



## LEITURA DOCUMENTÁRIA: ASPECTOS PRAGMÁTICOS DA RECEPÇÃO E CONDENSAÇÃO DE TEXTOS PARA INDEXAR E RESUMIR

Nair Yumiko Kobashi<sup>1</sup> e Cibele Araújo Marques dos Santos<sup>2</sup>

<sup>1</sup>Professor Doutor (Livre-Docente) – Escola de Comunicações e Artes – USP – Brasil

<sup>2</sup>Doutora em Ciência da Informação. Bibliotecária da Faculdade de Medicina – USP – Brasil

### RESUMO

Reflexões sobre Leitura documentária problematizada à luz das teorias sobre a recepção de textos, com base em conceitos derivados da Semiótica da recepção, da Estética da recepção e da Linguística textual. Nessa medida, são discutidas as relações dialéticas entre autor-texto-leitor e os problemas da interpretação de textos. Conclui-se, com base nessas teorias, que na Leitura documentária, realizada sempre sob coerções institucionais, a *intentio lectoris* requer o respeito à *intentio operis*. É o respeito a esse princípio que poderá garantir a produção de índices e resumos aderentes ao conteúdo informacional dos textos, como também a relevância na recuperação de informações.

**Palavras-Chave:** Leitura Documentária; Organização da Informação; Recepção e Condenção de Textos; Indexação; Resumos.

### ABSTRACT

Reflections on reading in documentation highlighted for theories of text reception, based on concepts derived from the Semiotics of reception, Aesthetics of reception and Textual linguistics. In this sense, the dialectical author-text-reader relationships and the problems of interpretation are discussed. We conclude, based on these theories, that reading in documentation is always intentional and performed under institutional constraints. It requires that the *intentio lectoris* respect the *intentio operis*. This principle can guarantee the production of indexes and abstracts adherent to the informational content of the texts and be helpful to retrieve precise information.

**Keywords:** Reading in Documentation; Information Organization; Text Interpretation; Indexing; Abstracting.

## 1 INTRODUÇÃO

A organização e a recuperação de informações são processos comunicacionais que envolvem diferentes atores, métodos, ferramentas e contextos. Os processos centrais da organização da informação – indexar, resumir e elaborar linguagens documentárias – são complexos, exigindo pesquisas contínuas que contribuam para o aprimoramento dos sistemas que abrigam as representações documentárias. Sob essa ótica, o campo da Organização da informação e do conhecimento, em sua vertente didático-pedagógica, tem procurado construir métodos que possam ser apropriados pelos profissionais da informação para aplicação nas tarefas de elaborar e avaliar os diferentes tipos de representações documentárias.

A organização da informação e do conhecimento é, por outro lado, processos institucionalizados que requerem “[...] a reconstrução de um complexo cenário onde sejam agregadas as populações de fontes e canais de informação, de modo a permitir processos seletivos, organizados e econômicos de busca e recuperação” (GONZÁLEZ DE GOMEZ, 1993, p.271). Impõem-se, portanto, a aderência das representações construídas à compreensão e recepção dos que consultam os repositórios informacionais.

A busca de alternativas para tornar cada vez mais potentes os sistemas de informação incluem as pesquisas de construção de índices e resumos automáticos, desde meados do século passado. As pesquisas mais sistemáticas sobre o tema foram feitas por Luhn (1958, 1959). Esse ato inaugural inspirou diferentes pesquisadores, em diferentes países, a proporem sistemas baseados em métodos matemáticos, estatísticos ou apoiados em modelos da Inteligência artificial.

O fato concreto é que esses métodos não conferem, ainda, a performance desejável aos sistemas para recuperar informação com precisão. Com efeito, os critérios estatísticos de Luhn, ou mesmo o modelo vetorial de Salton (1989), Salton; Buckley (1988), Salton e McGill (1983) são econômicos, porém, pecam pela ausência de critérios semânticos que permitam representar informação para recuperação dentro dos critérios requeridos pela demanda. No caso dos sistemas baseados em Inteligência artificial, que fundamenta a formalização das operações de representação na lógica clássica, como afirma Sowa (2000), há ainda um longo

caminho a ser percorrido para que eles se tornem efetivamente úteis.

Por outro lado, os estudos sobre o texto/discurso desenvolvidos no âmbito das Ciências da linguagem, tanto nas abordagens semiológicas quanto semióticas e psicolinguísticas, têm tido importantes reflexos nos estudos do processamento da linguagem natural. A compreensão do texto como um todo de sentido - distante portanto da concepção sintática do texto, que o considera como um artefato composto de palavras justapostas-, foi responsável pela virada pragmática que orienta as pesquisas contemporâneas sobre o processamento de textos e sua recepção.

Diversos grupos de pesquisa de processamento automático de textos têm procurado incorporar as questões textuais aos sistemas. No entanto, parece ser necessário ter compreensão teórica mais profunda das várias concepções sobre a interpretação de textos/discursos para que elas possam ser simuladas computacionalmente. Em outras palavras, a visão pragmática da textualidade parece ser imprescindível para avançar em direção à implementação da web semântica.

É este contexto que justifica a perspectiva que desenvolvemos neste trabalho, no qual têm preponderância as reflexões sobre o papel da recepção de textos na elaboração de representações documentárias. A abordagem aqui proposta fundamenta-se na ideia de que a elaboração de semelhantes representações requer a problematização das noções de texto/discurso e sua recepção. Dito de outra forma, as pesquisas sobre a elaboração de informações documentárias requerem a problematização do ato de interpretar textos.

O texto de Cintra (1987) é, ainda hoje, referência fundamental para a compreensão do ato de ler para indexar e resumir. Pode-se afirmar que esse texto promoveu, no Brasil, o desenvolvimento das pesquisas sobre a Leitura documentária. Fujita, (2006) e Kobashi (1994) são alguns dos autores que, com base na proposta de Cintra, desenvolveram suas próprias abordagens sobre a leitura com fins de indexação e elaboração de resumos documentários.

### **1.1 A Recepção de Textos/Discursos**

Na obra “O príncipe de Maquiavel e seus leitores”, Cortina (2000, p. 10) indaga: “Em que medida é possível falar de uma leitura possível/correta ou

impossível/incorreta de um texto escrito? Quem determina a possibilidade de leitura de um texto é seu autor ou o leitor? O processo de leitura é sempre igual para todo tipo de texto? Quais os motivos que podem desencadear a leitura? (p. 20).

As respostas a essas questões não são triviais, como demonstra Umberto Eco em vários textos paradigmáticos. Na obra *Os limites da interpretação*, Eco (2000) discute de forma instigante o papel do leitor, no quadro da Semiótica da recepção. Em sua opinião, o paradigma estruturalista foi um marco importante nas teorias sobre a análise de textos. Na perspectiva estruturalista, os textos “são passíveis de serem descritos através de um formalismo rigoroso” (ECO, 2000, p.1). Essa teoria aborda o texto como estrutura que pode ser abstraída dos contextos de recepção e uso, não havendo, nesse esquema, lugar para o leitor.

A partir dos anos 1960, diversos teóricos incluem as questões pragmáticas nas pesquisas sobre a leitura, em oposição aos parâmetros instituídos pelo estruturalismo linguístico. Dentre os que criticam o formalismo estruturalista pode ser citado Iser (1979), que desenvolveu suas teorias a partir de uma visão hermênutica. Este autor introduz, no contexto da Estética da recepção, a ideia de leitura como interpretação nascida da interação texto-leitor.

A Estética da recepção tem origem em um movimento de revisão dos Estudos literários, tendo, portanto, como objeto privilegiado de análise o texto ficcional. Esse ponto de partida promoveu também os estudos sobre a natureza do texto ficcional em contraposição aos outros tipos de textos “utilitários”, que Stierle (1979) denomina de texto pragmático. Nas perspectivas de Iser e Stierle, a significação, ou o valor atribuído à informação textual, nasce das relações comunicativas delimitadas por um “horizonte de expectativas”. Surge com força, portanto, a figura do leitor como parte integrante e imprescindível das relações dialéticas entre autor-obra-leitor (ISER, 1976, apud ECO, 2000).

A abordagem hermenêutica da Estética da recepção fundamenta, em larga medida, a Semiótica da recepção de Eco. De fato, pode-se afirmar que as diferenças entre interpretação de textos e uso de textos, bem como a interpretação como processo que deve considerar a tríade *Intentio auctoris, intention operis e intentio lectoris* (ECO, 2000), restabelece, nas teorias sobre a leitura, as necessárias relações dialéticas entre autor-obra-leitor, já enunciadas por Iser (1979).

Na visão de Eco (2000, p.7), um texto “pode suscitar infinitas ou indefinidas

interpretações”. No entanto, deve haver algum critério para validá-las. Nessa medida, deve-se compreender dialeticamente a relação entre a liberdade do intérprete e a natureza do texto. Segundo o autor, “a iniciativa do leitor consiste em fazer uma conjectura sobre a *intentio operis*”, estabelecendo, esta última, os limites da interpretação.

Cortina (2000) retoma de modo original a tríade intenção do autor, intenção da obra e intenção do leitor, proposta por Eco. Segundo Cortina, a leitura como procura da intenção do autor foi uma perspectiva dominante, por muitos séculos, tendo sido modificada somente a partir do início do Século XX, quando emergem novos estudos sobre o texto e a leitura. O “autor é destronado e, em seu lugar instala-se o texto. Agora, é ele o todo poderoso”. Porém, ultimamente, questiona-se o papel do leitor como o “árbitro de toda interpretação” (CORTINA, 2000, p. 37).

A partir do pragmatismo, torna-se pouco defensável a ideia de que qualquer leitura seja válida. Na perspectiva de Cortina (2000, p. 40): “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”.

Introduz-se, aqui, uma questão polêmica: há leituras certas e leituras erradas? Porém, que parâmetros podem servir para avaliar a correção ou a incorreção da leitura?

A motivação para ler, com base na tríade - *Intentio auctoris, intention operis e intetio lectoris* - de Eco (2000), pode ser uma das balizas de avaliação. A diferença entre texto ficcional e texto pragmático, de Stierle (1979), é outro parâmetro que pode ajudar a elucidar a questão. Com efeito, salvo na leitura crítica feita por especialistas, os textos ficcionais ou figurativos são lidos, via de regra, com intenções de fruição. Nesses textos predomina o aspecto poético, ou seja, importa não o que se diz, mas como aquilo é dito.

Cortina apresenta outro tipo de intenção do leitor: a leitura investigativa. O leitor do texto técnico-científico busca novos conhecimentos nesses textos, ou então, os lê com o objetivo de comparar abordagens. O texto científico é, por definição, a apresentação de uma verdade, baseada em argumentos empíricos ou discursivos. Nessa medida, a compreensão do texto requer o respeito ao próprio texto. Um texto pode ser tomado, também para falar sobre outro texto. Neste caso, não há propriamente interpretação do texto, mas este último é apenas um pretexto não

importando o seu conteúdo informacional.

As questões sobre a intencionalidade permitem introduzir a questão: Pode-se afirmar que em leitura documentária há leitura correta e leitura incorreta? Para desenvolver esta questão é necessário delimitar o contexto em a leitura se realiza.

A leitura documentária insere-se no contexto da Organização da informação. Lê-se para condensar textos e representá-los com a finalidade de inserir as representações em sistemas que serão interrogados. Isto é, o usuário interroga os sistemas de informação esperando localizar nessas memórias os documentos que possam satisfazer suas necessidades de informação. Nessa medida, parece evidente que o indexador ou resumidor deve elaborar representações condensadas aderentes à *intentio operis*.

O indexador ou organizador de informação é um leitor que interpreta dentro dos limites impostos por coerções institucionais. Dito de outro modo, a *intentio lectoris* atua no eixo da interpretação, respeitando a *intentio operis*.

O uso dos textos, por outro lado, é uma decisão do usuário. Desse modo, a ideia de que um texto possa ser indexado tendo em vista o uso que o leitor dele fará é uma meta em princípio inalcançável. Da mesma forma, é praticamente impossível adivinhar a *intentio auctoris* em face de uma obra, sendo também utópica a possibilidade de parametrizar a intenção do usuário ao indexar ou resumir textos para armazenar em um sistema de informação.

## 2 LEITURA E TIPOS DE TEXTOS

No item acima, sobre a recepção de textos, afirmamos que a interpretação nasce do equilíbrio delicado entre intenção da obra e intenção do leitor. A interpretação pode, além disso, depender do tipo de texto. Nesse sentido, Cortina (2000, p. 83) afirma que os textos podem ser classificados em mono-isotópicos e pluri-isotópicos. O texto científico, de caráter mono-isotópico, se apresenta como objeto que propõe uma verdade. O texto ficcional, ao contrário, tem compromisso com o verossímil e não com a veridicção. Ou seja, o texto científico se configura como texto que expressa à busca da verdade, colocando ênfase na veridicção. O discurso figurativo, ao contrário, opera no eixo estético, colocando-se como discurso verossímil.

Não abordaremos, neste trabalho, as possibilidades de análise do texto figurativo. Trata-se de um tipo de texto pouco afeito às possibilidades de interpretação com fins documentários. Sendo a análise documentária um conjunto de operações para identificar o conteúdo informacional de textos e condensá-los, ignoram-se, via de regra os aspectos estéticos e poéticos de um texto.

A identificação de informação em textos pode ser facilitada pelo domínio de superestruturas, que se expressam como esquemas canônicos que garantem a coerência global e temática de um texto (CINTRA, 1987). Koch e Fávero (1987), na perspectiva da Linguística textual, propõem a caracterização dos textos segundo a dimensão estrutural. Nesse sentido, classificam os textos em narrativos, descritivos, expositivos e argumentativos. Os textos argumentativos podem apresentar diversas configurações superestruturais ou esquemas canônicos (KOBASHI, 2008, p. 61-62), como segue:

Problema	Indagação
Hipótese	Conjetura
Metodologia	Observação
Resultados	Interpretação
Conclusão	Comentário final

**Figura 1: Texto Argumentativo 1.**

Tese	Ponto de vista
Argumentos	Provas
Conclusão	Confirmação do ponto de vista

**Figura 2: Texto Argumentativo 2.**

Problema	Questão que requer solução
Causas do problema	Aspectos do problema
Solução para o problema	Proposta de resposta à questão

**Figura 3: Texto Argumentativo 3.**

Os textos narrativos e descritivos podem ser esquematizados como segue:

Quem	Actantes
O quê	Acontecimento
Quando	Aspecto temporal
Onde	Aspecto espacial
Como	Modo
Por quê	Razão

**Figura 4: Texto Narrativo.**

Os textos descritivos, por sua vez, apresentam os seguintes constituintes:

Referente	Objeto do mundo
Características	Predicados/atributos físicos, subjetivos ou funcionais

**Figura 5: Texto Descritivo.**

A esquematização dos textos, apresentado nas Fig. 1 a 5, acima, responde à própria definição de textos apresentada por diferentes autores, tal como o faz Guimarães (1990, p.14-15):

[...] vê-se o texto [...] como um sistema concluído, um conjunto hierarquizado de configurações estruturais interna; de outro lado, como um objeto aberto, plural, dialogante, ligado ao contexto extra verbal. Dessa face dúplice do texto deduz-se sua significação global emergente das relações fonológicas, morfológicas, sintáticas, semânticas e pragmáticas que estão na base desse complexo sistema – a língua. Concebe-se, assim o texto ou discurso como uma “totalidade em funcionamento”.

Na concepção de Ducrot e Todorov (1972, p.87), “[...] o texto é um todo de sentido”. Ducrot complementa a definição, afirmando que um texto apresenta:

- a) a condição de progressão, já que a ausência de informação nova implica paráfrases perpétuas;
- b) a condição de coesão e coerência, expressa pela reiteração e repetição de certos conteúdos que reaparecem ao longo do texto.

A ideia de que os textos são totalidades em funcionamento, constituídos de partes solidárias que apresentam coesão e coerência, progressão e reiterações, permite apresentar um esquema de Leitura documentária baseada em estratégias. Com efeito, os segmentos da estrutura articulam-se para conferir sentido ao todo e, ao mesmo tempo, o seu fechamento. Nessa medida, cada texto, ao ser reconhecido tipologicamente, pode ser desconstruído, condensado e reconstruído como representação. Mais especificamente, a cada constituinte do texto podem ser aplicadas as regras de compreensão de textos, propostas por Van Dijk e Kintsch (1983): apagar o que é redundante, selecionar o que é pertinente, generalizar informações, substituindo termos específicos por termos genéricos e, finalmente, combinar os segmentos retidos, articulando-os em um novo texto ou em um conjunto de termos de indexação.

### 3 CONSIDERAÇÕES FINAIS

Procuramos, neste texto, apresentar reflexões sobre a Leitura documentária, recorrendo a algumas teorias sobre a recepção de textos. Nessa perspectiva, apresentamos as propostas de Eco, Cortina, Iser, Stierle e Koch e Fávero que foram discutidas criticamente, procurando-se observar sua aplicabilidade à Leitura documentária, na perspectiva da Organização da informação, mais especificamente para as finalidades de indexar e resumir.

As discussões da Estética da recepção, da Semiótica da recepção, da Linguística do texto e da Psicolinguística, apesar de sua importância para a Documentação, e também para as diversas áreas que lidam com o texto, têm ficado restritas ao campo dos estudos da linguagem. As reflexões apresentadas neste trabalho foram motivadas pela convicção de que as contribuições das áreas acima apontadas são imprescindíveis para aprimorar a produção de representações documentárias, sejam elas elaboradas por indexadores e resumidores humanos, quanto por processos automáticos. Com efeito, práticas irrefletidas têm sido responsáveis pela criação de sistemas de informação pouco eficazes. Esses dispositivos impedem que a informação circule e seja apropriada pelos distintos segmentos da sociedade.

Uma questão polêmica perpassou o trabalho: a indagação sobre os limites da interpretação. Este tema foi discutido por Eco em várias obras, tendo como centro a pergunta: existem leituras certas e leituras erradas? Essa questão é pertinente também para o campo da Documentação. Basta lembrar que ao se indexar e resumir, muitos se perguntam se a indexação e o resumo devem ser versões condensadas do original ou, ao contrário, se devem ser incorporadas a essas representações termos que respondam aos possíveis usos do documento pelo usuário. A análise e discussão das propostas semióticas, semiológicas e linguísticas apontam que não se pode confundir interpretação de textos pragmáticos, portanto mono-isotópicos, com seus possíveis usos. A tríade intenção do autor-intenção do texto-intenção do leitor, esclarece as condições e contextos da leitura e da significação e as condições em que a leitura pode ser considerada verdadeira ou pertinente.

A problematização das noções de texto/discurso e sua recepção, no âmbito das ações de informação, permitiu avançar na compreensão dos processos

comunicacionais operados pelo texto e os diferentes atores envolvidos no processo. As reflexões apontam também para a ideia de que a operacionalização da Leitura documentária é dependente de regras explícitas, baseadas em teorias e métodos rigorosos. Nesse sentido, as tipologias textuais propostas pela Linguística do texto podem ser associadas à tríade autor-texto-leitor e serem apropriadas para analisar textos e condensá-los. Da mesma forma, a concepção de Van Dijk e Kintsch sobre a compreensão de textos podem ser compreendidas como regras aplicáveis à condensação de textos para fins documentários.

Em tempos de proliferação de sistemas de informação orientados pelas filosofias de acesso aberto e autoarquivamento, parece ser necessário dedicar especial atenção à qualidade das representações documentárias. A adoção de mecanismos rigorosos, baseados em teorias, métodos e regras pode promover a construção de repositórios capazes de cumprir suas funções principais, quais sejam as de promover o acesso à informação a públicos cada vez mais amplos. Acredita-se que o domínio de metodologias adequadas de tratamento de informação (intelectual ou por máquina) será o antídoto capaz de garantir a qualidade dos repositórios e, em consequência, o fluxo social da informação.

## REFERÊNCIAS

- CINTRA, A. M. M. Estratégias de leitura em documentação. In: Grupo Temma. **Análise documentária: A análise da síntese**. Brasília: IBICT, 1987. p.27-35
- CORTINA, A. **O Príncipe de Maquiavel e seus leitores: uma investigação sobre o processo de leitura**. São Paulo: Editora Unesp, 2000.
- ECO, U. **Os limites da interpretação**. São Paulo: Perspectiva, 2000.
- ESCARPIT, R. **Théorie générale de l' information et de la communication**. Paris: Hachette, 1976.
- FUJITA, M. S. L.; RUBI, M. P. Um modelo de leitura documentária para a indexação de artigos científicos: princípios de elaboração e uso para a formação de indexadores. **Datagramazero**, Rio de Janeiro, v.7, n.3, jun. 2006.
- GONZÁLEZ DE GÓMEZ, M. N. A representação do conhecimento e o conhecimento da representação: algumas questões epistemológicas. **Ciência da Informação**, Brasília, v.22, n.3, p.271-222, set./dez. 1993.
- GUIMARÃES, E. **A articulação do texto**: São Paulo: Ática, 1993.
- ISER, W. A interação do texto com o leitor. In: LIMA, L. C. (Org.). **A literatura e o leitor: textos de estética da recepção**. Rio de Janeiro: Paz e Terra, 1979. p.83-132
- Revista EDICIC, v.1, n.4, p.130-140, Oct./Dic. 2011. Disponible en: <<http://www.edicic.org/revista/>>.

KOBASHI, N. Y. **A elaboração de informações documentárias: em busca de uma metodologia**. São Paulo: USP, 1994. Tese (Doutorado) – Programa de Pós-Graduação em Ciências da Comunicação - Escola de Comunicações e Artes - Universidade de São Paulo.

\_\_\_\_\_. Linguística textual e elaboração de informações documentárias. In: GASPAR, N. R; ROMÃO, L. M. S. (Orgs.). **Discurso e texto: multiplicidade de sentidos na Ciência da Informação**. São Carlos: EDUFScar, 2008.

KOCH, I. G. V; FÁVERO, L. L. Contribuição a uma tipologia textual. **Letras e letras**, v.3, n.1, p.3-10, jun. 1987.

LUHN, H. P. Keyword-in-context for technical literature. (KWICK Index). New York: **ASDD Report** . Yorktown Heights: IBM Advanced system Development Division, 1959.

LUHN, H. P. The automatic creation of literature abstracts. **IBM Journal of Research and Development**, v.2, p.159-165, 1958.

SALTON, G. **Automatic text processing: The transformation, analysis, and retrieval of information by computer**. New York: Addison-Wesley, 1989.

SALTON, G.; BUCKLEY, C. J. Term-Weighting approaches in automatic text retrieval. **Information Processing and Management**, v.24, n.5, p.513-523, 1988.

SALTON, G.; MCGILL, M. J. **Introduction to Modern Information**. New York: McGraw-Hill, 1983.

SOWA, J. **Knowledge representation: logical, philosophical, and computational foundations**. Pacific Grove: Brooks/Cole, 2000.

STIERLE, K. Que significa a recepção dos textos ficcionais. In: LIMA, L. C. (Org.). **A literatura e o leitor: textos de estética da recepção**. Rio de Janeiro: Paz e Terra, 1979. p.133-187

VAN DIJK, K. **Strategies of discourse comprehension**. Orlando: Academic Press, 1983.