descritores, pactuando com as necessidades dos usuários (BLAIR, 2003).

Fica claro, portanto, que a indexação é uma atividade de representação e que, por sua vez, está atrelada ao conceito de substituição, de “estar no lugar de outra coisa”. Para fins de realização dessa atividade, o indexador se coloca entre o objeto informacional a ser representado e o usuário final. Por esse motivo, essa representação não consiste na apresentação do registro do conhecimento propriamente dito.

Seguindo essa linha de raciocínio, Pato (2015, p. 73) atenta que:

[...] quem ou aquele que representa se apresenta no lugar de outrem. Ao se colocar no papel de alteridade, do autor, aquele que agrega polifonicamente ao seu texto os textos de vários outros sujeitos, o esforço do indexador será no sentido de traduzir, concorrer com as representações desse(s) outrem(ens) e alimentar um sistema de organização da informação.

Sobre esse prisma, a indexação consiste em uma prática de interpretação e, ao mesmo tempo de representação, a qual, por meio de ações seletivas, privilegia certos elementos acerca de objetos informacionais e oculta outros. Nesse cenário, o indexador exerce o seu papel crucial de mediação entre o item informacional e o usuário final, buscando apurar termos representativos mediante a pluralidade de sentidos existentes no registro do conhecimento. Isso ocorre pelo fato de que, ao interpretar o conteúdo intelectual deste, são acionadas variadas leituras e esse profissional é quem apresenta o que há de informação nesse registro, captando alguns aspectos e ignorando outros.

Concorda-se com o pensamento de Pato (2015, p. 73) quando o mesmo enfatiza a importância do indexador, considerando o fato de que esse profissional é coparticipante do processo de atribuição de significados dos itens informacionais, os quais se relacionam com o contexto social, cultural e econômico do qual emerge, pois, “a leitura de um texto [ou registro do conhecimento] depende da interação entre leitor, texto, autor e mundo socialmente estruturado”.

Essa assertiva também se aproxima das ideias de Lucas (2000, p. 49) quando essa autora reflete que a indexação “[...] encaminha o leitor por vários percursos, enriquece ou confunde os possíveis olhares que serão lançados à obra”, uma vez que o olhar leitor do profissional silencia ou expõe diferentes leituras.

Sobre esse prisma, os traços semânticos existentes no material informacional e que são selecionados pelo indexador revelam a visão de mundo desse profissional, por isso essa atividade é considerada difícil de ser ensinada, haja vista que não há um manual a ser seguido que inclua os aspectos subjetivos para o desenvolvimento da operação.

Em síntese, a indexação assume uma posição crucial e condicionante na recuperação da informação, ela “é um fator de auxílio na comunicação entre o sistema de recuperação da informação e o usuário que deseja satisfazer sua necessidade informacional” e deve ser encarada como a materialização de uma dinâmica de natureza ética, refletindo universos culturalmente distintos (FRANCELIN; PINHO, 2011, p. 61).

Ainda de acordo com esses autores, as competências específicas para os profissionais da informação, em especial aqueles que lidam com o tratamento temático da informação, não se esgotam no “como fazer”, mas encontram terreno fértil de indagação do “por que” fazer e no “para que” fazer. Principalmente porque é necessário considerar que tanto o produto quanto o processo de indexação devem estar imbuídos da preocupação com a sua utilidade e com a sua adequação para determinada comunidade ou grupo-alvo.

Nesse entremeio o sujeito cognoscente, ou seja o interagente, conforme sinaliza Capurro (2003, *online*), é essencial neste processo, pois o processo de tradução é algo que a final de contas, terá a sua razão de existir quando direcionado a seu público, pois, em essência se apresenta para organizar e se organiza para recuperar a informação seja qual for o suporte, a mídia, a ambiência e a maneira como a comunicação é empregada.

3 SEMÂNTICA DISCURSIVA E ANÁLISE DOCUMENTÁRIA

A representação da informação consiste em uma prática de enunciação das propriedades de um objeto informacional (unidade de informação organizável) ou das relações desse objeto com outros que o identificam, através de operações que visam a sua descrição.

Conforme Fogl (1979, p. 21) a informação se configura como uma unidade de três elementos, são eles:

1. Conhecimento (conteúdo da informação);
2. Linguagem (instrumento de expressão de itens de informação);
3. Suporte (objetos materiais ou energia).

Para esse autor, a fonte de origem da informação é o conhecimento, portanto, ela pode ser analisada e representada segundo pontos de vistas semânticos (cognitivos), pragmáticos (reais), inferenciais (juízos de valor) e de acordo com o formato utilizado (textos, imagens, áudios, vídeos, etc.), levando em conta as propriedades do seu conteúdo e a sua função social.

Nesse processo, vê-se que o registro informacional, independente do meio, suporte ou ambiente, é o aspecto central da representação da informação, colocando tanto o produtor, quanto o indexador e o interagente (público-alvo), em uma situação de constante vigilância, pois os estoques oriundos desse processo desafia constante e cotidianamente o profissional da informação, uma vez que a relação dialógica do acesso e uso da informação é cada vez mais dependente do ato de representar. Por isso, Gaudêncio e Albuquerque (2014, p. 47), afirmam que “a representação da informação faz parte da vida do profissional da informação que lida diariamente com as questões documentárias e que busca recuperar com mais eficácia determinados conteúdos”.

A análise documentária diz respeito a uma operação associada à semântica que se concentra nas informações extraídas de um recurso informacional e dos referentes internos (cognição) do(s) sujeito(s) que o analisa. Nesse entendimento, independente da configuração desse recurso – seja de caráter visual, textual, sonoro, etc. ele pode ser lido e analisado. Para a realização dessa operação é importante que o sujeito tenha conhecimento prévio acerca do conteúdo presente no item informacional a ser indexado. Todavia, o conhecimento prévio do sujeito não deve ser fator condicionante para a execução dessa operação, uma vez que, para auxiliar na compreensão do documento a ser analisado podem ser extraídas informações extrínsecas em outras fontes de informação.

Sob esse viés, nesta pesquisa busca-se estudar as contribuições da semântica no nível discursivo, contemplando os procedimentos de tematização e figurativização como subsídios para leitura documentária de objetos informacionais.

A semântica, enquanto área do conhecimento, emergiu no final do século XX no intuito de responder as demandas da sociedade científica por meio do estudo do significado ou teoria da significação (FIORIN, 1999). Desse modo, o seu foco não está ligado apenas às palavras, mas

também às orações, gestos, símbolos, imagens e entre outros elementos de significação.

De acordo com Greimas (1966, p. 11) a significação consiste em elemento central das preocupações da semântica, uma vez que essa estuda os problemas relacionados ao significado de todas as coisas que são pensadas, produzidas e transmitidas.

Desse modo, a atividade de representação advém do processo de significação e esse, por sua vez, contribui para a atividade de organização - o que justifica a relevância e necessidade das investigações de natureza semântica nos processos de organização e tratamento da informação na Ciência da Informação.

A semântica atribui sentido aos signos, esclarecendo o que é apreendido no mundo das coisas, e se configura como um componente semiótico-discursivo (ALBUQUERQUE, 2011).

A semiótica, por sua vez, consiste no estudo da significação, a qual é compreendida como a relação dependente entre o conteúdo e a expressão de signos verbais e não-verbais (BATISTA, 2001). Greimas e Coutés (2008, p. 447) afirmam que a semiótica realiza uma “operação que, ao instaurar uma relação de pressuposição recíproca entre a forma e a expressão e a do conteúdo” e entre “o significante e o significado”, “[...] produz signos, resultando em uma semiose²⁵.

Santaella (1993) infere que a semiótica estuda todos os tipos possíveis de ações sígnicas, ou seja, o centro da sua teorização são as representações de todas as linguagens, uma vez que estuda os meios pelos quais o homem se comunica, sendo eles verbais ou não (gestos, ideias, símbolos, imagens, etc.).

Para Albuquerque (2011), a semiótica de origem francesa, também conhecida como semiótica greimasiana, parte da concepção de signo criada por Hjelmslev (1975), a qual enriquece o conceito de signo delineado por Saussure (2004, p. 80) como “uma entidade psíquica” que se associa a dois elementos: o significante (conceito) e o significado (imagem acústica).

²⁵ De acordo com Batista (2001, p. 146), a semiose é o “[...] o processo de produção, acumulação e transformação da função semiótica”. Assim, a semiose se constrói e se manifesta ao longo do discurso e só atinge a sua completude no percurso sintagmático do discurso por inteiro.

Enquanto abordagem metodológica, a Semântica Discursiva foi idealizada por Algirdas Julien Greimas e busca atribuir sentido ao discurso com base no modelo teórico-metodológico denominado Percurso Gerativo da Significação (PGS) ou Percurso Gerativo do Discurso (PGD), cujo foco é o estudo da produção e interpretação de sentidos. Embora a noção de signo seja considerada, Greimas não parte do signo para estruturar e propor a sua metodologia, uma vez que são levadas em consideração as denominadas “figuras”.

Para Fiorin (1999, p. 38) “[...] a noção de percurso gerativo de sentido constitui um simulacro metodológico para explicar o processo de entendimento, em que o leitor precisa fazer abstrações, a partir da superfície do texto, para poder entendê-lo [...]”. Portanto, esse modelo teórico-metodológico representa um processo que parte do nível simples ao mais complexo, são eles: *profundo (ou fundamental), narrativo e discursivo*.

Cada nível supracitado contém um elemento sintático e um elemento semântico, considerando que a sintaxe busca estudar questões relativas à estrutura do vocabulário e a semântica busca focar “[...] nas regras que presidem às relações entre os vocábulos, as construções das orações e as relações interacionais.” (ANTÔNIO, 2008, p. 12).

Seguindo essa linha de raciocínio, aqui são estudadas as contribuições da semântica de nível discursivo, no intuito de considerar os procedimentos de tematização e figurativização como subsídios para a construção da metodologia de indexação de imagens, em especial de xilogravuras.

A Semântica Discursiva integra duas formas básicas de discursos: os discursos concretos e os discursos abstratos. Eles não são antagônicos entre si, pois coexistem e estão interligados apresentando um caráter lógico que percorre do concreto (figuras) para o abstrato (temas). Desse modo, o discurso concreto construído com figuras é denominado figurativo. E o discurso abstrato edificado por temas é definido como temático.

Para Fiorin e Savioli (2003, p. 89) os textos figurativos “[...] produzem um efeito de realidade, e por isso representam o mundo, criam uma imagem do mundo, com seus seres, seus acontecimentos.”. Os textos temáticos, por sua vez, “explicam as coisas do mundo, ordenam-nas,

classificam-nas, interpretam-nas, estabelecem relações e dependências entre elas, fazem comentários sobre suas propriedades”.

Nesse prisma, os textos figurativos dirigem-se a algum aspecto da realidade, criando imagens da mesma, bem como dos seres, das coisas, das ações, acontecimentos e qualidades associados a ela. Para Antônio (2008, p. 66) “[...] quando fala-se mundo real devemos pensar também no mundo construído.”.

Tal como as figuras, os temas “[...] também se encadeiam em percursos, isto é, em conjuntos organizados. São os percursos temáticos. Para apreender o tema geral, é preciso perceber esse encadeamento dos temas e depreender a unidade subjacente à diversidade.” (FIORIN; SAVIOLI, 2003, p. 101).

Assim, os temas expressam elementos abstratos por meio de uma natureza conceitual representada por classes e categorizações, conforme explica Barbalho (2006, p. 88):

A tematização expressa elementos abstratos buscando explicar a realidade e representar o mundo através de um investimento conceptual. Os temas organizam, categorizam e ordenam a realidade significante de modo a permitir sua interpretação. Figuras e temas são para manter a coerência interna do texto, necessitam seguir um percurso ou encadeamento lógico de modo a gerar sentido. As figuras devem ser vistas através do conjunto por elas composto e não isoladamente.

Albuquerque (2011, p. 173, grifo da autora), por sua vez, apresenta as relações entre os níveis temático e figurativo da Semântica Discursiva:

A *semântica discursiva* tem como componente a *tematização* – elementos abstratos presentes no texto – e a *figurativização* – elementos concretos presentes no texto – que dão concretude ao tema. As figuras do texto formam uma rede, uma trama que, para entendê-las, é necessário conhecer o primeiro nível temático assim como o nível figurativo em que palavras e expressões apresentam traços comuns de significação, que podem ser agrupados.

Com isso, verifica-se que a tematização e a figurativização são componentes da Semântica Discursiva e consistem em procedimentos semânticos da discursivização. No Quadro a seguir são apresentadas as principais características dos textos figurativos e temáticos:

Quadro 1 - Principais características dos textos figurativos e temáticos

TEXTOS FIGURATIVOS	TEXTOS TEMÁTICOS
*Apresentam função descritiva ou representativa	*Elementos abstratos no texto cuja função é explicar a realidade
*Representam o mundo	*Classificam e ordenam a realidade
*Criam imagem do mundo	*Estabelecem relações e dependências
*Criam imagem dos seres	*Apresentam função predicativa ou interpretativa
*Criam os acontecimentos do mundo	*Apresentam comentários sobre as propriedades do mundo
*Referem-se ao concreto presente no texto	*Apresentam temas genéricos (classes) que contempla assuntos principais
*São palavras ou expressões correspondentes ao mundo natural	*Apresentam coerência ao texto principal
*Constroem simulacro da realidade	

Fonte: Adaptado de Antônio (2008).

A relação entre as figuras apresentadas para a construção do sentido e, conseqüentemente, a identificação do tema atinente a elas é denominada por Albuquerque (2011) de “encadeamento de figuras”. Portanto, a sequência dessas figuras resultante do seu agrupamento produz uma rede relacional, denominada no sistema semiótico como *percurso figurativo*, e expressa os temas subentendidos no texto - tornando viáveis as possibilidades significativas em um dado contexto.

Depreende-se, portanto, que os estudos que possuem maior conexão com a Semântica Discursiva e que contribuem significativamente para a Ciência da Informação são aqueles relacionados à Organização e Representação da Informação e do Conhecimento, especificamente sobre a análise documentária. Esses estudos, em conjunto, podem

colaborar para o processo de análise e síntese do sentido das informações contidas em textos verbais e não verbais, considerando que os sujeitos possuem mentes interpretadoras que executam processos semióticos, interpretam e ressignificam diferentes visões de mundo apresentadas por estruturas conceituais.

Diante do exposto, as metodologias de análise documentária, ao considerarem a Semântica Discursiva, devem contemplar a discursivização semântica, privilegiando os processos de figurativização e tematização. Nesse limiar, a análise dos objetos informacionais deve estar pautada em níveis descritivos (figuras) e interpretativos (temas). As figuras consideram os aspectos denotativos, ou seja, aquilo que o item informacional (imagem, texto, áudio, vídeo, etc.) mostra de forma explícita (elementos concretos); e os temas levam em conta os aspectos conotativos (elementos abstratos), ou seja, aquilo que pode ser simulado pelo objeto informacional – associando-se ao processo de tematização.

Os temas não necessariamente precisam refletir o contexto ou significado “real” do material informacional, mas o que esse pode (res) significar na concepção de quem a analisa. Eles são extraídos de classes ou categorizações, podendo ser provenientes de instrumentos de controle terminológicos como tesouros, ontologias, taxonomias, etc. ou até mesmo da linguagem natural como as *tags* (etiquetas) atribuídas pelos usuários em um dado sistema colaborativo.

Pesquisas na Ciência da Informação, em especial na área de Organização e Representação da Informação e do Conhecimento, evidenciam a pertinência e adequação do percurso figurativo e do percurso temático de Greimas, onde ocorrem os processos de tematização e figurativização, para a análise documentária de objetos informacionais como textos e imagens, apontando diretrizes e metodologias para a indexação desses recursos. Tal aspecto contribui para a interdisciplinaridade das pesquisas sobre a área supracitada e a linguística.

Albuquerque (2011) em sua pesquisa constata a eficácia da Semântica Discursiva para a indexação textos, estabelecendo como recorte os folhetos de cordel, a partir da identificação de figuras e temas que subsidiam a indexação dessa espécie documental. Além disso, essa pesquisadora propôs temas e 27 classes temáticas para a organização de acervos que possuem folhetos de cordel. De outro modo disposto,

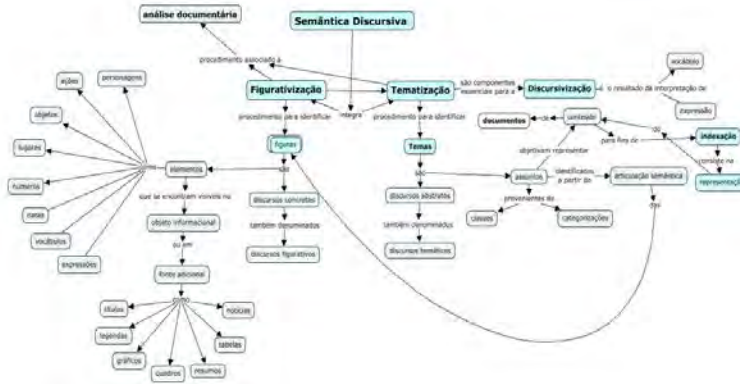
Gandier e Pinho (2018) ressaltam a importância da Semântica Discursiva para a análise documental de textos ficcionais, especificamente crônicas, constatando que esse método pode ser satisfatoriamente aplicado a essa tipologia textual.

De maneira complementar, Santos (2019) em sua pesquisa doutoral constata as contribuições da Semântica de nível discursivo para a indexação de imagens, estabelecendo como recorte as xilogravuras de cordel. Essa autora enfatiza que os métodos e técnicas de indexação de imagens existentes na literatura pautam-se, principalmente, na análise descritiva de fotografias em diferentes níveis de detalhamento, não contemplando elementos intrínsecos de imagens específicas como as xilogravuras, por exemplo. **Diante disso, surge a necessidade de encontrar alternativas que auxiliem o trabalho do profissional da informação na prática de indexação desses recursos imagéticos, fornecendo caminhos** para que os temas desses objetos informacionais atendam os parâmetros conceituais das imagens, reduzindo o nível de dispersão interpretativa do profissional indexador - aspecto inerente à operação.

Diante do exposto, percebe-se que os percursos temático e figurativo possuem qualidades metodológicas a serem aplicadas na leitura documentária. Para tanto, faz-se necessário que o indexador reconheça as superestruturas, ou seja, as estruturas globais responsáveis, independentemente do conteúdo, pela caracterização do tipo do objeto informacional, considerando as suas especificidades. Tal processo contribui para que esse profissional reconheça quais os pontos de acesso no objeto informacional que facilitam a identificação das figuras e dos temas para fins de indexação, podendo utilizar fontes adicionais como subsídios para o reconhecimento do(s) tema(s) a ser(em) utilizado(s) na indexação.

A Figura a seguir apresenta de maneira sistematizada os contributos da Semântica Discursiva, através dos procedimentos de figurativização e tematização, para a análise documentária a partir dos aspectos discutidos anteriormente.

Figura 2 - Contributos da Semântica Discursiva para a análise documental



Fonte: Elaborado pelas autoras com base nas ideias de Santos (2019).

Conforme elucidado anteriormente, a Semântica Discursiva integra os discursos figurativos (concretos) e os discursos abstratos (temáticos). Aqui os discursos figurativos são tratados como figuras e correspondem aos elementos que se encontram visíveis ou descritos no item informacional, como por exemplo: personagens, ações, objetos, lugares, números, vocábulos, expressões, etc. Seguindo essa lógica, o procedimento de análise e identificação dessas figuras no objeto informacional é denominado de figurativização.

Já os discursos abstratos correspondem aos temas, os quais são identificados a partir da confluência ou da imbricação das figuras identificadas em um dado recurso informacional, no intuito de representar o seu conteúdo por meio de conceitos oriundos de classes e categorizações. Tal procedimento é denominado de tematização.

As figuras e os temas também podem ser identificados e extraídos em fontes adicionais relacionadas ao recurso informacional, como por exemplo: títulos, legendas, gráficos, quadros, resumos, tabelas, sinopses, etc. Nesse caso, deve-se considerar a relação semântica entre as figuras extraídas dessas fontes e o material informacional analisado.

Portanto, a figurativização e a tematização são procedimentos que pertencem à Semântica Discursiva e são componentes essenciais ao processo discursivização - resultado da interpretação de um vocábulo

ou expressão com base em tais procedimentos. Por exemplo, o vocábulo “*tortura*” figurativiza o tema “*sofrimento*”, que significa sensação consciente ou inconsciente de dor, mal-estar ou infelicidade. Portanto, “*tortura*” é uma figura associada ao tema “*sofrimento*”, cujo significado apontado anteriormente corresponde à discursivização.

Desse modo, considera-se também a importância de o indexador conhecer as características do sistema, bem com as funções e usos do recurso a ser indexado no contexto da unidade de informação. Esses aspectos também podem ser aplicados e/ou aperfeiçoados para contribuir no processo de análise documentária de diferentes tipologias de recursos (imagens, textos, áudios, vídeos, etc.).

Portanto, observa-se que a Ciência da Informação não pode se isentar dos estudos sobre as tendências contemporâneas de representação da informação que visam contribuir para o desenvolvimento da competência dos profissionais da informação, pesquisadores e usuários que lidam com as práticas de representação temática de recursos informacionais.

4 CONSIDERAÇÕES FINAIS

A análise documentária e a linguística integram ferramentas que podem fornecer subsídios teóricos e práticos para a correta interpretação do conteúdo de itens informacionais, objetivando compreender os sentidos e significados dos assuntos explorados nos mesmos para fins de indexação, recuperação, acesso de uso da informação.

Sujeitos que fazem parte de determinadas culturas podem não conseguir (ou apresentar dificuldades para) reconhecer, assimilar e interpretar os personagens, elementos e ações retratadas em objetos informacionais pelo fato de não possuírem conhecimentos adequados para entendê-los. Isso se deve ao fato de que os indivíduos atribuem sentido às coisas pela maneira como as utilizam e associam em suas práticas cotidianas. Por isso a importância de considerar as novas metodologias que contribuem para a análise documentária, no intuito de contribuir para a produção significativa de metadados semânticos.

Com base nessa perspectiva, a análise documentária deve estar pautada, primeiramente, no objeto informacional e na articulação

semântica do mesmo com o conteúdo das fontes adicionais relacionadas ao recurso (como o título, legendas, resumos, sinopses, quadros, tabelas, *tags* atribuídas pelos usuários em um dado sistema, etc.), contemplando os procedimentos de figurativização e tematização da Semântica Discursiva. Portanto, esta proposta está diretamente associada à indexação, não incluindo procedimentos voltados para a representação descritiva desses recursos informacionais.

Diante do exposto, acredita-se que os contributos da Semântica Discursiva para a análise documentária resultam em uma nova maneira de extrair os conteúdos de recursos informacionais, trazendo como diferencial a redução do nível de subjetividade, tornando a indexação mais próxima possível de atender os parâmetros conceituais dos documentos. Para tanto, a criação de uma política de indexação em consonância com as especificidades e singularidades dos recursos informacionais a serem tratados, consiste em um fator importante para a execução de um trabalho eficiente em bancos de dados e sistemas de recuperação da informação.

REFERÊNCIAS

- ALBUQUERQUE, M. E. B. C. *Literatura popular de Cordel: dos ciclos temáticos à classificação bibliográfica*. 2011. Tese (Doutorado em Letras) – Universidade Federal da Paraíba, João Pessoa, 2011. Disponível em: <https://repositorio.ufpb.br/jspui/handle/tede/6183>. Acesso em: 26 abr. 2019.
- ANTONIO, D. M. *O Percurso Gerativo de Sentido aplicado à análise documental de textos narrativos de ficção: perspectivas de utilização em bibliotecas universitárias*. 2008. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2008. Disponível em: http://bdtd.ibict.br/vufind/Record/UNSP_af8c4a251e7e06516d941efe1b9e8c22. Acesso em: 26 abr. 2019.
- BARBALHO, C. R. S. Fazer semiótico: subsídios para exame do espaço concreto. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, Florianópolis, 2. n. esp., p. 79-96, 2006.
- BATISTA, M. F. B. de M. O discurso semiótico. In: ALVES, E.; CHRISTIANO, M. E. (org.). *Linguagem em foco*. João Pessoa: Editora Universitária; Ideia, 2001. p. 120-139.
- BLAIR, D. C. Information retrieval and the philosophy of language. *Annual Review of Information Science Technology*, Medford, v. 37, p. 3-50, 2003.

- CAMPOS, A. T. A indexação. *Revista de Biblioteconomia de Brasília*, Brasília, v. 15, n. 1, p. 69-72, jan./jun. 1987. Disponível em: <http://www.brapci.ufpr.br/documento.php?dd0=0000003009&dd1=33e85>. Acesso em: 26 abr. 2019.
- CAPURRO, R. Epistemologia e ciência da informação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 5., 2003, Belo Horizonte. *Anais [...]*. Belo Horizonte: ANCIB, 2003. Disponível em: <https://bit.ly/2qQTUZZ>. Acesso em: 26 abr. 2019.
- CHAUMIER, J. Indexação: conceito, etapas, instrumentos. *Revista Brasileira de Biblioteconomia e Documentação*, São Paulo, v. 21, n. 1/2, p. 63-79, jan./jun. 1988.
- FIORIN, J. L. *Elementos de análise do discurso*. São Paulo: Contexto, 1999.
- FIORIN, J.; SAVIOLI, F. *Para entender o texto*. São Paulo: Ática, 2003.
- FOGL, J. Relations of the concepts 'information' and 'knowledge'. *International Fórum on Information and Documentation*, The Hague, v. 4, n. 1, p. 21-24, 1979.
- FRANCELIN, M. M.; PINHO, F. A. *Conceitos na Organização do Conhecimento*. Recife: Editora Universitária da UFPE, 2011.
- GANDIER, A. M.; PINHO, F. A. A importância da semântica discursiva para a análise documental: um estudo em texto ficcional. *Brazilian Journal of Information Science*, Marília, v. 12, n. 2, p. 13-23, 2018. Disponível em: <http://www2.marilia.unesp.br/revistas/index.php/bjis/article/view/7897>. Acesso em: 26 abr. 2019.
- GAUDÊNCIO, S. M.; ALBUQUERQUE, M. E. B. C. Diálogos teóricos em representação da informação. In: ALVES, E. C. et al. (org.). *Práticas e abordagens contemporâneas em ciência da informação*. João Pessoa: Ed. da UFPB, 2014. p. 57-78.
- GREIMAS, A. J. Condições de uma semântica científica. In: GREIMAS, A. J. *Semântica estrutural*. Tradução Haquira Osakabe. 2. ed. São Paulo: Cultrix; USP, 1966. p. 28-53.
- GREIMAS, A. J.; COURTÉS, J. *Dicionário de semiótica*. São Paulo: Contexto, 2008.
- HALL, S. *Cultura e Representação*. Tradução Daniel Miranda e Wiliam Oliveira. Rio de Janeiro: PUC-Rio; Apicuri, 2016.
- HJELMSLEV, L. *Prolegômenos a uma teoria da linguagem*. São Paulo: Perspectiva, 1975.
- LANCASTER, F. W. *Indexação e resumos: teoria e prática*. 2. ed. Brasília: Briquet de Lemos, 2004.
- LUCAS, C. R. *Leitura e interpretação em Biblioteconomia*. Campinas: Ed. da Unicamp, 2000.
- MEDEIROS, M. B. B. M. Terminologia Brasileira em Ciência da Informação: uma análise. *Ciência da Informação*, Brasília, v. 15, n. 2, p. 135-142, jul./dez. 1986.

Disponível em: <http://ridi.ibict.br/bitstream/123456789/283/1/BRASCHERCI1986.pdf>. Acesso em: 26 abr. 2019.

MOREIRA, W. *Sistemas de Organização do Conhecimento: aspectos teóricos, conceituais e metodológicos*. 2018. Tese (Livre-Docência em Sistemas de Organização do Conhecimento) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2018. Disponível em: http://repositorio.unb.br/bitstream/10482/7465/1/2010_ElianaCarlan.pdf. Acesso em: 26 abr. 2019.

MOTTA, D. F. *Método Relacional como nova abordagem para a construção de tesaurus*. Rio de Janeiro: SENAI/DN/DPEA, 1987.

PATO, P. R. G. *Imagens: polissemia versus indexação e recuperação da informação*. 2015. Tese (Doutorado em Ciência da Informação) – Faculdade de Ciência da Informação, Universidade de Brasília, Brasília, DF, 2015. Disponível em: <http://repositorio.unb.br/handle/10482/19050>. Acesso em: 26 abr. 2019.

SANTAELLA, L. *A percepção*. São Paulo: Experimento, 1993.

SANTOS, R. F. *Modelos colaborativos de indexação social e a sua aplicabilidade na Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (BRAPCI)*. 2016. Dissertação (Mestrado em Ciência da Informação) - Universidade Federal de Pernambuco, Recife, 2016. Disponível em: <https://repositorio.ufpe.br/handle/123456789/17218>. Acesso em: 26 abr. 2019.

SANTOS, R. F. *Indexação de xilogravuras em versos: a representação entre o real e o imaginário coletivo*. 2019. Tese (Doutorado em Ciência da Informação) - Universidade Federal da Paraíba, João Pessoa, 2019.

SAUSSURE, F. de. *Course in general linguistics*. Paris: Payot, 2004.

ZENG, M. L. Knowledge Organization Systems (KOS). *Knowledge Organization*, Baden-Baden, v. 35, n. 2/3, p. 160-182, 2008.

6

A LEITURA DOCUMENTÁRIA E AS PERSPECTIVAS DE APLICAÇÃO EM CONTEXTO DA *WEB*

Gercina Ângela de LIMA

RESUMO: A mudança do texto impresso para o texto eletrônico criou uma grande transformação na maneira como se armazena e acessa a informação. O hipertexto, fruto da tecnologia informacional das últimas décadas, é a forma texto que estimula e torna possível o desenvolvimento do universo digital. No ambiente Web, o processo de indexação mostra-se útil e eficiente para criar a estrutura conceitual com *links* para as partes relevantes do texto. Neste capítulo, apresenta-se um estudo sobre a contribuição da leitura documentária na conversão de textos em hipertextos no contexto da Web, do ponto de vista do autor. Os estágios da leitura documentária, no processo de *análise de assunto*, constituem uma abordagem importante a ser considerada na estruturação de hipertextos. Conclui-se que os estágios previstos na leitura documentária realizados no texto linear podem auxiliar o autor na estruturação de hipertextos, desde que sejam adaptados para a representação conteúdo do texto não linear, garantindo a interatividade entre módulos de informação, que garante a compreensão global do texto.

PALAVRAS-CHAVE: Leitura documentária. Hipertexto. Web.

ABSTRACT: The change from printed text to electronic text has created a major transformation in the way information is stored and accessed. Hypertext, fruit of the information technology of the last decades, is the text form that stimulates and makes possible the development of the digital universe. In the Web environment, the indexing

process is useful and efficient to create the conceptual framework with links to the relevant parts of the text. In this chapter, we present a study about the contribution of documentary reading in the conversion of texts into hypertexts in the context of the Web from the author's point of view. The stages of documentary reading, in the process of subject analysis, constitute an important approach to be considered in the structuring of hypertexts. It is concluded that the stages provided in the documentary reading in the linear text can help the author in the structuring of hypertexts, as long as they are adapted to the nonlinear text content representation, guaranteeing the interactivity between information modules, which guarantees the global comprehension of the text.

KEYWORD: Documentary reading. Hypertext. Web.

1 INTRODUÇÃO

A necessidade de representação da informação para sua posterior recuperação começou desde quando o papel foi introduzido no mundo ocidental, em meados do século XII, e difundiu-se na Europa, entre os séculos XIII e XV, com a necessidade do aumento da quantidade de novos originais e cópias de livros. À medida que crescia o aumento de exemplares disponíveis, aumentava também o estímulo ao registro do conhecimento, o que influenciou muito o clima intelectual na Idade Média. Assim, a escrita cria a figura do leitor, para o qual a realidade passa por um filtro muito mais refinado. Ampliou-se o número de editores e possíveis leitores. A partir daí, a leitura e a interpretação do texto adquiriram caráter mais individualizado, com particularidades e exigências diversas. As obras começaram a incluir representações gráficas mais precisas e específicas, como tabelas, gráficos e desenhos, incorporando inventos e artifícios anteriores à tipografia. Formas mais sofisticadas de leitura não linear começaram a surgir, como as notas de rodapé, sumários, por exemplo, que direcionavam o leitor a outras partes do mesmo livro ou a outras obras relacionadas, como índices e bibliografias.

Desde então, a comunicação escrita e o modo de transmissão dos textos sofreram profundas mudanças. A escrita em suportes artesanais foi a tecnologia de comunicação mais avançada, desde o quarto milênio a.C. até a invenção da imprensa e da tipografia na Mogúncia (c.1450), quando Johann Gutenberg compôs, com tipos móveis, o texto da Bíblia, o primeiro livro a passar do manuscrito para o papel impresso (McGARRY, 1999).

A informática provocou muito mais do que uma revolução nas formas e nos métodos de geração, armazenamento, processamento e transmissão da informação. Pode-se dizer que o desenvolvimento das tecnologias relacionadas com o computador compara-se à revolução causada com a invenção dos tipos móveis de Gutenberg. A mudança do texto impresso para o texto eletrônico criou uma grande transformação na maneira como armazenamos e acessamos a informação. Os textos eletrônicos, que podem ter as mesmas características dos textos lineares, bem como dos não lineares, são armazenados ou disponibilizados em HDs (*Hard Disk* ou Disco Rígido), disquetes, *CD-ROM*, *LANs*, ou redes digitais. O formato eletrônico pode ser considerado como um formato adicional, em vez de substituição dos formatos existentes. Porém, a possibilidade de explorar o texto de maneira interativa introduziu o conhecimento por simulação.

A internet tem poucos pontos de entrada, uma estrutura pobre e uma forte ênfase na busca por assunto. A teoria e processos da indexação subsidiam diversas atividades em uma teoria comum e diversas técnicas, porém suas aplicabilidades na Web precisam ser adaptadas. Por exemplo: no ambiente Web, o processo de indexação pode ser útil e eficiente para criar estrutura conceitual com termos que podem ser armazenados na base de dados e servir como *links* para as partes relevantes do texto. Somado a isso, o uso de um Sistema de Organização do Conhecimento, do tipo tesouro, pode auxiliar no controle do vocabulário para melhorar a recuperação na Web, reduzindo a frustração e ajudando a assegurar que a informação relevante não se perca.

Lima (2004) aponta que o hipertexto, fruto da tecnologia informacional das últimas décadas, é a forma texto que estimulou e tornou possível o desenvolvimento do universo digital. O hipertexto, como nova tecnologia, também desterritorializa o texto, tornando-o sem fronteiras, o que remete, de certa forma, à transmissão da informação na tradição oral, em que o elemento da aleatoriedade é inerente à comunicação. Porém, o texto eletrônico acelera o tempo e concatena contextos; encadeia e justapõe diferentes documentos, num compartilhamento do mesmo espaço de produção e, às vezes, alterando sua compreensão.

Os hipertextos já eram utilizados, em modo off-line, para representação de informações, imagens, sons, organizados de forma a

promover uma leitura (ou navegação) não linear, baseada em indexações e associações de ideias e conceitos, sob a forma de *links*. No entanto, essa concepção de textualidade possibilitou a surgimento da *World Wide Web* (WWW), nascida em 1991, que é a parte da Internet baseada nos princípios do hipertexto. Uma característica marcante do ambiente Web é o fato de ser uma organização descentralizada, o que a torna um sistema ágil, diferente do ambiente físico (ou real), com seus tradicionais sistemas organizados de forma centralizada. As linguagens de marcação de textos *HTML* (*Hypertext Markup Language* ou Linguagem de Marcação de Hipertexto) e *XML* (*eXtensible Markup Language* ou Linguagem de Marcação Estendida) são consideradas a linguagem básica de representação do hipertexto na Web. Sendo que a linguagem *XML* surgiu como recurso diferenciado na autoria de hipertextos, pois permite aos autores a marcação de conteúdo semântico dentro do próprio texto e o acréscimo de comandos, explicitando decisões de fragmentação e interconexão segundo configurações mais convenientes para aquele documento específico.

De acordo com Mai (2004, p. 93), a maioria das pesquisas da área de organização da informação e do conhecimento concentra-se mais na melhoria da representação do item bibliográfico na WWW do ponto de vista descritivo, por meio de seus metadados, do que a representação de seu conteúdo, e exemplifica esta realidade com uma citação: “Weinberg (1999) calls it “*the stepchild of the metadata literature*.”, que, apesar de ser de 1999, tem essa perspectiva mantida até o momento atual.

No entanto, o processo de indexação, especificamente a leitura documentária na etapa da *análise de assunto*, para modelagem conceitual de um hipertexto, dá-se de forma diferenciada por suas características não lineares, o que torna mais complexo realizar a leitura documentária com perspectivas de sua representação na Web. Agregam-se a este fato as diversas tipologias documentais existentes e suas estruturas, que exigem conhecimentos diferenciados, no que tange à identificação de seus conceitos, tornando, assim, um procedimento ainda mais complexo para sua representação.

O principal papel da leitura documentária é a identificação de conceitos de um determinado documento para que seu conteúdo possa ser representando visando a sua recuperação pelo leitor em uma unidade de informação. Porém, existem variáveis que irão permear esta leitura textual,

que é realizada diferentemente da convencional, que integra o conhecimento profissional do leitor-indexador, ao realizar este procedimento, perpassando pelos aspectos linguísticos, cognitivos e lógicos que são associados a este procedimento, e as variáveis: o texto, o leitor e o contexto.

Neste capítulo, não se pretende esgotar a temática sobre hipertextos como um documento Web tanto na sua construção quanto na sua navegação e recuperação, mas abordar as perspectivas da contribuição da leitura documentária na etapa da *análise de assunto*, para modelagem conceitual de hipertextos, com vistas à sua disponibilização na Web, do ponto de vista do autor conteudista.

2 HIPERTEXTO

A inquietude para o surgimento de uma nova forma textual já se delineava no início da década de 1970. Ironicamente, a necessidade de mudança nos conceitos das funções do autor e do leitor, do livro e do texto, foi trazida à tona por expoentes da teoria crítica literária como Foucault (1979), Derrida (1977) e Barthes (1979), cujas posturas filosóficas tratavam do texto tradicional, e não do hipertexto. Para Foucault (1979, p. 102) “[...] o texto é como uma rede formada por interconexões, na qual as fronteiras não são bem definidas, mas captadas em um sistema de referências e outros livros, outros textos.”. Derrida (1977, p. 173) aponta que “[...] o texto ideal é composto por unidades que, apesar de separadas, podem se complementar na construção do sentido. Cada unidade pode ‘quebrar’, modificar o contexto e abrir para uma infinidade de novos contextos.”. Já Barthes (1979, p. 8) descreve o texto ideal como “[...] um composto de blocos de palavras (ou imagens), perpetuamente ‘não finalizados’, descritos a partir de termos como nós, networks, redes e caminhos.”

Permeando a reflexão desses três filósofos, está a ideia de um rizoma que se materializaria na forma de texto hipertextual, na qual não haveria uma hierarquia de início ou fim, nem de acima ou abaixo, nem de antes ou depois. Foucault fala da interrelação, invisível a muitos, entre obras separadas fisicamente. Derrida reconhece a fragmentação inerente ao texto, mas também sua mobilidade e potencial de transformação como uma entidade viva. Já Barthes aborda o texto como fenômeno em processo, não como matéria acabada.

Segundo Koch (2002), se considerarmos que a linguagem é uma atividade interativa que conduz à concepção processual da construção do sentido e que todo texto é constituído por uma proposta de múltiplos sentidos, como querem Foucault, Derrida e Barthes, pode-se afirmar que todo texto é um hipertexto. Para que o leitor possa construir um sentido, que nem sempre se dá de maneira linear e sequencial, é necessário realizar um constante movimento em variadas direções, recorrendo a diversas fontes de informação textuais e extratextuais. Dessa forma, o hipertexto é uma forma de estruturação textual que permite ao leitor ser uma espécie de coautor do texto, oferecendo-lhe a possibilidade de opções entre caminhos diversificados, de modo a permitir diferentes níveis de desenvolvimento e aprofundamento do tema. Esta variabilidade do grau de aprofundamento é notável em textos impressos. Povoado de referências, citações, notas de rodapé, quadros, tabelas, exemplos etc., sua concentração de estímulos convoca o leitor a pensar sobre outros textos, consultá-los e até mesmo suspender sua leitura, em um processo de alteração de curso para obter outras informações ou fazer anotações. Como o hipertexto oferece uma multiplicidade de caminhos a seguir, compete ao leitor decidir qual o fluxo de informações será incorporado em sua trajetória de leitura.

A definição de texto virtual de Bolter (1991), duas décadas depois de Foucault, Derrida e Barthes, também confirma a inquietude já presente naqueles autores:

[...] texto é um contínuo de parágrafos, dispostos na tela do computador para o leitor ler através de um caminho tradicional. Algumas palavras estão marcadas em negrito; estilo que indica que há uma observação para aquela palavra ou frase, [...] uma segunda janela pode também conter frases em negrito que podem levar o leitor para outros parágrafos. [...] O processo pode continuar indefinidamente [...] formando uma rede que é chamada de hipertexto. (BOLTER, 1991, p. 15).

Bolter (1991) nos mostra, também, que a noção de texto acompanhou o desenvolvimento da tecnologia de transmissão da informação:

Todas as formas de escrita são espaciais [...] Cada tecnologia dá-nos um espaço diferente. Para certas escritas antigas, o espaço [bidimensional] era a superfície interna do rolo contínuo [de pergaminho], que o escriba dividia em colunas. Para a escrita medieval e a impressão na era moderna, o espaço é a superfície branca de uma página, especialmente em volumes encadernados. Para a escrita eletrônica, o espaço é a tela do monitor de um computador, onde o texto é exposto e armazenado em memória eletrônica. (BOLTER, 1991, p. 11).

No contexto histórico, o documentalista e advogado belga Paul Otlet e seus colegas se anteciparam muito da funcionalidade do hipertexto e desenvolveram um complexo de organizações funcionalmente similares a ele, conforme uma série de trabalhos publicados entre 1892 e início deste século. Na obra *Traité de Documentation* (1934), Otlet identificou máquinas intelectuais e processos contendo dispositivos de som e conectados ao telefone, telegrafia e televisão, com o armazenamento baseado principalmente em cartões e folhas de papel de tamanho padrão. A invenção destas máquinas ajudaria a realizar um novo tipo de enciclopédia e a tornaria muito próxima a um anexo do cérebro, um substrato da memória. O autor vislumbrou um sistema internacional de manipulação de informações abrangendo desde a criação de uma entrada em um catálogo até as novas formas de publicação, do gerenciamento de bibliotecas, arquivos e museus como agências de informação inter-relacionadas até o desenvolvimento colaborativo de uma enciclopédia universal codificando todo o conhecimento humano. No centro dessas ideias, estavam o sistema de Classificação Decimal Universal, um novo tipo de agência de informação (*Office of Documentation*), um novo princípio de armazenamento e indexação da informação (Princípio Monográfico) e o microfilme.

Em 1945, Vannevar Bush (1987a) publica o artigo *As we may think*²⁶, no qual previu um sistema muito parecido com o hipertexto, com a ideia básica de organizar documentos em trechos e combiná-los conforme as necessidades de compreensão e organização, o qual denominou de *Memex*. Esse sistema teria a capacidade para armazenar um grande volume

²⁶ Publicado originalmente em: *The Atlantic Monthly*, v. 176, n.1, p.101-108, July 1945.

de informações e, por meio de determinado mecanismo, estabeleceria associações entre diferentes pontos da base de dados, propiciando a sua consulta com rapidez e flexibilidade. Assim, o *Memex* serviria como suplemento para a memória dos usuários de um banco de dados de literatura científica. Ele propôs a organização associativa que tentava imitar a estrutura da memória humana.

Além de Bush e Otlet, Theodor Nelson (1965) e Douglas Engelbart (1962) foram outros pesquisadores que compartilharam a mesma visão de que seria estabelecida nova forma de interação homem-computador. Theodor Holm Nelson foi quem criou o termo hipertexto, definindo-o como “[...] uma combinação de texto em linguagem natural com a capacidade do computador de fazer pesquisa interativa e exibição dinâmica de um texto não linear.” (CONKLIN, 1987, p. 17). O sistema *Xanadu*, desenvolvido por ele e sua equipe de pesquisadores, possuía um gerenciador de armazenamento que permitia a criação de ligações entre os tópicos correlatos e mantinha informações sobre a origem, as variações e as interconexões dos textos. Na década de 1960, Nelson começou a estruturar alguns fundamentos do hipertexto e apontou as limitações da informática em face da nova ferramenta. Paralelamente aos estudos de Nelson, Douglas Engelbart, do Instituto de Pesquisa Stanford, EUA, estudou a utilização do computador como instrumento representacional e ampliador do intelecto humano (MARQUES, 1995). Foram apresentados novos conceitos relativos à interação homem-máquina para a construção de sistemas mais dinâmicos e flexíveis. Juntamente com outros pesquisadores, implementou o sistema *On Line System* (OLS), conhecido mundialmente como sendo a primeira implementação que seria conhecida como um hipertexto, estruturado por uma rede de *links* que interligavam blocos pequenos de informação. Esse sistema foi o primeiro a introduzir noções como edição estruturada, uso do *mouse*, correio eletrônico, janelas múltiplas e tela compartilhada. Em 1969, Bush publicou o artigo *Evolution of an information society* (republicado em 1987b),²⁷ concluindo que o *Memex* podia, finalmente, ser viabilizado, porém a um alto custo, exigindo também manipulação de alta tecnologia e uma enorme dose de paciência dos usuários.

²⁷ BUSH, Vannevar. MEMEX revisited. In: CALKELL, A. E. (ed.). *Evolution of an information society*. London: ASLIB, 1987. p. 179-191. Publicado originalmente em *Science is not enough*. New York: Apollo, 1969. p. 75-101.

Os sistemas de hipertexto são fruto da tecnologia informacional das últimas décadas e causaram uma verdadeira revolução nos conceitos de armazenamento e recuperação de informações. As rupturas com as formas tradicionais de organização da informação vão desde as características de entrada da informação e a criação das estruturas de armazenamento, até o desenvolvimento radical no processo de recuperação da informação, possibilitado pela interação homem–máquina.

Com o surgimento da Web, no início da década de 1990, os sistemas de hipertextos alcançaram maior popularidade e visibilidade mundial. Hoje o sistema de hipertexto mais conhecido é a própria rede Web, na qual o navegar no hipertexto, dado seu alcance mundial, é uma realidade para milhões de pessoas em todo o mundo (NONATO, 2009). A mudança das páginas estáticas disponíveis na internet para páginas interativas com conexões, utilizando o protocolo HTTP (*Hypertext Transfer Protocol*), alterou a forma de acesso da informação na Internet. Gomes (2011, p. 23) afirma que “[...] podemos dizer que a Web, hoje, é um hipertexto aberto, pois quase tudo o que há nela está interconectado, e documentos podem ser acessados e acrescentados a partir de qualquer ponto (mesmo simultaneamente).”.

A maioria das estratégias de estruturação de hiperdocumentos se apoia nas linguagens de marcação de textos como *HTML* (*Hypertext Markup Language* ou Linguagem de Marcação de Hipertexto) e *XML* (*eXtensible Markup Language* ou Linguagem de Marcação Estendida). Segundo o W3C (2010), a linguagem de marcação de hipertexto evoluiu para as versões HTML+, HTML2.0 e HTML3.0, de 1993 a 1995, quando várias mudanças foram propostas para melhoria da linguagem e, somente em 1997, trabalhou na versão 3.2 da linguagem, fazendo com que ela fosse tratada como padrão de linguagem. Também para facilitar a manipulação dos elementos, possibilitando, assim, ao desenvolvedor, modificar as características dos objetos, de forma que sejam transparentes para o usuário final, criando novas *tags* e permitindo modificar a função das demais. Em 2007, a W3C buscou soluções para implementação de uma versão da linguagem HTML moderna e compatível com as versões existentes. Para tanto, a linguagem deve ser única, podendo ser escrita tanto com sintaxe HTML como XML (*eXtensible Markup Language*), denominada HTML5.

Essas linguagens são gerenciadas por uma série de comandos que, uma vez inseridos em um documento tradicional, identificam sua estrutura lógica em títulos, seções e parágrafos. Esses tipos de componentes de um documento servem de guias para o processo de conversão de textos em hipertextos, pois sugerem maneiras de fragmentá-los e de conectar algumas de suas partes entre si. Estas divisões lógicas em títulos, seções e parágrafos podem ser suficientes para determinar uma boa configuração de um hipertexto, no caso de documentos cuja organização interna é bem-definida e rígida, a exemplo de dicionários e catálogos. Todavia, podem ser pouco eficientes em textos livres com subdivisões menos sugestivas, para os quais os algoritmos de conversão gerariam redes não muito adequadas à estruturação de seu conteúdo semântico. Entre as linguagens de marcação, o *XML* surgiu como recurso diferenciado na autoria de hipertextos, pois permite aos autores a marcação de conteúdo semântico dentro do próprio texto e o acréscimo de comandos, explicitando decisões de fragmentação e interconexão segundo configurações mais convenientes para aquele documento específico.

Encontram-se na literatura vários estudos e definições sobre hipertexto advindas de diferentes áreas do conhecimento, tais como: Ciência da Computação, Linguística, Educação, Ciências Cognitivas, Ciência da Informação, Gestão da Informação e Comunicação. Cada uma dessas perspectivas abordam o hipertexto sob um ponto de vista que vai ao encontro de sua aplicabilidade em cada área especificamente. Salienta-se, também, que alguns autores não fazem distinção entre hipertexto e hiperímídia. Entretanto, a hiperímídia ultrapassa o hipertexto por considerar qualquer tipo de mídiã, não apenas texto, e com o suporte de qualquer tecnologia, e não apenas de computadores (RADA, 1991b).

Uma das definições mais citadas sobre hipertexto é a de CONKLIN (1987, p. 17), que o define como “[...] janelas na tela [que] são associadas a objetos na base de dados e ligações [que] são estabelecidas entre estes objetos, tanto graficamente, na forma de marcas rotuladas, como na base de dados, na forma de ponteiros.”.

Conklin (1987) apresenta, ainda, três elementos que podem ser considerados hipertextos: (1) *um componente de banco de dados*, em que objetos na tela são associados a objetos armazenados em um banco de dados. As conexões entre estes objetos podem ser definidas não apenas graficamente, mas também por ponteiros em uma base de dados; (2)

um esquema de representação, no qual a informação textual informal é misturada com processos e operações mais formais e mecanizados. Para ser qualificado como “hipertexto”, o sistema deve proporcionar um suporte eletrônico constante para suas inter-relações; (3) *uma interface com o usuário*, geralmente janelas representando a correspondência “um-para-um” com os nós/nodos na base referencial de dados. A interface gráfica, com o auxílio de *browsers* e diagramas, auxilia o usuário a navegar através de uma grande quantidade de informações, ativando as ligações e acessando o conteúdo dos nodos.

Para Snyder (1997 p. 126), o hipertexto é como “[...] uma estrutura composta por blocos de textos conectados por *links* eletrônicos, os quais oferecem diferentes caminhos para os usuários.”.

Levy (1993, p. 33), por sua vez, define o hipertexto como

[...] um conjunto de nós ligados por conexões. Os nós podem ser palavras, páginas, imagens, gráficos, sequências sonoras, documentos complexos que podem eles mesmos ser hipertextos. Os itens de informação não são ligados linearmente, como uma corda com nós, mas cada um deles, ou a sua maioria, estende suas conexões em estrela de modo reticular.

No Brasil, um dos primeiros autores a estudar sobre o hipertexto foi Vilan Filho (1994), que o definiu como “[...] uma rede de nós e ligações entre documentos, onde documentos são os nós e as ligações são referências cruzadas.” (VILAN FILHO, 1994, p. 297). Esse autor destaca ainda que

Os sistemas de hipertextos são conjuntos de programas de computador (suporte lógico), [com] suas tabelas e dados de controle necessários para a operação de um sistema construído para operar com hiperdocumentos, segundo a filosofia de hipertexto. Em geral, esses sistemas são compostos de: (a) um subsistema de autoria, e (b) um subsistema de navegação (VILAN FILHO, 1992, p. 61).

Lima (2004, p. 23) considera que o hipertexto pode ser

[...] tanto uma ferramenta para o autor quanto para o leitor. Para o autor, funciona como ferramenta que ajuda a organizar, estruturar e armazenar informações. Para o leitor, ele permite, através da navegação, a recuperação da informação organizada.

De acordo com Lima (2004), os sistemas de hipertexto traduzem uma abordagem de estruturação e manipulação de textos caracterizada pela não linearidade, tendo como seus componentes básicos:

a) *Nó* (ou *Nodo*): em cada uma das unidades de informação na base hipertexto, os nós correspondem a uma ou mais exibições de tela. Essas unidades de informação em um hiperdocumento podem conter diferentes tipos de dados, textos, figuras, fotos, sons e são conectadas por *links* a uma variedade de estruturas. Geralmente, o nó descreve um único conceito ou tópico, de modo que pode ser caracterizado como autocontido, não dependendo da leitura prévia de outros nós.

b) *Link*: conceito mais importante do hipertexto, o *link* é também denominado elo, ligação, vínculo, âncora ou botão, de acordo com o sistema de hipertexto. *Links* são marcas que conectam um nodo com outro. A ativação desses *links* implica abertura de nova janela, contendo o documento referenciado. Assim, é possível deslocar-se fácil e rapidamente, de um ponto a outro, no conteúdo das páginas. Esses *links* podem ser representados por palavras ou frases em destaque (negrito, itálico, cores), ou, ainda, por figuras ou ícones.

Frei e Stieger (1995) dividem os *links* em dois tipos: *links* referenciais (ou organizacionais) e *links* semânticos (ou de conteúdo). A finalidade do *link* referencial é tornar a leitura do documento mais confortável. Já a função do *link* semântico é apontar para um detalhamento da informação ou uma informação similar, ou, ainda, uma informação adicional em tópico específico.

Em relação à sua estrutura, o hipertexto determina e descreve um sistema de ligações (*links*) ou relacionamentos entre os nós ou unidades de informação, sendo um fator decisivo na facilidade de criação, uso e atualização do hiperdocumento.

De acordo com os nós conceituais que estão sendo ligados, RADA (1991a) classifica o hipertexto conforme as tipologias microtexto e macrottexto. Microtexto é um neologismo que significa um hipertexto de pequeno volume, com ligações explícitas entre seus componentes. Aparentemente, os microtextos são muito apropriados para navegação. Uma rede de microtextos pode ser vista como uma rede semântica e, por isso, ser tomada como modelos da memória humana. Através dos recursos

computacionais, evidenciam-se o caminho ou os caminhos em que o texto foi internamente construído, e que podem variar de acordo com pontos de vista lógicos. Macrotexto é definido como um sistema hipertextual em larga escala, composto de vários documentos, e cuja ênfase está em *links* que os conectam entre si, em vez de conectar internamente apenas um documento.

Corroborando com essa ideia, Lévy (1993, p. 25) considera o hipertexto como “[...] uma metáfora cálida para todas as esferas da realidade em que significados estejam em jogo.”, e aponta seis princípios básicos de sua lógica, que permitem visão panorâmica, a qual organiza, resume e amplia a ideia de rede que se pretende construir: (1) Princípio de metamorfose: processo de constante construção e renegociação de sentidos, que se dá nos hipertextos. O princípio de metamorfose explicita a ideia de que a rede de significados que constitui o conhecimento está em permanente transformação; (2) Princípio de heterogeneidade: a variedade e diferença entre as informações organizadas em determinada seção do hipertexto, como os nós e as conexões que se estabelecem entre as diversas partes da rede hipertextual. Existem, na memória, informações cognitivas que passam por vários tipos de associações entre estes elementos; (3) Princípio de multiplicidade e de encaixe das escalas: organização “fractal” do hipertexto, ou seja, qualquer nó ou conexão, quando analisado, pode revelar-se como sendo composto por toda a rede, e assim por adiante, indefinidamente, ao longo da escala dos graus de precisão; (4) Princípio de exterioridade: a rede não possui unidade orgânica, nem um “motor interno”. Seu crescimento, diminuição ou recomposição permanente dependem de ação exterior indeterminada. O princípio de exterioridade pretende caracterizar a permanente abertura da rede hipertextual e do conhecimento em construção; (5) Princípio de topologia: nos hipertextos, tudo funciona por proximidade, por vizinhança. Neles, o curso dos acontecimentos obedece a uma topologia, na qual os caminhos seguem a ideia de proximidade entre significados; (6) Princípio de mobilidade dos centros: a rede não tem centro fixo, mas possui diversos centros que são como pontas luminosas perpetuamente móveis, saltando de um nó a outro, trazendo ao redor de si uma ramificação infinita de pequenas raízes, como finas linhas brancas esboçando por um instante um mapa de informações.

Essas seis características da rede hipertextual propostas por Lévy (1993) constituem uma teia básica a partir da qual pode ser tecida a ideia de rede para a representação do conhecimento. Assim, as características próprias do hipertexto e as novas relações entre autor e leitor que este sistema está introduzindo estão transformando radicalmente a comunicação. O hipertexto é capaz de sedimentar uma nova forma de literatura, conforme previsto por alguns autores: Landow (2006), Murray (1997) e Bolter (1991). No cenário atual, vislumbram-se alguns elementos básicos, como: (1) a escrita e a leitura não sequencial; (2) a interatividade possibilitada pelo meio digital; (3) a existência de elos de ligação (*links*) – textuais ou não.

Pode-se dizer, portanto, que o hipertexto não pode ser considerado um gênero textual, mas documentos eletrônicos, organizados em forma de rede, com *links* eletrônicos que possibilitam a interatividade entre as relações inter e intratextuais de diferentes segmentos textuais.

Conclui-se, assim, que o hipertexto é tanto uma ferramenta para o autor (designers/programadores) quanto para o usuário (leitor). Para o autor, funciona como ferramenta que ajuda a organizar, estruturar e armazenar informações. Para o leitor, permite, por meio da navegação, a recuperação da informação organizada. Araújo (2006, p. 38) acrescenta que “[...] se pensarmos que o autor é quem decide como e onde disponibilizar os *links* no hipertexto, não haveria uma liberdade incondicional do leitor e sim uma liberdade, até certo ponto de vista vigiada.” Já Ensslin (2007) registra que existem estudos sobre como os princípios de organização estrutural podem moldar os processos de autoria e leitura do hipertexto. Acrescenta que os papéis de autor e leitor têm, frequentemente, sido revertidos, principalmente no que tange à relativa liberdade de escolha do usuário entre diferentes caminhos de leitura e suposto “desempoderamento” simultâneo do autor.

Para Vilan Filho (1992), a estruturação da informação em sistemas de hipertexto é similar à de uma rede semântica, ou seja, um modelo de representação do conhecimento. Os nós representam conceitos e ideias, e as ligações representam as relações entre eles. As relações devem refletir a estrutura organizacional do assunto relacionado ou, em outras palavras, uma rede semântica de uma área do conhecimento. Se a informação for disposta de maneira organizada, será fácil para o usuário recuperar o que

estiver procurando. Por outro lado, se a estrutura apresentar uma trilha de navegação desorganizada e uma estrutura cognitivamente ineficiente, a busca e o folheio (*browsing*) por parte do usuário será difícil ou ineficiente.

Para este estudo, considerou-se a primeira abordagem, a do autor, apresentada por Pierre Lévy (1993), na qual o hipertexto é utilizado como ferramenta de organização da informação por meio da sua característica de escrita não sequencial e da associação dessas informações através de nós, representados por conceitos, possibilitando assim a construção de uma rede navegacional em documentos completos sem perda significativa do conteúdo semântico após a fragmentação. Como o processo de organização hipertextual é muito parecido com a indexação tradicional, são apresentados, a seguir, os princípios da teoria da indexação e suas possíveis aplicações aos sistemas de hipertexto.

3 INDEXAÇÃO

Indexar é representar um documento por meio de uma descrição abreviada de seu conteúdo essencial. Embora algumas vezes a complexidade desse processo possa ser minimizada, essa é, sem dúvida, umas das atividades mais subjetivas do cotidiano do bibliotecário. A capacidade intelectual de reconhecer o conteúdo de um documento em análise é a questão central do procedimento de indexação. Essa representação é realizada a partir da análise do conteúdo do texto-fonte, com base em metodologias e procedimentos particulares, sempre levando em consideração uma clientela específica.

Para a execução da atividade, faz-se necessário ter uma formação especializada na área de tratamento de conteúdos de documentos. A experiência e os conhecimentos prévios desse profissional diferenciarão o resultado do tratamento informacional realizado por outro indexador menos preparado. Naves (2004, p. 8) afirma que:

O profissional da informação que desenvolve a atividade de indexar assuntos de documentos é chamado de *indexador*, catalogador de assuntos ou classificador. A maioria desses profissionais é graduada em Biblioteconomia, e deve conhecer os fundamentos teóricos e técnicos do tratamento temático da informação.

Por se tratar de uma atividade intelectual, é natural que, no cotidiano dos indexadores, sejam percebidas divergências entre termos atribuídos a um mesmo documento por profissionais de diferentes instituições e contextos. Assim, uma mesma publicação poderá apresentar conjuntos diferentes de termos de indexação, dependendo do grupo de usuários ao qual se destina e dos interesses particulares desse grupo, ou seja, há várias formas ‘corretas’ de se indexar.

De acordo com a ISO 5963, a indexação é vista como “[...] a representação do conteúdo dos documentos por meio de símbolos especiais, quer retirados do texto original, quer escolhidos numa linguagem de informação ou de indexação.” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 1985). Segundo Silva e Fujita (2004, p. 136-137), “[...] o conceito de indexação surgiu a partir da elaboração de índices e atualmente está mais vinculado ao conceito de *análise de assunto*.” De acordo com o UNISIST (1981, p. 8), “[...] indexação é o processo utilizado para descrever e identificar um documento de acordo com seu assunto.”

Na Biblioteconomia e Ciência da Informação (BCI), a *análise de assunto* é parte do processo de indexação que envolve procedimentos cujo número de etapas varia conforme diferentes autores. Jens-Eric Mai (2001) ressalta que o processo de indexação é algumas vezes apontado, na literatura, como tendo duas, três ou mesmo quatro etapas. Conforme as diretrizes da *International Standards Organization-ISO 5963*, do UNISIST (1981) e conforme os autores Foskett (1973), Chaumier, (1988); Fugmann (1993), Taylor (2004) Lancaster (2004), Dias e Naves (2007), esse processo é realizado em duas etapas, sendo a primeira a da *análise de assunto*, ou extração de conceitos; e a segunda etapa a de *tradução* dos conceitos nos termos de uma linguagem de indexação. Já os autores Hutchins (1978), Frohmann (1990) e Blair (1990), Farrow (1991), Taylor (2004) e Petersen (1994) e a ABNT 12.676 (1992) adicionam mais uma etapa, que, apesar de, também, considerarem como primeira a *análise de assunto*, incluem uma etapa intermediária, que reformula o assunto em uma frase de indexação na linguagem natural, para depois fazer a tradução; enquanto Chaumier (1988), Langridge (1989), Van Slype (1991), Chu e O’Brien (1993) e Robredo (2005) consideram quatro etapas. Contudo, todas têm basicamente as mesmas operações. O foco principal deste capítulo é, na primeira etapa, a *análise de assunto*, especificamente, o processo da leitura

documentária e as perspectivas de aplicação em contexto da Web, tendo o hipertexto como objeto de estudo. Sendo assim, na próxima seção, passa-se a descrever mais detalhadamente essa etapa, com ênfase nesse processo da leitura documentária.

3.1 ANÁLISE DE ASSUNTO E A LEITURA DOCUMENTÁRIA

A etapa de *análise de assunto* determina de que trata um documento, isto é, qual é o seu assunto. De acordo com Cesarino e Pinto (1980, p. 32, grifo do autor), “[...] a *análise de assunto* é a operação-base para todo o procedimento de recuperação de informação.”. Isso justifica o fato de todo Sistema de Recuperação da Informação (SRI) utilizar essa técnica, tanto na etapa de entrada de dados no sistema quanto no momento de busca da informação pelo usuário final. A *análise de assunto* pode ser considerada a etapa mais importante do processo de indexação.

Segundo Lancaster (1993, p. 20), “[...] ao indexador raramente é dado o luxo de poder ler um documento do começo ao fim.”, realidade que pode ser comprovada no cotidiano de indexadores de instituições nacionais e internacionais. Para tanto, a leitura e a compreensão do texto são primordiais, porém, o tempo restrito do indexador e a quantidade cada vez maior de documentos que demandam tratamento são fatores preocupantes, porque podem comprometer a qualidade da atividade realizada.

Fujita (2003, p. 69) acrescenta que

[...] o processo de análise de assunto reveste-se (*sic*) de uma subjetividade característica, dadas as circunstâncias e elementos envolvidos, pois, a partir da leitura do documento pelo indexador, é realizado um processo de comunicação interativo entre três variáveis: leitor, texto e contexto. Cada uma dessas variáveis estará sujeita a diferentes condições, mas é o indexador como leitor a variável mais influente nessa interação para análise de assunto, porque precisa realizar a compreensão da leitura mediante sua cognição.

Hjørland (1992) agrega que, para a execução dessa primeira etapa, é preciso considerar o domínio no qual o documento está inserido,

identificando as características específicas do campo de conhecimento, sejam elas de ordem cultural, terminológica, histórica ou linguística. Para tanto, o conhecimento do indexador sobre esse domínio é importante para a qualidade da análise. Assim, a atividade poderá ser feita de acordo com o contexto, pois o documento não será considerado algo isolado, mas parte de um todo.

Segundo Kobashi (1994, p. 112), “[...] o texto, em sentido mais amplo, designa uma unidade de comunicação organizada sintagmaticamente e dotada de coesão e coerência.”. O texto pode ser classificado, quanto à sua estrutura interna, como descritivo, narrativo, dissertativo ou, quanto à sua finalidade, como texto técnico, científico, didático, jornalístico, jurídico, político, de lazer e etc. A grande variedade de tipos de documentos pode ser diferenciada pelas características físicas e intelectuais: livros de leitura corrente, teses, dissertações, livros de referência, publicações periódicas e impressos diversos.

A leitura é um processo interativo de compreensão que depende das variáveis do texto (intenção do autor, forma e conteúdo), do leitor indexador (estruturas e processos), e do contexto (social, psicológico e físico) no qual a leitura está sendo realizada (GIASSON, 1993). O conhecimento prévio do indexador é fundamental para a compreensão do texto e identificação de seu conteúdo, para selecionar os conceitos que serão representados com o uso de vocabulário controlado, para que este possa ser recuperado pelos usuários da unidade de informação. Naves (2004, p. 5) considera:

O texto é o ponto de partida para operações analítico-documentárias, sendo constante a presença dos elementos conteúdo e forma, como partes essenciais do mesmo. Para ter uma competência textual, é preciso que, além de conhecer o texto que tem em mãos para análise sob todos os aspectos, o indexador faça dele uma leitura adequada [...].

Assim, pode-se destacar que, para a assimilação e a compreensão do conteúdo do texto, é necessário o processamento de informações na mente do indexador. De acordo com Naves (2004, p. 8), “[...] parece ser consenso entre os especialistas em leitura que o processamento do ato de ler se dá interativamente, dependendo dessa interação a compreensão ou

a não compreensão de um texto.”. No entanto, a leitura documentária difere da leitura comum, porque esta é realizada pelo leitor indexador fazendo a mediação entre as ideias do autor com as do usuário, visando à representação de conteúdo por meio de conceitos. A leitura documentária pode, então, ser considerada a atividade inicial da etapa da *análise de assunto*, que permitirá identificar o conteúdo de um documento a partir de seus aspectos intrínsecos e extrínsecos.

De acordo com os Princípios do Unisist (1981, p. 8), os três estágios para a etapa *análise de assunto* são: (1) compreensão do conteúdo do documento como um todo; (2) identificação dos conceitos que representam esse conteúdo e (3) seleção dos conceitos válidos para recuperação, sendo que se destaca que “na prática, esses três estágios se superpõem». Para Fujita (2003), esta superposição ocorre no momento da leitura do documento. Essa atividade é subjetiva, o que a torna complexa, exigindo, então, do indexador conhecimentos cognitivos, lógicos e linguísticos, e, conseqüentemente, interdisciplinares, para fazer uma análise coerente.

No primeiro estágio, *compreensão do conteúdo do documento como um todo*, o indexador deve fazer a leitura do documento, embora a leitura integral ideal seja impraticável. Destacam-se as partes entendidas como mais importantes para que o indexador possa se orientar durante a leitura; entre elas: título; introdução e as primeiras frases de capítulos e parágrafos; ilustrações; tabelas; diagrama e suas explicações; conclusão; e palavras ou grupos de palavras sublinhadas ou impressas com tipo diferente. Ressalta-se a importância de o indexador não se dedicar exclusivamente às partes preliminares do texto, uma vez que as intenções do autor se encontram no conteúdo principal, enquanto as partes finais indicam se essas intenções foram alcançadas.

No segundo estágio, *identificação dos conceitos que representam o conteúdo do texto*, o indexador deve abordar o texto de maneira lógica, selecionando os conceitos que representem adequadamente o assunto do documento. Para isso, recomenda-se que seja obedecido um esquema de categorias da área coberta pelo documento, como o indicado a seguir: o fenômeno; o processo; as propriedades; as operações; o material; o equipamento, entre outros aspectos (UNISIST, 1981).

Faz-se necessário aqui apresentar uma definição que será considerada sobre o termo *conceito*. Dahlberg (1978, p. 147) considera que

A formação de um conceito é feita através da identificação, no universo de itens (ideias, objetos, fatos, leis, etc.), de um item de referência (entidades, propriedades, atividades e dimensões). Enunciados verdadeiros acerca desse item são formulados e sintetizados sob forma verbal (termo, palavra) que será utilizada no universo do discurso. Os conceitos são, portanto, definidos como unidades do conhecimento, identificadas através de enunciados verdadeiros sobre um item de referência e representados por uma forma verbal.

Para Dahlberg (1978), quando caminhamos rumo a sintetizar, compilar e reunir enunciados verdadeiros sobre um objeto, estamos caminhando para a formação de um conceito. Assim, pode ser definido como sendo a compilação de enunciados verificáveis sobre determinado objeto, fixado por um símbolo linguístico.

Nesse segundo estágio, o processo cognitivo está intrinsecamente ligado ao processo de identificação da ideia principal apontada em um texto:

O processo de análise de assunto reveste-se de uma subjetividade característica, dadas as circunstâncias e elementos envolvidos, pois, a partir da leitura do documento pelo indexador, é realizado um processo de comunicação interativo entre três variáveis: leitor, texto e contexto. Cada uma dessas variáveis estará sujeita a diferentes condições, mas é o indexador como leitor a variável mais influente nessa interação para análise de assunto, porque precisa realizar a compreensão da leitura mediante sua cognição. (FUJITA, 2003, p. 69).

Sendo o indexador um “sujeito”, essa subjetividade é inerente a ele. O indexador, apesar de ser um profissional com ideologias próprias, é o mediador entre as ideias do autor com as necessidades dos usuários, mantendo-se imparcial.

O terceiro estágio da *análise de assunto*, chamado de *seleção dos conceitos válidos para recuperação*, corresponde à seleção daqueles conceitos que são julgados como úteis para a recuperação do documento indexado, tendo em vista os objetivos para os quais as informações são indexadas. Assim, nem todos os conceitos identificados serão necessariamente selecionados (UNISIST, 1981).

Há pesquisadores, como Lancaster (2004) e Naves (2004), que definem a atividade de seleção de conceitos como a *determinação da atinência*. Em tradução do inglês, a expressão corresponde ao termo *aboutness*, porém, encontram-se na literatura outras traduções: concernência, sobrecidade, temacidade.

O final do estágio é indicado com a definição da chamada *frase de indexação*. Essa é elaborada pelo indexador em Linguagem Natural (LN). Após todo o processo intelectual de leitura e compreensão do texto, de identificação e seleção de conceitos representativos do documento em foco, o indexador deve afirmar: “*Este documento trata de...*” A partir dessa definição, o indexador pode passar para a etapa final do processo de indexação, a tradução da *análise de assunto* em termos de indexação.

Tessier (1992) afirma que o processo de desenvolvimento de classes e relações de conceitos, que ocorre no processo de indexação, no qual o indexador pode utilizar diversas estratégias ao mesmo tempo, é um processo semelhante ao de ligação dos segmentos de textos, em um sistema de hipertexto.

Na próxima seção, passa-se a descrever sobre a perspectiva do uso da leitura documentária na estruturação hipertextual de um documento.

4 A LEITURA DOCUMENTÁRIA NA CONVERSÃO DE TEXTOS EM HIPERTEXTOS

Esclarece-se que o processo de leitura aqui referido estará direcionado ao procedimento da leitura documentária realizada pelo autor, com vistas à estruturação e conversão de hipertextos, e não a leitura que é realizada pelo usuário, no momento da navegação.

Nota-se, na literatura, que os teóricos em hipertexto têm dado maior ênfase a ele enquanto forma de leitura, interessando-se menos pelas suas características de ferramenta de auxílio ao autor e ao texto. No entanto, Landow (2006, p. 4, tradução nossa) ressalta que

Tanto como uma ferramenta para o autor, ou quanto um meio de leitura, um sistema de hipertexto permite que autores ou grupos de autores juntos vinculem informações, criem caminhos por meio de blocos de informações relacionados, anotem nos textos existentes e

criem notas que apontam os leitores para dados bibliográficos, ou para o corpo do texto referenciado... Os leitores podem navegar pelos textos anotados com referência cruzada vinculados de uma maneira ordenada, mas não sequencial.²⁸

A importância fundamental do hipertexto como texto eletrônico é ter o potencial de implementar sua característica crucial – a capacidade do leitor/usuário adicionar *links*, comentários ou ambos, realizando o processo de leitura–escrita como um processo interativo comunicacional. No presente estudo, interessa mais este segundo aspecto: os detalhes que permitem criar um texto dinâmico, estruturado semanticamente em seus aspectos cognitivos e seu papel na navegação. No entanto, a função de leitor se funde com a do autor; e a divisão entre os dois, muitas vezes, é indistinta.

De acordo com Lima (2004), a criação de um hiperdocumento bem estruturado a partir do documento em papel é um processo de conversão que ainda exige um trabalho de reelaboração do texto. A passagem do “texto” ao “hipertexto”, como o acréscimo do prefixo gramatical sugere, é a hiperbolização da forma, dos contornos, dos limites da palavra e do texto. Mas se há mudança pela interrupção, há de se ter também uma continuidade. E esse ultrapassar de fronteiras do suporte impresso para o virtual adquire sentido se forem consideradas as características da textualidade precedente, averiguando a preservação de algumas de suas características, quais limitações são superadas e quais novos domínios são potencializados.

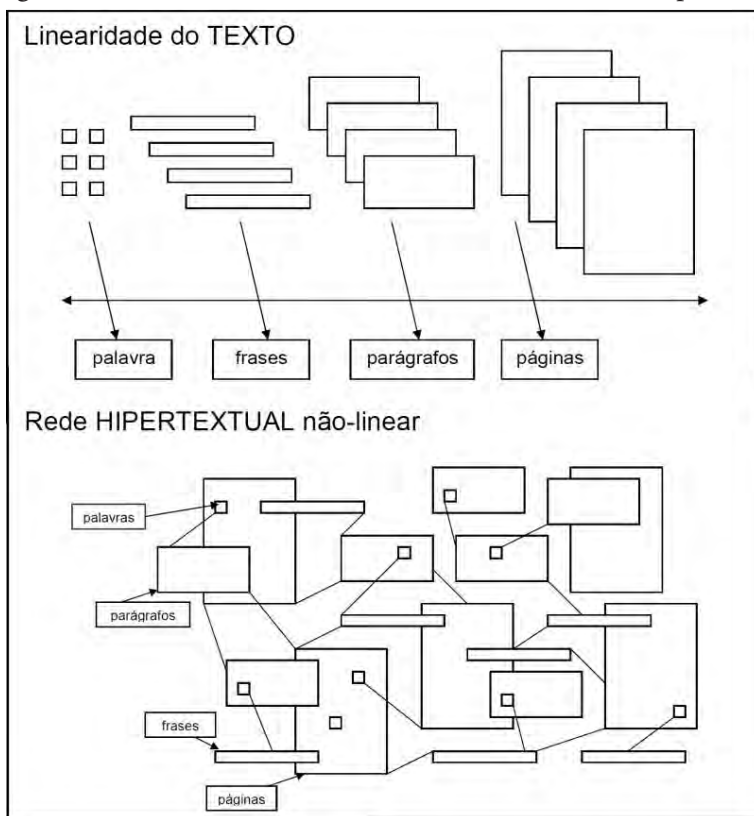
Enquanto o texto tradicional se caracteriza pela univocidade, a certeza, a linearidade, a perenidade, no hipertexto, o documento é dividido em pequenos módulos padronizados, com acesso seletivo e descontínuo, conforme ilustrado na Figura 1; porém deve-se assegurar ao usuário coerência e consistência internas, semelhantes à estrutura tradicional do documento. Assim, apesar de a informação hipertextual ser multissequencial e possibilitar uma leitura dinâmica e randômica, ainda

²⁸ *Both an author's tool and a reader's medium, hypertext document system allows authors or groups of authors to link information together, create paths through a corpus of related material, annotate existing texts, and create notes that point readers to either bibliographic data or a body of the referenced text... Readers can browse through linked, cross-referenced, annotated texts in an orderly but nonsequential manner.*

persiste, em sua estrutura, um mínimo de linearidade, de sequência para que se processe a inteligibilidade do material a ser lido.

O primeiro desafio na criação de um hipertexto é estruturar o conhecimento, de maneira que uma visão geral possa ser apresentada ao leitor no nó inicial, ou raiz. Essa visão deve identificar a ideia principal e a amplitude da cobertura, para que os leitores formem um modelo mental dos tópicos cobertos. Dependendo da aplicação e do tipo de informação em foco, um hipertexto estruturado pode ter o formato hierárquico de uma rede ou de filas (no mesmo nível). Os maiores problemas ocorrem nas fases de escolha e estruturação das informações relevantes e criação de *links* pertinentes.

Figura 1- A linearidade do texto e não linearidade do hipertexto



Fonte: Fachinetto (2005, p. 13).

Nessa conversão do texto linear para o texto não linear, faz-se necessária a realização prévia de um projeto para organização deste conteúdo textual, levando em consideração o contexto da obra analisada, a natureza do texto, perspectivas do usuário e o trabalho com a relação das ideais secundárias com a principal, para compreender a estrutura como um todo. Nesse processo, deve-se levar em consideração os *links* referenciais (ou organizacionais) e os *links* semânticos (ou de conteúdo). Os *links* referenciais são criados para a organização da superestrutura textual e o *link* semântico é utilizado para a criação da estrutura semântica hipertextual, direcionando o usuário para o conteúdo do nó. Pode-se utilizar o procedimento *bottom-up*, associando os *links* em estruturas, ou o *top-down*, no qual pegam-se as ideias gerais e as decompõem em detalhes, ou seja, em módulos.

Enquanto, no texto linear, o leitor indexador pode contar com a coerência textual e a estrutura hierárquica da informação, no caso do hipertexto, o autor tem que considerar que nem todos os parágrafos podem ter a mesma importância em todo texto a ser convertido. Na estruturação de um hipertexto, o autor deverá analisar os blocos de informação que terão seus conteúdos representados por unidades de informação/conceitos, que visam a uma representação com coerência e coesão em sua estrutura semântica, que abarque um conteúdo específico, e não dependendo da leitura prévia de outros nós para seu entendimento. Salienta-se que é por meio dos *links* que o autor constrói a coerência semântica do hipertexto, produzindo o significado dentro do contexto.

A questão da fragmentação é um dos maiores problemas que ocorre na estruturação de um hipertexto, pois não se pode comparar o parágrafo com um nó de informação. Segundo Heinonen (1998), a fragmentação é um problema de escolha dos limites dos parágrafos que melhor podem representar o conteúdo semântico fragmentado. Ela se dá na fase de escolha e estruturação das informações relevantes e criação de *links* pertinentes. Esses fragmentos, ou seções, são conectados e, quanto mais *links* existir em um hiperdocumento, mais fragmentado ele será.

Uma possível solução para a questão da fragmentação está no conceito de modularidade, que é a formação de blocos de informação e a estruturação do documento em módulos, permitindo definir os blocos de informação, os seus conteúdos, as ligações entre eles e a sua organização

coerente. O módulo é a unidade básica na construção do sistema. Modularidade, que será definida aqui como a divisão de documentos completos em pequenos módulos lógicos, é vista como importante conceito para as ferramentas biblioteconômicas que representam os conceitos de uma área, proporcionando a divisão mais específica dos esquemas dos documentos.

Cada módulo pode ser composto de um ou mais parágrafos que representam seu conteúdo, mantendo-o independente de outros módulos, sem a necessidade de relação entre eles, podendo ser representado por um único conceito. Isso permite que o autor do hipertexto use o conceito para representar o conteúdo do módulo, em vez de mostrar todas as sentenças e períodos que o compõem. Em cada módulo de informação, três aspectos são considerados essenciais: clareza, foco e completude. Esta interatividade se realiza por meio dos *links* determinados pelo autor do hipertexto, em níveis de modularidade que caracterizam essa navegação.

Dentre os procedimentos realizados na primeira etapa da indexação, *análise de assunto*, a leitura documentária tem como propósito principal a compreensão do texto para sua representação. Para fins de estruturação de um hiperdocumento, este procedimento não é exceção. Porém, o autor do hipertexto tem que ter em mente que o texto será refletido em uma tela, e os recursos de interatividade e multimodalidade devem ser levados em consideração.

De modo geral, o indexador realiza uma leitura rápida em diagonal no texto, analisando título, concentrando-se no início e no fim de capítulos, seções, resumos, índices dos livros e artigos. Analisam-se as partes mais importantes para extrair os conteúdos importantes do texto para, então, traduzir a informação que considera relevante para uma linguagem documentária. A leitura documentária tradicional com objetivo de compreender o conteúdo do documento e representá-lo não segmenta o texto em módulos para representá-los, mas sim, entende o texto como um todo com o intuito de fazer a organização estrutural do hipertexto, em blocos de informação, chamados nós, que serão inteligidos pelos *links* (frase indexatória) escolhidos posteriormente, com vistas à necessidade do usuário para recuperação do documento.

Assim, pode-se realizar o procedimento da leitura documentária para a representação do conteúdo do hipertexto seguindo as mesmas

orientações para sua realização no texto linear, mas deve-se adequá-la de acordo com as características que serão exigidas para representar um texto não linear. Deve-se levar em consideração a macroestrutura e a superestrutura textual, tendo em mente o sentido global do texto e sua estrutura. Aliado a isso, ressalta-se a mesma necessidade do conhecimento prévio do autor do hipertexto, as questões cognitivas, o contexto e o usuário.

O autor do hipertexto pode seguir os mesmos estágios previstos na *análise de assunto* realizada no texto linear. Começando pela leitura documentária, que perpassa todos os estágios, o autor (1) compreenderá o conteúdo do documento como um todo, analisando as partes entendidas como mais importantes, conforme descritas anteriormente, para que possa nortear-se durante a leitura, e entender as intenções do autor. No caso da compreensão do texto para conversão em hipertexto, o autor não precisa ir nas partes tão detalhadas, como sugerida pelo UNISIST (1981), podendo dividir o texto em quatro partes: o título, o resumo, o conteúdo principal do texto e as referências. Diferentemente da *análise de assunto* do texto linear, interessa mais para o autor a macroestrutura do texto, para determinar quais serão os blocos de informação que comporão a estrutura semântica, demarcando os limites textuais e transformando-os em pequenos módulos lógicos e coerentes, que comporão a organização estrutural deste hipertexto, completando, assim, o primeiro estágio da *análise de assunto*.

Na sequência, o autor do hipertexto passa para o segundo estágio, no qual irá (2) identificar os conceitos/*links* que representarão os conteúdos dos módulos selecionados de uma maneira lógica; porém, neste estágio, o autor deve ter em mente que ele não estará selecionando conceitos que irão representar o documento no todo. Como descrito anteriormente, cada módulo selecionado é transformado em um conteúdo independente; portanto deve-se pensar no conteúdo do módulo, porém sem perder a ideia do documento no todo; fazendo a mediação entre as ideias do autor com as necessidades dos usuários, inseridos dentro de um determinado contexto. Esses módulos irão formar a estrutura semântica do documento, na qual o usuário poderá selecionar, fazer associações, que podem ser sequenciais ou não, sendo distribuídos em níveis hierárquicos. Sobrepor diferentes camadas ou vários documentos suplementares sobre um único assunto pode simular um hipertexto estratificado, como se fossem vários textos em um só documento.

No terceiro estágio, o autor irá (3) selecionar os conceitos válidos para representação do conteúdo de cada módulo, transformando-os em *links*, que irão constituir a rede de conexões do hipertexto. Nesse estágio, o autor deve considerar, também, os aspectos conceituais envolvidos na determinação dos *links*, que constituem uma unidade de informação, que levará o usuário ao acesso do conteúdo do nó. Este procedimento de seleção dos conceitos relaciona-se com a escolha dos conceitos julgados relevantes para o acesso ao módulo de informação, tendo em vista os objetivos a serem alcançados. Ressalta-se aqui que, no caso dessa atividade no hipertexto, escolhe-se somente um conceito que representará o módulo de informação, seja por meio de um termo ou uma frase de indexação.

O Quadro 1 apresenta os aspectos das atividades da leitura documentária na conversão do texto em hipertexto, tendo por base os estágios dessa leitura sugeridos pelo Unisist (1981). Ao final desse processo intelectual, passa-se para a implementação da estrutura do hipertexto, combinando os módulos de informação, levando em consideração o processamento cognitivo e lógico do usuário.

Quadro 1 - Atividades da leitura documentária realizadas durante o processo de conversão de texto em hipertexto

ATIVIDADES DA LEITURA DOCUMENTÁRIA- UNISIST (1981).	ATIVIDADES DA LEITURA DOCUMENTÁRIA REALIZADAS DURANTE O PROCESSO DE CONVERSÃO DO TEXTO EM HIPERTEXTO.
Compreensão do conteúdo do documento como um todo.	Compreensão da macroestrutura do texto para determinar quais serão os blocos de informação que comporão a estrutura semântica, demarcando os limites textuais e tornando-os pequenos módulos lógicos e coerentes, que comporão a organização estrutural deste hipertexto.
Identificação dos conceitos que representam esse conteúdo.	Identificação dos conceitos/ <i>links</i> que representarão os conteúdos de cada módulo selecionados, sem perder a ideia do documento no todo.

Seleção dos conceitos válidos para recuperação.	Determinação dos <i>links</i> que representarão os conteúdos dos módulos na estruturação do hipertexto.
---	---

Fonte: Elaborado pela autora.

5 CONSIDERAÇÕES FINAIS

A mudança ocorrida no mundo ocidental desde os meados do século XII até a chegada da informática transformou e evoluiu as formas de expressões do conhecimento. A alteração do texto impresso para o texto eletrônico criou uma grande transformação na maneira como armazenamos e acessamos a informação. Surgem os textos eletrônicos, que podem ter as mesmas características dos textos lineares; e podem também ser textos não lineares, representando uma nova forma de escrita, chamados de hipertexto.

A forma não linear do hipertexto traz uma fragmentação textual como um problema, requerendo, assim, uma estrutura de relacionamentos configuradores de uma rede semântica que assegure a organização e recuperação deste conhecimento de acordo com o raciocínio humano. A modularidade apresenta-se como uma solução para a fragmentação, em que se subdividem os textos em unidades menores com conteúdos completos e mais focalizados, facilitando a organização do próprio autor do hipertexto e daqueles encarregados de sua manutenção. Além disso, o fato de serem esquemáticos os torna mais cognitivos, facilitando sua absorção e retenção na memória.

Na construção de um hipertexto, mantém-se a noção de autoria, porém alteram-se as relações entre o livro, o autor e o leitor. Assim, as fronteiras entre ler e escrever se tornam mais tênues. Apesar de o leitor ter a possibilidade de escolher uma trilha para navegar, de uma certa forma, ele continua lendo o texto linearmente, porque a estrutura foi proposta pelo autor, ou seja, as escolhas das trilhas a serem disponíveis são controladas pelo autor do hipertexto. Assim, o autor, o texto e o leitor têm uma interação a partir da característica intertextual do hipertexto.

O conhecimento do autor sobre o conteúdo, o usuário e o contexto é de suma importância, pois ele poderá pressupor quais as trilhas o usuário possa seguir, fazendo com que ele se sinta como se estivesse lendo

um texto tradicional. Por isso, é importante o autor manter a coerência textual, evidente no texto linear, propondo uma representação por meio de nós e *links* que considera relevante para a representação do conteúdo do texto.

Os processos de indexação podem subsidiar a conversão do texto em hipertexto para disponibilizá-lo na Web, porém exigem adaptação. Nesse contexto, o processo de indexação, especificamente a primeira etapa, a *análise de assunto*, pode ser apontado como uma solução para auxiliar nessa estruturação. A leitura documentária, realizada pelo autor do hipertexto durante estes três estágios, possui um papel interativo e mediador entre as ideias do autor, o usuário e o contexto. Ela irá auxiliar na compreensão do texto, na identificação dos módulos a serem representados e na seleção dos conceitos/*links*, contruindo uma estrutura semântica hipertextual com coerência e coesão.

O uso de método estruturado em conceitos, em vez de termos/palavras que são extraídos dos documentos, para representação dos *links* que farão a ligação entre os nós na estrutura hipertextual, possibilita criar uma rede semântica que considera o conteúdo do documento, o contexto e o usuário do hipertexto, evitando problemas semânticos, tais como a polissemia e a sinonímia. A construção de uma rede semântica coerente e o uso de vocabulário controlado podem tornar mais eficiente a recuperação na Web, assegurando que a informação relevante não se perca. Nesse contexto, na primeira etapa da indexação, os procedimentos da *análise de assunto* realizados no momento da leitura documentária se mostram de suma importância, a ser considerada na estruturação de hipertextos. Os *links*, como os conceitos, são unidades de informação, que devem representar o conteúdo do nó, dando sentido na trilha escolhida pelo usuário, formando uma rede de conteúdo que garante a compreensão global do texto. O ideal seria que a organização semântica do texto, que pode ser denominada de estrutura conceitual, refletisse as relações semânticas entre os assuntos que o compõem.

O autor do hipertexto, que, muitas vezes, cria estas estruturas intuitivamente, poderá, a partir dessas reflexões, utilizar as abordagens linguística, cognitiva e lógica, inerentes à leitura documentária para auxiliá-lo na conversão de textos em hipertextos. Espera-se que este estudo

sirva como ponto de partida para futuras reflexões da perspectiva da Biblioteconomia e Ciência da Informação.

REFERÊNCIAS

- ARAÚJO, M. A. *Compreensão de hipertexto: sob a perspectiva da teoria da mesclagem*. 2006. Dissertação (Mestrado em Estudos Linguísticos) - Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2006.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *NBR 12676*. Métodos para análise de documentos: determinação de seus assuntos e seleção de termos de indexação. Rio de Janeiro, 1992.
- BARTHES, R. *From work to text in textual strategies: perspectives in post-structuralist criticism*. Ithaca: Cornell University Press, 1979.
- BLAIR, D. C. *Language and representation in information retrieval*. Amsterdam; New York: Elsevier Science, 1990.
- BOLTER, J. D. *Writing space: the computer, hypertext, and the history of writing*. Hillsdal: Lawrence Erlbaum, 1991.
- CESARINO, M. A. N.; PINTO, M. C. M. F. Análise de assunto. *Revista de Biblioteconomia de Brasília*, Brasília, v. 8, n. 1, p. 32-43, jan./jun. 1980.
- CHAUMIER, J. Indexação: conceito, etapas, instrumentos. *Revista Brasileira de Biblioteconomia e Documentação*, São Paulo, v. 21, n. 1/2, p. 63-79, jan./jun. 1988.
- CHU, C. M.; O'BRIEN, A. Subject analysis: the critical first stage in indexing. *Journal of Information Science*, London, v. 19, n. 6, p. 439-54, 1993.
- CONKLIN, J. Hypertext: an introduction and survey. *Computer*, New York, v. 20, n. 9, p. 17-41, 1987.
- DAHLBERG, I. A referent-oriented, analytical concept theory of Interconcept. *International Classification*, Baden-Baden, v. 5, n. 3, p. 122-151, 1978.
- DERRIDA, J. Signature event context. In: MARIN, S.; SUSSMAN, H. *Glyph*. Baltimore, Maryland: Johns Hopkins University, 1977. p. 172-197.
- DIAS, E. W.; NAVES, Madalena M. L. *Análise de assunto: teoria e prática*. Brasília: Thesaurus, 2007.
- ELGELBART, D. *Augmenting human intellect: a conceptual framework*. Chapel Hill: Ibiblio, [1962]. Disponível em: <http://www.ibiblio.org/pioneers/englebart.html>. Acesso em: 07 fev. 2019.
- ENSSLIN, A. *Canonizing Hypertext: explorations and constructions*. London: Continuum, 2007.

- FACHINETTO, E. A. O hipertexto e as práticas de leitura. *Letra Magna: Revista Eletrônica de Divulgação Científica em Língua Portuguesa, Linguística e Literatura*, Brasil, ano 02, n. 3, 2. sem. 2005.
- FARROW, J.F. A cognitive process model of document indexing. *Journal of Documentation*, London, v. 47, n. 2, p. 149-166, 1991.
- FOSKET, A. C. *A abordagem temática da informação*. Tradução de Agenor de Briquet de Lemos. São Paulo: Polígono, 1973.
- FOUCAULT, M. What is an author? In: RABINOW, P. (ed.). *The Foucault Reader*. London: Penguin Books, 1979. p. 101-120.
- FREI, H. P.; STIEGER, D. The use of semantic links in hypertext information retrieval. *Information Processing & Management*, Tarrytown, v. 31, n. 1, p. 1-13, 1995.
- FROHMANN, B. Rules of indexing: a critique of mentalism in information retrieval theory. *Journal of Documentation*, London, v. 46, n. 2, p. 81-101, June 1990.
- FUGMANN, R. *Subject analysis and indexing: theoretical foundation and practical advice*. Frankfurt/Main: Indeks Verlag, 1993.
- FUJITA, M. S. L. A identificação de conceitos no processo de análise de assunto para indexação. *Revista Digital de Biblioteconomia e Ciência da Informação*, Campinas, v. 1, n. 1, p. 60-90, jul./dez. 2003.
- GIASSON, J. *A compreensão na leitura*. Lisboa: Asa, 1993.
- GOMES, L. F. *Hipertexto no cotidiano escolar*. São Paulo: Cortez, 2011.
- HEINONEN, O. Optimal multi-paragraph text segmentation by dynamic programming. In: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 36., INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS, 17., 1998, Montreal. *Proceedings* [...]. Québec: L'Université de Montréal, 1998. Disponível em: <http://citeseer.ist.psu.edu/heinonen98optimal.html>. Acesso em: 13 mar. 2019.
- HJORLAND, Birger. The concept of “subject” in information science. *Journal of Documentation*, London, v. 48, n. 2, p. 172-200, June 1992.
- HUTCHINS, W. J. The concept of “aboutness” in subject indexing. *Aslib Proceedings*, Bingley, v. 30, n. 5, p. 172-181, May 1978.
- INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. *Documentation: methods for examining documents, determining their subjects, and selecting indexing terms*. Suíça: ISO, 1985. (ISO 5963-1985 (E)).
- KOBASHI, N. Y. *A elaboração de informações documentárias: em busca de uma metodologia*. 1994. 195 f. Tese (Doutorado em Ciências da Comunicação) – Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 1994.

- KOCH, I. G. V. *Desvendando os segredos do texto*. São Paulo: Cortez, 2002.
- LANCASTER, F. W. *Indexação e resumos: teoria e prática*. Brasília: Briquet de Lemos, 2004.
- LANCASTER, F. W. *Indexação e resumos: teoria e prática*. Brasília: Briquet de Lemos, 1993.
- LANDOW, G. P. *Hypertext 3.0: critical and new media in a era of globalization*. Baltimore: The Johns Hopkins University Press, 2006.
- LANGRIDGE, D. W. *Subject analysis: principles and procedures*. London: Bowker-Saur, 1989.
- LIMA, G. Â. B. O. *Mapa hipertextual (MHTX) um modelo para a organização hipertextual de documentos*. 2004. 199 f. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2004.
- MAI, J. E. Semiotics and indexing: an analysis of the subject indexing process. *Journal of documentation*, London, v. 57, n. 5, p. 591-622, 2001.
- MARQUES, E. V. Introdução aos sistemas de hipertexto. *Revista da Escola de Biblioteconomia da UFMG*, Belo Horizonte, v. 24, n. 1, p. 85-111, jan./jun. 1995.
- McGARRY, K. *O contexto dinâmico da informação: uma análise introdutória*. Brasília: Briquet de Lemos, 1999.
- MURRAY, J. H. *Hamlet on the Holodeck: the future of narrative in cyberspace*. New York: Free Press, 1997.
- NAVES, M. M. L. *Curso de indexação: princípios e técnicas de indexação, com vistas à recuperação da informação*. Belo Horizonte: UFMG; Biblioteca Universitária, 2004. Material didático.
- NELSON, T. The hypertext. In: SMITH, J. B.; HALASZ, F.; YANKLOVICH, N. (ed.). *Proceedings International Documentathm Federation: Hypertext '87 Proceedings*. New York: Association for Computing Machinery Presses, 1965. p. vii.
- NONATO, R. S. *Teoria do conceito e hipertextos: uma proposta para determinação de relacionamentos em links conceituais*. 2009. 122 f. Dissertação (Mestrado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2009. Disponível em: http://www.bibliotecadigital.ufmg.br/dspace/bitstream/handle/1843/ECID-7V2GYA/dissertacao_nonato_2009.pdf?sequence=1. Acesso em: 16 mai. 2016.
- OTLET, P. *Traité de documentation: le livre sur le livre: théorie et pratique*. Bruxelles: Mundaneum, 1934.

- PETERSEN, T. Introduction. In: PETERSEN, T.; BARNETT, P. J. (ed.). *Guide to indexing and cataloging with the Arts and Architecture Thesaurus*. New York: Oxford University Press, 1994. p. xiii-xvi.
- RADA, R. Focus on links: a holistic view of hypertext. *International Classification*, Frankfurt, v. 1, n. 18, p. 13-18, 1991a.
- RADA, R. Trends in... A critical review. *Information Processing & Management*, Tarrytown, v. 27, n. 6, p. 659-617, 1991b.
- ROBREDO, J. *Documentação de hoje e de amanhã: uma abordagem revisitada e contemporânea da Ciência da Informação e de suas aplicações biblioteconômicas, documentárias, arquivísticas e museológicas*. 4. ed. Brasília: Editora do Autor, 2005.
- SILVA, M. R; FUJITA, M. S. L. A prática da indexação: análise da evolução de tendências teóricas e metodológicas. *Transinformação*, Campinas, v. 16, n. 2, p. 133-161, mai./ago. 2004.
- SNYDER, I. *Hypertext: the electronic labyrinth*. New York: New York University Press, 1997.
- TAYLOR, A. G. *The organization of information*. 2. ed. Westport, Connecticut: Libraries Unlimited, 2004.
- TESSIER, J. A. Hypertext linking as a model of expert indexing. *Advances in Classification Research*, Medford, p. 1991. Trabalho apresentado no 2º ASIS SIG/CR Classification Research Workshop, 1991. Disponível em: <https://journals.lib.washington.edu/index.php/acro/article/view/12555/11068>. Acesso em: 13 mar. 2019.
- UNISIST. Princípios de indexação. Tradução de Maria Cristina M. F. Pinto. *Revista da Escola de Biblioteconomia*, Belo Horizonte, v. 1, n. 10, p. 83-94, mar. 1981. Título original: Indexing principles.
- VAN SLYPE, G. Languages de indización: concepción, contrucción y utilización en los sistemas documentales. Tradução de Pedro Hípola e Félix de Moya. Madrid: Fundación Germán Sánchez Ruipérez, 1991.
- VILAN FILHO, J. Hipertexto: visão geral de uma nova tecnologia de informação. *Ciência da Informação*, Brasília, v. 23, n. 3, p. 295-308, set./dez. 1994.
- VILAN FILHO, J. *Sistemas hipertexto para microcomputadores: uma aplicação em informação científica e tecnológica*. 105p. Dissertação (Mestrado em Biblioteconomia e Documentação) - Faculdade de Estudos Sociais Aplicados da UNB, Brasília, 1992.
- WEINBERG, B. H. Improved internet access: guidance from research on indexing and classification. *Bulletin of the American Society for Information Science*, Maryland, v. 25, n. 2, p. 26-29, 1999.
- WORD WIDE WEB CONSORTIUM. *Semantic web*. 2010. Disponível em: <http://www.w3.org/standards/semanticWeb/>. Acesso em: 12 dez. 2019.

A LEITURA DOCUMENTÁRIA E A INDEXAÇÃO DE DOCUMENTOS DA ÁREA DE SAÚDE

Maria Cristiane Barbosa GALVÃO

RESUMO: Tem-se por objetivo apresentar os processos de leitura documentária e indexação na área da saúde seguindo uma abordagem conectada com os demais processos de trabalho do campo da saúde. Para tanto, inicialmente, apresenta-se o conceito de saúde, as diversas áreas que a compõem e as instituições que atuam no setor saúde. Em seguida, apresenta-se o *modus operandi* do sistema de saúde brasileiro e os padrões terminológicos empregados no país para a indexação da informação produzida no contexto da assistência em saúde. Posteriormente, o capítulo apresenta exemplos de conteúdos informacionais produzidos por instituições de saúde do Brasil e o modo de indexação desses conteúdos realizados por tais instituições. Finalmente, na conclusão, discutem-se alguns delineamentos da leitura documentária e indexação no campo da saúde que possam ser integrados aos conhecimentos mais tradicionais sobre a temática. Conclui-se que o setor da saúde é um contexto que pode trazer muitas oportunidades para os profissionais da informação que possuem conhecimento sobre os processos de organização da informação, bem como sobre saúde, o sistema de saúde, sobre o complexo industrial da saúde e seus padrões.

PALAVRAS-CHAVE: Saúde. Informação clínica. Leitura documentária. Indexação.

ABSTRACT: The objective is to present the processes of documentary analysis and indexing in the health area following an approach connected with the work processes of the health field. Thus, the chapter starts from more general concepts to more specific concepts.

Initially, it presents the concept of health, the several areas that compose it and the institutions that work in the health sector. Next, the modus operandi of the Brazilian health system and its terminological standards are presented. Subsequently, the chapter presents examples of information contents produced by health institutions in Brazil, as well as the process of indexing these contents. Finally, some aspects of documentary analysis and indexing in the field of health that can be integrated with the more traditional knowledge about the subject are discussed. It is concluded that the health sector is a context that can bring many opportunities for information professionals who have knowledge about the processes of information organization, as well as about health, the health system, industrial complex of health and health standards.

KEYWORDS: Health. Clinical information. Documentary analysis. Indexing.

1 INTRODUÇÃO

Muitas vezes, para fins acadêmicos e didáticos, os processos de leitura documentária e indexação são apresentados aos discentes de graduação e pós-graduação de forma pasteurizada, ou seja, atenuando-se as potenciais dificuldades de sua realização nos diferentes campos de atuação profissional. A idealização na apresentação desses processos pode estar associada a múltiplos fatores como: exiguidade de tempo nos currículos para cobrir uma enorme gama de conteúdos; a ausência de pesquisas mais aprofundadas sobre o funcionamento dos processos de leitura documentária e indexação nos diferentes contextos; um pressuposto da área da ciência da informação de que o mesmo modelo teórico pode ser empregado em múltiplos contextos, apesar de suas especificidades; ou, um real desconhecimento de como as diferentes áreas técnicas ou científicas funcionam. Independente dos fatores envolvidos na construção dessa abordagem mais idealizada, observa-se que os graduados na área da ciência da informação possuem alguma dificuldade em transpor os modelos teóricos apreendidos academicamente para os diversos e heterogêneos campos de atuação profissional, incluindo-se aqui a área da saúde.

Considerando-se situações como a descrita acima, este capítulo tem por objetivo apresentar os processos de leitura documentária e indexação na área da saúde seguindo uma abordagem, na medida do possível, mais conectada com os demais processos de trabalho do campo da saúde. Para tanto, o capítulo parte de conceitos mais gerais para conceitos mais específicos. Assim, na primeira parte, apresenta o conceito de saúde, as diversas áreas que a compõem e as instituições que atuam no setor saúde.

Em seguida, apresenta-se o *modus operandi* do sistema de saúde brasileiro e os padrões terminológicos empregados no país para a indexação da informação produzida no contexto da assistência em saúde. Posteriormente, o capítulo apresenta exemplos de conteúdos informacionais produzidos por instituições de saúde do Brasil e o modo de indexação desses conteúdos realizados por tais instituições. Finalmente, na conclusão, discutem-se alguns delineamentos da leitura documentária e indexação no campo da saúde que possam ser integrados aos conhecimentos mais tradicionais sobre a temática.

2 O CONCEITO DE SAÚDE E ALGUMAS DE SUAS IMPLICAÇÕES

Segundo a constituição da Organização Mundial da Saúde (OMS), publicada em 1948, saúde é um estado de completo bem-estar físico, mental e social, e não consiste apenas na ausência de doença ou de enfermidade (WORLD HEALTH ORGANIZATION, 1948). Dada essa delimitação conceitual, pode-se derivar que diversas áreas estudam saúde e fornecem assistência aos aspectos físicos, mentais e sociais dos seres humanos, estando aqui incluídas a medicina, a enfermagem, a psicologia, a nutrição, a biomedicina, a educação física, a fisioterapia, a terapia ocupacional, a odontologia, a farmácia, a assistência social, a fonoaudiologia, entre outras.

Além disso, cada uma dessas áreas possui especialidades. Assim, por exemplo, algumas especialidades da medicina são: angiologia, imunologia, anestesiologia, cancerologia, cardiologia, cirurgia, clínica geral, dermatologia, endocrinologia, gastroenterologia, genética, geriatria, hematologia, infectologia, medicina do trabalho, medicina do viajante, medicina esportiva, medicina legal, nefrologia, neurologia, obstetrícia, oftalmologia, ortopedia, otorrinolaringologia, patologia, pediatria, pneumologia, psiquiatria, radiologia, reumatologia, urologia etc. Observando a enfermagem, algumas de suas especialidades seriam: enfermagem em saúde da mulher, enfermagem em saúde do homem, enfermagem em saúde do trabalhador, enfermagem em saúde indígena; enfermagem em saúde mental, enfermagem aeroespacial, enfermagem aquaviária; enfermagem em acesso vascular e terapia infusional; assistência de enfermagem em anestesiologia; enfermagem em assistência

domiciliária; enfermagem em captação, doação e transplante de órgãos e tecidos; enfermagem em cardiologia; enfermagem em central de material e esterilização; enfermagem em centro cirúrgico; enfermagem em cuidados paliativos; enfermagem dermatológica etc.

Nesse momento, o mais importante a ser destacado é que cada uma das áreas da saúde e cada uma de suas especialidades, por sua vez, possuem múltiplos propósitos comunicacionais que são desdobrados em diferentes mensagens em linguagem verbal (oral ou escrita) e em mensagens que podem empregar linguagem visuais e mesmo sinais sonoros. Disso resulta que uma mensagem em oftalmologia difere enormemente de uma mensagem em ortopedia, ou em patologia ou em cirurgia, pois os objetos de estudo dessas especialidades diferem, bem como seus processos de trabalho, seus instrumentos e formas pelas quais coletam e compartilham dados e informações, conforme representado na figura 1.

Voltando ao conceito de saúde, a constituição da OMS também esclarece que a saúde constitui um dos direitos fundamentais de todo o ser humano, sem distinção de raça, de religião, de credo político, de condição econômica ou social (WORLD HEALTH ORGANIZATION, 1948). Neste ponto é importante esclarecer que o fato de o acesso à saúde ser um direito de todos não significa que todas as pessoas serão tratadas de igual modo dentro do sistema de saúde. Vários estudos indicam que, por exemplo, a religião, o território onde a pessoa habita ou trabalha, sua cultura etc podem interferir na condição de saúde. Assim, a assistência em saúde oferecida para católicos pode ser diferente da assistência oferecida para testemunhas de Jeová, pois nesta religião não se aceita a doação nem a transfusão de sangue. Já os adventistas do sétimo dia, por exemplo, não podem comer carne de porco e alguns espíritas se recusam a comer carne vermelha. Logo, quando internados, precisarão receber uma dieta diferenciada. Em decorrência da heterogeneidade humana, os dados e as informações coletadas, por profissionais de saúde e por unidades de saúde, podem abordar as mais diferentes searas a fim de proverem uma assistência mais adequada para cada indivíduo.

Figura 1 - Exemplos de recursos empregados para coleta de dados e informações de pacientes, respectivamente, pela oftalmologia, ortopedia, patologia e pela cirurgia.



Fonte: <https://pixabay.com/pt/>

Voltando ao conceito de saúde, a constituição da OMS também esclarece que a saúde constitui um dos direitos fundamentais de todo o ser humano, sem distinção de raça, de religião, de credo político, de condição econômica ou social (WORLD HEALTH ORGANIZATION, 1948). Neste ponto é importante esclarecer que o fato de o acesso à saúde ser um direito de todos não significa que todas as pessoas serão tratadas de igual modo dentro do sistema de saúde. Vários estudos indicam que, por exemplo, a religião, o território onde a pessoa habita ou trabalha, sua cultura etc podem interferir na condição de saúde. Assim, a assistência em saúde oferecida para católicos pode ser diferente da assistência oferecida para testemunhas de Jeová, pois nesta religião não se aceita a doação nem a transfusão de sangue. Já os adventistas do sétimo dia, por exemplo, não podem comer carne de porco e alguns espíritas se recusam a comer

carne vermelha. Logo, quando internados, precisarão receber uma dieta diferenciada. Em decorrência da heterogeneidade humana, os dados e as informações coletadas, por profissionais de saúde e por unidades de saúde, podem abordar as mais diferentes searas a fim de proverem uma assistência mais adequada para cada indivíduo.

A constituição da OMS também esclarece outro ponto importante para compreensão da área da saúde ao afirmar que os países têm responsabilidade pela saúde dos seus povos, a qual só pode ser assumida pelo estabelecimento de medidas sanitárias e sociais adequadas (WORLDHEALTH ORGANIZATION, 1948). Em outras palavras, a forma de organização da área de saúde pode variar de país para país, conforme as especificidades sociais, econômicas, religiosas e políticas. Disso decorre que alguns países, por exemplo, podem ter uma assistência em saúde onde o médico tem um papel mais importante que os demais profissionais (abordagem médico-centrada) ou ter uma assistência onde a equipe multiprofissional de saúde é formada por profissionais com papéis complementares. Além disso, alguns países podem focar mais a prevenção de doenças, como é o caso de Cuba, em contraposição a países com foco maior na atenção de doenças com maior complexidade, como é o caso dos Estados Unidos. De mesmo modo, os países podem apresentar uma organização da área da saúde com maior enfoque no financiamento público ou com mais enfoque no financiamento individual e privado. Assim, a organização da área da saúde assume diferenças significativas nos diferentes países, gerando, por sua vez ciclos informacionais diferenciados, com impactos significativos para a produção dos documentos, bem como para os processos de leitura, síntese e indexação documentária.

Pelo exposto, dadas as diferentes especificidades nacionais, imagine-se que não se pode pretender criar uma visão generalista e universal sobre a “*leitura documentária e a indexação de documentos da área de saúde*”. Mas que, talvez, seja possível elaborar abordagens com foco nas especificidades de cada país. Este capítulo, por exemplo, focará no contexto brasileiro.

Outra delimitação conceitual importante é que, do ponto de vista econômico, a área da saúde é também conhecida como um complexo industrial, onde atuam diferentes instituições, quais sejam:

- Prestadores de serviços, incluindo instituições que desenvolvem atividades de atenção primária, secundária e terciária, instituições que realizam serviços diagnósticos e serviços terapêuticos. Esses prestadores organizam a cadeia de suprimento dos produtos industriais em saúde, articulando o consumo por parte dos cidadãos no espaço público e privado;
- Indústrias de base química e biotecnológica, abrangendo as indústrias farmacêuticas, de vacinas, hemoderivados e reagentes para diagnóstico;
- Indústrias de base mecânica, eletrônica e de materiais, englobando as indústrias de equipamentos e instrumentos mecânicos e eletrônicos, órteses, próteses e materiais para uso no campo da saúde;
- Instituições de pesquisa e desenvolvimento, englobando instituições públicas e privadas, institutos de pesquisa, universidades, as demais indústrias, as instituições que atuam nas transferências de tecnologia e propriedade intelectual;
- Instituições de governança e regulação, englobando os ministérios, as agências de saúde, as secretarias de saúde e demais instituições que zelam pelo bem-estar da saúde da população em geral (GADELHA, 2003; MOREL, 2005).

A partir do conceito do complexo industrial da saúde fica mais fácil compreender que se pode encontrar, por exemplo, um médico pediatra (profissional com formação na área da medicina) que atua diretamente com o paciente, ou que atua na indústria farmacêutica no desenvolvimento ou avaliação de medicamentos para crianças e adolescentes, ou que atua como secretário de saúde em algum município, ou que atua como pesquisador em uma universidade pública. Dessa forma, além das diferentes áreas e profissionais que integram a saúde, essas áreas e profissionais produzirão informações diferenciadas de acordo com os contextos nos quais estão inseridos. Assim, por exemplo, o pediatra no contexto da assistência ao paciente, poderá estar focado na descrição de um novo caso clínico em pediatria. Na indústria farmacêutica, poderá ser responsável pela revisão do bulário dos medicamentos para crianças e adolescentes. No município, poderá ser o responsável por produzir alguma norma técnica em saúde

para higienização das mãos em creches e escolas infantis. Na universidade, será demandado a escrever um artigo científico.

Quando se apresenta o complexo industrial da saúde, pode-se também perceber que na área da saúde há múltiplos espaços para a atuação do profissional da informação e para o desenvolvimento dos processos de leitura documentária e indexação. No entanto, neste capítulo, focar-se-á na leitura documentária e na indexação de documentos nos contextos onde ocorrem a assistência direta em saúde, ou seja, nos prestadores de serviços como unidades de atenção básica, secundária e terciária, conforme será detalhado a seguir.

3 O SISTEMA DE SAÚDE NO BRASIL, SUA ESTRUTURAÇÃO E SEUS PADRÕES

A concepção do sistema de saúde brasileiro, cujo nome oficial é Sistema Único de Saúde (SUS), sua estrutura, forma de atuação e seus princípios foram estabelecidos pela Constituição Brasileira de 1988. Diferentemente de muitos países, o SUS prioriza três níveis de assistência. São elas: a atenção básica ou primária, a atenção secundária e a atenção terciária (BRASIL, 1988).

A atenção básica é um conjunto de ações, de caráter individual e coletivo, voltadas para a promoção da saúde, a prevenção de doenças, o tratamento e a reabilitação. Uma das ações bem conhecidas da atenção básica brasileira é a vacinação da população. No Brasil, a atenção básica é de responsabilidade dos municípios e as unidades básicas de saúde (UBS) possuem equipes multiprofissionais que atuam em um território geográfico definido, prestam ações de saúde individuais e coletivas, trabalham com prevenção, diagnóstico e tratamento, realizam visitas domiciliares periódicas e ações com a participação popular e ações intersetoriais. Por exemplo, se os profissionais da UBS percebem que em um bairro há muitos adolescentes e jovens usando drogas, a UBS pode atuar frente ao município solicitando ações conjuntas com escolas primárias e secundárias a fim de que mais informações sobre os riscos associados ao uso de drogas ilícitas cheguem a esse público.

A atenção secundária é formada pelos serviços especializados empregando tecnologias intermediárias entre a atenção primária e a terciária. No Brasil, a atenção secundária é de responsabilidade, sobretudo, dos estados

da federação. As unidades de atenção secundária abarcam procedimentos especializados e realizam cirurgias ambulatoriais especializadas, procedimentos traumatológico-ortopédicos, ações especializadas em odontologia, patologia clínica, citopatologia, radiodiagnóstico, ultrassonografias, fisioterapia, terapias especializadas, próteses, órteses etc. Um exemplo de unidade de saúde que oferece atenção secundária no Brasil são as unidades de pronto atendimento (UPA) que estão abertas todos os dias, inclusive nos feriados e finais de semana, quando as UBS estão fechadas.

As unidades de atenção terciária, como os hospitais, trabalham com procedimentos de média e de alta complexidade, bem como com procedimentos de alto custo, como assistência aos pacientes oncológicos, cirurgias cardiovasculares, neurocirurgias, assistência aos portadores de queimaduras, transplantes, partos de alto risco, cirurgias reprodutivas, cirurgias reparadoras, cirurgias bariátricas, além de procedimentos ambulatoriais de alta complexidade como a quimioterapia. No Brasil, uma das fontes para manutenção da atenção terciária é o governo federal.

O SUS, para funcionar em sua plenitude, possui fluxos informacionais bem definidos e, na medida do possível, informações indexadas segundo terminologias regulamentadas. Em 2011, o Brasil estabeleceu que, no âmbito do SUS, incluindo os níveis municipal, distrital, estadual e federal, passariam a ser empregadas terminologias como: a Nomenclatura Sistematizada de Medicina – Termos Clínicos (SNOMED-CT), a Classificação Estatística Internacional de Doenças e Problemas Relacionados com a Saúde (CID), a Classificação Internacional de Atenção Primária (CIAP), e a Tabela de Procedimentos, Medicamentos, Órteses, Próteses e Materiais Especiais do SUS (BRASIL, 2011). Além dessas terminologias, as diferentes áreas da saúde também empregam padrões para a indexação para os dados e informações que produzem, como a Classificação Internacional de Funcionalidade, Incapacidade e Saúde (CIF), empregada no âmbito da fisioterapia, fonoaudiologia, terapia ocupacional e reabilitação, e a Classificação Internacional da Prática de Enfermagem (CIPE), empregada no contexto da enfermagem.

A SNOMED-CT é mantida pela *International Health Terminology Standards Development Organisation* (IHTSDO) e tem por objetivo ser a mais extensa e abrangente terminologia clínica multilíngue do mundo; contribuir para a melhoria da assistência ao paciente; facilitar

a recuperação da informação registrada no prontuário eletrônico; viabilizar o acesso efetivo à informação necessária à tomada de decisão; e facilitar a comunicação no contexto da saúde. O conteúdo coberto pela SNOMED CT abarca: achado clínico; procedimento; entidade observável; estrutura corporal; organismo; substância; produto farmacêutico/biológico; espécime; conceito especial; objeto físico; força física; evento; ambientes e localizações geográficas; contexto social; situação com contexto explícito; estágio e escala; relação conceitual; qualificador de valor; artefato de registro (INTERNATIONAL ..., 2019).

A CID é mantida pela OMS e empregada por mais de 100 países, sendo a terminologia de base para a identificação global de tendências e estatísticas de saúde e o padrão internacional para relatar doenças, condições de saúde, distúrbios, lesões e outras condições relacionadas à saúde. A CID permite: o fácil armazenamento, recuperação e análise de informações de saúde para tomada de decisões; o compartilhamento e a comparação de informações de saúde entre unidades de saúde, regiões e países; e permite comparações de dados e informações no mesmo local em diferentes períodos de tempo. Seus usos incluem: o monitoramento da incidência e prevalência de doenças, observação de reembolsos e tendências de alocação de recursos, e acompanhamento das diretrizes de segurança e qualidade. Permite também a contagem de mortes, doenças, lesões, sintomas, razões para o uso das unidades de saúde, fatores que influenciam o estado de saúde e causas externas da doença (WHO, 2019).

A CIAP integra a família de classificações adotadas pela OMS e tem por objetivo sistematizar a demanda dos pacientes pela atenção primária, pois nem sempre o paciente que procura esse nível de atenção tem ainda um diagnóstico ou receberá um diagnóstico, nos moldes estabelecidos pela CID. Assim, a CIAP está preocupada, principalmente, com compreensão do contexto social do paciente e em como as unidades básicas de saúde podem melhor atendê-lo (COMITÊ INTERNACIONAL DE CLASSIFICAÇÃO DA WONCA, 2009).

A Tabela de Procedimentos, Medicamentos, Órteses, Próteses e Materiais Especiais do SUS é mantida pelo Ministério da Saúde do Brasil e congrega todos os procedimentos autorizados e oferecidos pelo SUS em território nacional, bem como os respectivos valores monetários a eles associados. Por meio do uso dessa Tabela terminológica, é possível que

uma unidade de saúde dimensione quais são os procedimentos que mais realiza, bem como, os custos envolvidos na assistência que oferece. Essa Tabela também é empregada em múltiplas situações de caráter econômico que ocorrem internamente no SUS (BRASIL, 2019).

A CIF integra a família de classificações da OMS, sendo uma classificação com múltiplas finalidades que fornece uma base científica para a compreensão e o estudo dos determinantes sociais da saúde. A CIF contempla códigos relacionados às funcionalidades biológicas, intelectuais, emocionais e sociais do ser humano (ORGANIZAÇÃO MUNDIAL DE SAÚDE, 2004).

A CIPE é uma das classificações empregadas no âmbito da enfermagem e também integra a família de classificações da OMS. Por meio de seu uso, é possível realizar de forma sistemática a análise dos cuidados de enfermagem, recursos de enfermagem e resultados obtidos na assistência de enfermagem (ORDEM DOS ENFERMEIROS, 2016).

Tendo apresentado como o sistema de saúde é estruturado no Brasil, bem como os padrões terminológicos regulamentados pelo Ministério da Saúde, pode-se imaginar erroneamente que todos os documentos produzidos na assistência em saúde do Brasil são bem analisados e indexados. Para dissolver essa ilusão, é importante ressaltar que os profissionais de saúde do Brasil nem sempre recebem uma educação formal sobre o uso dos padrões terminológicos. Também, nas unidades de saúde, diferentes profissionais, não necessariamente treinados, recebem a tarefa de executar a indexação das informações produzidas. Logo, existe uma grande defasagem entre aquilo que o Ministério da Saúde gostaria que as unidades de saúde fizessem em termos informacionais e aquilo que, realmente, as unidades de saúde realizam ou são capazes de realizar no que se refere à padronização dos dados e das informações. Dito de outra forma, para que padrões terminológicos sejam empregados, a motivação precisa ser maior que a existência de uma portaria ministerial.

Adicionalmente, alguns padrões terminológicos recomendados pelo Ministério da Saúde do Brasil se sobrepõem semanticamente. Por exemplo, a terminologia SNOMED-CT se propõe a ser a terminologia mais abrangente, ou seja, com o maior número de termos padronizados, do setor saúde em todas as suas dimensões (física, mental e social). Logo, outras terminologias não seriam necessárias, se a SNOMED-CT fosse

efetivamente empregada no território nacional. No entanto, a grande maioria dos profissionais de saúde do Brasil desconhecem essa terminologia, pois é ainda muito recente na área. Em uma visão panorâmica, são as empresas de software para o setor da saúde que mais têm empregado a SNOMED-CT no processo de estruturação semântica dos sistemas que desenvolvem a fim de criar sistemas mais interoperáveis.

De forma geral, a terminologia mais empregada no Brasil, na atenção primária, secundária e terciária, ainda é a CID, pois vários sistemas epidemiológicos de caráter nacional e internacional demandam seu uso para contabilização estatística de doenças e demais condições de saúde. Contudo, isto não significa dizer que o emprego da CID é sempre apropriado, uma vez que há profissionais de saúde que não assimilaram a importância de seu uso para a padronização dos dados e informações em saúde, em grande parte porque não tiveram uma educação formal que priorizasse a dimensão informacional da assistência em saúde, como já explicado anteriormente.

Uma segunda terminologia bastante empregada no Brasil, sobretudo pela atenção terciária, é a Tabela de Procedimentos, Medicamentos, Órteses, Próteses e Materiais Especiais do SUS, pois alguns repasses de verba do governo federal e estadual para unidades de saúde implicam em seu uso. Logo, além da padronização em si, seu uso possui uma motivação financeira.

Embora algumas das constatações apresentadas possam ser frustrantes para os brasileiros, elas evidenciam que os profissionais da informação, conhecedores dos processos de leitura documentária e indexação, possuem um vasto campo para atuação em organização da informação na área da saúde, seja nas unidades de assistência em saúde, seja em disciplinas de graduação e pós-graduação do campo da saúde, seja na gestão municipal, estadual e federal onde são tomadas as decisões sobre o uso de padrões terminológicos em saúde.

4 A LEITURA DOCUMENTÁRIA E A INDEXAÇÃO NA ASSISTÊNCIA EM SAÚDE

Para ilustrar o processo de leitura documentária e indexação na assistência em saúde serão apresentados a seguir dois casos clínicos, um do contexto da atenção primária e um do contexto da atenção secundária.

Geralmente, a atenção primária analisa todo o contexto de vida do paciente, sua interação com as demais pessoas e o ambiente em que vive. Para exemplificar a leitura documentária e a indexação nesse nível de assistência à saúde, focar-se-á no caso clínico de Dona Margarida, um personagem fictício. Este caso foi originalmente desenvolvido para alunos de um curso de especialização em atenção primária (UNIVERSIDADE FEDERAL DE SÃO PAULO, [20--]).

Caso clínico 1: Registro de visita domiciliar realizada pela médica Joana à Dona Margarida Silva de 75 anos. Dona Margarida é viúva de Seu Quinzinho. Ele morreu de derrame aos 85 anos. Ela continua apaixonada pelo marido. Como diz sua filha Jandira, “a mamãe vive no passado”. Dona Margarida teve três filhos com o finado Quinzinho e mora agora na companhia de sua filha mais nova, Jandira, de 54 anos. Dona Margarida apresenta quadro de disúria, algúria e dor suprapúbica com sete dias de evolução, sem febre. Apresenta: baixa acuidade visual, má higiene bucal, lesões nos lábios e comissura labial, incontinência urinária há 5 anos, teste Mini-mental alterado (20 pontos) e Escala de Lawton para atividades instrumentais de 14/27. Além disso, a filha refere que “deve ter uns cinco anos que a mamãe começou a caducar. No começo, ela esquecia coisas assim, que tinha feito há pouco tempo, do tipo onde pôs as chaves, se deu ou não comida para os cachorros. Com o tempo foi piorando, esquecendo o nome dos netos e filhos, das coisas. Mas é impressionante: ela lembra de coisas da sua infância e adolescência com uma nitidez (UNIVERSIDADE FEDERAL DE SÃO PAULO, [20--]).

Dado o caso clínico 1, no processo de leitura documentária, é importante identificar as informações e dados que se referem, de fato, ao paciente que está sendo atendido, pois, muitas vezes, o relato clínico pode abordar pessoas com as quais o paciente se relaciona, mas que não são o foco da assistência. Por exemplo, Seu Quinzinho, marido de Dona Margarida, morreu de derrame, mas Dona Margarida continua viva e, pelo relato, nunca teve derrame. Logo, o termo derrame não faz parte do perfil clínico de Dona Margarida. Assim, seria inadequado colocar uma padronização via CID para o relato de derrame que aparece neste caso clínico. Além disso, no processo de indexação, é importante ter a certeza que o paciente tem o problema que será indexado, bem como se todos os seus problemas foram identificados e indexados. Tanto a inclusão quanto a

exclusão de termos equivocados no processo de indexação poderão causar erros durante as futuras assistências ao paciente.

No quadro 1, foram então apresentados os termos extraídos do caso clínico 1 passíveis de indexação segundo alguns padrões terminológicos empregados na saúde.

Quadro 1 - Indexação do caso clínico de Dona Margarida segundo diferentes terminologias em saúde (continua)

TERMOS PARA INDEXAÇÃO	CIAP (2010)	SNOMED-CT (2019)	CID (2018)	CIF (2004)	TABELA DE PROCEDIMENTOS DO SUS (2019)
Visita domiciliar		50357006 Evaluation and management of patient at home (procedure)			03.01.05.014-7 Visita domiciliar realizada por profissional de nível superior.
Mulher		224526002 Woman (person)			
75 anos		105438007 Elderly woman (person) 102528008 Age more than 65 years (finding)			
Viuvez	Z15 Perda ou falecimento do parceiro	3071008 Widow (finding)		d770.2 Relacionamentos íntimos. Possui problema moderado.	
Teve três filhos e mora com sua filha mais nova		72705000 Mother (person) 224135000 Lives with daughter (finding)		d760.0 Relacionamentos familiares. Não apresenta problemas.	

Disúria, algúria e dor suprapúbica há 7 dias	U01 Disúria/ micção dolorosa U13 Sinais/ sintomas da bexiga e outros;	49650001 Dysuria (finding) 162053006 Suprapubic pain (finding)	MF50.7 Dysuria	b6200.3 Micção. Possui problema grave.	
Lesões bucais	D82 Doença dos dentes e gingivas; D83 Doença da boca/ língua/ lábios	1089981000119106 Lesion of oral mucosa (disorder)		s320.2 Estrutura da boca. Possui problema moderado	

Fonte: Elaborado pela autora.

Quadro 1 - Indexação do caso clínico de Dona Margarida segundo diferentes terminologias em saúde (continuação)

TERMOS PARA INDEXAÇÃO	CIAP (2010)	SNOMED-CT (2019)	CID (2018)	CIF (2004)	TABELA DE PROCEDIMENTOS DO SUS (2019)
Teste Mini-mental alterado (20 pontos).	P20 Alterações da memória; P70 Demência; A91 Investigação com resultado anormal;	273617000 Mini-mental state examination (assessment scale) 52448006 Dementia (disorder)	6D8Z Dementia, unknown or unspecified cause	b117.2 Funções intelectuais. Possui problema moderado. Inclui: Demência	
Baixa acuidade visual	F28 Limitação funcional/incapacidade	13164000 Reduced visual acuity (finding)	9D40 Impairment of visual acuity	b210.3 Funções da visão. Possui problema grave.	

Incontinência urinária aos esforços	U04 Incontinência urinária	165232002 Urinary incontinence (finding)	MF50.2 Urinary incontinence	b6202.2 Continência urinária. Possui problema moderado.	
Escala de Lawton para atividades instrumentais de 14/27	A91 Investigação com resultado anormal A28 Limitação funcional/ incapacidade	273534002 Instrumental activities of daily living (assessment scale) 404981008 Ability to perform instrumental activities of daily living (observable entity)			
Má higiene bucal	P29 Sinais e sintomas psicológicos. Inclui falta de higiene	110299009 Poor oral hygiene (finding)		d5201.2 Cuidar dos dentes. Possui problema moderado.	

Fonte: Elaborado pela autora.

Como se pode observar no quadro 1, cada instrumento terminológico prioriza uma dimensão do caso clínico, sendo a SNOMED-CT a terminologia que permite uma melhor representação do caso clínico. Um esclarecimento importante é que no caso clínico da atenção primária, incluiu-se como instrumento de indexação a Tabela de Procedimentos, Medicamentos, Órteses, Próteses e Materiais Especiais do SUS. No entanto, esta não tem sido empregada na atenção primária pois o repasse de verbas do governo federal para esse nível de atenção têm ocorrido de acordo com o número de habitantes do município. Logo, seu uso não traz implicações econômicas para as unidades de saúde de atenção primária.

Geralmente, quando o paciente possui uma condição de maior complexidade ele é encaminhado para ser assistido por unidades de atenção secundária. Para ilustrar essa situação, será apresentado um caso clínico 2 no qual o paciente demanda atendimento de fisioterapia. O caso

em questão foi apresentado por Medeiros e Bertoldo (2011) e foi aqui adaptado para ilustrar o processo de leitura documentária e indexação.

Caso clínico 2: Paciente Márcio Silva, sexo masculino, branco, nascido no dia 11 de fevereiro de 2008, 11 anos. Avó relata que não houve problemas durante a gestação, porém a mãe da criança não fez acompanhamento pré-natal, histórico de prematuridade, nasceu de 24 semanas, após o nascimento permaneceu 3 meses na incubadora, fez uso de ventilação mecânica, foi diagnosticado como quadriplegia. Começou fazer acompanhamento fisioterapêutico aos 6 meses. Apresentou atraso no desenvolvimento neuropsicomotor (DNPM), conquistou controle cervical aos 6 meses, sentou com 1 ano e 2 meses, engatinhou com 1 ano e 8 meses e adquiriu a marcha com 2 anos e 4 meses. Atualmente, seu diagnóstico é seqüela de paralisia cerebral do tipo diplegia. Faz uso de órteses em ambos os pés. As últimas sessões de fisioterapia foram realizadas duas vezes por semana, totalizando 10 sessões com duração de 40 minutos cada uma. O paciente continua em tratamento fisioterapêutico (MEDEIROS; BERTOLDO, 2011).

Conforme pode ser observado no quadro 2, para realizar a indexação do caso clínico da atenção secundária, excluiu-se a CIAP, pois esse padrão terminológico se aplica apenas à atenção primária. Em relação às especificidades do caso, não se considerou as condições de saúde da mãe do paciente, como é o caso da gestação e ausência de acompanhamento pré-natal.

Quadro 2 - Indexação do caso clínico de Márcio Silva segundo diferentes terminologias em saúde.

TERMOS PARA INDEXAÇÃO	SNOMED-CT (2019)	CID (2018)	CIF (2004)	TABELA DE PROCEDIMENTOS DO SUS (2019)
Homem	339947000 Man (person)			
11 anos	67822003 Child (person)			
Prematuridade	138091000119101 History of prematurity (situation)	KA21.4 Preterm newborn		

Quadriplegia	11538006 Quadriplegia (disorder)	8D20.10 Spastic quadriplegic cerebral palsy		
Atraso no desenvolvimento neuropsicomotor	65401001 Psychomotor development, function (observable entity) 248290002 Developmental delay (disorder)	MB23.N Psychomotor retardation		
Controle cervical				
Engatinhou	282606002 Ability to crawl (observable entity)		d4550.0 Gatinhar. Não possui problema	
Diplegia	6481005 Diplegia (disorder)	MB52 Diplegia of lower extremities		
Órteses em ambos os pés	266732009 Orthosis for lower limb (physical object)		d465.2 Deslocar-se utilizando algum tipo de equipamento. Possui problema moderado	
Sessões de fisioterapia	91251008 Physical therapy procedure (regime/therapy)			03.02.06.003-0 Atendimento fisioterapêutico nas desordens do desenvolvimento neuromotor.

Fonte: Elaborado pela autora.

5 CONCLUSÃO

Este capítulo partiu de uma compreensão de que a análise documentária e a indexação são processos que sofrem a influência dos contextos no quais estão inseridos. Em relação à área da saúde, destacou que o conceito de saúde abrange a dimensão física, mental e social dos indivíduos,

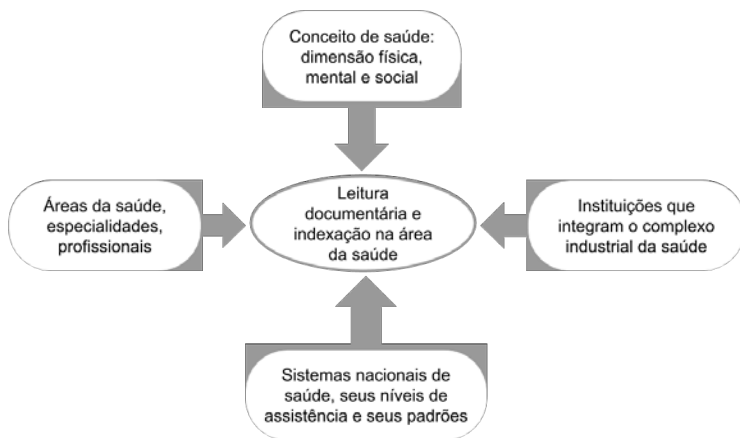
sendo várias as áreas e especialidades que estudam aspectos do objeto saúde. Tais áreas, por sua vez, possuem especificidades comunicacionais. O capítulo destacou que, partindo do conceito de saúde, os diferentes países possuem autonomia para criar sistemas de saúde nacionais que reflitam os interesses religiosos, culturais, econômicos e políticos de suas populações. Dito de outro modo, os países também podem adotar diferentes fluxos informacionais no setor da saúde, incluindo a adoção de diferentes padrões terminológicos. No que se refere ao contexto brasileiro, o Ministério da Saúde, em portaria ministerial, definiu alguns padrões a serem empregados nos sistemas de informação em saúde. No entanto, grande parte desses padrões são pouco conhecidos dos profissionais de saúde brasileiros, o que dificulta seu uso. O capítulo elencou ainda algumas instituições que integram o complexo industrial da saúde, mas que não foram abordadas em profundidade como as instituições de pesquisa, a indústria farmacêutica e as agências regulatórias.

De forma geral, os elementos elencados, ao longo do capítulo, permitem esboçar algumas variáveis que interferem no processo de leitura documentária e indexação no setor da saúde, conforme sistematizado na figura 2.

Em relação aos casos clínicos apresentados para ilustrar a leitura documentária e a indexação na assistência em saúde, pode-se observar que a SNOMED-CT é a terminologia mais abrangente para a representação dos conteúdos informacionais em saúde. Contudo, as unidades de saúde têm optado por empregar os padrões terminológicos mais tradicionais como a CID e Tabela de Procedimentos do SUS. Imagina-se que estes casos também esclareceram que unidades de assistência em saúde produzem informações bem diferenciadas das universidades e dos textos técnico-científicos, demandando a utilização de terminologias clínicas para sua indexação. Assim, não se recomenda o emprego de terminologias que foram desenvolvidas para fins acadêmicos, como é o caso dos Descritores em Ciências da Saúde (DECS), para a indexação de informações clínicas.

Como conclusão pode ser reiterado que o setor da saúde é um contexto que pode trazer muitas oportunidades para os profissionais da informação que possuem conhecimento sobre os processos de organização da informação, bem como sobre saúde, o sistema de saúde, sobre o complexo industrial e padrões demandados.

Figura 2 - Variáveis que interferem nos processos de leitura documentária e indexação na área da saúde



Fonte: Galvão (2019).

REFERÊNCIAS

- BRASIL. *Constituição da República Federativa do Brasil*. Brasília: Senado, 1988.
- BRASIL. Ministério da Saúde. Portaria nº 2073, de 31 de agosto de 2011. Regulamenta o uso de padrões de interoperabilidade e informação em saúde para sistemas de informação em saúde no âmbito do Sistema Único de Saúde, nos níveis Municipal, Distrital, Estadual e Federal, e para os sistemas privados e do setor de saúde suplementar. *Diário Oficial da União*, seção 1, Brasília, p. 63, 01 set. 2011.
- BRASIL. *Tabela de procedimentos, medicamentos, órteses, próteses e materiais especiais do Sistema Único de Saúde*. Brasília: DATASUS, 2019. Disponível em: <http://sigtap.datasus.gov.br/tabela-unificada/app/sec/inicio.jsp>. Acesso em: 02 fev. 2019.
- COMITÊ INTERNACIONAL DE CLASSIFICAÇÃO DA WONCA. *Classificação Internacional de Atenção Primária (CIAP 2)*. 2. ed. Florianópolis: Sociedade Brasileira de Medicina de Família e Comunidade, 2009. Disponível em: https://www.sbmfc.org.br/wp-content/uploads/media/file/CIAP%202/CIAP%20Brasil_atualizado.pdf. Acesso em: 2 fev. 2019.
- GADELHA, C. A. G. O complexo industrial da saúde e a necessidade de um enfoque dinâmico na economia da saúde. *Ciência e Saúde Coletiva*, Rio de Janeiro, v. 8, p. 521-535, 2003. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-81232003000200015&lng=en&nrm=iso. Acesso em: 02 fev. 2019.

INTERNATIONAL Health Terminology Standards Development Organisation. SNOMED CT. Dinamarca: IHTSDO, 2019. Disponível em: <http://www.ihtsdo.org>. Acesso em: 02 fev. 2019.

MEDEIROS, D. L.; BERTOLDO. M. G. W. Fisioterapia na paralisia cerebral: um relato de caso. *EFDeportes*, Buenos Aires, v. 16, n. 161, p. 1, 2011. Disponível em: <http://www.efdeportes.com/> Acesso em: 02 fev. 2019.

MOREL, C. *et al.* Health innovation networks to help developing countries address neglected diseases. *Science*, Washington, v. 39, n. 15, p. 401-403, 2005. Disponível em: <http://science.sciencemag.org/content/309/5733/401.full>. Acesso em: 02 fev. 2019.

ORDEM DOS ENFERMEIROS. *CIFE: Classificação Internacional para a Prática de Enfermagem*. Lisboa: OE, 2016.

ORGANIZAÇÃO MUNDIAL DE SAÚDE. *CIF: Classificação Internacional de Funcionalidade, Incapacidade e Saúde*. Lisboa, 2004. Disponível em: http://www.inr.pt/uploads/docs/cif/CIF_port_%202004.pdf. Acesso em: 02 fev. 2019.

UNIVERSIDADE FEDERAL DE SÃO PAULO. *Módulo de casos complexos: Dona Margarida*. São Paulo: UNIFESP, [20--]. Disponível em: https://www.unasus.unifesp.br/biblioteca_virtual/esf/1/casos_complexos/Caso_6.pdf. Acesso em: 02 fev. 2019.

WORLD HEALTH ORGANIZATION (WHO). *Constitution*. New York, 1948. Disponível em: http://www.who.int/governance/eb/who_constitution_sp.pdf. Acesso em: 02 fev. 2019.

WORLD HEALTH ORGANIZATION (WHO). *ICD 11 is here*. América: WHO. c2019. Disponível em: <https://www.who.int/classifications/icd/en/#>. Acesso em: 02 fev. 2019.

A TEMATICIDADE TEXTUAL NA ANÁLISE DE ASSUNTO DA INDEXAÇÃO: ASPECTOS INERENTES À ATUAÇÃO PROFISSIONAL DO INDEXADOR

Roberta Cristina Dal'Evedove TARTAROTTI

RESUMO: No domínio da Organização do Conhecimento (OC), a abordagem teórica da indexação objetiva a representação temática em sistemas de recuperação da informação. No entanto, o processo de indexação realizado pelo indexador reveste-se de uma complexidade característica notadamente em dois momentos distintos: a análise de assunto/identificação de conceitos (tematicidade intrínseca) e a seleção de conceitos (tematicidade extrínseca). Nesse sentido, o trabalho tem como objetivo trazer algumas considerações teóricas em torno da tematicidade textual na análise de assunto do processo de indexação, particularmente sobre a tematicidade extrínseca textual da etapa de seleção de conceitos. Neste viés, busca-se contribuir para o entendimento dos fatores interferentes no processo de indexação e da melhoria da atuação do bibliotecário indexador como um leitor profissional. Conclui-se que o estabelecimento de interfaces interdisciplinares da Organização do Conhecimento (OC) com outros campos científicos contribui não apenas para a resolução de problemas teórico-práticos em torno da representação e recuperação temáticas da informação como também para seu fortalecimento como campo científico.

PALAVRAS-CHAVE: Indexação. Análise de assunto. Tematicidade textual. Indexador.

ABSTRACT: In the Knowledge Organization (OC) domain, the indexing theoretical approach aims at thematic representation in information retrieval systems. However, the indexing process carried out by the indexer has a complexity characteristic, especially in two different moments: the subject analysis/concept identification (intrinsic thematicity) and the concept selection (extrinsic thematicity). Therefore, the work aims to bring some theoretical considerations around the textual theme in the subject analysis of the indexing process, particularly on the extrinsic textual theme of the concepts selection stage. In this bias, it was sought to contribute to the understanding of the interfering factors in the indexing process and the improvement of the indexer librarian's role as a professional reader. It is concluded that the establishment of interdisciplinary interfaces of the Knowledge Organization (OC) with other scientific fields contributes not only to the resolution of theoretical-practical problems around the representation and the information thematic retrieval but also to its strengthening as a scientific field.

KEYWORDS: Indexing. Subject analysis. Textual thematicity. Indexer.

1 INTRODUÇÃO

Diversos autores têm se debruçado nas questões teórico-metodológicas visando à consolidação da Organização do Conhecimento (OC) como disciplina científica e domínio de conhecimento. Por definição, o campo da Organização do Conhecimento (OC) é “[...] um sub-domínio-chave da Ciência da Informação, que é dedicado à ordem conceitual do conhecimento” (SMIRAGLIA, 2011, p. 1, tradução nossa) e “[...] um domínio inter e multidisciplinar indispensável para as atividades científicas, que também é praticada dentro da moderna Ciência da Informação” (BABIK, 2014, p. 328, tradução nossa).

Na Ciência da Informação, a leitura documental é pressuposto básico das atividades desempenhadas pelos profissionais que lidam com o tratamento temático da informação. O objetivo deste tratamento temático é possibilitar a representação por assunto de recursos informacionais e sua posterior recuperação pelos usuários nos sistemas de recuperação da informação: catálogos online, bases de dados ou mais atualmente em repositórios institucionais ou repositórios de dados de pesquisa no contexto de bibliotecas acadêmicas, dentre outros. Assim, uma das características imprescindíveis em um sistema de recuperação da informação é representar o conteúdo intelectual dos documentos para prover acesso aos mesmos baseando-se no seu **assunto**.

A análise de assunto, que se caracteriza pelo conjunto dessas subetapas, é considerada na literatura e na prática profissional como a mais importante de todo o processo de indexação, pois resultará na representação do assunto principal do recurso informacional, tornando-se a “operação base para todo o procedimento de recuperação de informações” (CESARINO; PINTO, 1980, p. 32). A importância da análise de assunto é reforçada por Fujita (2003, p. 78), para quem “[...] o assunto ou tematicidade do documento é o cerne principal e mais carente de esclarecimentos dentro dos estudos em análise documental”. Entretanto, as variáveis na indexação são muitas e pouco conhecidas, considerando-se as inúmeras possibilidades de significados e ideias contidos nos recursos informacionais em cada elemento no processo, além da presença da subjetividade, que depende particularmente do indexador (MAI, 1997a, p. 65).

A abordagem teórica da indexação, ao adaptar ou propor uma metodologia que vise uma análise de assunto adequada dos recursos informacionais deve admitir que esta ocorre na leitura documental. Neste sentido, abordagens interdisciplinares com outros campos científicos tal como a Linguística que têm o texto como objeto de estudo contribuem para um maior entendimento desta atividade (TARTAROTTI; BOCCATO, 2013, p. 43). Nesse viés, busca-se tecer algumas considerações teóricas em torno da tematicidade textual da análise de assunto da indexação, especialmente no que tange à **tematicidade extrínseca textual da etapa de seleção de conceitos**.

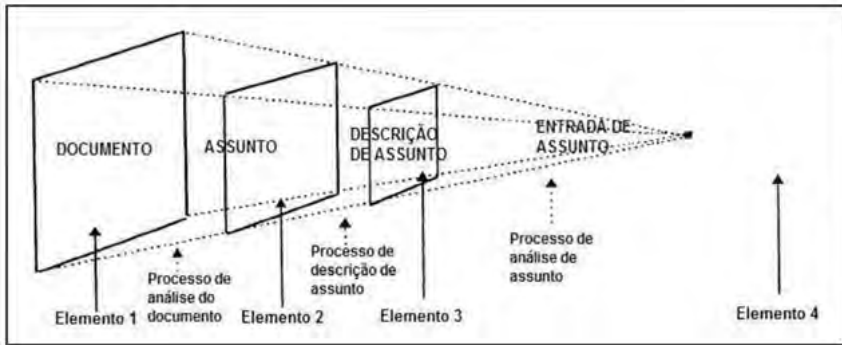
2 A ANÁLISE DE ASSUNTO PARA IDENTIFICAÇÃO E SELEÇÃO DE CONCEITOS NO PROCESSO DE INDEXAÇÃO

Conceitualmente, a indexação de assuntos caracteriza-se como um processo formado por subprocessos ou etapas que identificam o conteúdo de um recurso informacional por meio de uma metalinguagem construída – a linguagem documental –, visando a efetiva recuperação da informação (TARTAROTTI; BOCCATO, 2013, p. 40). Em termos teóricos, qualquer recurso informacional pode ser indexado/representado por meio de conceitos em busca da satisfação de necessidades informacionais específicas (GIL LEIVA, 2008, p. 64). Logo, a indexação vai além da representação temática da produção científica – foco dos produtos e serviços – podendo

ser aplicada, em um contexto mais amplo, a qualquer recurso informacional ou objeto passível de ser denominado documento.

Embora não consensuais entre os autores da literatura especializada no temário, para Mai (2000, p. 294) o processo de indexação pode ser desconstruído revelando três etapas: *processo de análise do documento*, *descrição do assunto* e *processo de análise de assunto* e quatro elementos: *documento*, *assunto*, *descrição do assunto* e *entrada do assunto*, assim representada geometricamente (Figura 1):

Figura 1. O processo de indexação de assunto



Fonte: Mai (1997a, p. 61) adaptado de Miksa (1983)²⁹.

Nesta **abordagem interpretativa** do processo de indexação, o autor considera a indexação como uma atividade intelectual que requer a interpretação do texto do recurso informacional por parte do indexador, indo além de uma atividade puramente técnica. Ao imprimirmos tais etapas e elementos no domínio dos modernos sistemas de recuperação da informação tais como bases de dados, catálogos online, repositórios institucionais ou repositórios de dados de pesquisa, na primeira etapa do processo de indexação é realizada a análise do recurso informacional digital ou, na terminologia do autor - objeto digital - visando à descrição temática, denominada de *processo de análise*. O primeiro elemento é o *objeto digital* que está sendo analisado. A segunda etapa é a formulação de uma frase de indexação ou descrição de assunto, denominada de *processo de descrição*, uma formulação mental ou escrita do assunto pelo indexador, tendo como

²⁹ MIKSA, F. *The subject in the dictionary catalog from Cutter to the present*. Chicago: American Library Association, 1983.

segundo elemento o *assunto* do documento ou do objeto digital, que pode estar presente apenas na mente do indexador. Já na terceira etapa ocorre a tradução da descrição de assunto em uma linguagem de indexação ou esquema de classificação, denominada de *processo de análise de assunto*. O terceiro elemento nesta etapa é a *descrição formal do assunto*, que pode ser escrita. Já o quarto elemento, denominado de *entrada de assunto*, é o produto da tradução da descrição formal do assunto em uma determinada linguagem de indexação do repositório institucional (MAI, 1997b, p. 55).

Cabe esclarecer que o autor propõe o uso do termo “análise do documento” para a primeira etapa do processo e “análise de assunto” para a última etapa do processo. Na literatura percebe-se a existência de mais de uma denominação para o termo *análise de assunto*, podendo ser denominada ainda de *análise temática* ou *análise conceitual*. Entretanto, entendemos que o termo **análise de assunto** é o que melhor representa o processo de análise do conteúdo temático do recurso informacional, considerando-se seu sólido estabelecimento e aceitação da terminologia na literatura científica do campo da Organização do Conhecimento (OC).

Conforme Fujita, Rubi e Boccato (2009, p. 24), as etapas do processo de indexação não precisam ser necessariamente realizadas de maneira sequencial, visto que o indexador profissional já familiarizado com o processo pode realizá-las simultaneamente e até de maneira automática, situação corroborada na prática em estudo realizado com catalogadores e indexadores por Tartarotti (2014, p. 198), para quem quanto mais familiarizado com a área científica especializada no qual atua, mais estas etapas são superpostas durante o processo de indexação pelo profissional. Além disso, de acordo com Mai (2000, p. 280), a ação tomada por indexadores com diferentes níveis de experiência pode não ser exatamente a mesma, mas as etapas e elementos do processo de indexação de assunto são considerados fundamentais para qualquer processo de indexação.

No processo de indexação, a **análise de assunto** é considerada a etapa mais complexa e mais importante da indexação, pois determina os resultados das estratégias de busca realizadas pelos usuários do sistema de recuperação da informação, no momento da recuperação da informação. Durante a análise de assunto, o indexador examina o documento, procurando compreender o texto por meio da **identificação** do assunto, com o objetivo posterior de **selecionar** os conceitos indexáveis. Chu

e O'brien (1993) consideram a análise de assunto como fase inicial do processo de indexação que decidirá sobre os principais tópicos do assunto de um documento, precedendo a fase de tradução desses tópicos de acordo com a linguagem documental adotada pelo sistema.

Segundo Taylor e Joudary (2009, p. 305), a “[...] análise de assunto é a parte do processo de criação de metadados que identifica e articula o assunto do recurso informacional que está sendo descrito”. O processo de análise de assunto abarca três subetapas; 1) análise conceitual para determinar de que trata o item; 2) descrição da tematicidade em uma declaração por escrito; e 3) utilização dessa declaração de tematicidade para atribuir termos de vocabulário controlado e/ou notações de classificação. Já para Raju e Raju (2006, p. 14), teóricos da abordagem teórica da catalogação de assunto, a análise de assunto consiste no “processo de analisar o assunto contido em um item antes da tradução desta análise conceitual em uma linguagem”. Em síntese, a análise de assunto implica em determinar a tematicidade do documento mediante a identificação e seleção dos conceitos que comporão o assunto ou temas principal e secundários.

No contexto da análise de assunto realizada pelo indexador, a leitura documental possui um fundamental, pois “[...] para o indexador, seu cotidiano é concentrado no ato da leitura, de forma a viabilizar o acesso à informação contida nos documentos aos usuários dos sistemas de informação” (NEVES; DIAS; PINHEIRO, 2006, p. 141). Conforme Fujita (2003, p. 94), é a leitura do documento que dá início à atividade de indexação, diferindo de uma leitura normal ao exigir procedimentos diferentes, “[...] ainda que os conhecimentos necessários para um bom entendimento de um texto sejam comuns a ambas”. Aqui, “[...] esse leitor-indexador tem objetivo definido: identificação e seleção de conceitos que representem o conteúdo do texto e que coincidam com as necessidades informacionais da comunidade usuária do sistema de informação” (LARA, 1993, p. 50).

Todavia, é notável a dificuldade apresentada pelo indexador ao ler o documento com o objetivo de identificar e selecionar os conceitos representativos do assunto do documento, considerando-se que, na maioria das vezes, não é um especialista sobre o assunto que indexa. Por outro lado, o indexador bibliotecário pode ser tornar um especialista na

área em que atua devido à experiência adquirida ao longo do tempo na prática na atividade de indexação, além de cursos especializados (FUJITA, 2003, p. 84).

Considerando-se que a construção de qualquer texto de um recurso informacional é baseada em um ou mais temas, é por meio da **leitura profissional** na representação temática da informação que ocorre a busca pela **tematicidade** documental, revelada por meio da **análise de assunto** realizada pelo indexador - leitor profissional que interage com o texto para cumprir o objetivo da indexação. Sendo realizada durante a leitura documental, a análise de assunto é subdividida em outras três etapas: **compreensão** do conteúdo do recurso informacional; **identificação dos conceitos** que representam este conteúdo; e **seleção dos conceitos** válidos para recuperação. Se a análise de assunto, como primeira etapa da indexação, é realizada através da leitura - uma leitura documental -, é no momento da leitura em que os conceitos tratados em um documento são **identificados** e **selecionados**, para posterior representação em termos de indexação. Nesse sentido,

[...] a **representação por conceitos** na **análise de assunto** para identificação e seleção de conceitos é diferente da **representação realizada na tradução** porque, em primeiro lugar, é realizada durante a análise de conteúdo do documento, ou seja, com o documento na íntegra e em segundo lugar, porque utiliza conceitos com os quais são representados os termos e assim identificada a tematicidade intrínseca do documento. Na tradução, de outro modo, a **representação** se dá com os **termos extraídos do documento analisado**, portanto, **fora do contexto documental** em que o autor desenvolveu o conteúdo significativo e utiliza uma linguagem *documental* que representa o vocabulário de áreas de especialidade com a finalidade de compatibilizar os termos identificados com os termos buscados pelo usuário (FUJITA, 2013, p. 50, grifo nosso).

Conforme apresenta Dal'Evedove (2002, p. 58), em uma abordagem ideal da prática da indexação, as etapas de leitura documental, análise de assunto/identificação de conceitos e seleção de conceitos precisam ocorrer durante a leitura, enquanto que a etapa de tradução dos conceitos que representam os conceitos em descritores da linguagem do

sistema deve ser feita após a leitura do documento. Esta visão permite que a análise seja realmente conceitual e compreensiva, pois a preservação do conteúdo do documento é uma garantia de relevância de recuperação, objetivo da eficiente indexação de conteúdo.

Referente às subetapas da análise de assunto, a **compreensão de leitura** é considerada uma condição necessária à leitura, ou seja, “[...] não existe leitura sem compreensão. Então, quando falamos em leitura para indexação, podemos dizer que o indexador necessita compreender o texto para identificar e selecionar conceitos, pois somente o fará a contento se houver compreensão” (FUJITA, 2003, p. 82). Sobre isso Farrow (1991, p. 151) adverte que o indexador compreende o texto basicamente do mesmo modo que um leitor fluente, porém, sob a influência de condições de tempo, objetivo definido, modelo a ser produzido e áreas temáticas definidas com estrutura textual padronizada dos recursos informacionais, podendo induzi-lo a um processo repetitivo e automático.

Na **identificação de conceitos**, o indexador, “após o exame do texto, passa a abordá-lo de uma forma mais lógica a fim de selecionar os conceitos que melhor representem seu conteúdo”. Esta etapa depende “[...] da tematicidade do texto e está atrelada à leitura do indexador e às suas concepções de análise de assunto adquiridas pela sua formação, objetivos e políticas de indexação” (FUJITA, 2003, p. 64-85). Deste modo, a identificação dos conceitos tratados num documento envolve como aspecto principal a compreensão do conteúdo do documento pelo leitor, de forma a representar e propiciar a seleção dos conceitos válidos para recuperação.

Já a etapa de **seleção de conceitos**, que ocorre após a identificação dos conceitos, está diretamente relacionada aos “objetivos para os quais as informações são indexadas”, sendo que “nem todos os conceitos identificados serão necessariamente selecionados” (FUJITA, 2003, p. 64).

3 A TEMATICIDADE TEXTUAL NA ANÁLISE DE ASSUNTO: DA TEMATICIDADE INTRÍNSECA À TEMATICIDADE EXTRÍNSECA

Como visto, a análise de assunto é subdividida nas etapas de **compreensão** do conteúdo do recurso informacional; **identificação dos conceitos** que representam este conteúdo; e **seleção dos conceitos**. Todo

texto possui uma **tematicidade intrínseca** – que podemos denominar de **tematicidade básica**, ou seja, um tema nuclear inerente ao texto e ligado à **produção** do mesmo, identificada na etapa de análise de assunto ou identificação de conceitos. De acordo com Todd (1992, p. 102), o nível de relação entre tematicidade e significado varia, pois depende “do uso que a pessoa pode encontrar da tematicidade do documento numa certa época, e o mesmo documento pode vir a ter diferentes significados para o mesmo leitor em diferentes épocas, entretanto o documento possui uma atinência fundamental”. Esta etapa se refere à **identificação de conceitos**.

Por outro lado, a **tematicidade extrínseca** – que podemos denominar de **tematicidade profunda** – está diretamente relacionada com o **contexto** de produção e de recepção do texto, com a necessidade informacional dos usuários do sistema de recuperação da informação e com a política de indexação da instituição, ligada à **recepção** do texto. Estes fatores, quando conhecidos pelos profissionais que executam a atividade de indexação, são decisivos para a adequação temática ao contexto e respectiva recuperação temática. Esta etapa se refere à seleção de conceitos.

Na tematicidade do documento, destaca-se o conteúdo relativamente **permanente do documento** (*aboutness*), o que nos leva à **tematicidade intrínseca textual**, enquanto que o **significado compreendido pelo usuário** (*meanings*) refere-se à **tematicidade extrínseca textual**. De acordo com Fujita (2003, p. 80), Begthol (1986)³⁰ fez distinção entre *aboutness* e *meanings*: “o *aboutness* é o conteúdo intrínseco do documento, que independe do uso temporal que um indivíduo possa fazer do mesmo em análise e que o faz possuir uma tematicidade relativamente permanente”; já o *meanings* (significados) pode ser medido de acordo com o uso particular do documento tendo em vista os usuários (FUJITA, 2003, p. 80).

Ainda para Fujita (2003, p. 79), não existe um consenso entre os autores em torno do termo *aboutness* em português, cunhado em 1969 por Fairthorne (1969). Enquanto para alguns autores significa “do que trata um texto”, outros utilizam “atinência” ou ainda “**tematicidade**”. Wellish (2000, p. 5) define *aboutness* como “[...] o conjunto de assuntos ou tópicos tratados em um documento, inclusive a intenção do autor e o

³⁰ BEGHTOL, C. Bibliographic classification theory and text linguistics: aboutness analysis, intertextuality and the cognitive act of classifying documents. *Journal of Documentation*, London, v. 42, n. 2, p. 84-113, 1986.

possível uso pelos leitores” Para Albrechtsen (1993, p. 220), tal conceito passou a ser pesquisado em substituição ao conceito de *subject* ou “tema”. Nesse sentido, conceitualmente podemos considerar a tematicidade como sinônimo de tema. De acordo com Fujita (2001, p. 61), o tema possui uma estrutura temática composta por conceitos ou categorias ou facetas. A identificação das categorias ou conceitos na estrutura textual do recurso informacional ocorre por meio da análise conceitual e o conjunto das categorias identificadas formulará o tema do documento. Segundo Guimarães (2009, p. 35), o tema “[...] é uma representação abstrata da estrutura global do significado de um texto”, compreendido ainda como um

[...] núcleo informativo fundamental ou elemento em torno do qual se estrutura a mensagem, sua identificação permite ao receptor considerar “entender” o texto, daí poderá partir para a elaboração de resumo do texto, ou para exercício de retextualização sob diversas formas (paráfrases, comentários, resenhas, resenhas, resenhas, etc.). O processo seguido pelo receptor é, de certo modo, oposto ao utilizado pelo emissor. Se este procede desenvolvendo o tema, o receptor deve, reduzindo as informações que lhe são transmitidas, limitar-se ao fundamental, até chegar a esse núcleo informativo (GUIMARÃES, 2009, p. 35).

De forma geral, enquanto o *aboutness* refere-se ao “conteúdo relativamente permanente do documento”, o *meanings* é entendido como “o significado compreendido pelo usuário” do sistema de recuperação da informação (FUJITA, 2003, p. 80). A literatura científica sinalizava que o interesse do indexador deve limitar-se ao *aboutness*, ou seja, à tematicidade textual, tendo em vista que o indexador lida com os conteúdos permanentes dos recursos informacionais (com as ideias propostas pelo autor), representadas através da linguagem textual.

Entretanto, os estudos mais contemporâneos defendem o estabelecimento de um novo paradigma no processo de análise de assunto, pois o *meanings*, ou seja, o contexto, também deve ser considerado pelo indexador durante a atividade de indexação, considerando-se que embora a tematicidade seja o conteúdo relevante do documento no contexto próprio do autor, algumas variáveis irão influenciar na determinação

desse conteúdo. Logo, a determinação do(s) assunto(s) do recurso informacional também está ligada a estes fatores. Segundo Cavalcanti (1989) a tematicidade intrínseca refere-se ao tema importante para o autor, enquanto que a tematicidade extrínseca refere-se ao tema importante do ponto de vista do leitor. Isso é denominado pela autora como “saliência autor – relevância leitor”.

Em síntese, no processo de representação da informação existem dois tipos de tematicidade: a **tematicidade intrínseca** (resultado da fase de **identificação de conceitos** na análise de assunto) e a **tematicidade extrínseca** (resultado da fase de **seleção de conceitos** na análise de assunto), conforme apresentado no Quadro 1:

Quadro 1 - A tematicidade textual na análise de assunto do processo de indexação

ETAPA DA INDEXAÇÃO	SUBETAPA DA INDEXAÇÃO	TIPO DE TEMATICIDADE
Leitura documental	-	-
Análise de assunto	Compreensão de leitura	-
	Identificação de conceitos	<i>Tematicidade intrínseca</i>
	Seleção de conceitos	<i>Tematicidade extrínseca</i>
Tradução dos conceitos	-	-

Fonte: Elaborado pela autora.

Considerando-se que a tematicidade intrínseca e a tematicidade extrínseca são aspectos norteadores do processo de indexação, entendemos que a prática ideal desta atividade realizada pelo indexador deve abarcar tanto o *aboutness* (contexto de produção) quanto o *meaning* (contexto de recepção) dos recursos informacionais.

4 A TEMATICIDADE EXTRÍNSECA DA ETAPA DE SELEÇÃO DE CONCEITOS NA ANÁLISE DE ASSUNTO

Na prática, a indexação é analisada sob três diferentes concepções teóricas ou perspectivas de análise que refletem na atuação profissional do indexador: indexação orientada para o documento (ênfase no documento), indexação orientada para o usuário (ênfase nos usuários) e indexação orientada para o domínio (engloba o **contexto**, o documento e os usuários) (GIL LEIVA, 2008, p. 62). De acordo com Fujita (2003, p. 72), a concepção de leitura orientada para o conteúdo deve orientar a identificação de conceitos e a concepção orientada para a demanda, a seleção de conceitos, sendo que a concepção do indexador está “diretamente vinculada com sua formação educacional (concepção orientada pelo conteúdo) e com a postura do sistema de informação (concepção orientada pela demanda) e não pelo fato de ele ser um leitor menos ou mais habilitado”. Por abarcar uma visão mais ampla, a abordagem da indexação centrada no domínio representa o ideal em termos de indexação, pois considera outros elementos além do documento ou o usuário.

Sob a ótica de Smiraglia (2009, p. 674), a interpretação de um recurso informacional envolve o contexto em que o recurso informacional é interpretado, visto que anteriormente ao processo de indexação houve a criação do recurso informacional por uma determinada **comunidade discursiva**, resultado de diversos **atos de interpretação** por parte de seus atores sociais. Uma vez concluído, publicado e analisado por meio do processo de indexação possibilitando que o recurso informacional seja acessível, o objetivo é que este seja recuperado e utilizado pelos usuários da informação, estejam ou não inseridos dentro da própria comunidade discursiva que o elaborou (MAI, 2000, p. 272; MAI, 2001, p. 604).

No processo de indexação, a seleção de conceitos corresponde à etapa da análise de assunto que resultará na extração da **tematicidade extrínseca textual** do recurso informacional em análise pelo profissional indexador. Logo, a etapa de seleção de conceitos é influenciada pela variável **contexto**, que refere-se à análise de assunto em contextos de informação especializados e todos os elementos influentes no momento da leitura profissional, isto é, “[...] todas as condições nas quais se encontra o leitor quando entra em contato com um texto.” (GIASSON, 1993, p. 40). Além disso, “[...] representa não apenas o **conhecimento prévio profissional** e

os **objetivos da atividade em questão**, mas o **ambiente** em que é realizada esta atividade profissional em leitura documental” (DAL’EVEDOVE, TARTAROTTI; FUJITA, 2015, p. 610).

Nessa perspectiva, para que o significado do conteúdo informacional de um documento seja o menos ambíguo possível, o mesmo deve ser analisado de acordo com o **contexto** no qual está inserido, uma vez que tal variável tende a agregar valor em determinados conceitos, os quais passam a assumir novos significados apenas mudando-se o ambiente de análise (PINTO MOLINA, 1993). Desse modo, “o conceito somente assume sentido quando levado em consideração o contexto de situação, em que o bibliotecário indexador deve ajustá-lo as suas decisões conforme a necessidade exigida pelo meio” (DAL’EVEDOVE; TARTAROTTI; FUJITA, 2015, p. 610).

Entretanto, como observa Olson (2007, p. 536), a maioria dos relacionamentos na **abordagem temática da informação** como os tesouros (produtos da concepção teórica da indexação), listas de cabeçalhos de assunto (produtos da concepção teórica da catalogação) ou esquemas de classificação (produtos da concepção teórica da classificação) está limitada às relações paradigmáticas, ou seja, são **relações intrínsecas**; não dependendo do contexto. Nesta **abordagem interpretativa** do processo de indexação, os principais problemas da representação de documentos estão relacionados ao significado e à linguagem, conseqüentemente que o **indexador** é o fator mais interferente do processo de indexação, pois os referentes de cada elemento no processo são entendimentos **subjetivos** e, portanto, que os resultados do processo são bastante incertos, e dependendo em grande parte de quem realiza a análise (MAI, 2000, p. 270; MAI, 1997a, p. 61).

Na opinião de Neves (2006, p. 43), o momento mais crucial do processo de indexação é a análise de assunto, pois reveste-se de uma subjetividade inerente à política de indexação, que determina prioridades a alguns assuntos. Para minimizar esta subjetividade na indexação, a formação e capacitação profissional desempenham um papel-chave. Entretanto, conforme alertam Neves, Dias e Pinheiro (2006, p. 151), “[...] apesar desse treinamento, em sua busca por termos para representar o assunto tratado no documento, o indexador baseia-se no discernimento próprio e na prática adquirida no exercício da profissão, o que, além

de tornar o processo de indexação **subjetivo**, questiona a eficácia da formação profissional”.

No que concerne aos aspectos cognitivos do indexador enquanto leitor, ao considerarmos que todo fato cultural, atividade ou prática social constituem-se como práticas significantes, ou seja, práticas de produção de linguagem e de sentido (SANTAELLA, 2007, p. 10), isto nos leva ao campo da Linguística (a ciência da linguagem verbal), mais especificamente à abordagem teórica da **Linguística Textual**, considerando-se a relevância do objeto texto como *materializador de sentidos* no processo de indexação no âmbito da Organização do Conhecimento (OC).

De acordo com Koch (2009, p. 154), a multiplicidade cultural, histórica e funcional que caracteriza a moderna evolução do conhecimento seria impossível sem a existência de textos. Para Discini (2007, p. 14), “[...] o texto é uma unidade de sentido, dada por recorrência daquilo que é dito e de um modo próprio dizer”.

Cavalcanti (1989, p. 48) considera a interação leitor-texto como um processo caracterizado por quatro principais aspectos: **ativo** (que envolve não compreensão, mas interpretação e exige competência comunicativa); **construtivo** (baseia-se em interpretação das expressões indexicais sinalizadas no texto); **interativo** (influenciado por variáveis de desempenho); e **reconstrutivo** (permite a introdução de vieses tanto das variáveis de desempenho como dos sistemas de valores do indexador). Neste contexto, a interação entre o indexador e o texto “[...] consiste numa amplitude de procedimentos coordenados e de operações perceptíveis, linguísticas e cognitivas, em que as características da estrutura textual fornecem subsídios ao conhecimento do bibliotecário indexador, aumentando a compreensão do assunto abordado” (DAL’EVEDOVE; TARTAROTTI; FUJITA, 2015, p. 609).

Nos estudos mais recentes sobre o texto e o discurso da Linguística Textual, recai uma ênfase na concepção de cognição humana de forma integrada, “[...] onde a linguagem, percepção, afeto, atenção, memória, estrutura cultural e outros componentes do sistema cognitivo encontram-se definitivamente inter-relacionados”, assim como uma retomada do interesse pelas tipologias e gêneros textuais (BENTES, 2011, p. 282). No âmbito da Ciência da Informação, a cognição é percebida em todas as atividades realizadas pelos profissionais da informação, tanto no

tratamento descritivo ou temático quanto na recuperação da informação. Nesse sentido, “[...] a cognição implica que cada ato de processamento da informação, seja ele perceptivo ou simbólico, é mediado por um sistema de categorias e conceitos os quais, para o mecanismo de processamento da informação, constituem um modelo de mundo” (NEVES, 2006, p. 42).

Entre as diversas concepções de texto que fundamentaram os estudos em Linguística Textual, destacam-se: **concepção gramatical**: texto como frase complexa ou signo linguístico mais alto na hierarquia do sistema linguístico; **concepção semiótica**: texto como signo complexo; **concepção semântica**: texto como expansão tematicamente centrada de macroestruturas; **concepção pragmática**: texto como ato de fala complexo; **concepção discursiva**: texto como discurso ‘congelado’, como produto acabado de uma ação discursiva; **concepção comunicativa**, texto como meio específico de realização da comunicação verbal; **concepção cognitivista**: texto como processo que mobiliza operações e processos cognitivos e, por fim, **concepção sociocognitiva-interacional**: texto como lugar de interação entre atores sociais e de construção interacional de sentidos (KOCH, 2009, p. 12). Na concepção sociocognitiva-interacional, “a compreensão deixa de ser vista como simples “captação” de uma representação mental ou como a decodificação de mensagem resultante de uma codificação de um emissor”, mas entendida como uma “atividade interativa altamente complexa de produção de sentidos”, realizada “com base nos elementos linguísticos presentes na superfície textual e na sua forma de organização, mas que requer a mobilização de um vasto conjunto de saberes (enciclopédia) e sua reconstrução no interior do evento comunicativo” (KOCH, 2009, p. 18).

Ao considerarmos o processo de indexação com base em Koch (2009, p. 19), o sentido de um texto de um determinado recurso informacional é construído na interação recurso informacional-indexador (texto-sujeito) com três elementos: o **autor (produtor)**, que procura viabilizar o seu “projeto de dizer”, recorrendo a diversas estratégias de organização textual e orientando o indexador profissional por meio de marcações textuais para a construção dos (possíveis) sentidos; o **recurso informacional (texto)**, organizado estrategicamente de dada forma, em decorrência das escolhas feitas pelo autor entre as inúmeras possibilidades de materialização oferecida pela língua e estabelecendo limites quanto

às leituras possíveis; e o **indexador (leitor)**, que, a partir do modo como o texto se encontra linguisticamente construído, das sinalizações que lhes oferece, bem como pela mobilização do contexto relevante à interpretação, vai proceder à construção dos sentidos a fim de cumprir o objetivo da indexação.

De acordo com Koch (2009, p. 14-15), a concepção de sujeito – ou mais especificamente, a de *sujeito da linguagem* (neste caso, de autor e indexador) varia de acordo com a concepção de língua adotada. Em uma concepção de **língua como estrutura**, caracteriza-se um autor/indexador “não consciente”, inerte, “assujeitado” pelo sistema, onde o comportamento individual é apenas um reflexo do comportamento coletivo, seja este linguístico ou social. O autor/indexador é inserido numa ideologia, numa instituição da qual é apenas porta-voz: é um discurso anterior que fala através dele. Ainda para a autora, se pensarmos na **língua como representação do pensamento**, percebe-se um autor/indexador individual, dono de sua vontade e de suas ações, social na medida em que se constrói em sociedade e com isto adquire a habilidade de interagir, que constrói uma representação mental e deseja que esta seja “captada” pelo interlocutor da maneira como foi mentalizada. Já na concepção de **língua como lugar de interação**, percebe-se o autor/indexador como ativo na produção do social e da interação. O autor/indexador reproduz o social na medida em que participa ativamente da definição da situação na qual se acha engajado, sendo um ator na atualização das imagens e das representações, sem as quais a comunicação não poderia existir. Desse modo, a concepção de língua como lugar de interação e do autor/indexador como um elemento ativo neste processo caracteriza o **contexto sociocognitivo**.

Conforme a autora, atualmente na Linguística Textual o **contexto** abrange não apenas o co-texto, como a situação de interação imediata, a situação mediata (entorno sociopolítico-cultural) e também o contexto sociocognitivo dos interlocutores (que integra os demais), mas também todos os tipos de conhecimentos arquivados na memória dos atores sociais (neste caso, indexadores e atores durante o processo de indexação). Estes conhecimentos são classificados como: o **conhecimento linguístico** propriamente dito; o **conhecimento enciclopédico** (declarativo ou episódico, frames e *scripts*); o **conhecimento da situação comunicativa** e

de suas ‘regras’ (situacionalidade); o **conhecimento superestrutural** (tipos textuais); o **conhecimento estilístico** (registros, variedades de língua e sua adequação às situações comunicativas); o **conhecimento sobre os variados gêneros** adequados as diversas práticas sociais; o **conhecimento de outros textos** que permeiam nossa cultura (intertextualidade) (KOCH, 2009, p. 24).

Esta interação ocorre com o uso de estratégias cognitivas (inferências, focalização, busca de relevância, etc.), estratégias sociointeracionais (preservação das faces, polidez, atenuação, atribuição de causas a (possíveis) mal-entendidos etc.) e estratégias textuais (formas de organização do texto com vistas à produção de determinados sentidos) (VILELA; KOCH, 2001, p. 24, p. 464). Assim, todos estes conhecimentos são importantes para o indexador realizar uma análise conceitual e compreensiva no **processo de indexação**. Percebe-se que a literatura da área tem se voltado principalmente para as estratégias cognitivas e metacognitivas do indexador. Do ponto de vista da Linguística Textual, o processo de indexação é um **processamento textual**, e, portanto, torna-se relevante mais estudos sobre estas outras estratégias de interação descritas por este campo.

Se um texto escrito não é recebido ativamente pelo indexador, pressupondo energia de processamento cognitivo por parte do indexador. Desse modo, a cognição será mais eficaz se completada pela **metacognição**, isto é, “pelos princípios que regem a desautomatização consciente das estratégias cognitivas ou o pensamento sobre nosso próprio pensamento”. Assim, “[...] um texto não tem sentido por si mesmo, mas graças à interação que se estabelece entre o conhecimento apresentado no texto e o conhecimento de mundo armazenado na memória do interlocutor” (GUIMARÃES, 2009, p. 17). De acordo com Vilela e Koch (2001, p. 559), enquanto os **modelos cognitivos** (frames, *scripts*, cenários, modelos episódicos, esquemas, modelos mentais) controlam o conteúdo semântico do texto, isto é, o que está sendo ou será dito, os **modelos de contexto** controlam a maneira como os interlocutores vão formular tais conteúdos em função do contexto em que a interação se realiza.

Somada às características inerentes à etapa de identificação de conceitos da análise de assunto que trazem uma complexidade própria à atividade de indexação por envolver aspectos cognitivos, a etapa de seleção

de conceitos é influenciada notadamente pelos elementos de **política de indexação** e pelo perfil dos **usuários**, e em uma abordagem mais ampla, pela política de informação presentes no contexto institucional em que o indexador atua. Embora o assunto seja a informação relevante abordada no recurso informacional, a seleção do assunto sofre a influência da política de indexação do sistema de informação à qual pertence, pois a instituição decidirá, dentre outros elementos, se o tema extraído do documento terá um nível mais específico ou mais genérico com base no perfil da comunidade usuária. Nesse cenário,

A atuação profissional no tratamento temático da informação deve ser regida por uma política própria e única, que considere a instituição como um todo, seus usuários, colaboradores, estrutura física e financeira, entre outros. Como um processo que vai muito além de um fazer técnico – mas sim, uma atividade altamente cognitiva e humana, sendo, portanto, passível de influências sociais - a qualidade do tratamento temático da informação no contexto de bibliotecas universitárias depende ainda da coleção e da comunidade usuária local. [...] São diversos os fatores que afetam o processo de indexação, já que não existe uma única maneira de indexar corretamente um documento (TARTAROTTI, 2014, p. 228).

Assim, ao refletir os fundamentos teórico-metodológicos do campo da Organização do Conhecimento e, ao mesmo tempo, as necessidades profissionais, o estabelecimento de uma política de indexação formalizada em um manual de indexação contribui com o processo de melhorias contínuas na atuação profissional do indexador, pois fornece as diretrizes de indexação ao profissional indexador em seu contexto de atuação.

5 CONSIDERAÇÕES FINAIS

Como considerações finais, verifica-se que os aportes teóricos da Linguística Textual são relevantes e podem contribuir, em **termos teóricos**, para o fortalecimento da abordagem teórica da indexação do campo da Organização do Conhecimento (OC) e em **termos práticos**, para a atuação do indexador como um leitor profissional, reforçando ainda mais

a relação interdisciplinar entre a Ciência da Informação e a Linguística, em torno da representação e materialização de sentidos.

A complexidade inerente ao processo de análise de assunto – realizada durante a leitura documental, ou seja, com o texto “em mãos” – é ampliada ao lançarmos o olhar para a etapa de **seleção de conceitos**, que ocorre após a leitura do documento, ou seja, não mais com o texto. Isto porque na seleção de conceitos são vários os fatores interferentes. Desse modo, a etapa de seleção de conceitos é tão ou ainda mais complexa que a etapa de identificação de conceitos, por conter ainda mais elementos e fatores interferentes no processo, **fatores extrínsecos** que independem da leitura documental e, portanto, do contexto onde o profissional atua. Reforça-se ainda a visão de Dal’Evedove, Tartarotti e Fujita (2015, p. 612) sobre a importância de considerarmos tanto a abordagem sociocognitiva como a abordagem sociocultural nos estudos em torno da indexação, mais especificamente na etapa de seleção de conceitos que abriga a tematicidade extrínseca textual da análise de assunto, tanto na formação quanto na capacitação do indexador.

Diante destas reflexões em torno da tematicidade extrínseca textual da análise de assunto para indexação e do texto como objeto de estudo da Linguística Textual, recomenda-se estudos comparativos em torno dos fatores interferentes na análise de assunto, especialmente sobre as convergências e divergências entre os fatores interferentes na etapa de identificação de conceitos e os fatores interferentes na etapa de seleção de conceitos, bem como a relação entre a seleção de conceitos e a recuperação da informação. Além disso, pesquisas interdisciplinares que versem sobre a abordagem teórica da análise do discurso, tal como a abordagem bakhtiniana, podem contribuir com aportes teórico-práticos da tematicidade intrínseca e da tematicidade extrínseca na representação temática da informação no processo de indexação

Finalmente, ao *incorporar novas, diferentes e futuras vozes* (McTAVISH, 2014, p. 330) nos estudos sobre a análise de assunto do processo de indexação, acredita-se que a representação temática da informação nesta abordagem discursiva e contextual como um modelo alternativo permita realmente a recuperação dos recursos informacionais nos sistemas de recuperação da informação, contribuindo para o entendimento

dos fatores interferentes no processo de indexação e da melhoria da atuação do bibliotecário indexador como um leitor profissional.

REFERÊNCIAS

- ALBRECHTSEN, H. Subject analysis and indexing: from automated indexing to domain analysis. *The indexer*, London, v. 18, n. 4, p. 219-223, Oct. 1993.
- BABIK, W. *et al.* ISKO and Knowledge Organization's 25th anniversary: the future of Knowledge Organization and ISKO Panel Discussion. Reported by Rebecca Green. *Knowledge Organization*, Wurzburg, v. 41, n. 4, p. 327-331, 2014.
- BENTES, A. C. Linguística textual. *In*: MUSSALIM, F.; BENTES, A. C. (org.). *Introdução à linguística: domínios e fronteiras*. 9. ed. São Paulo: Cortez, 2011. v.1.
- CAVALCANTI, M. do C. *Interação leitor-texto: aspectos de interpretação pragmática*. Campinas: Ed. da Unicamp, 1989.
- CESARINO, M. A. N., PINTO, M. C. M. F. Análise de assunto. *Revista de Biblioteconomia de Brasília*, Brasília, v. 8, n. 1, p. 32-43, jan./jun.1980.
- CHU, C. M.; O'BRIEN, A. Subject analysis: the critical first stage in indexing. *Journal of Information Science*, Cambridge, v. 19, n. 6, p. 439-454, 1993.
- DALEVEDOVE, P. R.; TARTAROTTI, R. C.D.; FUJITA, M. S. L. Reflexões acerca do desenvolvimento de metodologias para análise de assunto. *In*: GUIMARÃES, J. A. C.; DODEBEI, V. (org.). *Organização do conhecimento e diversidade cultural*. Marília: ISKO-Brasil; FUNDEPE, 2015. p. 607-614. (Série: Estudos Avançados em Organização e Representação do Conhecimento, v. 3).
- DALEVEDOVE, R. C. *Compreensão de leitura em análise de assunto para identificação e seleção de conceitos*. 2002. 154 f. Trabalho de Conclusão de Curso (Graduação em Biblioteconomia) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2002.
- DISCINI, N. *A comunicação nos textos*. São Paulo: Contexto, 2007.
- FAIRTHORNE, R. A. Content analysis, specification, and control. *Annual Review of Information Science and Technology*, Medford, v. 4, p. 73-109, 1969.
- FARROW, J. F. A cognitive process model of document indexing. *Journal of Documentation*, London, v. 47, n. 2, p. 149-66, 1991.
- FUJITA, M. S. L. A identificação de conceitos no processo de análise de assunto para indexação. *Revista Digital de Biblioteconomia e Ciência da Informação*, Campinas, v. 1, n. 1, p. 60-90, dez. 2003.
- FUJITA, M. S. L. *Leitura em análise documentária: uma contribuição à formação do indexador*. Marília: UNESP; CNPq, 2001. 190 p. (Relatório Parcial de Pesquisa).

- FUJITA, M. S. L. Representação documentária no processo de indexação com o modelo de leitura documentária para textos científicos e livros: uma abordagem cognitiva com protocolo verbal. *Ponto de Acesso*, Salvador, v. 7, n. 1, p. 42-66, abr. 2013.
- FUJITA, M. S. L.; RUBI, M. P.; BOCCATO, V. R. C. As diferentes perspectivas teóricas e metodológicas sobre indexação e catalogação de assuntos. In: FUJITA, M. S. L. (org.). *A indexação de livros: a percepção de catalogadores e usuários de bibliotecas universitárias*. São Paulo: Cultura Acadêmica, 2009. p. 19-42.
- GIASSON, J. *A compreensão na leitura*. Lisboa: Asa, 1993.
- GIL LEIVA, I. *Manual de indización: teoría y práctica*. Gijón: Trea, 2008.
- GUIMARÃES, E. *Texto, discurso e ensino*. São Paulo: Contexto, 2009.
- KOCH, I. G. V. *Desvendando os segredos do texto*. 6. ed. São Paulo: Cortez, 2009.
- LARA, M. L. G. *A representação documentária: em jogo a significação*. 1993. 133 f. Dissertação (Mestrado em Ciência da Comunicação) – Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo, 1993.
- MAI, J. E. Deconstructing the indexing process. *Advances in Librarianship*, New York, v. 23, p. 269-298, 2000.
- MAI, J. E. Semiotics and indexing: an analysis of the subject indexing process. *Journal of Documentation*, London, v. 57, n. 5, p. 591-622, 2001.
- MAI, J. E. The concept of subject: on problems in indexing. *Knowledge Organization for Information Retrieval: Proceedings of the 6th International Study Conference on Classification Research*, v. 6, p. 60-67, 1997a.
- MAI, J. E. The concept of subject in a semiotic light. Digital Collections: Implications for Users, Funders, Developers and Maintainers. *Proceedings of the ASIS Annual Meeting*, v. 34, p. 54-64, 1997b.
- McTAVISH, J. ISKO and Knowledge Organization's 25th anniversary: the future of Knowledge Organization and ISKO Panel Discussion. Reported by Rebecca Green. *Knowledge Organization*, v. 41, n. 4, p. 327-331, 2014.
- NEVES, D. A. B. Ciência da Informação e cognição humana: uma abordagem do processamento da informação. *Ciência da Informação*, Brasília, v. 35, n. 1, p. 39-44, jan./abr. 2006.
- NEVES, D. A. B.; DIAS, E. W.; PINHEIRO, Â. M. V. Uso de estratégias metacognitivas na leitura do indexador. *Ciência da Informação*, Brasília, v. 35, n. 3, p. 141-152, set./dez. 2006.
- OLSON, H. How we construct subjects: a feminist analysis. *Library Trends*, Champaign, v. 56, n. 2, p. 509-541, 2007.

- PINTO MOLINA, M. *Análisis documental: fundamentos y procedimientos*. 2. ed. rev. aum. Madrid: EUDEMA, 1993.
- RAJU, J.; RAJU, R. *Descriptive and subject cataloguing: a workbook*. Oxford: Chandos Publishing, 2006.
- SANTAELLA, L. *O que é semiótica*. 13. ed. São Paulo: Brasiliense, 1995.
- SMIRAGLIA, R. P. Bibliocentrism, cultural warrant, and the ethics of resource description: a case study. *Cataloging & Classification Quarterly*, New York, v. 47, n. 7, p. 671-686, 2009.
- SMIRAGLIA, R. P. Domain coherence within Knowledge Organization: people, interacting theoretically, across geopolitical and cultural boundaries. In: ANNUAL CAIS/ACSI CONFERENCE, 39., 2011, Canada. *Proceedings* [...]. Canada: University of New Brunswick, 2011. p. 1-6.
- TARTAROTTI, R. C. D. *Atuação bibliotecária no tratamento temático da informação em unidades informacionais: um estudo comparativo qualitativo-quantitativo*. 2014. 277 f. Dissertação (Mestrado em Ciência, Tecnologia e Sociedade) – Universidade Federal de São Carlos, São Carlos, 2014.
- TARTAROTTI, R. C. D., BOCCATO, V. R. C. A abordagem teórica bakhtiniana e o processo de indexação: diálogos. *Versão Beta: sob o signo da palavra*, São Carlos, ano 11, v. 74, p. 33-47, 2013.
- TAYLOR, A. G.; JOUDAREY, D. N. *The organization of information*. 3. ed. Wesport: Libraries Unlimited, 2009.
- TODD, R. J. Academic indexing: what's it all about? *The Indexer*, London, v. 18, n. 2, p. 101-104, 1992.
- VILELA, M.; KOCH, I. G. V. *Gramática da Língua Portuguesa*. Coimbra: Livraria Almedina, 2001.
- WELLISH, H. H. *Glossary of terminology in abstracting, classification, indexing and thesaurus construction*. 2. ed. Medford: American Society of Indexers, 2000.

ABORDAGEM TEÓRICA DA SEMIÓTICA PEIRCEANA PARA IDENTIFICAÇÃO DE ASSUNTOS EM MÚSICA E SUAS REPRESENTAÇÕES

Camila Monteiro de BARROS

Lígia Maria Arruda CAFÉ

RESUMO: Conhecer e mapear termos e conceitos relacionados com a música para fins de indexação perpassa pela compreensão da amplitude do processo de significação que tal tipo de informação pode suscitar. A discussão apresentada tem como base a Semiótica de Peirce, considerando a música como signo e os níveis de interpretante (emocional, energético e lógico) como indicadores de potenciais descritores da informação musical, aplicáveis à indexação. Propomos, assim, uma matriz conceitual semiótica como modelo de leitura do fenômeno da informação musical para fins de indexação. A matriz é dividida em três grandes dimensões, correspondentes às categorias fenomenológicas e níveis de interpretante: primeiridade, secundidade e terceiridade. Cada dimensão, de acordo com as características das categorias fenomenológicas, apresenta um grau diferente de precisão de representação e um grau diferente de relevância, dependendo do contexto em que ocorre a representação.

PALAVRAS-CHAVE: Informação musical. Indexação. Semiótica.

ABSTRACT: Knowing and mapping terms and concepts related to music for indexing purposes claims for the understanding of the extend of the signification process that this kind of information can evoke. The discussion is based on peircean Semiotics considering music as sign and the levels of interpretant (emotional, energetic and logic) as indicators of potential music information descriptors applied to indexing process. We propose a semiotic conceptual framework as a music phenomena reading model for indexing purposes. The framework is organized in three dimensions that are correspondents to the peircean phenomenological categories and interpretant levels: firstness, secondness and thirdness. Each dimension, according to the phenomenological categories, presents a specific precision level of representation, and a specific level of relevance, depending on the representation context.

KEYWORDS: Music information. Indexing. Semiotics.

1 INTRODUÇÃO

A informação não é algo exterior às pessoas ou que se encontre apenas dentro de suas mentes, é um processo construtivo que engloba relações entre diversos elementos possibilitadas pela percepção sensível. E, “[...] se a primeira fonte do nosso conhecimento é a sensibilidade, a segunda é o entendimento, poder de julgar, poder de conhecer não sensível.” (JAPIASSU; MARCONDES, 2006, p. 86). O não sensível refere-se à reflexão, ao pensamento, à experiência (sensível) levada à razão. É nesse espaço que surge a proposta de pensarmos analiticamente como ocorre nossa interação com o fenômeno musical e, a partir daí, buscarmos elementos que possam ser úteis à tarefa de representação da informação. Esse exercício confere um olhar informacional sobre a música - daí a noção de “informação musical”, inserindo-a no âmbito de um dos propósitos mais elementares da Ciência da Informação (CI): a representação de assunto por meio do processo de indexação para fins de recuperação da informação.

Conforme veremos a seguir, a noção de “assunto” no contexto da informação musical é diferente do contexto da informação verbal. Da mesma forma, o que diferencia elementos de representação descritiva daqueles de representação temática também não apresentam delimitação clara. Assim, conhecer e mapear termos e conceitos relacionados com a música perpassa pela compreensão da amplitude do processo de significação que tal tipo de informação pode suscitar, justamente pela forma peculiar com que a música pode desencadear significados. A natureza expressiva e não necessariamente referencial da música enaltece a hipótese de que

entender o que ela pode significar é premissa para a construção de uma base teórica capaz de sustentar outras discussões no contexto da indexação, como adotada pela CI. Essa situação tem impacto principalmente sobre aqueles conceitos que têm sua gênese na emoção experimentada por um indivíduo, isto é, quando as emoções são determinantes na significação de certo tipo de informação, como é o caso da informação musical.

A abordagem da música como recurso passível de ser inserido em um sistema de informação não é um assunto novo. Tal temática já está consolidada como âmbito de pesquisa haja vista, por exemplo, a realização anual, desde 2000, da conferência da *International Society for Music Information Retrieval* (ISMIR), principal evento da área de *Music Information Retrieval* (MIR). A conferência tem característica multidisciplinar, sendo que, de acordo com o site do evento, envolve estudos de processamento de sinais sonoros e textuais, comportamento e perfil de usuários, propriedades e características da música, classificação automática, transcrição musical, sistemas de recuperação e recomendação de músicas, entre outros.

Na medida em que a CI busca trabalhar no sentido de aprimorar o processo de mediação da informação, de comunicação de ideias, depara-se com questões relacionadas à organização social, aos sujeitos, às instituições, questões de ordem linguística, instrumental, semiótica, entre outros aspectos que interferem na comunicação.

Nesse contexto, inevitavelmente, o panorama epistemológico da CI se expande em uma interação com diferentes disciplinas, cada qual evidenciando determinado aspecto do objeto central da área. A Semiótica de Peirce surge, então, como teoria possivelmente capaz de fornecer um aparato conceitual para o estudo da dimensão do significado da informação corporificada em qualquer tipo de signo.

Para Peirce (1995), a semiose, que é o processo de produção de significado, ocorre com base em três correlatos: o signo que, nesse caso, é a própria música; o objeto, aquilo a que o signo se refere e representa; e o interpretante, que é o significado criado na mente que interpreta o signo. Os três níveis de ocorrência do significado, ou seja, do interpretante, são: emocional (sensações, sentimentos, emoções), energético (experiências particulares, respostas físicas) e lógico (convenção, cognição). Cada nível de interpretante carrega características (mais ou menos definidas)

das categorias fenomenológicas peirceana, quais sejam: primeiridade, secundidade e terceiridade, conforme explicaremos mais adiante. No caso da música, os níveis emocionais e energéticos desempenham um papel mais evidente que na significação dos signos verbais de nível lógico, já que estes últimos são, por natureza, signos de convenção social e linguística.

A CI tem cada vez mais admitido a profunda característica sociológica da sua constituição como área científica. Nas áreas de organização do conhecimento (OC) e organização da informação (OI), esse aspecto fica ainda mais evidente quando se percebe que os significados, de fato, não são estáticos, pois perpassam por uma série de experiências fenomenológicas experimentadas por indivíduos continuamente em contato com a realidade. A leitura de um documento ou a análise do conjunto conceitual de uma área de conhecimento para fins de sua representação lidam não só com estruturas da própria linguagem, representada pela sistematização de diversos signos, mas com o contexto de significação, com os objetos a que os signos se referem e, em um prospecto de acesso futuro à informação, com os possíveis significados que os signos terão em uma mente interpretante.

Com base nesse contexto desafiador e estimulante, o objetivo deste capítulo é propor parâmetros de fundamento Semiótico para o que chamaremos de “modelo de leitura do fenômeno”, aproximando esse processo dos objetivos da indexação.

2 CONSIDERAÇÕES SOBRE “ASSUNTO” NO CONTEXTO INFORMAÇÃO MUSICAL

Em Café e Barros (2015) já foram discutidas algumas especificidades da informação musical quando se pretende delimitar o que seria o “assunto” da música. De acordo com as autoras, aspectos como dimensão emocional, percepção social da música, uso, gênero musical, entre outros (SELFRIDGE- FIELD, 2006; LAPLANTE; DOWNIE, 2006; LAPLANTE, 2010, 2011; HU; DOWNIE; EHMAFNN, 2006; HU; DOWNIE, 2007) são passíveis de serem representados dependendo do objetivo do sistema em questão (recomendação de músicas, recuperação precisa, descoberta, etc.). Muitos desses aspectos estão relacionados à experiência e comportamento do ouvinte, questão também já admitida como fundamental em estudos da informação musical (ZANGERLE;

PICHL, 2018; FONT; SERRÀ; SERRA, 2015; LAPLANTE, 2015; LEE; DOWNIE, 2004; LEE, 2010; CRUZ, 2011; LEE; CUNNINGHAM, 2013).

Pesquisas envolvendo perfil de usuários da informações musical, devem ter em vista, segundo Lee (2010), os desafios transcultural e multilíngue, para os quais os modelos³¹ de comportamento de busca e recuperação da informação se mostram demasiadamente genéricos e não servem como parâmetro para conhecer o comportamento de usuários de um tipo de informação peculiar como a música. Lee (2010) empreendeu um amplo levantamento das questões de busca expressadas em linguagem natural pelos usuários no *Google Answer* (site de interação entre usuários, do tipo pergunta-resposta) e, por meio da categorização dos aspectos da música apontados nas expressões de busca, atestou que a maior parte das questões se referia a aspectos muito pessoais e centrados no usuário, quando em comparação com a descrição bibliográfica habitual. Como recomendação para o aprimoramento de sistemas de MIR, a autora sugere que sejam incorporados termos familiares, bem como o contexto dos usuários para empreender a recuperação da informação musical.

O desafio transcultural (*crosscultural*) e multilíngue foi tema de pesquisas anteriores. Lee, Downie e Cunningham (2005) trazem à tona o problema do desenvolvimento de uma biblioteca digital de música em nível mundial, já que esse empreendimento implica lidar com distintas culturas e idiomas, além da problemática da ampla ênfase dada à música ocidental nas pesquisas em MIR. Seu estudo de caráter exploratório buscou mapear o comportamento de busca de usuários que procuram músicas de diferentes culturas em idiomas estrangeiros. Os autores apresentam como exemplo a questão abaixo exposta, retirada de um site coreano do tipo pergunta-resposta:

Primeiro de tudo, é muito longo. Apenas o primeiro movimento parece ir além de 10 minutos. O estado de ânimo [mood] é, em geral, obscuro e triste. A melodia principal é interpretada por violino e o piano aparece brevemente. Eu sei apenas que o nome do compositor tem quatro caracteres quando expressado em coreano. Alguém sabe o

³¹ Lee (2010) cita como exemplos o modelo *sense-making* de Dervin, o modelo comportamental de Ellis, o comportamento informacional humano descrito por Wilson, entre outros.

título dessa obra? (LEE; DOWNIE; CUNNINGHAM, 2005, p. 2, tradução nossa³²).

Em nota de rodapé, Lee, Downie e Cunningham (2005) relatam que a resposta a essa questão é “*The Devil’s Trill*”, do compositor italiano Tartini, cujo nome em coreano tem quatro caracteres: 타르티니. Por meio da coleta de expressões de busca inseridas no site *Google Answers* e no site coreano *Naver* 지식 (knowledge) iN (da mesma natureza interativa de pergunta-resposta), os autores fizeram comparações entre o comportamento de busca dos coreanos e dos americanos. Em resumo, as conclusões a que os autores chegaram apontam para a necessidade de incluir a possibilidade de busca por meio de características da música que são relativamente mais neutras culturalmente, ou seja, que são possivelmente universais, como *query-by-humming* (já que a sonoridade da música ultrapassa barreiras linguísticas) e o gênero do artista (masculino/feminino).

Casey et al. (2008) realizaram um levantamento sobre os métodos utilizados na representação da informação musical, cada qual baseando-se em determinadas características da música. Entre outras questões, os autores definem uma escala de três níveis de especificidade de sistemas de MIR, listados abaixo, de acordo com a precisão com a qual as características da música podem ser extraídas para fins de representação e de acordo com a precisão com que essas características podem ser utilizadas no momento da recuperação, do ponto de vista da relação entre questão de busca e correspondência com o documento recuperado.

- 1) Sistemas de alta especificidade: aqueles que utilizam instâncias do próprio áudio, como conceitos musicais, melodia, harmonia, timbre e outros aspectos da estrutura da música. Referem-se, portanto, à descrição exata de cada música individualmente.
- 2) Sistemas de baixa especificidade: utilizam conceitos mais gerais como gênero musical, estilo, emoção, entre outros. Nesse caso, são recuperadas músicas não necessariamente com conteúdo musical em comum com a questão de busca, mas que compartilham algumas características globais.

³² “*First of all, it’s very long. Just the first movement seems to go over 10 minutes. The mood is dark and sad overall. The main melody is played by violin and the piano briefly appears. I only know that the composer’s name is four characters when it is spelled out in Korean. Does anybody know the title of this work?*” (LEE; DOWNIE; CUNNINGHAM, 2005, p. 2).

- 3) Sistemas de média especificidade: utilizam métricas extraídas do sinal de áudio para recuperação de músicas. As características extraídas pelos sistemas de média especificidade podem ser convertidas em informações para representações tanto de alta quanto de baixa especificidade.

Existem muitas pesquisas que testam aplicações automatizadas que lidam com a representação da dimensão emocional da música, algumas dessas aplicações, utilizadas para fins de classificação, recomendação ou ainda organização de coleções pessoais estão disponíveis no site do ISMIR³³. Testes e resultados de variados tipos de aplicações voltados à representação e recuperação da informação musical são apresentados e confrontados nos encontros do *Music Information Retrieval Evaluation eXchange* (MIREX), organizado pelo *The International Music Information Retrieval Systems Evaluation Laboratory* (IMIRSEL), da Escola de Biblioteconomia e Ciência da Informação da Universidade de Illinois em Urbana-Champaign. O IMIRSEL possui ainda outros projetos relacionados como o *The Virtual Research Labs* (VRL) *using Music-to-Knowledge* (M2K) e o *The Human Use of Music Information Retrieval Systems* (HUMIRS)³⁴.

A eficiência da representação e recuperação da informação musical também engloba a forma como a interface do sistema será apresentada ao usuário. Para Casey et al. (2008), essa questão é fundamental para promover aos usuários finais do sistema uma efetiva interação e compreensão da forma como a visualização da informação musical é apresentada, aprimorando, principalmente, a tarefa de navegação (*browsing*).

Se adotarmos a visão de Brascher e Café (2010) sobre os processos de organização da informação, ou seja, representação descritiva (ligada à descrição física, à catalogação) e representação temática (ligada à indexação, classificação e condensação) veremos que não é possível executar a simples transposição da abordagem dedicada ao texto verbal para a informação musical.

Ocorre que alguns aspectos da música citados anteriormente (emoções, conceitos gerais) não estão associados à representação descritiva da informação, já que não podem ser extraídos do próprio

³³ Disponível em: <http://ismir.net/resources.html>.

³⁴ Disponível em: <http://www.music-ir.org/evaluation/>.

documento musical. Nesse sentido, podemos aproximá-los da atividade de representação temática. Entretanto, esses aspectos também escapam da análise que o indexador realizaria na intenção de extrair conceitos relevantes a respeito da temática do documento (LANCASTER, 2004). Não se busca, exatamente, responder à questão “a respeito do que a música está tratando?”, mas “a respeito do quê a música passa a tratar quando interpretada por um ouvinte?” e, nesse caso, o termo “assunto” não reflete a amplitude necessária para a indexação da música.

Em outras palavras, a noção de assunto é mais difusa no contexto da informação musical. Um exemplo que pode ilustrar essa situação é a pesquisa dos termos constantes no campo “assunto” da música (gravação sonora) “Garota de Ipanema”, composta por Vinicius de Moraes e Antônio Carlos Brasileiro de Almeida Jobim “Tom Jobim”. No quadro abaixo podemos observar a variação gramatical e semântica dos termos de alguns desses registros bibliográficos.

Quadro 1 - Termos constantes no campo “assunto” da música Garota de Ipanema.

UNIVERSIDADE	TERMOS DO CAMPO "ASSUNTO"
1	Música popular - Brasil
2	Jobim, Tom, 1927-1994. Música popular - Brasil
3	Quartetos de cordas
4	Música de câmara Duo (Violino e piano) Música brasileira
5	Jobim, Tom, 1927-1994. Música Música popular - Brasil Bossa-nova

Fonte: Elaborado pelas autoras.

O quadro mostra a descrição de aspectos de naturezas distintas relacionados à noção de assunto, como definição de instrumentos,

do gênero musical, do tipo de documento, de aspectos geográficos e históricos.

Da mesma forma, a noção de “resumo” também é diferente dependendo do tipo de informação. O resumo no contexto da informação musical pode ser entendido como aquela pequena amostra da música que é disponibilizada para que o ouvinte a “experimente” antes de decidir se vai ou não ouvi-la na íntegra. Geralmente, são disponibilizados alguns segundos do início da música ou do refrão. Huron (2000), considerando resumo da informação musical um extrato melódico, afirma que nem todas as partes da música são igualmente representativas e que “a prática de extrair os segundos iniciais (*incipit*) não é ótima para identificar ou reconhecer uma obra” (HURON, 2000, p. 2, tradução nossa³⁵). O autor defende que seria mais eficiente uma síntese de duas ou mais passagens da música, editadas em uma única amostra de alguns segundos de duração. Ensaçando uma aproximação à abordagem mais habitual da CI, pode-se pensar o resumo da música também como um texto que pode descrever de forma verbal características da estrutura musical, apresentar uma crítica musical ou informe de lançamento (os *releases*).

O próprio termo “tema” tem um significado institucionalizado na área da Música, e se refere à melodia principal que a fundamenta e caracteriza. É um entendimento distinto daquele usualmente trazido pela CI, de que “tema” e “assunto” compartilham significados muito próximos.

No processo de indexação de documentos textuais, podemos executar movimentos cognitivos deliberados buscando, por exemplo, um assunto geral e exercendo um movimento em direção à sua especificação. Pode-se também definir um assunto central e depois elencar os aspectos secundários. Seja qual for o modelo de leitura adotado, a finalidade é a análise de assunto, identificação, seleção de conceitos e definição de termos de representação que assegurem o equilíbrio entre exaustividade e precisão (LANCASTER, 2004).

Já na música, diferentes dimensões do seu significado se justapõem podendo ser igualmente centrais, dependendo do contexto. Portanto, o movimento cognitivo da análise da informação musical não é vertical, ou seja, não busca a definição de conceito geral/conceito particular, conceito

³⁵ “the practice of extracting the initial few seconds (*incipit*) is not optimum for identifying or recognizing a work.” (HURON, 2000, p. 2).

central/ conceito secundário. É um movimento horizontal que se direciona para uma e outra dimensão de significado, avaliando sua relevância para o contexto que se apresenta. Por exemplo, utilizar o termo “pop para malhar” ao descrever uma música em um site de *streaming* é tão relevante quanto utilizar o termo “sonata” na descrição voltada a músicos especialistas. Nesse contexto, a mediação daquele que realiza a indexação se dá no sentido de pensar como facilitar a comunicação das experiências dos usuários para outros usuários, ou seja, não é o “conteúdo” da música que é descrito, mas a experiência/entendimento do usuário. Nesse caso, a noção de leitura técnica pode ser substituída por uma abordagem que olha para as expressões de emoções e opiniões como elementos potencialmente descritivos. Dessa forma, a definição do “assunto” da música precisa, necessariamente, da contribuição dos próprios usuários, seja no fornecimento ou na validação constante de termos e dimensões de representação.

Assim, convidamos o leitor a abrir sua mente para uma nova noção de “assunto” que ainda está em construção e, portanto, inacabada. Trata-se de ampliar a visão de indexação para além do delineamento já conhecido pela CI, aceitando dimensões da informação que nem sempre estarão em acordo com os preceitos de precisão, consistência e constância da indexação.

Sabe-se que a indexação é um processo interpretativo. Essa intrínseca característica interpretativa que compõe a efetivação da indexação é abordada por Mai (2001) como um atributo que está presente em todas as fases desse processo. Em função dessa característica, a indexação traz consigo uma frágil certeza de seu resultado pelo principal fato de que a interpretação está longe de ser neutra, independentemente da vontade do indexador de sê-lo. Na CI, sabemos que a não existência da neutralidade na leitura documental se deve, em primeira instância, à orientação do processo de indexação para um grupo de usuários definido. Entretanto, mesmo com a adoção de certos parâmetros, a interpretação *per se* não é neutra, pois ela está entrelaçada com o contexto do indexador, responsável pela busca dos possíveis significados. Assim, Mai (2001) defende que a indexação, seja qual for a quantidade de etapas que se considere que tenha, é um processo que, como um todo, é de natureza interpretativa e, portanto, se dá por meio de signos. Ou seja,

[...] apesar de cada um dos elementos da indexação de assunto – documento, assunto, descrição de assunto e entrada de assunto – serem signos, cada um é um tipo diferente de signo, que fará uma considerável diferença na forma de abordá-los e interpretá-los. (MAI, 2001, p. 611, tradução nossa³⁶).

Sousa e Almeida (2012) expõem a questão de como se constitui o referente que o indexador adota no momento de representar o assunto de um documento. A principal questão levantada pelos autores é que o indexador não compartilha da mesma realidade que o usuário, ou seja, não tem a mesma experiência que o usuário atinge na sua realidade, aspecto de especial relevância para a informação musical. Dessa forma, para alcançar certos significados, o indexador tem como referente as delimitações apresentadas nos Sistemas de Organização do Conhecimento e na sua própria vivência. “O indexador só tem a experiência sociocognitiva gerada a partir do contato com as obras, mas não possui a experiência de um especialista da área dos documentos que tem de indexar.” (SOUSA; ALMEIDA, 2012, p. 30) e é por meio dessa experiência com as obras que é extraída a constituição do referente e, por decorrência, do significado.

Para além da leitura documentária (FUJITA, 2004, 2013, KOBASHI; SANTOS, 2011), a informação musical exige um processo ainda mais flexível de interpretação, daí nossa proposta da abordagem Semiótica, teoria capaz de possibilitar a análise da expressão e da experiência musical (DOUGHERTY, 1994). Assim, nos parece que não são as teorias utilizadas no estudo da representação temática que são insuficientes, mas que a música é um objeto que ultrapassa essas teorias na medida em que suscita significados peculiares. A perspectiva semiótica, portanto, vem contribuir para a ampliação desses estudos, com o intuito de acrescentar novos olhares às abordagens teóricas já adotadas, e não substituí-las.

A música pode *carregar* certos significados, representações e relações intencionadas pelo compositor, expressa Svenonius (1994), mas, por outro lado, tem o poder de *suscitar* sensações e relações (similares ou distintas daquelas intencionadas pelo compositor). Aí a questão da indexação se envolve

³⁶ “[...] although each of the elements in the subject indexing process – document, subject, subject description and subject entry – are signs, that each is a different kind of sign will make a considerable difference in how one approaches them and interprets them.” (MAI, 2001, p. 611).

com uma especificidade peculiar desse tipo de informação, diferentemente da linguagem discursiva, que é relativamente mais objetiva e referencial. Situação similar ocorre com outros materiais que não são utilizados para fins de documentação (como poesia, escultura etc.).

3 INTRODUÇÃO À SEMIÓTICA DE PEIRCE

Para Peirce (1995), a concepção de representação é fundamental ao próprio pensamento e ocorre em uma relação triádica entre três correlatos: signo ou representâmen, objeto e interpretante.

O signo, para que assim possa ser entendido, precisa estar no lugar de outro, referir-se a outro, representar. Essa referência ao objeto significa que deve ser possível expor, logicamente, algum atributo ou contexto que explique essa representação. Esse atributo ou contexto pode ser entendido por meio do fundamento do signo, ou seja, por meio de determinado aspecto do objeto dinâmico emanado no signo que o constitui e empregasse a possibilidade de representação. Em outras palavras, algo precisa conectar o signo ao objeto e esse “algo” é o fundamento.

Para Peirce (1995), a forma como se dá a experiência com o fenômeno na realidade desencadeia e influencia o desenvolvimento dos processos significativos. Esse é um princípio que fundamenta a noção de pragmatismo adotada na semiótica do autor. Dessa forma, para Peirce existe de fato uma realidade última, que independe da nossa vontade (uma tendência ao realismo), porém, o entendimento dessa realidade é sempre aproximativo na medida em que tal entendimento perpassa pelos processos semióticos que envolvem toda e qualquer relação do ser com o mundo (SILVEIRA, 2007). Considerando não uma dicotomia, mas um *continuum* entre mente e matéria, esse entendimento da realidade é, então, novamente confrontado com a experiência, reforçando ou formando novos significados. Dentro da semiótica, a pertinência do pragmatismo peirceano para a abordagem da informação musical está justamente na relevância da própria experiência fenomenológica, em qualquer nível, como base para análise desse tipo de informação. É preciso observar como de fato ocorre a percepção do fenômeno, quais elementos concorrem para a interpretação da música no real momento em que a experiência ocorre. Considerar a interpretação e representação da música no sentido semiótico é operar

com os significados pragmaticamente atualizados pelos usuários, ou seja, os interpretantes.

Para Peirce (CP 1.184) , a Fenomenologia se apresenta como ciência fundamental da sua Filosofia na medida em que “descreve o universo da experiência” (SILVEIRA, 2007, p. 38) tendo, portanto, como central interesse, os “[...] componentes elementares que caracterizarão os fenômenos que preenchem o universo da experiência.” (idem). Peirce chamou esses componentes elementares de “categorias”. Às categorias fundamentais dos fenômenos Peirce (1995) denominou primeiridade (*firstness*), secundidade (*secondness*) e terceiridade (*thirdness*). A atualização do interpretante acompanha a lógica das categorias, podendo ser um interpretante emocional (primeiridade), energético (secundidade) ou lógico (terceiridade).

A primeiridade está relacionada com o sentimento, “[...] um sentir meramente passivo que não atua e não julga” (PEIRCE, 1995, p. 14), a consciência de um instante no tempo, “[...] consciência passiva da qualidade, sem reconhecimento ou análise.” (PEIRCE, 1995, p. 14). Na primeiridade, dá-se apenas a sensação imediata e espontânea. No caso da música, o interpretante emocional pode ser exemplificado pela ocorrência de um som ambiente, em que a percepção do som ocorre, mas esse não é identificado. O som permanece apenas uma qualidade sonora, sem uma apreciação direta. Nessa categoria, o interpretante pode também ser um tipo de emoção, ou seja, a percepção de diversas qualidades e sensações que se traduzem em um sentimento mais complexo (MARTINEZ, 2001, p. 156), como alegria, inveja, raiva, etc.

A secundidade, segundo Peirce (1995, p. 14), compreende um “[...] sentido de resistência [da consciência], de um fato externo ou outra coisa”. Assim, o reconhecimento da realidade é presença da secundidade na medida em que “[...] a realidade é aquilo que insiste, nos força a reconhecer um *outro* diferente do espírito.” (PEIRCE, 1983, p. 90, grifo do autor). O interpretante energético da música está relacionado à ressonância física do som no corpo do intérprete, de forma mais ou menos consciente. Das situações menos conscientes, depreende-se a ocorrência de movimentos instintivos, como o arrepio. Os movimentos conscientes são interpretações que resultam em ações como cantar e dançar. A secundidade também pode ocorrer somente no âmbito mental, mas com a característica de haver

um esforço como resposta ao estímulo externo do som. Como exemplo, podemos pensar no empenho do ouvinte em tentar compreender a letra ou a estrutura da música.

Aterceiridade é a “[...] consciência sintética, reunindo tempo, sentido de aprendizado, pensamento.”, envolve, portanto, cognição (PEIRCE, 1995, p. 14). Na categoria da terceiridade, por meio do pensamento, representamos e reconhecemos o mundo de forma deliberada. É o pensamento deliberado que diferencia o interpretante lógico do emocional, que tem a consciência passiva como traço definidor ou do interpretante energético, que se baseia na repercussão e resposta como efeitos interpretativos. O interpretante lógico da informação musical pode ser visualizado na interação do ouvinte com os aspectos textuais da música como o conteúdo da letra, análise e reconhecimento (mesmo que hipotético) de instrumentos e vocais ou ainda o estabelecimento da relação da música com outros aspectos da memória como outros momentos e locais.

É importante ressaltar que na medida em que se experiencia um fenômeno de terceiridade, aspectos de primeiridade e secundidade terão também existência, ainda que de modo menos evidente. O mesmo ocorre com um fenômeno de secundidade, que terá em si a primeiridade também presente na experiência. Peirce considera o princípio da continuidade como fundamento de especial relevância para qualquer elemento da realidade. Assim, compreende que as regularidades não são absolutas, mas estão em evolução para certo grau de aproximação de como a mente ou o mundo orgânico geral deve se comportar. Em outras palavras, os significados mudam, são flexíveis, se modificam e se atualizam conforme se dão nossas experiências. Especialmente no âmbito da música, essa característica precisa ser considerada quando pensamos nos processos de organização da informação.

Essa brevíssima introdução à Semiótica de Peirce tem o objetivo de apresentar os níveis de interpretante, ou seja, as possibilidades de significação da música, como base para análise da informação musical. Tal base, nos fornece um percurso analítico útil à descrição dos assuntos relacionados a esse tipo de informação, já que a lógica da linguagem musical é diferente daquela da linguagem verbal. O significado da música e os possíveis assuntos a ela conectados não provém, necessariamente, da

lógica da sua sintaxe ou da estrutura da sua textualidade (combinação de notas, sons, etc.), mas das percepções imputadas à música pelos ouvintes.

Também é relevante sinalizar que estamos tentando um tipo de tradução por meio da verbalização de uma informação não verbal e de significados que não têm início na linguagem verbal (como teria a leitura de um texto, por exemplo). É uma adaptação, ainda que experimental, da teoria semiótica com vistas ao uso de palavras para representação e objetivação de outros sistemas comunicativos que não se baseiam fundamentalmente em palavras (sentimentos, ações, metáforas...).

4 MODELO DE LEITURA DO FENÔMENO MUSICAL PARA FINS DE INDEXAÇÃO

Com base na Semiótica de Peirce e nas considerações a respeito das características da informação musical discutidas anteriormente, o modelo de leitura que propomos tem como base um tipo de classificação ou associação dos elementos percebidos em cada nível de interpretante com os elementos cotados como potenciais descritores desses mesmos níveis, aplicáveis à indexação. A partir dessa análise classificatória de elementos, podemos criar parâmetros para representação de assunto da música. É relevante, no entanto, nos mantermos atentos e críticos já que a proposta implica na transposição dos conceitos coletados de uma teoria geral dos signos, a Semiótica, para uma teoria muito específica, com finalidades bem particulares, que é a representação da informação. Por outro lado, considerando-se que toda experiência fenomenológica, em todos os seus níveis, implica na interação dos e com os signos, é razoável que os princípios teóricos da Semiótica nos guiem no desvelamento dos possíveis significados da música. Desses significados, a Semiótica também nos mostra quais elementos são mais constantes, mais estáveis, com maior nível de generalidade e quais são mais variáveis, precisando de atualização recorrente quando da sua “tradução” e registro para uma forma verbal, como o termo.

Vale registrar que em Barros, Café e Laplante (2019), a abordagem semiótica é aplicada na análise do autorrelato dos usuários da informação musical. Os usuários narram, com suas próprias palavras, experiências intensas com a música e, a partir dessas narrações, as autoras extraem expressões que se referem à descrição do signo (a música), dos objetos e

dos interpretantes alcançados na semiose relatada. A pesquisa, apesar de apresentar objetivos distintos da discussão em andamento aqui, tem importante contribuição para a proposta que ora apresentamos.

Propomos, assim, uma matriz conceitual como modelo de leitura da informação musical para fins de indexação, dividida em três grandes dimensões, correspondentes às categorias fenomenológicas e níveis de interpretante, conforme abaixo.

Quadro 2 - Matriz Semiótica de leitura da informação musical para fins de indexação.

CATEGORIA FENOMENOLÓGICA	NÍVEL DE INTERPRETANTE	NATUREZA DO ELEMENTO A SER DESCRITO	EXEMPLOS
Primeiridade	Emocional	Sentimentos mais ou menos definidos	Alegria, tristeza, raiva (...)
		Ambiência	Calma, divertida (...)
		Apreciação crítica	Boa, bonita (...)
Secundidade	Energético	Uso da informação musical	Para estudar, para dançar, para fazer exercícios físicos, para concentração (...)
Terceiridade	Lógico	Estrutura musical	Instrumentos (violão, música vocal, música instrumental, vocal feminino/masculino, flauta). Tonalidade. Arranjo (redução para piano) (...)
		Histórico	Biografia do artista, história da banda, história da canção (...)
		Letra	Letra da música (...)
		Gênero musical	Samba, partido alto, rock, reggae (...)

Fonte: Elaborado pelas autoras.

Da perspectiva Semiótica, a leitura do fenômeno que implica a informação musical tem como base o usuário, o momento da sua experiência e a interpretação resultante da experiência musical. Por isso, para além da leitura, essa matriz propõe uma compreensão do fenômeno musical nas suas três dimensões.

Cada dimensão apresenta um grau de relevância diferente, dependendo do objetivo da representação. Conforme apresentamos anteriormente, tal objetivo pode estar relacionado ao contexto do usuário no que se refere a seu objetivo no uso da informação musical; ao seu grau de conhecimento formal de música ou de temas a ela relacionados (músicos, historiadores, antropólogos, repórteres, ouvinte leigo, etc.); seu grau de engajamento com a música (tempo dedicado à atividade de escuta ou outras atividades que envolvam música, como participação em shows, grupos de fãs, etc.). O objetivo da representação, naturalmente, também está relacionado ao objetivo do serviço de informação como recomendação, descoberta, etc. Cabe ao profissional da informação definir os elementos de descrição que melhor atenderão ao contexto de representação.

Cada dimensão, de acordo com as características das categorias fenomenológicas, apresentam um grau diferente de precisão de representação. Bem como possibilidades diferentes de que os elementos correspondentes sejam determinados pelo profissional que realiza a representação ou pelo próprio usuário, conforme veremos a seguir.

Na representação da informação musical considerando o nível emocional, a participação do usuário é fundamental. Seja no fornecimento de inputs para aprendizagem do sistema de descrição automática, seja para fornecer os próprios termos e expressões de descrição. Na verbalização destes significados é muito comum o uso de gírias, metáforas, analogias e outras expressões linguísticas. Como Félix Mendehlsson (1809-1847), compositor do início do período romântico já afirmou: “Os pensamentos que me são expressos pela música que eu amo não são muito indefinidos para serem colocados em palavras, mas, pelo contrário, são definidos demais.” (MENDEHLSSON apud SHORT, 2007, p. 204).

A característica principal deste nível é a pluralidade de significados, já que representa a experiência particular, passada ou presente do ouvinte. Sentimentos e emoções não estão relacionados à música em si, mas a lembranças, momentos, sensações, etc. Neste sentido, observa-se

que, apesar de relevante para a recuperação, a precisão e o controle de vocabulário (univocidade) não são o objetivo principal no caso do nível emocional da música, sobressaindo-se a noção de variação de vocabulário.

A heterogeneidade do vocabulário é o que o aproxima mais de representar “fielmente” o significado neste nível. Nesse contexto, a tarefa de tradução da linguagem natural para uma linguagem documental (CINTRA, 2002), como é considerada nas teorias de indexação, é uma tarefa que pode implicar na perda da riqueza semântica da expressão das emoções. Termos que inicialmente parecem opostos (como alegria/tristeza), podem ocorrer para representar a mesma música, não sendo essa situação um caso de contradição ou falta de consistência de indexação. O mesmo não ocorre com o nível lógico de significados, pois nesse nível certas referências a objetos mais ou menos fixos precisam ser mantidas já que representam convenção (por exemplo, a ideia de autoria, gravadora, ano, etc.).

A representação da informação musical considerando o nível energético se caracteriza principalmente pela recomendação de uso. Nesse caso, características da música em si como bpm (batidas por minuto), instrumentos, tonalidade podem ser fontes para a descrição automática da música por meio de algoritmos que executem a análise, relacionem essas características e classifiquem automaticamente a música. Entretanto, a participação do usuário na confirmação dessa análise automática é desejável. Essa participação pode se dar por meio do mapeamento do perfil de escuta, da avaliação do ouvinte ao escutar a música, ou outras formas de compartilhamento de experiência.

No nível lógico, os elementos de descrição são mais estáveis, ou seja, têm maior probabilidade de alcançar uma certa generalidade na forma como são percebidos. A letra da música, sua estrutura (como tonalidade) ou a biografia dos artistas são elementos que têm menor dependência da percepção do ouvinte, porém, têm maior dependência das condições prévias do ouvinte para percebê-los. Por exemplo, não é a percepção do ouvinte que vai determinar se a canção está em tonalidade de “lá maior”, já que o que define a tonalidade é uma série de regras convencionais do âmbito da música. Entretanto, se o ouvinte não tem conhecimento formal de música, não poderá perceber esse significado.

De qualquer forma, os conceitos referentes ao nível lógico de interpretação são aqueles com maior possibilidade de serem extraídos

automaticamente ou definidos pelo indexador, precisando de menos intervenção humana para serem endossados.

Sendo o gênero musical uma forma muito comum de as pessoas se referirem à música, cabem algumas considerações específicas sobre esse tema. Os gêneros musicais não são demarcados somente pela forma ou estilo de um texto musical em sentido estrito mas, sim, “[...] pela percepção de suas ‘formas’ e ‘estilos’ pela audiência através das performances pressupostas pelos gêneros.” (JANOTTI JUNIOR, 2003, p. 37). De acordo com Janotti Jr. (2003), o gênero musical é definido por três elementos: o princípio mercadológico representado pelas produtoras/gravadoras, a crítica musical exercida pela mídia e pelos grupos reais e potenciais de consumidores, fãs e ouvintes. Assim, na medida em que essas referências se modificam histórica e geograficamente, as definições de gênero musical também se modificam. Por exemplo, os parâmetros culturais que definiam o gênero “rock” na década de 1950, período de músicos como Elvis Presley e Chuck Berry, passaram por modificações substanciais tendo-se, na década de 1990, a band The Strokes como um exemplar desse gênero.

Abrahamsen (2003), quando discute a descrição dos gêneros musicais, afirma que a música clássica parece ter uma melhor diferenciação nas categorias que representam o gênero musical do que a música popular. Esse aspecto revela um tipo de visão de mundo, de desenvolvimento e construção de discurso, pois, de fato, o aparato teórico que acompanha o desenvolvimento da música clássica de origem europeia fornece parâmetros que delimitam tecnicamente a diferenciação entre seus subgêneros. Diferentemente da música popular em que elementos muito além dos técnicos e estruturais competem para formar a ideia não terminada do que se entende que seja determinado gênero musical.

Em resumo, o modelo de leitura propõe uma sequência de análise que começa por observar o fenômeno musical em seus aspectos de primeiridade, secundidade e terceiridade e os possíveis interpretantes de cada nível (cujos exemplos estão no quadro 2). A segunda etapa, após o mapeamento dos interpretantes, é buscar a melhor forma de expressá-los em linguagem verbal, sendo que a fonte de levantamento e coleta de termos varia conforme as características de cada nível. No nível emocional: menor controle de vocabulário, necessário maior participação do usuário; no nível energético: pode-se contar com mais apoio de sistemas automatizados e

participação do usuário no refinamento do sistema; no nível energético: possibilidade de maior uso de sistemas automatizados ou definições *a priori*, possibilidade de maior controle de vocabulário e, portanto, de uso de fontes formais de conhecimento para representação do domínio da música.

5 CONSIDERAÇÕES FINAIS

Todos os significados são corrigidos (no sentido de serem ajustados) conforme se dão nossas experiências. Isso é o que Mai (2011) chama de “distância semiótica” da realidade. Assim, a utilização da linguagem verbal para traduzir as concepções que têm origem na experiência com o som é, necessariamente, revista e modificada de acordo com as novas significações imputadas à música no contexto social - e sempre de forma meramente aproximativa.

Em um contexto complexo como o da música, que envolve elementos de naturezas diversas na sua caracterização, a tecnologia pode ser uma aliada importante na automatização de certas tarefas, na possibilidade de se trabalhar no cruzamento de uma quantidade maior de informações como fontes para representação da música mas, principalmente, na possibilidade de interação com o usuário. Especialmente no que se trata de representação de emoções (muito comum nos sites de serviço *streaming*) é fundamental que se abra espaço para o *input* humano de forma a enriquecer o âmbito da informação musical.

Quando se trata da música do ponto de vista da informação, o usuário não ocupa somente o papel de “utilizador” da informação. Esse ator do processo de mediação é fundamental na definição de qual a real dimensão que a música ocupa no âmbito social e, portanto, na definição do quê e de como se deve falar sobre música, especialmente no âmbito da CI.

REFERÊNCIAS

ABRAHAMSEN, Knut Tore. Indexing of musical genres: an epistemological perspective. *Knowledge Organization*, Baden-Baden, v. 30, n. 3/4, p. 144-169, 2003. Disponível em: <https://www.nomos-elibrary.de/10.5771/0943-7444-2003-3-4-144/indexing-of-musical-genres-an-epistemological-perspective-jahrgang-30-2003-heft-3-4>. Acesso em: 29 mar. 2019.

BARROS, Camila Monteiro de; CAFÉ, Lígia; LAPLANTE, Audrey. As contribuições da autonarrativa da experiência semiótica musical para a organização do conhecimento. *Informação & Sociedade*, João Pessoa, v. 29, n. 2, 2019. No prelo.

BRÄSCHER, Marisa; CAFÉ, Lígia. Organização da Informação ou Organização do Conhecimento? In: LARA, Marilda Lopes Ginez de; SMIT, Johanna Wilhelmina (org.). *Temas de Pesquisa em Ciência da Informação no Brasil*. São Paulo: Escola de Comunicação e Artes/USP, 2010. p. 87-103.

CAFÉ, Lígia; BARROS, Camila Monteiro de. Informação musical: aspectos de indexação. In: ARAÚJO JUNIOR, Rogério Henrique; BAPTISTA, Dulce Maria (org.) *Organização da Informação: Abordagens e Práticas*. Brasília: Thesaurus, 2015. p. 69-93.

CASEY, Michael A. *et al.* Content-based music information retrieval: current directions and future challenges. In: INTERNATIONAL SOCIETY FOR MUSIC INFORMATION RETRIEVAL CONFERENCE, 2008, Plymouth (EUA). *Anais [...]*. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.148.3288>. Acesso em: 29 mar. 2019.

CINTRA, Anna Maria Marques. *Para entender as linguagens documentarias*. 2. ed. rev. e atual. São Paulo: Polis, 2002.

CORTINHAS, Jorge Fernando Pereira. *O gênero musical na representação da informação*. 2016. Trabalho de Conclusão de Curso (Graduação em Biblioteconomia) - Centro de Ciências da Educação, Universidade Federal de Santa Catarina, Florianópolis, 2016.

CRUZ, Fernando William *et al.* Um modelo para mapeamento de necessidades e usos de informação musical. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 16, n. 2, p. 207-227, abr./jun. 2011. Disponível em: <http://www.scielo.br/pdf/pci/v16n2/13.pdf>. Acesso em: 29 mar. 2019.

DOUGHERTY, William P. The quest for interpretants: toward a Peircean paradigm for musical semiotics. *Semiotica*, Berlin, v. 99, n. 1/2, p. 163-184, 1994.

FONT, Frederic; SERRÀ, Joan; SERRA, Xavier. Analysis of the impact of a tag recommendation system in a real-world folksonomy. *ACM Transactions on Intelligent Systems and Technology*, New York, v. 7, n. 1, 2015. Disponível em: <https://dl.acm.org/citation.cfm?id=2743026>. Acesso em: 29 mar. 2019.

FUJITA, Mariângela Spotti Lopes. A leitura documentária na perspectiva de suas variáveis: leitor-texto-contexto. *DataGramaZero: Revista de Ciência da Informação*, Rio de Janeiro, v. 5, n. 4, 2004. Disponível em: <http://www.brapci.inf.br/index.php/article/view/0000007547/f439490af520c6a64bbdd99c4a74f163/>. Acesso em: 20 mar. 2019.

FUJITA, Mariângela Spotti Lopes. A representação documentária no processo de indexação com o Modelo de Leitura Documentária para Textos Científicos e Livros: uma abordagem cognitiva com protocolo verbal. *PontodeAcesso*, Salvador, v. 7, n. 1, p. 42-66, 2013. Disponível em: <http://www.brapci.inf.br/index.php/article/download/32618>. Acesso em: 20 mar. 2019.

- HU, Xiao; DOWNIE, J. Stephen. Exploring mood metadata: relationships with genre, artist and usage metadata. *In: INTERNATIONAL SOCIETY FOR MUSIC INFORMATION RETRIEVAL CONFERENCE*, 8., 2007, Viena. *Anais [...]*. Viena: ISMIR, 2007. Disponível em: http://ismir2007.ismir.net/proceedings/ISMIR2007_p067_hu.pdf. Acesso em: 29 mar. 2019.
- HU, Xiao; DOWNIE, J. Stephen; EHMANN, Andreas F. Exploiting recommended usage metadata: exploratory analyses. *In: INTERNATIONAL SOCIETY FOR MUSIC INFORMATION RETRIEVAL CONFERENCE*, 7., 2006, Victoria. *Anais [...]*. Victoria: ISMIR, 2006. Disponível em: http://ismir2006.ismir.net/PAPERS/ISMIR06157_Paper.pdf. Acesso em: 29 mar. 2019.
- HURON, David. Perceptual and cognitive applications in music information retrieval. *In: INTERNATIONAL SOCIETY FOR MUSIC INFORMATION RETRIEVAL CONFERENCE*, 1., 2000, Plymouth (EUA). *Anais [...]*. [S. l], 2000. Disponível em: https://www.researchgate.net/publication/220723259_Perceptual_and_Cognitive_Applications_in_Music_Information_Retrieval. Acesso em: 29 mar. 2019.
- JANOTTI JUNIOR, Jader. À procura da batida perfeita: a importância do gênero musical para a análise da música popular massiva. *ECO-Pós*, Rio de Janeiro, v. 6, n. 2, p. 31-46, ago./dez. 2003. Disponível em: [http://www.pos.eco.ufrj.br/ojs-2.2.2/index.php?journal=revista&page=article&op=view&path\[\]=212&path\[\]=207](http://www.pos.eco.ufrj.br/ojs-2.2.2/index.php?journal=revista&page=article&op=view&path[]=212&path[]=207). Acesso em: 29 mar. 2019.
- JAPIASSÚ, Hilton; MARCONDES, Danilo. *Dicionário básico de filosofia*. 4. ed. Rio de Janeiro: Jorge Zahar, 2006.
- KOBASHI, Nair Yumiko; SANTOS, Cibele Araújo Marques. Leitura Documentária: Aspectos Pragmáticos da recepção e condensação de textos para indexar e resumir. *Revista EDICIC*, Caribe, v. 1, n. 4, p. 130-140, 2011. Disponível em: <https://bdpi.usp.br/item/002256244>. Acesso em: 20 mar. 2019.
- LANCASTER, F. Wilfrid. *Indexação e resumos: teoria e prática*. Brasília: Briquet de Lemos, 2004.
- LAPLANTE, Audrey. Social capital and music discovery: an examination of the ties through which late adolescents discover new music. *In: INTERNATIONAL SOCIETY FOR MUSIC INFORMATION RETRIEVAL CONFERENCE*, 12., 2011, Miami. *Anais [...]*. Miami: ISMIR, 2011, p. 341-346 Disponível em: <http://ismir2011.ismir.net/papers/OS5-2.pdf>. Acesso em: 29 mar. 2019.
- LAPLANTE, Audrey. Tagged at first listen: an examination of social tagging practices in a music recommender system. *Encontros Bibli*, Florianópolis, v. 20, n. esp., p. 33-54, 2015. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2015v20nesp1p33/28636>. Acesso em: 29 mar. 2019.

- LAPLANTE, Audrey. Users' relevance criteria in music retrieval in everyday life: an exploratory study . *In: INTERNATIONAL SOCIETY FOR MUSIC INFORMATION RETRIEVAL CONFERENCE*, 11., 2010, Utrecht. *Anais [...]*. Utrecht: ISMIR, 2010, p. 601-606. Disponível em: <http://ismir2010.ismir.net/proceedings/ismir2010-103.pdf>. Acesso em: 29 mar. 2019.
- LAPLANTE, Audrey; DOWNIE, Stephen. Everyday life music information-seeking behaviour of young adults. *In: INTERNATIONAL CONFERENCE ON MUSIC INFORMATION RETRIEVAL*, 7., 2006, Victoria. *Anais [...]*. Victoria: ISMIR, 2006, p. 381-382. Disponível em: http://ismir2006.ismir.net/PAPERS/ISMIR06132_Paper.pdf. Acesso em: 29 mar. 2019.
- LEE, Jin Ha. Analysis of user needs and information features in natural language queries seeking music information. *Journal of the American Society for Information Science and Technology*, New York, v. 61, n. 5, p. 1025-1045, 2010. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.21302>. Acesso em: 29 mar. 2019.
- LEE, Jin Ha; CUNNINGHAM, Sally Jo. Toward an understanding of the history and impact of user studies in music information retrieval. *Journal of Intelligent Information Systems*, Boston, v. 41, p. 499-521, 2013. Disponível em: https://idp.springer.com/authorize/casa?redirect_uri=https://link.springer.com/article/10.1007/s10844-013-0259-2&casa_token=HMmZ8EQq2nsAAAAA:gi6370X6-9zmJa5M8y2_7BRaXVejZBr3JiHDEJvDvDGPehjks-4SAzZxVG_i7z3XoBqrI3guhMrLC2yl8A. Acesso em: 29 mar. 2019.
- LEE, Jin Ha; DOWNIE, J. Stephen. Survey of music information needs, uses, and seeking behaviours: preliminary findings. *In: INTERNATIONAL SOCIETY FOR MUSIC INFORMATION RETRIEVAL CONFERENCE*, 5., 2004, Barcelona. *Anais [...]*. Barcelona: ISMIR, 2004. Disponível em: <http://ismir2004.ismir.net/proceedings/p081-page-441-paper232.pdf>. Acesso em: 29 mar. 2019.
- LEE, Jin Ha; DOWNIE, J. Stephen; CUNNINGHAM, Sally Jo. Challenges in cross-cultural/multilingual music information seeking. *In: INTERNATIONAL SOCIETY FOR MUSIC INFORMATION RETRIEVAL CONFERENCE*, 6., 2005, London. *Anais [...]*. London: ISMIR, 2005, p. 1-7. Disponível em: <http://ismir2005.ismir.net/proceedings/1100.pdf>. Acesso em: 29 mar. 2019.
- MAI, Jens-Erik. Semiotics and indexing: an analysis of the subject indexing process. *Journal of documentation*, London, v. 57, n. 5, p. 591-522, set. 2001. Disponível em: <https://www.emeraldinsight.com/doi/abs/10.1108/EUM000000007095>. Acesso em: 29 mar. 2019.
- MARTINEZ, José Luiz. *Semiosis in Hindustani music*. Delhi: Montilal Banarsidass, 2001.
- PEIRCE, Charles Sanders. *Collected Papers of Charles Sanders Peirce*. HARTSHORNE, Charles; WEISS, Paul; BURKS, ARTHUR W. (ed.). Cambridge: Harvard University Press, 1931-1958.

PEIRCE, Charles Sanders. *Escritos coligidos*. 3 ed. São Paulo: Abril Cultural, 1983.

PEIRCE, Charles Sanders. *Semiótica*. 3. ed. São Paulo: Perspectiva, 1995.

SELFRIEDGE-FIELD, Eleanor. Social cognition and melodic persistence: where metadata and content diverge. In: INTERNATIONAL SOCIETY FOR MUSIC INFORMATION RETRIEVAL CONFERENCE, 7., 2006, Victoria, 2006. *Anais* [...]. Victoria: ISMIR, 2006. Disponível em: http://ismir2006.ismir.net/PAPERS/ISMIR0625_Paper.pdf. Acesso em: 29 mar. 2019.

SHORT, Thomas Lloyd. The development of Peirce's theory of signs. In: MISAK, Cheryl. (ed.). *The Cambridge Companion to Peirce*. Cambridge: Cambridge University Press, 2004. p. 200-238.

SILVEIRA, Lauro Frederico Barbosa da. *Curso de semiótica geral*. São Paulo: QuartierLatin, 2007.

SOUSA, Brisa Pozzi de; ALMEIDA, Carlos Cândido de. Um olhar semiótico sobre o processo de indexação: a questão da representação e do referente, *Informação & Sociedade: estudos*, João Pessoa, v. 22, n. 2, p. 23-34, maio/ago. 2012. Disponível em: <http://www.ies.ufpb.br/ojs2/index.php/ies/article/view/12211/7755>. Acesso em: 29 mar. 2019.

SVENONIUS, Elaine. Access to nonbook materials: the limits of subject indexing for visual and aural languages. *Journal of the American Society for Information Science and Technology*, New York, v. 45, n. 8, p. 600-606, set. 1994. Disponível em: http://polaris.gseis.ucla.edu/gleazer/462_readings/Svenonius_1994.pdf. Acesso em: 29 mar. 2019.

ZANGERLE, Eva; PICHL, Martin. Content-based user models: modeling the many faces of musical preference. In: INTERNATIONAL SOCIETY FOR MUSIC INFORMATION RETRIEVAL CONFERENCE, 19., 2018, Paris *Anais* [...]. Innsbruck: DBIS, 2018. Disponível em: <https://dbis-informatik.uibk.ac.at/sites/default/files/2018-12/ismir18.pdf>. Acesso em: 29 mar. 2019.