

ELEMENTOS DE LINGÜÍSTICA PARA ESTUDOS DE INDEXAÇÃO

Anna Maria Marques Cintra
IBICT

A operação denominada indexação é definida como a tradução de um documento em **termos documentários**, isto é, em descritores, cabeçalhos de assunto, termos-chave, que têm por função expressar o conteúdo do documento. A indexação assim definida é, pois; uma "tradução lexical" das unidades lexicais da língua em que está escrito o documento, para unidades lexicais de uma linguagem documentária. Em geral, as linguagens documentárias possuem uma gramática que corresponde a um conjunto de regras ou instruções (relações booleanas, indicadores de funções, etc.) e a indexação se faz levando em conta essas regras que expressam laços sintáticos e semânticos entre os termos.

Essas considerações evidenciam alguns aspectos da Lingüística Geral que merecem ser tratados num trabalho de lingüística aplicada à indexação.

É comumente admitido o interesse da Lingüística como área de apoio aos trabalhos de documentação, de modo especial aos procedimentos de indexação. Entretanto, não costuma ser claro o recorte que se deve fazer na teoria, a fim de atender ao interesse documentário.

Nossa preocupação nesse trabalho é tentar, didaticamente, localizar, dentro do âmbito da Lingüística Geral, aspectos auxiliares à indexação em língua portuguesa.

1 - LEITURA DO DOCUMENTO

Todo e qualquer trabalho de indexação supõe uma

RESUMO

Na conversão da linguagem natural para a linguagem documentária, intervêm vários fenômenos lingüísticos. Para um estudo conceitual desses fenômenos, procedeu-se a uma seleção de tópicos: processo de leitura e problemas relativos à semântica e à sintaxe.

Descritores: Linguagem natural; Linguagem documentária; Indexação; Semântica lexical; Leitura; Sintaxe.

leitura do documento e essa leitura tanto pode ser processada por cérebro humano, quanto por máquina.

Sabe-se que os custos da indexação automática ainda tornam o processo quase inviável. Frequentemente o documento é lido por cérebro humano, processado e só então entra em máquina.

Dada a rapidez com que um documento deve ser processado, não é admissível leitura lenta, nem mesmo do documento integral. Muitas vezes o indexador limita-se ao título, ao resumo e aos parágrafos iniciais e finais do documento, na suposição de que essas partes contenham as informações básicas para a sua "tradução" em termos documentários.

Seja numa leitura parcial, seja numa leitura global do texto, o fato é que há um trabalho de identificação de descritores ou palavras-chave que representem o documento.

Sabe-se, por pesquisas já realizadas*, que a leitura proficiente não se processa palavra por palavra, mas como um todo reconhecido instantaneamente, seja a nível de palavras, seja a nível de blocos ou segmentos maiores de informação.

* KATO (1982) refere-se a pesquisas nesta área e amplia a análise a partir de Smith (1978).

Para um trabalho de indexação, a primeira questão que se apresenta é a da própria proficiência. Será proficiente a leitura que faz o indexador, mesmo limitando-se a partes do documento?

A nosso ver, sim. Na medida em que o título é bem atribuído, o resumo representativo e os parágrafos de introdução e conclusão colocados de forma adequada, o indexador terá oportunidade de selecionar palavras-chave indicadoras do conteúdo do documento. Falhas nessas partes do texto poderão acarretar defeitos de indexação. No entanto, não afetam a leitura.

Há, pois, dois problemas envolvidos: o da representatividade efetiva das partes selecionadas em relação ao documento e o do processo de leitura que faz o indexador.

Não vamos discutir a representatividade das partes selecionadas. Elas costumam ser aceitas, até porque a imposição da rapidez num centro de informação não permitiria esse tipo de questionamento.

Quanto ao processo de leitura, entendemos que as maneiras de ler apontadas por Kato¹ estão, de certa forma, presentes no procedimento de indexação, como tentaremos apresentar.*

A leitura feita pelo indexador objetiva selecionar palavras-chave ou descritores, que traduzam o documento numa forma compatível com uma dada linguagem documentária, de sorte a possibilitar a sua recuperação.

Basicamente concorrem dois procedimentos para a apreensão dos descritores: a apreensão instantânea das unidades de informação e a apreensão por análise e síntese.

Determinadas unidades, porque fazem parte do glossário visual do indexador, são submetidas ao

* Observando a leitura de um texto novo, feita por um grupo de alunos de mestrado em Ciência da Informação, verificamos que ela se processa da macro estrutura do texto, para estrutura menor em níveis diferentes, mas não hierárquicos. A leitura integral do texto foi logo substituída por leituras parciais, especialmente das partes tidas como relevantes: introdução e conclusão. A busca dos termos representativos se consubstanciou na organização de unidades selecionadas para a elaboração de resumos do tipo indicativo e do tipo informativo. O procedimento de leitura não variou para um ou outro tipo, o que se explica pelo fato de o resumo indicativo corresponder à topicalização do texto, podendo os mesmos tópicos serem colocados em linguagem concatenada no resumo informativo.

processo de seleção sem dificuldades, independentemente de análise e síntese. São unidades já convencionalizadas pelo vocabulário específico da área que permitem a ativação da memória permanente do leitor.

Naturalmente, a prática do indexador numa determinada área do conhecimento amplia seu glossário visual e torna o processo de indexação muito mais rápido.

Outras unidades, entretanto, embora apreendidas em bloco, exigem análise e síntese. Ocorre, então, a ativação da memória temporária do leitor. São unidades desconhecidas que necessitam, para sua interpretação, da aplicação de regras lexicais, de um esforço a mais.

A rigor, esses dois tipos de apreensão deveriam cobrir o procedimento de leitura a nível de identificação e interpretação de blocos de informação. No entanto, uma terceira possibilidade deve ser considerada. Trata-se da possibilidade de ocorrerem seqüências de certa forma desordenadas para o leitor e por isso mesmo exigirem leitura palavra por palavra. Neste caso, a apreensão dos descritores é dificultada, ou até mesmo impedida, a menos que o leitor consiga superar esse estágio e atingir o da interpretação das unidades.

Um dos aspectos tidos como indicador das palavras-chave é a freqüência com que determinadas unidades ocorrem nas partes lidas. A leitura de uma unidade recorrente, tanto pode se dar por análise e síntese na primeira ocorrência e depois se processar por reconhecimento instantâneo, como através da memória temporária. Neste tipo de leitura dirigida para a execução de tarefa profissional, admitimos que a recorrência que normalmente ultrapassa o nível da sentença, chegando, por vezes, ao texto, deva ser restrita a blocos de informação. E neste caso a recorrência que em linguagem comum se identifica com "freqüência", tanto pode ser atribuída à memória temporária, quanto à permanente.

De qualquer forma, as unidades recorrentes permitem um reconhecimento instantâneo porque:

- o leitor já interpretou o bloco anteriormente, quando a unidade apareceu pela primeira vez;
- o conteúdo semântico da unidade estava no "estado de consciência" do leitor e a primeira ocorrência promoveu a interpretação.

Em qualquer das duas situações há a intervenção do

que Kato¹ chama de memória rasa; dá conta de algo já introduzido no texto e que mantém, dentro dele, o "status" de informação velha.

A atuação das três memórias justifica, a nosso ver, a especialização de indexadores por áreas de assuntos, não só por questões ligadas à rapidez do processo, como também por aquelas que marcam a sua eficiência.

Logo a leitura por cérebro humano será tanto melhor quando maior for o número de unidades de informação que possam ser processadas pela memória permanente do leitor.

2 - CONCEITOS GERAIS

2.1 Linguagem e língua

Embora os problemas de tradução do inglês e mesmo do francês cheguem, por vezes a dificultar a distinção entre linguagem e língua, em português podemos dispor de significados bastante distintos para esses termos.

Enquanto linguagem nomeia uma faculdade natural, língua refere-se a um caso particular de linguagem.

A linguagem é uma representação simbólica que expressa uma função psico-social complexa. Corresponde a uma manifestação intelectual e multiforme dos seres, que recobre inúmeras formas de significar: linguagem verbal (oral e escrita), a pictórica, a musical, a cinética, a mímica, a documentária etc.

A linguagem verbal, embora seja um caso particular, constitui, na verdade um sistema de signos de espectro tão amplo, que todos os outros sistemas de linguagem podem se repassar de língua. Daí, porque, freqüentemente, o termo linguagem seja usado por língua, ou a expressão linguagem natural, que normalmente se opõe a linguagem artificial ou controlada, seja usada também em lugar de linguagem verbal.

Antes de falar tal ou qual língua, o homem é um ser apto para estabelecer relações e é essa natureza relacional que está na base da linguagem, concretização da sua capacidade lógico-simbólica.

A linguagem é, pois, um sistema, uma organização relacional, onde cada elemento existe, na medida mesma em que se relaciona a outro ou a outros do mesmo conjunto. A linguagem documentária é pois linguagem, exatamente porque concretiza a capacidade simbólica do homem, através da

organização de seus termos e regras em sistema próprio.

No conjunto das linguagens, a linguagem documentária decorre das dificuldades que a linguagem natural oferece para operar com a descrição de documentos. De modo especial são entaves a polissemia, a sinonímia, a homonímia, a antonímia, os modos e expressões de relações complexas.

De fato, a linguagem verbal possui um caráter muito peculiar e complexo, enquanto que para uma estratégia de busca, todo o esforço é feito no sentido de representar o documento através de processos lógicos, nem sempre encontrados na linguagem natural.

Sob vários ângulos pode ser observada uma língua. Daí surgirem diferentes disciplinas Lingüísticas, dentre as quais destacamos as que decorrem dos três componentes básicos que integram uma dada língua: a fonologia, a sintaxe e a semântica.

Para estudos de indexação importam de modo especial a semântica, disciplina que se ocupa do sentido ou da significação dos elementos*, a sintaxe, disciplina que se ocupa das relações que se estabelecem a partir da organização sintagmática dos elementos e, de certo modo, a morfologia que, sintetizando parcialmente aspectos da semântica e da sintaxe, se encarrega da identificação das partes da palavra e de suas condições de ocorrência.

2.2 "Langue" e "parole"

Considerando que a língua é em si um sistema uno, a observação particular de suas partes e dos elementos que a organizam, constitui um esforço metodológico em busca de aprofundamento do todo.

A análise dos dados concretos, dadas as variações individuais, levaria a uma dispersão tal que dificilmente se chegaria a conferir à língua o atributo de objeto de uma ciência — a Lingüística, como se procedeu a partir da publicação póstuma do Cours de Linguistique Générale de Ferdinand de Saussure, em 1916. Basta observar as variações de sons entre pessoas e até numa mesma pessoa, as possibilidades de escolha lexical e as diversas atribuições de sentido que se pode dar às palavras.

* A Lexicologia está sendo tomada como disciplina que recobre parte da semântica e da sintaxe, já que trata da natureza e organização do vocabulário de uma língua.

De fato, a Lingüística busca traços invariantes que permitam estabelecer modelos. E a oposição Saussureana "langue/parole" persegue essa análise de invariantes.

"Langue", objeto específico da Lingüística, é a abstração que sustenta o processo de comunicação humana através de signos verbais, isto é, sinais que, partilhando do sistema, se submetem a um conjunto de regras (fonológicas, morfológicas, sintáticas e semânticas) que tornam, até certo ponto, indiferente a matéria dos sinais utilizados.

Já a "parole" diz respeito à realização concreta, aos enunciados produzidos, às manifestações individuais, nas quais ficam evidenciadas, inclusive, as expressões de criatividade do falante.

2.3 Signo lingüístico

Numa tentativa de síntese do fenômeno língua, podemos dizer que os atos concretos da fala demonstram que a linguagem verbal consiste num conjunto de signos verbais que se organizam conforme regras combinatórias.

O signo lingüístico, na concepção saussureana é entidade da "langue", constituído de um conceito (significado) e uma imagem acústica (significante). O conceito é coisa psíquica, abstrata, a imagem acústica é marca psíquica do som (não é o som material).

Na concepção saussureana, dois princípios básicos ao nível do signo orientam a organização das línguas; o da arbitrariedade do signo e o do caráter linear do significante.

A chamada arbitrariedade do signo responde, como veremos rapidamente, por problemas semânticos a nível lexical como a polissemia, a sinonímia, as relações textuais, por oposição à fixação dos significados nas linguagens documentárias.

O princípio da arbitrariedade, na sua formulação inicial, defendia que a seleção dos sons que compõem um signo não é imposta por nenhum traço do sentido, isto é, o sentido de *árvore*, por exemplo, não nos é imposto pela forma/*árvore*.

Posteriormente, Benveniste (1966)² demonstrou que o arbitrário não é a relação entre o significado e o significante, enquanto expressão, mas entre o significante e a "coisa" significada, ou seja entre o som de *árvore* e o objeto *árvore*.

Esta velha discussão que se originou na Grécia antiga, levou os lingüistas a admitirem que o atributo de arbitrariedade para a associação de um conceito a um significante, no mínimo não seria adequada, uma vez que ela era fruto de um conhecimento coletivo, numa dada língua. Seria, pois, mais apropriado dizer que trata-se de uma relação imotivada, já que se estabelece sob a forma de um contrato social que não pode ser arbitrariamente alterado pelo indivíduo. Assim, o signo lingüístico é arbitrário a priori, mas não a posteriori.

À primeira vista, esse princípio poderia levar a pensar na possibilidade de fixação da relação imotivada entre significado e significante, numa dada língua. Entretanto, tendo o signo uma existência apenas relacional e decorrendo o significado de dupla relação: significante/significado e signo/contexto, é a própria imotivação que explica toda a gama de significados para uma mesma seqüência de unidades do significante, nascendo daí, basicamente, a polissemia e a sinonímia.

Nas linguagens documentárias, os signos são também arbitrários a priori, na medida em que são utilizados significados da linguagem verbal e fixados, ou arbitrados seus significados no sistema construído.

O princípio da linearidade do significante manifesta-se na realização temporal do signo lingüístico. A observação nos mostra que um signo lingüístico qualquer, ao ser expresso, exige uma seqüência linear de produção que se efetua no tempo pela realização em língua oral de um som após o outro e em língua escrita, da esquerda para a direita, letra a letra.

Nas linguagens documentárias, o princípio da linearidade é mantido, até porque elas se valem, basicamente, de signos verbais que se transformam em signos documentários.

2.4 Léxico, vocabulário, palavra e termo

Sendo o signo unidade da "langue", julgaram os lingüistas mais adequado analisar essa mesma unidade em diferentes planos de ocorrência. Desta forma, o signo é, num nível mais abstrato de consideração, a unidade disponível no sistema e na própria consciência do falante. Compõem os signos, portanto, o Léxico da língua, enquanto sistema virtual e não quantificável, composto das unidades que alimentam o vocabulário ativo e passivo do indivíduo.

No momento em que uma unidade é atualizada,

tem-se um vocábulo. O vocabulário, então, pode ser analisado quantitativa e qualitativamente, enquanto ato realizado.

Palavra, por sua vez, é unidade de difícil conceituação, embora largamente usada na linguagem comum e por isso mesmo detentora de um significado ligado à organização gráfica da língua.

Pottier³ considera "palavra" a unidade formal, composta de morfemas e definida no quadro de uma língua determinada. Outros autores tomam palavra como unidade de texto.

De qualquer forma, o que se observa é que, na prática, lidamos com palavra e no trabalho de indexação, de modo especial, em que se opera com língua escrita, o termo "palavra" é largamente usado.

Para a organização de vocabulários, concorrem dois tipos de unidades: as que contêm o significado lexical, ou seja, aquelas que expressam o "suporte de conceito" do mundo bio-social, e as que contêm significado gramatical, por isso mesmo denominadas "indicadores de função". Essas unidades, respectivamente nomeadas morfema lexical e morfema gramatical, tanto podem coincidir com palavras da língua, quanto com unidades morfológicas mínimas, como mostram os exemplos:

morfemas lexicais	morfemas gramaticais
mar	que
leit-	-eria

Na verdade, não há entre lingüistas grande preocupação em conceituar **palavra** e muito menos **termo**, porquanto são denominações fortemente ligadas à língua escrita, código secundário para a Lingüística que trabalha prioritariamente com a língua falada.

Ainda com relação a **palavra** há tentativas de conceituação, embora, na prática, freqüentemente "palavra" e "vocábulo" sejam usados como verdadeiros sinônimos. **Termo**, entretanto, não costuma ser unidade de uso no vocabulário específico da Lingüística.

Mattoso Câmara Jr., no seu Dicionário de **Filologia e Gramática**, informa que "termo" é "em sentido gramatical estrito, vocábulo ou grupo de vocábulos que corresponde a uma unidade de significação ou de função, como elemento constitutivo para a inteligibilidade do que se enuncia".

Essa conceituação, de fato, autoriza o uso desse elemento como unidade do vocabulário específico das linguagens documentárias, ao mesmo tempo em que mostra a sua fragilidade para o uso especializado da Lingüística.

3 – LINGUAGEM NATURAL E LINGUAGEM DOCUMENTÁRIA: SEMELHANÇA E DIFERENÇAS

Observando as linguagens documentárias (LD) e o comportamento das línguas – linguagem natural (LN) – evidencia-se um aspecto de semelhança entre elas: as LD trabalham com descritores e termos-chave que são entidades bastante semelhantes aos lexemas da LN, na medida em que nos sistemas dos quais fazem parte, correspondem a unidades portadoras de forma e significado.

Ao nível formal, podemos analisar os descritores em termos de:

a. **sintagmas** de símbolos notacionais (números, letras, pontuação, marcas) isto é, unidades resultantes da combinação de formas menores em unidades de nível superior. Ex.: leit – eira → leiteira; o, vestido, verde, de, Lúcia → O vestido verde de Lúcia.

b. **lexemas** como combinação de fonemas, ou seja, como combinação de unidades capazes de promover a distinção entre signos da língua. Ex.: lago ≠ lato pela troca de um único elemento na seqüência, o fonema g por t.

Ao nível semântico, os descritores têm um sentido (designata*), semelhante aos sememas da LN, ordenados em semas, ou seja, conjuntos de unidades mínimas de significado.

Algumas diferenças básicas, entretanto, podem ser anotadas:

a. Na LN a escrita é secundária, enquanto que nas LD ela é fundamental.

b. Do ponto de vista semântico, as LD padronizam e simplificam o relacionamento entre lexema e o extrato semântico. Na LN há sinônimos e

* Na concepção semiótica de Ch. Morris, são necessários três componentes para que algo funcione como signo: ter um veículo, ter um sentido ou designatum que corresponde àquilo a que o veículo se refere e um interpretante que se refere ao efeito desta unidade sobre um intérprete.

homônimos, mostrando a não correspondência um a um entre significantes e sememas.

c. A LN pode funcionar como sua própria metalinguagem*. As LD, não. Elas necessitam da LN, ou de outra linguagem artificial para falar delas.

d. As LD têm um propósito específico de organização de funções e por isso mesmo pode-se dizer que são governadas por regras de "JURE", ou seja, por regras estabelecidas por um indivíduo ou grupo que cria o sistema. A LN, entretanto, embora portadora de organização, apresenta uma complexidade tal, que suas funções são governadas por regras de "FATO", não de direito, na medida em que não são criadas no todo por um indivíduo, mas atendem às diversas necessidades de comunicação do grupo social.

4 – SEMÂNTICA

Os estudos semânticos que ficaram mais ou menos estacionários até a década de 60, adquiriram uma dinamicidade muito grande, em especial nos últimos quinze anos.

Dentre os vários estudos de semântica, destacam-se, grosso modo, duas vertentes: a semântica estrutural e a semântica gerativa. A semântica estrutural ocupa-se, basicamente, com o estudo descritivo da natureza e funcionamento do significado do signo, visto como elemento do sistema lexical de uma língua. A semântica gerativa, de outra parte, se propõe a um estudo da competência do falante nativo, ao nível do significado. Segundo esse enfoque, o falante é portador de um "dicionário" que contém informações semânticas básicas e regras de projeção que lhe permitem produzir e reconhecer as frases de sua língua, bem como interpretar frases ambíguas e identificar anomalias sintático-semânticas.**

A Semântica como disciplina lingüística sempre se debateu com dificuldades de precisão metodológica, pois que a identificação do significado, quer a teoria

* Metalinguagem é a linguagem usada para falar sobre outra linguagem, a linguagem objeto, ou sobre si mesma. Ex.: a língua portuguesa pode ser usada como metalinguagem para falar sobre símbolos matemáticos: "o sinal + significa adição". Também pode ser usada para falar dela mesma. "a palavra menino é composta de três sílabas, seis fonemas e significa "ser humano", "jovem", "sexo masculino", etc.

** A simplificação que fizemos atende a objetivos práticos. De fato, tanto sob a égide da Lingüística Estrutural, quanto da Gerativo-transformacional e mais recentemente da Pragmática, o problema do significado lingüístico tem recebido orientações bastante diversas.

incorpore isso, ou não, resvala continuamente em fatores lógicos, antropológicos, sociológicos e psicológicos.

Assim, embora, por exemplo, a questão da verdade do significado interesse diretamente ao filósofo, ela não deixa de se emiscuir numa análise lingüística. Para o lingüista, no entanto, o que importa é circunscrever seu objeto de análise, no caso o significado, ao nível da palavra, da frase, ou do texto, dependente ou independentemente do contexto situacional, conforme a sua opção teórica.

Sem dúvida, é a questão do significado o problema de contorno mais complexo para as linguagens documentárias, no trato com termos da linguagem natural. Reconceituar ou delimitar a significação de um termo, mesmo que seja para incorporá-lo num sistema, como é o caso das linguagens documentárias, implica em escolha num quadro de possibilidades, por vezes, amplo e variável, de grupo para grupo, de região para região.

Diante do complexo universo da Semântica, entendemos que o apoio teórico para trabalhos de indexação teria de levar em conta, num primeiro momento, a Semântica Lexical, que trabalha ao nível da palavra e atinge, no máximo a frase.

Nesta linha, começamos por fazer notar que os estudos semânticos podem analisar o significado das unidades, buscando constantes e variáveis lexicais, a partir do mecanismo gerador (onomasiológico), centrado no emissor, ou a partir de mecanismo interpretador (semasiológico), centrado no receptor. No primeiro caso o foco fica sobre o significante e no segundo, sobre o significado.

Os signos lingüísticos, por simbolizarem ou representarem as "coisas", através de uma relação arbitrária entre significante e significado, ou imotivada em relação ao referente, alimentam toda a polissemia da linguagem natural, ao mesmo tempo em que propiciam construções denotativas e conotativas. A denotação decorre de considerações externas do signo, isto é, da consideração de seu caráter referencial. A conotação, ao contrário, nasce de uma relação simbólica.***

*** A linguagem denotativa é a que expressa o significado concebido "objetivamente", por oposição à linguagem conotativa que expressa valores "subjetivos" ligados ao próprio signo. O termo cloreto de sódio, provavelmente, denota um significado objetivo, o da própria substância que ele nomeia; o termo praia, no entanto, num dado contexto, pode conotar "descontração", "lazer", etc., embora não se anule seu significado objetivo.

Nas linguagens de indexação, no entanto, os signos em geral, recortam o âmbito significativo dos termos de uma língua particular, fixando significados, de forma a anular a arbitrariedade do signo lingüístico. São significados fixos, que por isso mesmo não permitem a conotação.

Desta forma, os signos, documentários, por sua natureza mais estática, exigem do indexador um trabalho de percepção do universo onde atua, para que não só a tradução do documento em termos da linguagem documentária se faça com fidelidade, mas para que essa tradução tenha um forte poder de partilha na comunidade à qual o documento se destina.

A redução dos significados, por ser uma operação de corte num universo praticamente ilimitado, poderá se valer de alguns conceitos correntes em linguagem natural. Dentre eles selecionamos a polissemia, a homonímia, a sinonímia e a antonímia.

4.1 Polissemia

A polissemia é o nome dado à pluralidade de sentidos de uma mesma forma. Estuda as várias significações da significação lingüística, significações essas que se definem e precisam num determinado contexto. Só o contexto poderá precisar se a palavra paixão, por exemplo, se refere a "sofrimento", a "sentimento imoderado", a "amor violento", a "falta de objetividade no julgar" etc.

A polissemia pode se dar por:

— extensão, como, por exemplo, em estação que pode significar: "parada", "épocas do ano", "cada uma das catorze paradas da Via Sacra", "temporadas da moda";

— metáfora que é um tipo de extensão, no qual atua um componente analógico. Em serra, por exemplo, o significado de "montes" decorreu da analogia com o instrumento "serra";

— restrição, como, por exemplo, abrir (do latim aperire) nos deu a palavra aperitivo, que por sua vez corresponde a dois termos restritos a dois significados: "purgativo" na linguagem médica e "beberete" na linguagem comum.

Enquanto na linguagem natural os mecanismos polissêmicos são promotores de larga margem de criatividade, nas linguagens documentárias eles

devem ser reduzidos ao máximo, em função da fixação dos significados.

4.2 Homonímia

Corresponde a homonímia à igualdade entre significantes de significados diferentes. É, pois, o estudo das formas que apenas se diferenciam pela significação ou função, já que a estrutura fonológica é a mesma.

A homonímia pode ser:

— total, como em fiar que tanto significa "tecer", quanto "confiar";

— parcial, como em coser e cozer.

É possível a um mesmo grupo de formas apresentar homonímia total e parcial. Exemplo: era = "época" e "1ª e 3ª. pessoas do singular do pretérito imperfeito do indicativo de ser; hera = "planta".

Nas linguagens de indexação, o homônimo parcial não gera nenhuma dificuldade, pois elas lidam com língua escrita e então a homonímia se desfaz. Já a homonímia total pode representar alguma dificuldade, contornada, provavelmente, pela escolha de sinônimo de maior poder de recuperação.

No caso de documentos especializados, é pouco provável a presença de homônimos totais, pois o recorte significativo da área dificulta a possibilidade de duplo significado.

4.3 Sinonímia

A sinonímia decorre de coincidência de significado entre diversas palavras. É o estudo da substituição de termos, sem prejuízo da comunicação. Exemplos: o significado de "mar" pode ser expresso através dos termos mar, oceano, pego, pélagos; e de "ser alado" por ave, pássaro.

A utilização do sinônimo é, provavelmente, uma dificuldade mais séria para as linguagens de indexação. Basta observar que em língua portuguesa, nossos padrões estéticos não permitem, mesmo em linguagem científica, a constante repetição de um mesmo termo. Isso gera no documento a presença da mesma idéia convertida em várias seqüências diferentes, isto é, em formas sinônimas. A escolha de uma destas formas torna, por vezes, muito complexa a situação do indexador, que acaba tendo

de se valer de sinônimos mesmo, onerando o sistema de indexação com muitas informações.

4.4 Antonímia

A antonímia decorre de significações contrárias de dois vocábulos ditos autônomos. É o estudo de palavras com significações opostas, como amor /ódio; bom/mau; moral/imoral, progrediu/regrediu.

Neste caso importa lembrar que o indexador, muito provavelmente, não usa o conceito de antônimo, ou dito de outra forma, ele incorpora a forma antônima ao sistema, como se fora um sinônimo, ou parte natural de um signo documentário. Assim, um documento que tratasse, por exemplo, de problemas de imoralidade, seria identificado pelo termo **moral** que recobriria o significado antonímico.

Nas áreas especializadas, entretanto, o problema não é tão simples. À primeira vista, poderíamos dizer que **oxidação** e **redução**, ou **aquecimento** e **resfriamento** expressam significações contrárias. No entanto, correspondem a processos autônomos, o que não nos permitiria, provavelmente, tratá-los como antônimos.

5 – CAMPOS SEMÂNTICOS E SINONÍMIA

A teoria dos campos semânticos nasceu em 1930, como uma corrente da onomasiologia sincrônica.

Segundo Jost Trier, o lingüista alemão que a concebeu, nenhuma palavra vive em estado de isolamento na consciência do falante, mas está rodeada de toda uma série de expressões ligadas pelo significado, formando no interior do conjunto lexical algo como um campo de palavras mais ou menos fechado.

O trabalho do semanticista corresponde, então, em efetuar um corte horizontal, sincrônico num determinado campo conceitual e identificar um conjunto de significações correlatas.

Uma das vantagens que essa teoria trouxe para a semântica foi o alargamento da concepção onomasiológica e semasiológica para estudos sincrônicos, antes restritas apenas a análises diacrônicas. Com a teoria dos campos semânticos, as palavras passaram a ser analisadas sincronicamente e os estudos semânticos se libertaram de certa limitação, decorrente de estudos dirigidos a conceitos concretos, a palavras do campo material.

A onomasiologia e a semasiologia deixaram de ser dois métodos diferentes e passaram a ser tomadas como dois enfoques, ou dois movimentos de observação dentro da teoria semântica.

No enfoque semasiológico a análise corresponde ao agrupamento das palavras em torno de um nó de significação, de um campo de semas. Exemplo:

espalhar — “separar a palha” dos cereais
“lançar para diferentes lados”
“dispersar”
“espargir”
“aspergir”
“divulgar”
“difundir”
“irradiar”
“tornar público”, etc.

O nó semasiológico, ao que tudo indica, é o significado original “separar algo de algum ponto”.

Constitui dificuldade para lidar com campo semântico a distinção entre significações correntes e significações esporádicas, além de interpretações diversas, tendo em vista o contexto lingüístico.

“Toda palavra é formada de um nó semântico mais ou menos denso, ou mais ou menos volumoso, cercado por uma auréola de associações secundárias, afetivas ou sociais.”
Guiraud⁴

De acordo com esse enfoque, o semanticista parte de elementos lingüísticos (formas) e verifica o campo que criam com suas conexões e oposições; em seguida, estabelece, previamente, certos campos de significação, em função da vida social e estuda os elementos lingüísticos que neles se dispõem.

Nas linguagens documentárias, esta concepção semasiológica é que determina a organização de conjuntos de termos em torno de nós semânticos, que por sua vez determinam os cortes no campo do significado.

Desta forma, só dominando o universo do significado é que se pode operar satisfatoriamente com restrições que cubram um dado campo semântico e organizar coleções finitas de palavras-chave.

O enfoque onomasiológico pode ser encontrado nos dicionários de sinônimos. Decorre das associações que uma palavra manifesta em relação a outras, graças a significações correlatas. Vale observar que, neste caso, há distinções entre vocábulos de uso

corrente e de uso esporádico, além de usos diversificados em função do contexto lingüístico. Exemplo: sol, astro, estrela, planeta; dinheiro, riqueza, fortuna, tutu, bufunfa.

O princípio básico é que o lugar semasiológico determina a posição da significação no campo onomasiológico.

Para o estabelecimento do campo semântico é prioritária uma análise sêmica, também chamada componencial pela Lingüística americana. Essa análise visa a apreender os traços semânticos mínimos (semas ou componentes) que relacionam e distinguem os elementos de um campo lexical e consiste em estabelecer, no plano do conteúdo, através de relações associativas, os semas constituintes de determinadas unidades.

Assim, um dado conjunto de palavras (vocábulos) constitui um campo semântico, na medida em que haja, pelo menos um sema que esteja presente em todas as unidades daquele conjunto.

Segundo Pottier⁵, a diversidade de relações dá ao falante (emissor) a possibilidade de, partindo de um ponto de vista onomasiológico, poder fazer escolhas em vários níveis:

a. entre variantes combinatórias ou livres;

Ex.: louro/loiro (variantes livres)
eu vou, eu irei, nós vamos (variantes combinatórias)
ler/leitura (lexemas – variantes combinatórias)

Nesses casos ocorre homossemia total, uma vez que as formas manifestam absoluta identidade de significado.

b. entre formas variadas para substância semântica bastante semelhante:

Ex.: faz dois dias/antes de ontem
fazer um erro/enganar-se
Pedro vende vinho/O vinho é vendido por Pedro

Nesses casos a homossemia é parcial, porque há diferenças de conotação na escolha de uma ou outra forma. Dependendo do foco que se quer ressaltar, da intensidade que se quer dar, usa-se uma ou outra.

c. entre forma especializada ou de uso corrente, conforme a região geográfica, a época, o nível social, o uso técnico, a faixa etária, etc.

Ex.: jerimum/abóbora (norte e sul)
vasca/tanque (passado e presente)
emagrecer/esmagrecer (nível social)
sal/cloreto de sódio (uso comum e uso técnico)
está bem/legal (velho e jovem)

Nesses casos a homossemia é dita diassistemática.

Observando a diversidade de comportamento semântico dos vocábulos, é possível identificar conceitos nos quais ocorre:

– delimitação bem precisa

Ex.: **segunda-feira** é diferente de **domingo** e de qualquer outro dia da semana.

– imprecisão expressa através de zonas fluídas de transição

Ex.: **dia e noite** — “dia” é também o período de 24 horas e inclui a noite;
frio e calor
grande e pequeno

A imprecisão e a sinonímia são, provavelmente os aspectos da linguagem natural que oferecem maiores dificuldades ao indexador. Para o lingüista esses problemas, embora não satisfatoriamente solucionados ao nível componencial (de modo muito especial a imprecisão) são de tratamento relativamente mais simples, na medida em que se tem como suposto que o significado decorre do relacionamento das unidades que organizam a sentença. Logo, qualquer alteração que se faça numa unidade acarretará modificações no significado.

Para o trabalho de indexação, o que nos parece claro é que uma análise componencial dos termos tornaria inviável o processo que exige eficiência num curto espaço de tempo. Desta forma, entendemos que a compreensão da problemática do significado, ao nível da “palavra”, oferece ao indexador muito mais a dimensão do problema, que a solução operacional. O trabalho de indexação corre por conta da seleção de palavras-chave num processo em que uma linguagem controlada se associa à intuição do indexador.

Na elaboração de um texto e mesmo na sua tradução, há situações de difícil troca de palavras e outras mais flexíveis.

O vocabulário técnico é, com certeza, o que exige maior precisão, o que apresenta o fator emocional do emissor em estado mais próximo ao “neutro”.

O vocabulário comum, entretanto, só perde em imprecisão, ambigüidade e intervenção do fator emocional para a linguagem literária.

Na observação de casos extremos cabe uma questão: a possibilidade de troca de palavras num mesmo contexto deve ser entendida como relação sinonímica ou de variantes lexicais?

Com efeito, a maioria das permutações de palavras gera alteração do significado objetivo, do tom sentimental, ou mesmo do valor evocativo.

No caso de termo técnico, a escolha, por exemplo entre Fonêmica e Fonologia, ou entre Ciência da Informação e Informática evidenciam, nas respectivas áreas, posições teóricas, linhas de orientação do emissor. Já a opção entre Usuário, Leitor e Consultante parece identificar, por um lado época (usuário e leitor/consultante) e de outro, uso comum e uso em situações de especialização (leitor/usuário), decorrentes da evolução tecnológica que vem tornando o acesso à informação cada vez mais dinâmico.

Os estudos lingüísticos levam a não admitir a existência do sinônimo perfeito. É o interesse ou a incerteza das apreciações que norteiam a escolha entre **justo e equitativo**, entre **punir e castigar**.

Por outro lado, há quem defenda a existência do sinônimo, respeitada a condição de que, ao nível do sentido objetivo, seja possível obter uma representação diferente para o mesmo significado.

Ex.: pediatra e médico de criança
Eu chegarei após sua partida e Você partirá antes de minha chegada.

Em termos de comunicação, não há prejuízo nem ganho com essas trocas. Com efeito, hoje, **pediatra** é termo de domínio comum; a mudança do tópico nos dois enunciados dá a mesma informação objetiva.

Em síntese, e de um ponto de vista estrutural, o que se observa numa análise semântica ao nível lexical é que:

— cada signo lingüístico se opõe a todos os outros da língua;

— todo signo está inscrito numa rede semântica e os trechos mais compactos formam famílias semânticas;

— só o contexto permite dominar e precisar o significado de um signo lingüístico;

— toda troca de vocábulo implica em troca de semema e esse processo acarreta alteração na comunicação, pela perda de alguns semas e/ou pela introdução de outros.

Já nas línguas documentárias observa-se que:

— cada signo documentário se opõe a todos os outros do sistema no qual se inscreve;

— cada signo está inscrito numa rede semântica que permite ordenar os termos em organizações do tipo de tesouros;

— o significado de cada signo está previamente delimitado e assume precisão na relação que estabelece com todos os outros do sistema;

— a troca de um termo não necessariamente implica em alteração no resultado comunicativo, quer pela utilização de sinônimos, quer pela inclusão dos antônimos.

6 – VOCABULÁRIO ESPECIALIZADO

Afirma Benveniste⁶ que a história de uma ciência se resume na história de seus termos próprios.

De fato no âmbito da linguagem natural, as línguas especializadas ou científicas apresentam como peculiaridade a existência de um vocabulário próprio que se caracteriza pela maior precisão de seus termos, já que a expressão técnica ou especializada é centrada na função referencial da linguagem.

Em geral a palavra, neste uso, tem um significado unívoco, daí podermos nos referir a ela como monovalente.

No entanto é preciso considerar que mesmo no vocabulário técnico ocorrem áreas de imprecisão. Como exemplo pode ser citado o espectro das cores, onde temos matizes com limites absolutamente fluídos: cinza/cinza claro/cinza escuro/cinza chumbo; verde/verde claro/verde escuro/verde musgo/verde água.

Freqüentemente, nas línguas técnicas, as palavras, por serem monovalentes, provocam uma hermetização, só rompida por processos de vulgarização, especialmente hoje, através dos meios de comunicação de massa (Ex.: acoplamento, Q.I.), pois há natural tendência para perder a monovalência. É o campo da imprecisão que se instaura.

Basicamente um vocabulário especializado se forma por um dos três procedimentos abaixo:

— a partir de vocábulos da linguagem comum que, redefinidos, assumem sentido especial.

Naturalmente esse procedimento é o mais sujeito a ambigüidades e a única forma de minimizar essa tendência indesejável na Ciência é proceder a constantes redefinições.

— pela criação de neologismos que se valem dos processos de formação vocabular da língua. Ex.: rádiodifusão.

— pela adoção de vocábulos estrangeiros, aportuguesados ou não. Ex.: xerox, feed-back.

6.1 À guiza de exercício

À guiza de exercício, escolhemos o vocabulário de Biblioteconomia e Ciência da Informação e tentamos organizá-lo em subáreas significativas.

Inicialmente, fizemos um levantamento dos termos em resumos e títulos de artigos publicados nas quatro revistas da área (São Paulo, Brasília (duas) e Belo Horizonte) e pudemos observar que boa parte do vocabulário é formada pela especialização de vocábulos da linguagem comum, ou por empréstimo a outras áreas. Ex.: recuperação da informação, sistema, rede.

Na tentativa de organizar núcleos semânticos dentro do vocabulário especializado da área, que de si já constitui um campo semântico, organizamos algumas classes a partir do seguinte raciocínio: a área lida com um **objeto**, apresentado em determinados **suportes materiais**, submetidos a determinadas **operações**, para certos **fins**.

A partir, pois desses quatro polos e sem nos limitarmos ao vocabulário retirado do corpus*, buscamos ordenar os termos, segundo categorias semânticas, como segue:

a. Objeto:	livro	acervo	informação
	periódico	documento	linguagem documentária
	jornal	coleção	autor

* Para um levantamento de vocabulário discriminante de área, organizamos um corpus a partir de 73 artigos (título e resumo) publicados entre 1978 e 1980, nas quatro revistas brasileiras de maior divulgação nas áreas de Biblioteconomia e Ciência da Informação.

revista	referência
manuscrito	base de dados
mapa	
livro raro	

b. Lugar:	
biblioteca	arquivo
biblioteca pública	estante
biblioteca universitária	fichário
biblioteca central	caixa-estante
biblioteca especializada	
seção de circulação	hemeroteca
consulta	áudio-visual
referência	base de dados
aquisição	banco de dados

c. Agente:	
bibliotecário	
auxiliar de biblioteca	
leitor	
usuário	
consultante	

d. Modo:	
recuperação da	biblioteconomia
informação	
catalogação	depósito legal
divulgação de serviços	
bibliotecários	
treinamento de pessoal	
sistema de biblioteca	
catalogação centralizada	
serviço de alerta	
indexação	
disseminação de informação	
processamento de informação	
rede	
acesso	
acesso on-line	

e. Instrumento:	
ficha	linguagem documentária
máquina leitora	áudio-visual
carro biblioteca	base de dados

f. Produto:	
catálogo	informação
índice	levantamento bibliográfico
sumário corrente	publicação

g. Finalidade:	
acesso	
busca	
busca retrospectiva	
recuperação	
disseminação da informação	

Como se observa, um mesmo termo pode, numa dada categoria, ter capacidade referencial tal, que recobre outros. É o caso de **documento** que pode recobrir o conjunto dos objetos com significado mais objetivo que **acervo**.

De outra parte, há elementos que podem participar de mais de uma categoria, como os referentes de "lugar onde se colocam os documentos", ou algo relativo a eles: **arquivo**, **estante**, **fichário**, que tanto podem ser focalizados como "objetos", quanto como "elementos do acervo".

7 - SINTAXE

A sintaxe, na mesma linha de simplificação que fizemos na parte de Semântica, pode ser enfocada a partir de duas vertentes: a sintaxe estrutural e a sintaxe gerativa-transformacional.

A sintaxe das linguagens documentárias, que com certeza não oferece o grau de dificuldade das línguas naturais, tanto tem se valido da visão estrutural, quanto da transformacional. A sintaxe estrutural pode servir de apoio aos trabalhos de análise da língua com objetivos à automatização, enquanto a sintaxe transformacional tem sido utilizada mais diretamente, através do modelo de Casos de Charles Fillmore, para fins de indexação pelo PRECIS*.

Dentre os modelos estruturais parece-nos ser a Sintaxe Distribucional o mais interessante para trabalhar com enunciados concretos, sem dar ao significado um valor central.

Observando os processos de indexação a partir de palavras-chave, algumas noções da sintaxe natural da língua, merecem análise. Dentre elas, destacam-se alguns aspectos da coordenação e da subordinação**.

* Preserved Context Index System. Sistema de indexação alfabética desenvolvido pela British Library, a partir de 1971.

** Noções como as de adição, disjunção, inclusão, exclusão etc. parecem básicas. A posição do adjetivo no Sintagma Nominal (SN) costuma incomodar os indexadores. Sabemos que vários estudos têm sido feitos a respeito da posição do adjetivo em português. Entretanto, a nosso ver, na base da problemática abordada por indexadores não está uma questão propriamente lingüística e sim o resultado de nossa dependência científica e tecnológica. Grande parte, senão a maioria dos documentos nas áreas especializadas de ciência e tecnologia são escritos em inglês, o que justifica a utilização de tesouros proveniente de países de língua inglesa, especialmente Estados Unidos. A utilização de uma margem insignificante de documentos em português não justifica a construção de linguagens controladas adequadas à realidade da língua portuguesa, permanecendo o desencontro entre a estrutura da língua portuguesa e a da linguagem documentária adotada.

Na concepção mais tradicional, à gramática cabe o estudo das relações sintagmáticas e associativas, por isso mesmo os estudos de Morfologia (forma) não se separam dos da Sintaxe (funções).

À Sintaxe, especificamente, compete estudar a combinatória de unidades da língua que permite, partindo de unidades menores, construir unidades maiores. Assim, de fonemas se constroem morfemas, de morfemas, vocábulos, de vocábulos, sintagmas e, por fim, de sintagmas se constroem frases.

Várias orientações ocorreram ao longo do tempo, tanto nos Estados Unidos, quanto na Europa. Mas, em linhas gerais, os métodos analíticos da sintaxe estrutural prevêm:

- a. a segmentação do texto em unidades sintaticamente irreduzíveis e a classificação das mesmas;
- b. a identificação dos princípios ou regras que determinam as relações que se estabelecem num enunciado.

Alguns pressupostos sustentam a análise sintagmática. São eles:

- a linearidade do discurso, decorrente da própria linearidade do significante. Ex.: a, bela, casa, verde → a bela casa verde;
- a consideração de níveis de análise que se montam numa hierarquia. O sintagma assim é formado por unidades do mesmo nível hierárquico. Ex.: minha, bolsa → minha bolsa; cas-inh-ola → casinholas
- a compatibilidade semântica entre o resultado sintagmático e os significados dos componentes. Ex.: des-fazer (neg. + base); -colar (neg. + base) A incompatibilidade semântica não autoriza segmentar no mesmo paradigma: de-mente e se-mente.

- a recorrência dos componentes do sintagma, expressa através da possibilidade de um componente integrar diferentes combinatórias. Ex.: fácil, facilidade, facilímo, facilmente, desfazer, descolar, desmentir.

 casa
minha casa
 a casa amarela
nesta casa grande

Embora as descrições sintáticas variem bastante, é possível, em linhas gerais, dizer que na Europa a

concepção predominante nos trabalhos de descrição era a de que "descrever é conhecer alguma coisa pelos seus acidentes, ou marcas".

A consequência disso é que toda análise sintática européia trabalhou menos na busca da essência universal dos fenômenos, na medida em que ficou, de preferência, em fenômenos individualizados, naquilo que distingue uma coisa de outra.

Nos Estados Unidos, a preocupação foi com a identificação de princípios, de regularidades subjacentes aos atos de fala e como consequência, buscou a Lingüística americana traços pertinentes que, ultrapassando o nível dos indivíduos, atingissem pragmaticamente, a todos os falantes, ao sistema enfim.

7.1 Análise Distribucional

Teoria sistematizada na década de 30, dominou a Lingüística americana até pouco mais de 1950.

Tendo por base a teoria psicológica behaviorista, tomam os distribucionalistas o ato de fala como um tipo especial de comportamento. A estrutura da língua corresponde a uma arquitetura, onde os elementos existem em estado de interdependência.

Disso decorre a possibilidade da linguagem a partir de situações concretas.

Então o estudo de uma língua tem de ser feito tomando-se um **corpus**, isto é, um conjunto de enunciados homogêneos, efetivamente emitidos, que independentemente da sua significação evidenciam as regularidades da língua (as próprias regras) num dado momento.

A análise, basicamente, sincrônica, realiza-se pela exclusão da função, da significação e da situação. A única noção levada em conta é a de contexto linear, por isso importa conhecer num enunciado, quais unidades seguem ou procedem um dado elemento.

Logo, o contexto é toda a série dos co-ocorrentes de um dado elemento. E a distribuição de uma unidade se define pelo conjunto de contextos em que ela é encontrada no corpus.

A segmentação do enunciado faz a sua decomposição em unidades sucessivas menores, do que nasce a noção de constituintes imediatos (CI), ou seja, de unidades de um nível inferior é que são organizadas

as unidades de um nível superior. Ex.: /d/ e /e/ → de; (a) e (casa) → a casa; (a casa) e (é grande) → a casa é grande.

As classes, então, são ordenadas a partir de CI idênticos, ou seja, a partir da distribuição dos elementos nos diversos níveis. Ex.: (Det) a e (Subs.) casa se distribuem numa certa ordem no Sintagma Nominal (SN); a (Det) / a (Pron) porque o primeiro ocorre no SN e o segundo no SV (Sintagma Verbal).

Também em função do Det. a observa-se que a e o determinam duas classes de morfemas nominais, que se distribuem completamente: o menino/a menina.

No sentido de obter um método mais econômico para a organização das classes, o distribucionalismo se vale da comutação que permite dizer, por exemplo, que dois fonemas são diferentes, na medida em que, pela troca dessas unidades entre si, são formadas, num segundo nível, unidades significativas diferentes.

Ex.: /l/ e /m/
l-ar l-ata
m-ar m-ata

Esse processo permite, ao nível dos morfemas e dos sintagmas, organizar as classes de Det. definidas pelas possíveis combinações com os nomes do mesmo tipo. Ex.:

a casa	era o popular Geraldo Boi
esta casa	era o Geraldo
aquela casa	era Geraldo BOI
certa casa	era Geraldo
mesma casa	
alguma casa	era o grande mentecapto

A substituição pode se dar entre seqüências. Ex.:
o vaso quebrou.
o vaso de porcelana chinesa quebrou.
o carro de minha filha quebrou.
o avião do presidente quebrou.

Pode-se observar também que, pelos exemplos, determina-se uma classe de nomes femininos que aceita combinar com os Det. A distribuição de um elemento é, pois, a soma de todos os seus contextos.

Basicamente se organizam dois tipos de classes:

a. as sintagmáticas, que correspondem à seqüência linear, horizontal. Ex.: um elemento qualquer A é definido pela sua relação na cadeia falada com elementos do tipo X e Y: a casa amarela casa a amarela

b. as paradigmáticas, que correspondem aos

segmentos possíveis num mesmo contexto. Assim A e B fazem parte da mesma classe, se forem encontráveis no mesmo contexto. Ex.:

- a casa amarela
- a pata amarela
- a rosa amarela

Esse método não exclui certas manifestações sintáticas do tipo: Pedro vende vinho/ O vinho é vendido por Pedro. Neste caso, a distribuição dos mesmos segmentos conduz a uma aproximação dos dois enunciados, de sorte a se poder estabelecer um paralelismo sistemático. O fato da forma passiva ser menos freqüente que a ativa, leva o lingüista a considerá-la como transformação desta. Logo, o próprio conceito de transformação é tirado da distribuição.

Além do contexto imediato, a análise distribucional leva em conta as co-ocorrências. Ex.: um SN sujeito plural implica em verbo plural. Portanto há co-ocorrência das marcas de plural.

Ao nível da comunicação essas marcas de co-ocorrência exercem forte poder de tornar a informação possível, pois como traços redundantes que são, minimizam ruídos na transmissão e até mesmo esquecimentos gerados pelas distâncias na cadeia expressiva.

O enunciado não é uma seqüência aleatória de elementos, mas a escolha de um segmento é determinada pelo segmento precedente, num conjunto limitado de elementos. É a concepção de que cada ato de fala se faz pela escolha dos membros de classes que regularmente ocorrem juntas numa ordem específica, que os estudos de distribuição estatística foram largamente amparados nos anos da Lingüística Distribucional.

Desta forma é possível agrupar os elementos de uma língua em classes e estabelecer, precisamente, a ocorrência relativa dessas classes. A ocorrência de um elemento de uma classe em relação a outra classe deve ser vista em termos de probabilidade, baseada na freqüência de ocorrência no corpus.

As restrições de ocorrência relativa de cada elemento devem ser descritas por uma rede de afirmações inter-relacionadas, de sorte que as restrições maiores precedam as menores.

Também um item particular da análise pode contribuir para a simplicidade do sistema. Ex.: a classificação de determinados elementos como

verbos permite sua subdivisão em transitivos e intransitivos, de acordo com a sua distribuição.

O mérito da Lingüística Estrutural, aqui representada pela análise distribucional, pode ser resumido no fato de ter:

- ampliado o conhecimento das estruturas lingüísticas;
- aperfeiçoado as técnicas de coleta de dados;
- demonstrado a possibilidade de generalização de determinadas estruturas.

7.2 Sintaxe Transformacional

Na década de 50, o modelo estrutural parecia esgotado para a Lingüística americana. E isso ocorreu porque, na medida em que se buscou construir uma gramática explícita, capaz de dar conta do funcionamento global da linguagem, não foi mais possível isolar a parte semântica da formal.

Os métodos estruturais, excessivamente presos a uma visão positivista de só operar com dados, não permitia a formulação de questões perturbadoras, como as que se faziam sobre a criatividade da linguagem.

Em 1957, N. Chomsky⁷ demonstrou a insuficiência dos modelos estruturais e formulou uma nova teoria, denominada gerativo-transformacional que colocava no centro das preocupações lingüísticas a sintaxe e procurava a todo custo encontrar critérios de adequação e simplicidade para as descrições lingüísticas*, de sorte a apreender quais se valem diuturnamente, na formulação de infinitos enunciados.

Livre das peias de um corpus, a gramática gerativo-transformacional usa processos de simulação na busca dos fatos da linguagem. Vale lembrar que em oposição aos dados de um corpus que são grandezas conhecidas e por isso mesmo quantificáveis, os fatos não necessariamente são conhecidos, mesmo porque podem ser reais ou potenciais.

O componente sintático, responsável pelo processo gerativo e criativo da linguagem passa a ser o centro

* A sintaxe estrutural não deu conta de problemas como concordância, constituintes descontínuos, ordem livre dos componentes, etc. Chomsky (1965)⁸ dá a necessidade de adequação da gramática de certa forma superada pelo portador explicativo que uma gramática gerativa deve ter.

de toda a análise lingüística que, contrariamente às posições anteriores, passa a se ocupar, prioritariamente com a conexão linguagem-pensamento. Segundo Chomsky (1975)⁹ tanto na percepção, quanto na aprendizagem, não se pode desconsiderar o pensamento que desempenha papel ativo na determinação da natureza do conhecimento adquirido.

A observação do fato de que o falante de determinada língua conhece essa língua, levou Chomsky a admitir que a experiência contínua e apropriada do indivíduo desenvolve nele uma gramática inata de sua língua. É esse sistema de regras que permite ao falante construir enunciados com propriedades formais e semânticas e é esse mesmo sistema que deve ser desvendado pela análise lingüística.

Para melhor compreender a teoria gerativo-transformacional, basta pensar numa criança exposta a determinada língua. Naturalmente ela manifesta uma capacidade para, num breve período, inferir regras que regem os enunciados que ouve ao seu redor. Então, unindo os estímulos à sua capacidade inata, a criança formula uma gramática. A esse conhecimento ativo, adquirido graças à faculdade inata da linguagem, Chomsky denominou **competência lingüística**. E é essa competência que autoriza a produção ao infinito de enunciados. À capacidade de produção denominou **desempenho** (performance).

A tarefa da Lingüística, então, é construir gramáticas formalizadas que através da lógica simbólica explicitem com maior clareza as estruturas lógicas do pensamento.

Assim os enunciados da língua são o resultado da aplicação pelo falante das regras sintáticas que devido à própria natureza humana ocorre em etapas.

É exatamente o domínio, ou a posse da gramática que o falante tem de sua língua que faculta a ele produzir e entender uma infinidade de enunciados, independentemente da sua capacidade consciente de refletir sobre as regras. De fato ao dizermos que o falante tem interiorizado o domínio da uma gramática gerativa que expressa seu conhecimento da língua, não estamos subentendendo que:

- a. ele tenha consciência das regras da gramática que usa;
- b. que ele possa chegar, sozinho ou intuitivamente a ter consciência das regras;

c. nem que suas manifestações sobre o conhecimento intuitivo que ele tem da língua tenham de ser exatas.

Isso porque a gramática gerativa trata de processos mentais que vão além do nível de consciência efetiva.

O falante sabe que "O governador eleito foi empossado" está correto e que não o está "o governador foi empossado eleito". Também é essa capacidade que permite perceber a ambigüidade de "Bateu no cachorro de seu pai", ou a "equivalência de:

Ao terminar as apurações o juiz anunciou os resultados.

Quando terminaram as apurações o juiz anunciou os resultados.

Terminadas as apurações o juiz anunciou os resultados.

Essas possibilidades todas devem estar presentes na teoria sintática da língua. Mas é preciso ressaltar que essa previsão teórica não implica em um levantamento da produtividade do falante e isto porque a descrição devera dar conta de todos os enunciados da língua a nível de regras poderosas. O problema da produtividade resulta da competência subjacente do falante e esta não pode ser vista por inteiro, já que o falante jamais terá oportunidade de usar sua competência lingüística inteiramente.

À vista disso, busca a Lingüística gerativo-transformacional um conjunto finito de regras que responda pela produção dos enunciados, ou dito de outra forma, é um conjunto de regras que faculta ligar o sentido à forma pela qual os enunciados se expressam. A forma pode ser observada em termos de constituintes; o sentido, no entanto exige um trabalho de abstração para o atingimento da estrutura subjacente que responde pela interpretação semântica dos enunciados.

Desta forma, a gramática da língua contém subjacentes as estruturas que submetidas à aplicação de regras de transformação convertem-nas em estruturas superficiais.

Em resumo, uma gramática gerativa deve ter capacidade para gerar todos os enunciados gramaticais de uma língua e somente eles; ser capaz de explicar a compreensão intuitiva que o falante tem dos enunciados. Assim, "O menino feriu-se" é o

- 6 **BENVENISTE, E. Problèmes de linguistique générale.** Paris, Gallimard, 1974.
- 7 **CHOMSKY, N. Syntactic Structures.** La Haye, Mouton, 1957.
- 8 _____ . **Aspectos da teoria da sintaxe.** Coimbra, Arménio Amado, 1975.
- 9 _____ . **Reflexões sobre a linguagem.** Lisboa, Edições 70, 1977.
- 10 **FILLMORE, C.** Em favor do caso. In: **LOBATO, L.M.P. A semântica na Lingüística moderna; O léxico.** Rio de Janeiro, Francisco Alves, 1977.
- 11 **BARTHES, R. Elementos de Semiologia.** S. Paulo, Cultrix, 1972.
- 12 **BORBA, F.S. Teoria sintática.** S.Paulo, Queiroz/EDUSP.
- 13 **CARVALHO, S.W.** Problemas e processos de formação de linguagem técnica. In: **TOLEDO E SILVA, M.R., org. Interrelacionamento das ciências da linguagem.** Rio, Gernasa, p.81-109.
- 14 **CINTRA, A.M.M. & SILVA, M.C.P.S.** Morfologia da língua portuguesa. **Revista da PUCSP**, 49 (97): 34-57, jan./jun. 1979.
- 15 **COYAUD, M. Linguistique et documentation.** Paris, Larousse.
- 16 **DUBOIS, J.** Grammaire distributionnelle. **Langue Française** 1, fev. 1969, Larousse, p.41-8.
- 17 **DUROT, O. & TODOROV, T. Dictionnaire encyclopédique des sciences du langage.** Paris, Seuil.
- 18 **GALMICHE, M. A Semântica gerativa.** Lisboa, Presença, 1979.
- 19 **HARRIS, Z.** Discourse analysis. In: **FODOR & KATZ, The Structure of language,** Englewood Cliffs, N.J., Prentice-Hall, 1964.
- 20 **HUTCHINS, W.J. Languages of indexing and classification.** Peter Peregrinus, UK.
- 21 **JACOBSON, R. Lingüística e comunicação.** São Paulo, Cultrix, 1969.
- 22 **KOCK, I.G.V. Os sintagmas preposicionais como modificadores nominais.** Dissertação de mestrado PUCSP, 1977.
- 23 **LOPES, E. Fundamentos da lingüística contemporânea.** São Paulo, Cultrix.
- 24 **LYONS, J. An introduction to theoretical linguistic.** Londres, Cambridge.
- 25 **MARTINET, A. Elementos de lingüística geral.** Lisboa, Sá da Costa, 1964.
- 26 **ULLMAN, S. Semântica.** Lisboa, Gulbenkian, 1967.

ABSTRACT

Several linguistic phenomena interfere in the conversion of natural language to documentary language. For a conceptual study of these phenomena some topics were selected: The reading process and problems related to semantics and syntax.