

# ANÁLISE DOCUMENTÁRIA E REPRESENTAÇÃO DA INFORMAÇÃO

NAIR YUMIKO KOBASHI

Davi Bertholdo, Lígia Arcanjo, Júlia Lopes, Milene  
Lee e Mirna Felix

Biblioteconomia – 1º Semestre Matutino



# INTRODUÇÃO

- **Teoria, Organização e Representação da Informação:**
  - ❖ *Documentos Mesopotâmicos;*
  - ❖ *Biblioteca de Alexandria.*
- **Documentação (Sistema Público de Informação):**
  - ❖ *Século XVII: A.F.C. (Resumo dos Trabalhos Científicos, Filosóficos, Artísticos, etc.);*
  - ❖ *Séculos XVIII e XIX: Semelhança ao Século XVII;*
  - ❖ *Atualidade: Base de Dados e Tecnologia Digital.*
- **Indexação e Elaboração de Resumos:** Modo **automático inexistente** pelo método de **exclusão** (Borko).



# A INTERFACE LINGUÍSTICA / ANÁLISE DOCUMENTÁRIA

- **Modelo de Luhn: Tesauros** (Valor Comunicacional e Interação Sistema/Usuário).
- **Análise Documentária: Base da Ciência da Informação**. Passagem de textos originais para a linguagem documentária, divididas em:
  - ❖ **Léxico;**
  - ❖ **Eixo Paradigmático;**
  - ❖ **Eixo Sintagmático;**
- **Linguística:** Contribuição dos **problemas de representação documentária** e dos **usuários** desse sistema.
- **Interface:** **Correlacionar** as **práticas empíricas** a fim de construir um **vocabulário documentário**.



# A INTERFACE ANÁLISE DOCUMENTÁRIA / INTELIGÊNCIA ARTIFICIAL (IA)

- **Sistemas Especialistas:** Orientação em **Bases de Catalogação**. **Benefícios** em **resumos** e **compreensão** de **textos jornalísticos, narrativas e mensagens de telex**.
  - ❖ **TOPIC:** Sistema de **recuperação de informações**. **Elaboração de resumos** dos mais variados. **Vocabulário controlado**.
- **Construção de Bancos de Dados:** Baseado em modelos enciclopédicos dos anos 60 e 70:
  - ❖ **INIST:** Construção de produtos documentários de maior valor agregado (*uso da linguística e estatística na recuperação de informações*).

# A INTERFACE ANÁLISE DOCUMENTÁRIA / INTELIGÊNCIA ARTIFICIAL (IA)

- Sistema de Juan Barceló:
  - ❖ *Conceitos relacionados a um problema;*
  - ❖ *Representação “ativa” (usuário/perguntas);*
  - ❖ *Conjunto de Regras;*
  - ❖ *Operação Lógica;*
  - **3 Meta-Conhecimentos:**
    - a) *Problema a ser resolvido;*
    - b) *Estrutura do Sistema;*
    - c) *Estratégia para resolver o problema.*

# A ORGANIZAÇÃO E A REPRESENTAÇÃO DE INFORMAÇÕES DOCUMENTARIAS

- Pesquisas sobre **organização e representação de informações documentárias**:
  - ❖ A **documentação** opera **sempre** com **grandes quantidades de textos**;
  - ❖ O **valor do produto documentário** está geralmente relacionado a **atualidade da informação**.
- Estes aspectos exigem a **adoção de princípios e mecanismos** que garantam **resultados** simultaneamente **rápidos e potentes**, aptos a produzir informações documentárias que observem os **parâmetros de qualidade e pertinência** exigidos pela demanda.



# A ORGANIZAÇÃO E A REPRESENTAÇÃO DE INFORMAÇÕES DOCUMENTARIAS

- A **elaboração de informações documentárias**, compreende em três fases:
  - ❖ *Ler um texto;*
  - ❖ *Selecionar o conteúdo informacional considerado **pertinente** para usos determinados;*
  - ❖ *Representar, ou seja, dar forma às informações selecionadas, de modo a torná-las **documentariamente manipuláveis**.*
- **Elaboração de Resumos:** Ato de **reduzir** textos.
- **Indexação:** Estabelecer **equivalência entre palavras** de uma Linguagem Documentária.
- **Todo ato documentário é um ato de comunicação** que tem a finalidade de **promover a circulação da informação**. Deve ter uma **visão clara dos objetos institucionais** como o **tipo de usuários, o de demandas e a estrutura organizacional**.



# A ORGANIZAÇÃO E A REPRESENTAÇÃO DE INFORMAÇÕES DOCUMENTARIAS

- A **análise documentária** *manipula e transforma* textos em dois tipos básicos de representação: o **resumo** e o **índice**.
  - ❖ No **resumo**, o **texto base** é **desestruturado** de modo a **permitir** que certos fragmentos, organizados sob forma de um novo texto, **reconstruam potencialmente o sentido original**.
  - ❖ Na **indexação**, procura-se obter um **grau ainda maior de compactação** do texto-base: do **processo de segmentação** resultam **fragmentos** que procuram **caracterizá-lo** por meio de **palavras ou sintagmas**.
- **A indexação e o resumo atuam de forma complementar.**
- O **processo global** de Análise Documentária, defronta-se com pelo menos dois objetos:
  - ❖ **O texto**, como sua unidade de análise
  - ❖ **A linguagem documentária** como instrumento comutador.

# A REPRESENTAÇÃO DOCUMENTÁRIA

- A palavra “**representação**” é utilizada para **evitar equívocos** que a polissemia pode causar.
  - ❖ **Senso Comum:** Reprodução daquilo que se pensa. **Representar é o ato de reproduzir.**
  - ❖ **Ciência da Comunicação:** Conceito **mediador** entre o **emissor e o receptor.**
  - ❖ **Documentação:** Conceito associado a **descrição de aspectos** que identifiquem materialmente os **documentos** e ao produto da **condensação de conteúdos de textos**, ou seja, **a indexação e a elaboração de resumos.**
- A **informação documentária** é uma **representação condensada** construída a partir de um objeto efetivamente presente – **o documento.**
  - ❖ **O texto é uma forma de representação simbólica.**

# A SIGNIFICAÇÃO NOS PROCESSOS DOCUMENTÁRIOS

- **Documentação + Ciência das Linguagens:**
  - ❖ **Semântica (Processo de Significação:** Conjunto semântico para a construção da coleta de dados);
  - ❖ **Semiologia;**
  - ❖ **Linguística.**
- **Décadas de 60/70:** Linguística + Semiologia decaem no sentido de “ter pouco a oferecer à Documentação” (Gardin);
- **Reaproximação das práticas (Linguagens científicas e Análise documentária) por meio da:**
  - ❖ **Análise do Discurso;**
  - ❖ **Linguística Textual;**
  - ❖ **Interseção com a Terminologia.**

# A SIGNIFICAÇÃO NOS PROCESSOS DOCUMENTÁRIOS

- **Semântica do Código:** “Discute as *relações* de significação entre as *unidades constitutivas das linguagens documentárias*”.
- **Semântica Textual:** Apresenta conceitos fundamentais para a utilização das paráfrases que caracterizam os resumos.
- **Três unidades que constituem as linguagens documentárias:**
  - ❖ *Linguagem de Especialidade;*
  - ❖ *Linguagem Natural;*
  - ❖ *Terminologia de Área.*

# A SIGNIFICAÇÃO NOS PROCESSOS DOCUMENTÁRIOS

- As **linguagens documentárias** não substituem nenhuma das três linguagens (todas possuem peculiaridades), mas são como uma linguagem “**intermediária**” que tem como **função** desempenhar tarefas específicas de **tratar e recuperar informação**.
- **Análise da Significação:** Feita no **âmbito do código** (convencionalidade que visa representar e transmitir informação) possibilitando a **caracterização** da **linguagem documentária**.
- Estudo do teor geral do texto, significado final: Semianálise (KRISTEVA).
  - ❖ **Granger:** Distinção entre “**sentido e significação**”.

# O TEXTO COMO OBJETO: ALGUMAS ABORDAGENS

- Compreende-se a **Análise Documentária** como uma disciplina de **natureza metodológica**, prezando o **tratamento do conteúdo do texto**, podendo assim, ser comparada com outras ciências que também operam com semelhante objeto:
  - ❖ *Análise de conteúdo;*
  - ❖ *Análise de discurso;*
  - ❖ *Linguística do texto;*
  - ❖ *Análise Logicista;*
  - ❖ *Análise de Textos segundo a Inteligência Artificial.*



# ANÁLISE DE CONTEÚDO

- "Conjunto de procedimentos de análise e interpretação de textos de natureza diversa [...] baseados na descrição e quantificação de palavras, temas ou frases desses textos (GARDIN, J.C 198. p. 54-55, BARDIN, L. 1988. p. 31)".
- **PROCEDIMENTO DE ANÁLISE DE CONTEÚDO:**
  - ❖ *Formulação de dados;*
  - ❖ *Redução de dados;*
  - ❖ *Inferência;*
  - ❖ *Análise.*
- Dados que são tratados em moldes formais para que seja possível a sua manipulação.
- "Toda linguagem de dados é um sistema intermediário entre o mundo dos fenômenos reais e o dos fatos científicos..." (KRIPPENDORF, K. 1990.p. 124).

# ANÁLISE DE CONTEÚDO

## PROPRIEDADES FUNDAMENTAIS DOS ELEMENTOS CONSTITUTIVOS:

- ❖ *Exaustividade;*
- ❖ *Mutuamente excludentes.*
- "A linguagem de dados, é, de fato, um código que dá forma ao universo objeto de análise. Enquanto Código, é um recurso destinado a aumentar o rendimento informativo de uma mensagem" e, como tal, deve apresentar "um conjunto de signos e suas regras de utilização ". (COELHO NETTO, J.T. 1980. p. 140)

## PROBLEMAS:

- ❖ *"Não define, com o devido rigor, as categorias utilizadas para descrever os dados coletados no texto";*
- ❖ *Falta de uniformidade dos dados.*
- **DIFERENÇAS ENTRE ANÁLISE DE CONTEÚDO E ANÁLISE DOCUMENTÁRIA**



# A ANÁLISE DO DISCURSO E A LINGUÍSTICA DO TEXTO

## Pluridisciplinar do Texto e do Discurso

- Discurso, definido como o lugar da subjetividade, a partir da oposição língua/fala; não se forma como um objeto observável, no interior da **teoria saussureana**.
- **Michel Pêcheux**, fundou a linha da análise de discurso; "apreender como no linguístico se inscrevem as condições sócio históricas de produção." (BRANDÃO, H. N. N. 1991. p. 83)
- Distinguir a Teoria do Texto (Linguística), da Análise do Discurso, a partir da descrição das tarefas que cada uma delas se atribui.
- No caso da **Linguística do Texto**:
  - ❖ **Determinar o texto;**
  - ❖ **Diferenciação das espécies dos textos.**



# A ANÁLISE DO DISCURSO E A LINGUÍSTICA DO TEXTO

- Já a **Análise do Discurso**, segundo Diana Barros:
  - ❖ A **relação do discurso** com sua enunciação, condições de produção e de recepção;
  - ❖ **Discurso** como o lugar, ao mesmo tempo, do social e do individual;
  - ❖ **Articulação** entre narrativa e discurso construído sobre estruturas narrativas que o sustentam.
- Os dois níveis de análise (**externo e interno**), propostos pela Análise do Discurso; **Não são aplicáveis aos objetivos da Análise Documentária.**
- **Enunciação e Sociolinguística:**
  - ❖ *Diferenciar as várias espécies de textos.*



# A ANÁLISE DO DISCURSO E A LINGUÍSTICA DO TEXTO

## Autonomia e seu Fechamento

- O quadro conceitual do texto é classificado em duas vertentes:
  - ❖ *Natureza estritamente linguística, centradas em **fenômenos "frásticos"**;*
  - ❖ *Pragmática textual, que procura captar, **a dimensão global da textualidade**, e não como a reunião de proposições.*
- Segundo Oswald Ducrot (1972. p. 87), todo texto apresenta simultaneamente:
  - ❖ A **condição de progressão**, já que a ausência de informação nova implica paráfrases perpétuas;
  - ❖ A condição de **coesão e de coerência**, expressas pela **reiteração e repetição** de certos conteúdos que reaparecem regularmente ao longo do texto;
  - ❖ A ideia de que esse texto poderá ser condensado, **suprimindo-se as redundâncias**.



# A ANÁLISE DO DISCURSO E A LINGUÍSTICA DO TEXTO

## O Conceito de Tipologia Textual

- A **superestrutura** designa uma estrutura convencional. Enquanto **paradigma** de organização textual, o esquema ou superestrutura fornece uma base para a interpretação do texto. O **texto-base** é resultante do produto entre conhecimento do esquema (superestrutura) e informação textual.
- Para Kintsch e Van Dijk, a compreensão do discurso é de **natureza estratégica**. Sendo parte de nosso conhecimento geral e precisam ser aprendidas antes de serem automatizadas.

## O que distingue um bom de um mau leitor?

- O **bom leitor**, segundo esses autores, é aquele capaz de **explorar o texto de forma mais direta**, utilizam melhor os **processos top-down**. Estratégias de processamento de informação e ordenação do conhecimento.
- Na prática, este esquema é rápido e possui consistência à coleta e organização de informações para a elaboração de resumos e índices.



# A ANÁLISE LOGICISTA COMO OBJETO DE DISCUSSÃO

- A **Análise Logicista**, proposta por *Jean-Claude Gardin*, é uma metodologia de análise de textos científicos produzidos pelas ciências ditas humanas.
- Os textos são **objetos construídos**, delimitados necessariamente pelos **pontos de partida e de chegada**, sendo os **raciocínios (inferenciais e/ou dedutivos)** "esquemáticos" sob a forma de proposições **descritivas e interpretativas**. A esquematização é, a representação condensada do objeto analisado, somente o essencial. Por meio de um **sistema simbólico**.
- Certos **aspectos** são **análogos aos processos documentários** de condensação e representação, mediados por um **código comutador**. Nos textos visando à sua condensação, sem se propor a avaliar a consistência ou validade das interpretações apresentadas no texto científico.
- A **Análise Documentária** justifica-se, por sua capacidade de **acompanhar a escala de produção de textos científicos e técnicos do mundo contemporâneo**.
- Suas **técnicas específicas** de manipulação de textos **não comportam**, pois, as operações comprometem a velocidade da análise para a produção de representações.



# A ANÁLISE LOGICISTA

- **Textos** são objetos construídos, ***delimitados pelos pontos de partida e chegada***, de ***raciocínios "esquemalizáveis"*** em proposições ***criativas*** e ***interpretativas***.
- ❖ ***Esquematização:*** *Representação condensada do objeto de análise, na qual apenas o essencial é mantido.*

# A ANÁLISE DE TEXTO NA PERSPECTIVA DA INTELIGÊNCIA ARTIFICIAL (IA)

- **Inteligência Artificial (IA):** Simulação computacional do comportamento inteligente.
- **Campo de pesquisa nascido nos anos 50**, porém as **ideias básicas** já existiam desde **1936**, com **Turing**.
- **Hipótese Fundamental da I.A.:** Processos de pensamento são mecanizáveis.

# LINGUAGEM NATURAL COMO CAMPO DE PESQUISA

- 1 – O que significa compreender a linguagem?
- 2 – Como um programa poderia reconhecer algo de sentido ambíguo ou sem sentido algum?
- 3 – Como um programa poderia distinguir entre metáfora e metonímia? Ou interpretar sentidos novos?



# UTILIDADES ESPERADAS – RESULTADOS

## UTILIDADES ESPERADAS

## RESULTADOS

### ➤ Interface de Linguagem de Texto e Análise do Discurso:

❖ *Tradução Automática;*

❖ *Elaboração de Resumos;*

❖ *Indexação;*

❖ *Ensino Assistido por Computador.*

### ➤ Estratégias de aprendizagem e modificação de comportamento:

❖ *Modificar estratégias de busca ou reindexar documentos em face a novos dados.*

### ➤ Catalogação

❖ *Regras Explícitas:* Sistemas automáticos.

### ➤ Análise e Indexação

❖ *Procedimentos Pouco Formalizados:* Problemas de automação.

