

## 2

### Quadro Teórico

*And every Man has so inviolable a Liberty, to make Words stand for what Ideas he pleases, that no one hath the Power to make others have the same Ideas in their Minds, that he has, when they use the same Words, that he does. (John Locke, Essay)*

#### 2.1

##### Introdução

O foco principal deste trabalho é o léxico e sua organização. O estudo do léxico é intimamente ligado ao estudo da mente e suas estruturas cognitivas, sendo abordado do ponto de vista de várias disciplinas: a psicolinguística, a lexicografia, a linguística teórica, a linguística de corpus, a linguística computacional, entre outras.

Na seção 2.2 deste capítulo discuto alguns modelos de organização do léxico dentro de perspectivas linguísticas às vezes conflitantes. No estruturalismo americano e na teoria gerativa o léxico era considerado idiossincrático demais para merecer uma sistematização semelhante às propostas sintáticas. Propostas dissidentes, entre elas modelos computacionais, incluem a integração do léxico a outras estruturas do conhecimento, seja por meio de interfaces ou em abordagens conexionistas.

Um dos problemas teóricos de maior relevância nesse trabalho é o da atribuição de significado a uma palavra e da participação desse significado na construção do sentido no texto. Para estabelecer as bases teóricas para a discussão desses problemas no caso das expressões em foco, apresento na seção 2.3 uma breve revisão das principais abordagens do fenômeno da polissemia.

#### 2.2

##### Modelos lexicais

O problema da delimitação de unidades lexicais tem se manifestado constantemente nas teorias linguísticas e lexicográficas. Não há uma definição da noção básica de palavra que se adeque às descrições das diversas línguas e nem

mesmo consenso sobre se a unidade mínima do léxico é a palavra. Mesmo assim, os lingüístas vêm trabalhando com definições operacionais, que priorizam determinados critérios de delimitação, tais como critérios fonológicos, sintáticos e semânticos. Bloomfield (1933; 1926) oferece a definição distribucionalista de palavra, baseada nos conceitos de *forma livre* e *forma presa*.

“A minimum form is a morpheme; its meaning a sememe. A form which may be an utterance is free. A form which is not free is bound. A minimum free form is a word. A non-minimum free form is a phrase.” (Bloomfield 1926, p. 155)

O morfema, e não a palavra, é a unidade mínima de análise lingüística e a questão de unidades maiores que a palavra não é contemplada.

Nas fases iniciais da teoria gerativa transformacional, (Chomsky 1965) conceitua o léxico como um conjunto de entradas lexicais, cada uma consistindo de uma matriz de traços distintivos e um conjunto de propriedades de vários tipos: traços sintáticos e semânticos, propriedades especificando que processos morfológicos e transformacionais são aplicáveis ao item, entre outras. Na teoria padrão, os processos derivacionais gerais eram tratados do ponto de vista sintático, através das regras transformacionais sintáticas.

Chomsky (1970) rejeita o tratamento transformacionalista do léxico, propondo a Hipótese Lexicalista. A partir daí, o léxico passa a ter uma importância fundamental na teoria gerativa, tornando-se responsável por explicitar as relações entre palavras e suas derivações. A idéia original de Chomsky, ainda que muito pouco desenvolvida, sugeria um modelo do léxico em que as entradas fossem neutras quanto à classe de palavras. A forma da palavra a ser utilizada na inserção lexical seria determinada pelos traços categoriais exigidos pelas regras sintáticas.

As teorias lexicais gerativas que se apresentaram no contexto da Hipótese Lexicalista, para cumprir o papel de componente lexical dentro do modelo da Teoria Padrão Estendida, pretendiam satisfazer três níveis de adequação à teoria (Jackendoff 1975):

1. **adequação observacional** - na teoria, a entrada lexical descreve completamente o comportamento do item na língua; palavras e não-palavras são identificáveis;
2. **adequação descritiva** - a teoria expressa as relações, regularidades e generalizações do léxico;

3. **adequação explicativa** - a teoria apresenta um mecanismo de avaliação para atribuir medidas a descrições lexicais concorrentes, em geral em termos de custos.

Halle (1973), já na Hipótese Lexicalista mas com resquícios da cultura estruturalista, acreditava em uma morfologia baseada em morfemas. Seu modelo do léxico é constituído por uma lista de morfemas e um conjunto de regras de formação de palavras a partir do material da lista, que gerariam o vocabulário completo da língua. O modelo proposto por Jackendoff (1975) estabelece uma lista completa de todas as palavras da língua, flexões incluídas, completamente especificadas, e um conjunto de regras de redundância, relacionando as palavras. Seu principal objetivo era garantir que as relações lexicais fossem representadas no léxico. Jackendoff propõe que as propriedades semânticas sejam separadas das morfológicas nas regras de redundância pois, como no caso das nominalizações, a conexão entre os sufixos nominalizadores e o significado das formas nominalizadas é imprevisível.

A teoria morfológica apresentada por Di Sciullo & Williams (1987) exclui radicalmente o léxico. Dentro do universo de palavras, Di Sciullo & Williams estabelecem três noções classificatórias fundamentais para a teoria.

**Objetos morfológicos** são os membros do conjunto de palavras definido recursivamente a partir de um conjunto básico de átomos – os morfemas – por meio de um conjunto de regras de afixação e composição. O principal objetivo da morfologia é caracterizar o conjunto de objetos morfológicos de uma língua.

**Átomos sintáticos** são as palavras no papel de unidades mínimas da análise sintática. Como átomos, não são divisíveis, analisáveis, nem possuem estrutura interna.

**Listemas** são as palavras enquanto itens de uma lista não caracterizável; contém as idiossincrasias da língua.

A gramática é essencialmente a teoria da formação de palavras e frases, abrangendo os objetos morfológicos e os átomos sintáticos, mas excluindo totalmente os listemas. A morfologia e a sintaxe diferem apenas na natureza de seu átomos e na formulação de suas regras de formação, possuindo uma estrutura teórica bastante similar. Quanto ao léxico, sua posição se tornou notória com a seguinte afirmação:

“The lexicon is like a prison: it contains only the lawless, and the only thing that its inmates have in common is lawlessness.” (Di Sciullo & Williams 1987, p.3)

Os principais argumentos usados por Di Sciullo & Williams para dissociar o léxico do conjunto de palavras da língua são: (i) existem objetos listados que são de natureza sintática; (ii) os objetos morfológicos não são listáveis.

**Objetos sintáticos listados** As expressões idiomáticas possuem uma estrutura sintática, porém possuem semântica imprevisível. O significado de uma expressão idiomática não pode ser computado composicionalmente, portanto é um listema. Sintagmas verbais como PUSH X TOO FAR e BITE X'S HEAD OFF são listemas com uma posição variável. Eles se comportam sintaticamente como sintagmas regulares, mas não semanticamente. Outro grupo de listemas sintáticos são as construções verbo-partícula do inglês, do tipo LOOK UP, THROW UP, etc.

**Objetos morfológicos não listados** A possibilidade de criação de palavras novas pelo falante evidencia a existência de objetos morfológicos não listados. A eventualidade de uma palavra nova vir a ser listada é função da transformação de seu significado, o que depende de fatores particulares a cada falante.

O modelo de Di Sciullo & Williams radicaliza o tratamento excepcional do léxico, que não apresenta interesse para a gramática, e as relações existentes entre itens lexicais com alguma irregularidade são totalmente desprezadas.

Num pólo oposto, destaco o modelo de Bybee (1988), digno de nota por apontar para um tratamento mais integracionista e poli-sistemático do léxico, onde as regras morfológicas e a representação do léxico são unificadas. O resultado pode ser visto como uma listagem extremamente sofisticada, onde os padrões morfológicos e morfofonêmicos emergem a partir da organização intrínseca do léxico.

Bybee baseia-se em estudos psicolinguísticos para destacar algumas características da aquisição de linguagem fundamentais em seu modelo: os mecanismos mentais de armazenamento e organização de itens lexicais possuem a capacidade de

- construir uma representação semântica e fonológica do material lexical;
- formar redes de itens, conectados por características comuns;
- registrar a frequência de itens e padrões e
- organizar estímulo sensorial em categorias.

Os fatos morfológicos de uma língua podem ser descritos em termos dessas habilidades. Bybee definiu dois conceitos que realizam a abstração desses mecanismos: **conexão lexical** e **força lexical**.

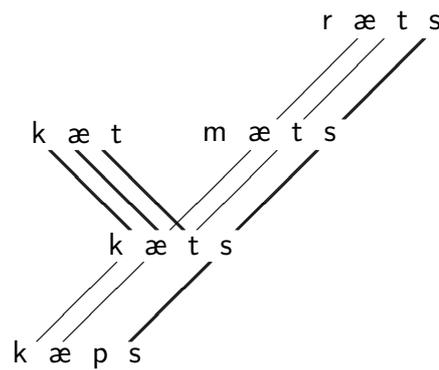


Figura 2.1: Rede de conexões para *cat*, *cats*, *rats*, *mats* e *caps*

Uma palavra armazenada é o pareamento de um conjunto de traços semânticos com um conjunto de traços fonológicos. Relações entre palavras ocorrem quando há compartilhamento de traços. Ocorre, na representação do léxico, uma conexão entre as palavras, no traço coincidente: a **conexão lexical**. Quando há identidade de conexões entre duas palavras, então elas são morfologicamente idênticas.

Por exemplo, considerando o grupo de palavras CAT, CATS, RATS, MATS e CAPS temos a rede da figura 2.1. As linhas escuras indicam conexões fonológicas e semânticas e as claras indicam conexões fonológicas apenas. Os S indicativos de plural são ligados por linhas escuras. Em particular, as conexões entre CAT e CATS são semânticas e fonológicas.

Uma rede de conexões lexicais é capaz de dar conta dos seguintes fatos morfológicos.

**Segmentação** A análise de palavras em morfemas é substituída por um processo de interconexão de padrões morfofonológicos. No exemplo da figura 2.1, o radical CAT fica identificado em CATS pelas conexões escuras. Quando uma nova palavra morfologicamente complexa é introduzida na rede, há a formação de conexões com material previamente existente, com base em significado e fonologia, de modo que os segmentos da palavra são identificados sem a necessidade de desmembramento. Vários casos de difícil segmentação morfêmica são resolvidos assim.

**Grau de proximidade** É possível no modelo estabelecer um grau de proximidade, dado pelo número e tipos de conexões semânticas e fonológicas entre duas palavras. Esse conceito é experimentalmente observável, já que falantes formam associações entre palavras relacionadas semântica e fonologicamente, de maneira consistente. Fatos diacrônicos e interlingüísticos também dão suporte ao conceito.

Para dar conta dos efeitos psicolinguísticos, históricos e interlinguísticos da frequência de uso de itens lexicais na morfologia, Bybee introduz o conceito de **força lexical**. Quanto mais usado for um item mais ele ganha força; formas pouco usadas perdem força.

Com essa medida, o modelo do léxico ganha dinamismo e passa a refletir a importância da frequência na teoria, o que já é evidenciado experimentalmente, principalmente no fato de que palavras mais fracas são armazenadas em termos de palavras mais fortes.

O conceito de força lexical é capaz de dar conta de dois processos lexicais importantes em qualquer sistema linguístico:

1. maior grau de irregularidade e supleção entre palavras e paradigmas mais frequentes;
2. a direção da relação

*básica* → *derivada*

se dá proporcionalmente a

*maior força* → *menor força*

Em suma, a unificação entre léxico e morfologia é extremamente profícua sob o aspecto descritivo da língua pois permite a identificação de padrões de regularidades não detectáveis nos modelos anteriores. No entanto, não há indícios de que uma instanciação do modelo tenha sido feita, o que seria um projeto extremamente ambicioso.

Voltando a uma linha histórica da abordagem do léxico no gerativismo, pode-se dizer que o reconhecimento do status do léxico no sistema gramatical se desdobrou em outras correntes gerativas (Newmeyer 1998). A dissidência se deu, em grande parte, pela rejeição ao modelo de uma estrutura profunda que passa a ser um enunciado, ou estrutura superficial, por meio de regras transformacionais. Em modelos gerativos não Chomskyanos, em geral, as construções deixam de ser epifenômenos e adquirem uma ligação mais direta com seu significado. Algumas destas teorias realizaram-se em sistemas computacionais, tais como Generalized Phrase Structure Grammar (Gazdar *et al.* 1985), Head-driven Phrase Structure Grammar (Pollard & Sag 1994) e a Gramática Categorial (Steedman 1993). Nessas teorias, cada vez mais as entradas lexicais são enriquecidas com informações que previamente eram codificadas em regras sintáticas.

Jackendoff (2002) motiva a questão das construções no léxico dentro do âmbito da questão: “what aspects of an utterance *must* be stored in long-term

memory, and what aspects *can* be constructed online in working memory”? Para ele, o **item lexical** é justamente essa unidade armazenada, enquanto a **palavra** é um objeto de natureza gramatical. O autor analisa as dificuldades decorrentes da proposta de tratamento de expressões multi-vocabulares como itens lexicais atômicos ou listemas, não construcionais, e a impossibilidade de processá-los como se fossem combinações livres de palavras. Exemplifico algumas construções desse tipo, apontando os problemas que apresentam.

**ex. 2.1** *Construção em que elementos internos são flexionados.*

1. *O próprio argumento de que não é dando comida que se resolverá o problema da fome no Brasil* FOI COLOCADO EM SEU DEVIDO LUGAR.
2. *Máquinas e equipamentos* FORAM COLOCADOS EM SEUS DEVIDOS LUGARES.

**ex. 2.2** *Construção que pode ser descontínua, com várias possibilidades de elementos que podem ser inseridos: LEVAR EM CONSIDERAÇÃO NP, LEVAR NP EM CONSIDERAÇÃO, LEVAR ADV EM CONSIDERAÇÃO NP.*

1. *Na opinião de ambos, essa seria uma análise rasa e apressada que* LEVARIA EM CONSIDERAÇÃO *apenas as aparências.*
2. *Se as regras da Argentina tomarem os mesmos caminhos que vêm tomando no Brasil, então nós passaremos a* LEVAR *esse país* EM CONSIDERAÇÃO *também.*
3. *...LEVANDO também* EM CONSIDERAÇÃO *a beleza que a iluminação trouxe à cidade.*
4. *...LEVANDO-se ainda* EM CONSIDERAÇÃO *as propriedades físicas dos materiais.*

**ex. 2.3** *Frase indivisível, um listema, cuja participação em enunciados seria problemática.*

...MAIS VALE UM PÁSSARO NA MÃO DO QUE DOIS VOANDO.

**ex. 2.4** *Construção NP A NP, bastante produtiva.*

1. *Afinal, é por causa dele que a indústria da falsificação de cassetes cresce* ANO A ANO.
2. *“A final vai ser disputada* JOGO A JOGO”, disse.

Na perspectiva de Jackendoff, para dar conta desses dados é necessário admitir que estruturas frasais são possíveis constituintes de itens lexicais, em um léxico que incorpora regras não só para a combinação de itens menores que a palavra, como também para lidar com o que chama de “construções idiomáticas”.

Essa visão de léxico enquanto “um sistema de produção, armazenamento e transmissão de formas simbólicas, isto é, formas que evocam significados ou dão acesso a estruturas conceituais” (Basilio 2005), que não tem na palavra morfológica sua unidade básica, vem ganhando espaço na pesquisa lingüística por se mostrar mais consistente e mais flexível na abordagem de fenômenos das línguas.

### 2.2.1

#### O léxico computacional

Todas as aplicações de Processamento de Linguagem Natural (PLN) envolvem, com maior ou menor grau de importância, um componente lexical. O léxico automático deve conter conhecimento de dois tipos: (i) informações necessárias para a análise e geração sintática e (ii) informações necessárias para interpretação semântica.

As diferenças entre o estudo computacional do léxico e as abordagens lingüísticas mais tradicionais são resumidas a seguir (Pustejovsky 1999).

- A representação lexical computacional deve ser explícita.
- Na computação, a estrutura global do léxico deve ser modelada. A complexidade do léxico não pode ser pressuposta como previamente conhecida pelo usuário computador, portanto as relações entre palavras são tão importantes quanto as relações entre componentes de palavras. Entradas compostas por mais de uma palavra ortográfica também devem ser representadas.
- O léxico computacional deve dar ampla cobertura de seu domínio. Na prática, léxicos computacionais podem chegar a 400.000 entradas, tipicamente divididas, na língua inglesa, em: verbos (5.000), substantivos (30.000), adjetivos (5.000), advérbios (<1.000), termos lógicos (<1.000), termos retóricos (<1.000), compostos (2.000), nomes próprios (300.000) e mais outros termos da língua.
- O léxico computacional deve ser avaliável, em termos de: (i) cobertura do domínio; (ii) extensibilidade, ou seja, com que facilidade o léxico pode ser ampliado; (iii) utilidade, ou seja, qual a contribuição do léxico para a aplicação.

Independentemente da aplicabilidade do estudo computacional do léxico, sua contribuição e relevância nas ciências cognitivas já é apreciável. Primeiramente porque as estruturas lexicais e interlexicais utilizadas em estudos computacionais geraram as descrições, até hoje, mais completas das bases lexicais das línguas naturais. Além disso, as escolhas feitas em projetos de léxicos automáticos têm tido impacto em estudos lingüísticos e psicolingüísticos. Finalmente, a representação explícita muitas vezes demonstra limitações de modelos teóricos.

Entre os modelos lexicais computacionais mais influentes podem-se citar o WordNet (Fellbaum 1998) e o FrameNet (Fillmore, Wooters, & Baker 2001). O projeto WordNet consiste de uma grande base de dados lexicais do inglês, onde substantivos, verbos, adjetivos e advérbios são agrupados em conjuntos de sinônimos cognitivos (os *synsets*) que expressam conceitos distintos. Os synsets são interligados por relações semântico-conceituais e lexicais, resultando em uma rede conceitual que pode ser consultada por meio hipertextual.

O projeto FrameNet trata de construir uma descrição de ampla cobertura para o léxico do inglês, de acordo com os pressupostos da **Semântica de Frames** (Fillmore 1976). É um projeto ambicioso, pois não apenas se preocupa com o aspecto da construção de um recurso computacional, mas também pretende validar experimentalmente o modelo de representação lexical de Fillmore.

De acordo com (Fillmore 1976), frames são arcabouços definidos como representações esquemáticas de situações envolvendo diversos participantes, coadjuvantes e outros papéis conceituais. Os dados são obtidos do British National Corpus, e vão sendo anotados, semântica e sintaticamente, e armazenados em um banco de dados organizado por itens lexicais e por frames. A figura 2.2 exibe um exemplo de frame do sistema: **Point\_of\_dispute**. Nele se encontram: uma definição da situação, seguida por alguns exemplos do corpus. A seguir, uma lista de **elementos de frame** (**frame elements**) descrevem os participantes da situação, ordenados como **centrais** (**core**) e **não centrais** (**non-core**). A seguir, são listados uma série de atributos que constróem uma rede de frames: heranças, ordenações e dependências. Ao final, há uma lista de itens lexicais que se ligam a essa frame, no exemplo CONCERN, ISSUE e QUESTION.

As principais áreas de aplicação do léxico automático podem ser subdivididas em quatro grandes grupos: processamento de textos, incluindo a análise sintática, geração de textos e tradução; processamento de fala, incluindo reconhecimento de fala e transformação de texto em fala; edição de texto, incluindo correção ortográfica e gramatical; e recuperação de informação, incluindo sis-

**Point\_of\_dispute**

**Definition:**

The answer to a Question is under discussion in a Group, which still has a difference of opinion among its members. The prominence of the Question relative to others can be indicated by a Status expression

Firstly, it is quite clear the ISSUE in this context was not so much apostasy as much as it was treason.  
 The QUESTION in Finucane's case is this: will an inquiry bring out the truth?  
 It remains an ISSUE of central importance to women .  
 Democracy in Serbia was blocked by the unresolved national QUESTION.

**FEs:**

**Core:**

Question [que] The Question is an open proposition the answer to which is under dispute.

The ISSUE is who will have control of Iraq's rich oil resources.

**Non-Core:**

Context [con] A state or event within which the Question arises as a problem for the Group.

Descriptor [des] The QUESTION in Argentina's case is what exchange rate is appropriate now.  
 A characterization of the Question, often with regard to its complexity or whether it has been resolved or not.

Domain [dom] Two unresolved ISSUES fuel speculation that he might not receive a fair trial.  
 The area of human experience which the Question concerns.

Group [gro] The third sector of the Master Plan explores the important economic ISSUE of tourism.  
 The Group are the persons or organizations who have different points of view on the Question.

**Semantic Type** Sentient

Point\_of\_view [poi] An individual or individuals from whose point of view the Question is a point of dispute within the Group.

Status [sta] For Elijah, the ISSUE was between two altars: worship of God and worship of Baal.  
 The Frame Element Status is used for expressions that indicate the relative importance of a Question with respect to the other topics that are disputed.

Time [tim] The main ISSUE is where the revenue is coming in from.  
 The time interval during which the Question exists as a problem for the Group.

**Semantic Type** Time

The ISSUE at the time was whether or not to relocate the Rennes-based part of the team to California.

Inherits From:

Is Inherited By:

Subframe of:

Has Subframes:

Precedes:

Is Preceded by:

Uses:

Be\_in\_agreement\_on\_assessment,Discussion

Is Used By:

Perspective on:

Is perspectivized in:

Is Causative of:

See Also:

**Lexical Units**

*concern.n, issue.n, question.n*

Created by josef on Mon Aug 15 11:12:35 PDT 2005

Figura 2.2: Frame para QUESTION no FrameNet

temas de indexação e recuperação de documentos.

O detalhamento das características do componente lexical, em termos das propriedades das entradas lexicais representadas, do conjunto de palavras listadas ou da abrangência das regras de formação de palavras, só pode ser definido no âmbito de uma aplicação. Hudson (1988) propõe uma espécie de lista de atributos constituintes de um hipotético “Léxico Exaustivo”, uma idealização do léxico computacional universal, provendo espaço para as seguintes informações:

1. Fonologia

- estrutura segmental subjacente;
- padrões prosódicos da palavra;

2. Morfologia

- estrutura em termos de morfemas;
- estruturas morfológicas irregulares vinculadas a traços morfossintáticos particulares
- similaridades parciais a outras palavras, de mesma base
- propriedades com respeito a cliticização

3. Sintaxe

- classe de palavra (ex. verbo)
- sub-classe (ex. auxiliar)
- traços morfossintáticos obrigatórios
- valência

4. Semântica

- nome da entidade referida (X)
- identidade de X (anafóricos)
- hiperônimos
- valência semântica de X
- entidades inerentes a X
- entidades implícitas (*default*) (ex. default de beber seria álcool)
- entidades que devem ser definidas por anáfora
- como os papéis semânticos são vinculados à valência

5. Contexto

- restrições relacionadas à estrutura social contextual
- restrições relacionadas a estilo
- restrições relacionadas à estrutura social mais ampla
- restrições relacionadas a estruturas de discurso

6. Grafia

- ortografia normal
- abreviações
- irregularidades flexionais da grafia

7. Etimologia e língua

- a língua a qual a palavra pertence
- a língua da qual foi emprestada
- a língua em que é baseada

- datação

#### 8. Uso

- frequência e familiaridade
- idade da aquisição
- ocasiões particulares em qua palavra foi utilizada
- clichés e colocações contendo a palavra
- tabu

Esse esquema é resultante de uma compilação de muitas propostas e modelos de léxicos computacionais, tendo em vista as aplicações e as descrições lingüísticas daquilo que constitui a unidade lexical. No entanto, esse tipo de empreitada será sempre suscetível a duas espécies recorrentes de críticas: se por um lado é amplo demais para que seja adequadamente instanciado, por outro certamente não inclui todos os recortes existentes para a descrição lexical.

## 2.3

### Polissemia

O poema “The Blind Men and the Elephant”, de Godfrey Saxe, conta uma estória do folclore indiano envolvendo seis cegos que buscam reconhecer um elefante, cada um apalpando uma parte diferente do corpo do animal.

It was six men of Indostan,  
To learning much inclined,  
Who went to see the elephant,  
(Though all of them were blind),  
That each by observation  
Might satisfy his mind.

The first approached the elephant,  
And happening to fall  
Against his broad and sturdy side,  
At once began to bawl:  
“God bless me! But the elephant  
Is very like a wall!”

The second, feeling of the tusk,  
Cried: “Ho! What have we here,  
So very round and smooth and sharp?  
To me ’tis very clear,  
This wonder of an elephant  
Is very like a spear!”

The third approached the animal,  
And happening to take  
The squirming trunk within his hands,  
Thus boldly up and spake:  
“I see,” quoth he, “the elephant  
Is very like a snake!”

The fourth reached out an eager hand,  
And felt about the knee.  
“What most this wondrous beast is like  
Is might plain,” quoth he;  
“’Tis clear enough the elephant  
Is very like a tree.”

The fifth, who chanced to touch the ear,  
Said: “E’en the blindest man  
Can tell what this resembles most:  
Deny the fact who can,  
This marvel of an elephant  
Is very like a fan.”

The sixth no sooner had begun  
About the beast to grope,  
Than seizing on the swinging tail  
That fell within his scope,  
“I see,” quoth he,  
“the elephant Is very like a rope.”

And so these men of Indostan  
Disputed loud and long,  
Each in his own opinion  
Exceeding stiff and strong.  
Though each was partly right,  
All were in the wrong.

O panorama das teorias do significado em muito se assemelha à estória do poema. A semântica ocupa um lugar de destaque principalmente na filosofia, na matemática e na lingüística e, por resistir a uma caracterização precisa em todas essas áreas, tem ocupado as cabeças de inúmeros pesquisadores

durante longo tempo. Dentre as linhas de investigação semântica majoritárias, Lyons (1995) lista (coincidentemente) seis correntes teóricas e suas respostas (simplificadas aqui) à pergunta “o que é significado?”

1. a teoria **denotacional** ou **referencial**: “o significado de uma expressão é aquilo que ela denota”;
2. a teoria **ideacional** ou **mentalista**: “o significado de uma expressão é a idéia ou conceito associado a ela, que está na mente do falante/ouvinte”;
3. a teoria **behaviorista**: “o significado de uma expressão é o estímulo que a evoca ou a resposta evocada por ela, ou uma combinação dos dois, em ocorrências particulares de um enunciado”;
4. a teoria **significado-é-uso**: “o significado de uma expressão é determinado pelo seu uso na língua, ou é o próprio uso”;
5. a teoria **verificacionista**: “o significado de uma expressão, se existente, é determinado pela verificabilidade das sentenças ou proposições que a contêm”;
6. a teoria de **condições de verdade**: “o significado de uma expressão é a sua contribuição para as condições de verdade das sentenças que a contêm”.

Do ponto de vista lingüístico, há uma unanimidade em torno do fato de que os focos determinados por essas teorias não conseguem individualmente delinear um conceito de significado que satisfaça os fenômenos empíricos percebidos nas línguas. Observando os abismos que dividem essas visões do significado, é interessante notar que a noção de **polissemia** é concebida em todas elas, de maneiras diferentes.

De modo geral, a polissemia é uma propriedade lexical de palavras que apresentam uma multiplicidade de significados. Aparentemente, essa propriedade não se manifesta de maneira estável, podendo surgir momentaneamente em uma situação pragmática onde as palavras são combinadas temporariamente de acordo com princípios lingüísticos gerais. De acordo com Geeraerts:

“The tremendous flexibility that we observe in lexical semantics suggests a procedural (or perhaps ‘processual’) rather than a reified conception of meaning; instead of meaning as things, meaning as a process of sense creation would seem to become our primary focus of attention.” (Geeraerts 1993, p. 259)

O célebre exemplo a seguir exemplifica essa dinamicidade do morfismo dos significados, com a seguinte situação:

**ex. 2.5** [uma garçonete diz para outra] *O SANDUÍCHE DE PRESUNTO ali no canto quer mais café.*

Na construção do significado do enunciado, o pressuposto nesse caso é de que a caracterização do referente é completamente irrelevante; o processo transforma o ‘cliente’ em seu ‘pedido’, utilizando uma informação compartilhada no contexto pragmático das garçonetes.

Essa indeterminação prévia do significado de uma parte do enunciado não parece ser um problema na comunicação humana, que resolve a seleção do sentido apropriado sem esforço aparente e de maneira inconsciente. Pelo contrário, quanto maior a frequência no discurso maior a variedade de sentidos adquiridos pela palavra, como atestam alguns dicionários como (Sinclair 2001) que registra a frequência das palavras no *British National Corpus*.

### 2.3.1

#### Entre o lógico e o psicológico

O significado lexical tem sido tratado classicamente por meio do estabelecimento definicional de condições necessárias e suficientes para a pertinência de um indivíduo à categoria conceitual nomeada pela palavra, como em uma semântica de modelos para a Lógica. Essa abordagem tem como consequência a existência de tantos sentidos para uma palavra quantas forem as diferentes configurações possíveis das condições. Além disso, não há espaço nessa visão para a influência do contexto no significado lexical.

Dentro dessa perspectiva, (Katz 1972) concebe a representação semântica como uma composição de **marcadores semânticos** que definiriam o significado central da palavra. Por exemplo, para a palavra CADEIRA alguns marcadores seriam ‘objeto’, ‘físico’, ‘artefato’, ‘móvel’, ‘tem pernas’, entre outros. Essas abstrações serviriam para construir uma hierarquia de conceitos com herança de propriedades.

Na teoria de Katz, a similaridade semântica é uma medida da coincidência de marcadores semânticos e a ambigüidade é a existência de mais de uma representação para a mesma palavra. O argumento crítico dirigido a esse modelo, de que a menor diferença conceitual entre duas ocorrências de uma palavra corresponderia a dois sentidos distintos da mesma, gerando uma explosão incontrolável de sentidos lexicais não relacionados, é assim respondido pelo autor.

“Meaning must be an abstraction from the variable features of the things referred to by the term: the meaning of a word must represent only the invariant features by virtue of which something is a thing, situation, activity, event, or whatever of a given type. Otherwise no word could ever be used again with the same meaning with which it is used at any one time, since there is always some difference in what is referred to from one time to the next.” (Katz 1972, apud Ravin & Leacock 2000, p. 10).

Mesmo ignorada na teoria de Katz, a polissemia regular é reconhecida por autores clássicos, como (Jackendoff 2002), em termos de processos produtivos sistematizados por regras e, portanto, previsível. Estão aí incluídas instâncias de fenômenos exemplificados em 2.6, como a) alternância de estrutura argumental em verbos, b) autohiponímia, c) relação entre recipiente e quantidade, entre outros.

#### ex. 2.6

1. [*Maria está cozinhando a carne*] e [*A carne está cozinhando*];
2. [*O gato é um animal doméstico*] e [*O gato está sobre a mesa*];
3. [*A colher está sobre a mesa*] e [*Leva duas colheres de açúcar*].

Já as polissemias mais imprevisíveis, resultantes de alguns tipos de metáforas e metonímias, não são facilmente explicadas nesse quadro teórico, sendo tratadas muitas vezes como sentidos não relacionados.

Uma perspectiva clássica mais inclusiva de classificação do fenômeno da polissemia é dada por (Weinreich 1964), apud (Pustejovsky & Boguraev 1996), que explora a distinção entre a **ambigüidade contrastiva**, similar ao conceito de **antagonismo** de (Cruse 1986) (discutido mais adiante), e **ambigüidade complementar**. O primeiro tipo corresponde à homonímia, o segundo, à situação em que leituras alternativas são manifestações do mesmo sentido nuclear em diferentes contextos.

Mais recentemente, as teorias lingüísticas de cunho cognitivista (Janssen & Redeker 1999) vêm enfatizando a integração de três fatores determinantes para o significado das palavras: as estruturas cognitivas ou lingüísticas inatas; a informação lingüística estabelecida por convenção; e o conhecimento de mundo. Da psicologia cognitiva, a noção de **prototipicidade** de conceitos, uma noção experimental, dá conta de que as pessoas categorizam objetos em termos da similaridade percebida entre o objeto e o membro prototípico da classe selecionada. Essa proposta foi utilizada no programa da lingüística cognitiva

como modelo para o significado lexical. Tem-se portanto a polissemia como processo central de construção do significado.



Figura 2.3: Linhas contínuas unindo entre homonímia/monossemia e polilexia/vagueza

A posição prevalente entre cognitivistas (Cruse 1995; Geeraerts 1993; Tuggy 1993) é a de não admitir limites fixos e bem demarcados entre os fenômenos de homonímia/polissemia e vagueza/polissemia, mas sim um contínuo de casos de polissemia, entre os dois extremos, como mostra a figura 2.3. A polissemia se apresenta quando os significados são desconexos por um lado e conexos por outro. Esses contrastes são discutidos a seguir.

### 2.3.2

#### Polissemia e homonímia

Na dimensão homonímia/polissemia, o contraste é feito com o objetivo de decidir quando os significados distintos podem ser explicados pela existência de duas palavras distintas que possuam a mesma forma: a **homonímia**.

Os exemplos tradicionais de homonímia no português incluem MANGA e BANCO:

#### ex. 2.7

1. *As células pétreas são os elementos formadores da casca de frutos duros, como noz, castanha-do-pará,...e do endocarpo de certos frutos do tipo drupa (pêssego, MANGA, azeitona, ameixa...)*
2. *Jorge Luis estava enforcado por uma camisa de MANGA comprida de algodão, amarrada à grade de ventilação da cela.*

**ex. 2.8**

1. *Um pequeno BANCO de madeira, que a Sotheby's avaliou em US\$ 100, saiu por inacreditáveis US\$ 33.350.*
2. *Alguns clientes não repararam que o BANCO estava sendo assaltado, apesar da gritaria.*

Alguns critérios de distinção dessas palavras de formas coincidentes recorrem à diacronia e são de difícil verificação. No entanto, é possível argumentar que, mesmo sem conhecer a etimologia das diferentes MANGAS, trata-se de um caso claro de homonímia pela distância dos significados percebidos pelo falante.

Essa distinção observada pelo falante motiva um conjunto de critérios baseados no grau de relacionamento dos diferentes significados da palavra para reconhecer a polissemia ou a homonímia. A conexão entre os significados de ORELHA no exemplo 2.9 parece clara: é uma extensão metafórica de uma parte do corpo para uma parte do livro.

**ex. 2.9**

1. *Por isso, também a ORELHA é formada de tecido cartilaginoso.*
2. *Está na ORELHA do livro cujo lançamento a revista foi cobrir.*

Entretanto, o que dizer das conexões entre as ocorrências de LADO no exemplo 2.10?

**ex. 2.10**

1. *No lugar do vidro da porta, ao LADO do piloto, apenas uma proteção de couro.*
2. *“Ele tem uma relação visceral com o Rio de hoje e de ontem”, diz Mayer, que garante que o LADO psicológico do personagem é o que menos importa na montagem.*
3. *Al Pacino desvenda o LADO obscuro da política.*

De acordo com (Lyons 1977), a possibilidade de que a polissemia para um falante seja julgada como homonímia por outro indica a indeterminação da fronteira entre os fenômenos, provavelmente devido ao caráter gradativo do conceito de relacionamento entre significados.

O contraste entre homonímia e polissemia motiva a questão da sistematicidade do relacionamento entre os significados polissêmicos, dos processos derivacionais desses significados e da organização desses significados para que reflita as regularidades existentes (Ravin & Leacock 2000).

### 2.3.3

#### Vagueza e ambigüidade

Enquanto a lexicografia tem se preocupado tradicionalmente com a distinção entre polissemia e homonímia, por ser de fundamental importância na quantificação do número de entradas de um dicionário, a lingüística vem cada vez mais enfocando a oposição entre **vagueza**<sup>1</sup> (também chamada **generalidade** ou **indeterminação**) e **ambigüidade**. Kilgarrieff elabora com clareza essas questões:

“First, the homonymy/polysemy question is ‘Do we have one word or two?’ We have been given two different senses to consider, and wish to determine whether they are senses of the same or of different words. For the vague/ambiguous distinction, the question is ‘Do we have one sense or two’? ” (Kilgarrieff 1992, p.47)

A questão, então, é a distinção entre vagueza denotativa e ambigüidade, que busca distinguir se um dado elemento semântico é inerente à palavra ou é fornecido pelo contexto (Dunbar 2001):

“A word is vague with respect to an interpretative element of meaning if it does not supply it; it is ambiguous between two elements if it supplies one or the other.”

Os enunciados do exemplo 2.11 se contrastam, ilustrando em 1. um contexto futebolístico em que o uso da palavra **ÁREA** é vago na seleção entre ‘pequena **ÁREA**’/‘grande **ÁREA**’, por exemplo.

#### ex. 2.11

1. *O Corinthians vai executar um bombardeio aéreo sobre a **ÁREA** do Palmeiras esta noite no Pacaembu, na primeira partida das finais do Brasileiro.*
2. *O Botafogo endoidou de vez, subiu todo e Gonçalves perdeu um gol na pequena **ÁREA**.*

<sup>1</sup>O termo “vagueza” é utilizado e definido de diferentes maneiras por diferentes autores. Em nosso texto, o mais relevante é o conceito de vagueza que se refere à ausência de especificação em certos tipos de substantivos. Mas o termo apresenta na literatura outras aplicações que não coincidem necessariamente com esta. Em (Moura 1999), por exemplo, o termo é usado sobretudo em relação a adjetivos e conceituado como “um dos tipos de indeterminação semântica, e ocorre quando não se pode determinar (em função do próprio conteúdo semântico) se uma determinada palavra se aplica ou não a determinados objetos, gerando proposições indefinidas quanto ao valor de verdade. Por exemplo, ‘Bill Clinton é gordo’ pode ser verdadeira ou falsa, dependendo da definição de ‘gordo’ ” (Moura 1999, p. 58).

Já entre os enunciados de 2.12 tem-se uma ambigüidade por haver uma extensão do sentido de um ‘evento’ para o ‘local no campo’ onde o evento ocorre.

**ex. 2.12**

1. *Minha namorada teve o desprante de me confessar que não viu o GOL de Marcelinho.*
2. *O resultado foi apenas um chute contra o GOL do Flamengo, enquanto Romário e Marques perderam boas oportunidades.*

Um aspecto fundamental na análise da distinção entre vagueza e polissemia é o contexto lingüístico, como enfatizam Ravin & Leacock, que utilizam o termo **indeterminação** no lugar de vagueza:

“The distinction between polysemy and indeterminacy is at the core of semantic theory as it defines the relation between the semantics of linguistic expressions and the extralinguistic entities to which these expressions refer.” (Ravin & Leacock 2000, p.3)

Segundo Geeraerts (1993), a distinção entre polissemia e vagueza envolve a determinação da proveniência das informações semânticas, entre a estrutura semântica subjacente à palavra ou a especificação contextual, no nível pragmático. Há uma grande diversidade de propostas para confrontar a questão de que porção do significado de uma palavra se mantém invariável em diversos contextos e - simetricamente - que outros aspectos se mantêm indeterminados até que a palavra seja contextualizada.

Por um lado, para Goddard (2000), grande parte dos problemas encontrados no tratamento da polissemia são devidos a métodos equivocados de definição lexical. Sob este ponto de vista, o autor acredita que, com a representação (definicional) adequada, a palavra teria o maior conteúdo semântico possível, e esse conteúdo permaneceria invariável nos diversos contextos. Na verdade, o contexto apenas enriqueceria o conteúdo semântico da palavra, mas nunca o alteraria.

No outro extremo, (Schütze 1998) dispensa essa porção invariante de uma definição de sentido, estipulando que o conteúdo semântico da palavra é reconhecido através da similaridade dos contextos em que ocorre. O ponto de vista do autor, que toma um modelo quantitativo do contexto como a representação do significado de uma expressão, é uma das bases do tratamento experimental dados aos substantivos de suporte no capítulo 5. Nos exemplos 2.13 e 2.14, a semelhança entre os contextos, pela presença de CORPO e INÉRCIA, agrupariam os sentidos das duas ocorrências de MATÉRIA.

**ex. 2.13** *A massa depende da quantidade de MATÉRIA do corpo, pois existe uma relação direta entre a quantidade de MATÉRIA e a inércia: os corpos que possuem grande quantidade de MATÉRIA possuem grande inércia, e vice-versa.*

**ex. 2.14** *Inércia, por sua vez, é a propriedade da MATÉRIA em virtude da qual um corpo tende a resistir a qualquer mudança em seu movimento (ou repouso).*

A posição intermediária rejeita tanto a relação rígida entre item lexical e sentido, como a idéia de que o sentido seja inteiramente proveniente do contexto, a partir de inferências pragmáticas ou textuais. É postulado um nível de interação entre o conteúdo lexical e o contexto (Moura 2002).

Como exemplo dessa visão de significado, Cruse (2000) concebe o significado de uma palavra como um agregado de conteúdo conceitual que se apresenta com o uso da palavra dentro do contexto. O que o autor chama de “nó de sentido” é uma unidade relativamente autônoma que desempenha um papel independente em vários processos semânticos, tendo comportamento estável com respeito a diferentes contextos. Assim, mesmo postulando a existência de propriedades semânticas invariantes, Cruse admite a impossibilidade de determinar qualquer dessas propriedades de maneira descontextualizada.

Para Cruse (2004), essa questão está localizada no quadro mais amplo da variedade contextual do significado lexical. O autor identifica duas propriedades de leituras variantes de uma palavra: a **discreção** (*discreteness*) e o **antagonismo**. Duas leituras são discretas se existe uma divisão semântica clara entre as duas; são antagônicas se competem pela escolha do falante/ouvinte, não admitindo que ambas sejam selecionadas ao mesmo tempo (nem mesmo temporariamente). O antagonismo seria o critério para o estabelecimento da ambigüidade. Com respeito à relação entre discreção e antagonismo, Cruse afirma:

“It may be presumed that antagonistic readings are *ipso facto* also discrete and therefore that antagonism represents the highest degree of distinctness.” (Cruse 2004, p. 106)

#### 2.3.4

##### Testes de ambigüidade

Os testes propostos para o diagnóstico da polissemia em contraste com a vagueza, dividem-se em três critérios metacognitivos: o critério lingüístico, o lógico e o definicional. Esses critérios são expostos e criticados em (Geeraerts 1993; Kilgarriff 1997; Dunbar 2001; Ravin & Leacock 2000). A seguir, discuto e exemplifico alguns desses testes.

**Critério lingüístico:** pressupõe que não é possível invocar vários sentidos de uma palavra polissêmica em uma única ocorrência desta palavra. Os testes agrupados sob esse critério buscam construções lingüísticas que exijam identidade denotacional entre dois elementos sintáticos, por exemplo, um elemento anafórico e seu antecedente. A aceitabilidade do enunciado resultante indica monossemia, como no caso do exemplo 2.15, onde a elipse de *ÁREA* na coordenação de *ÁREA COMERCIAL* com *ÁREA CULTURAL* não impossibilita a interpretação do enunciado. Já no exemplo 2.16, a zeugma do terceiro enunciado é inaceitável, indicando a polissemia.

**ex. 2.15**

1. *Restrito, por enquanto, às questões da ÁREA comercial, o juiz arbitral pode tornar os processos mais baratos e mais ágeis.*
2. *Para Nascimento Silva, a desativação do CNDA foi parte do desmonte da ÁREA cultural promovido pelo governo Collor.*
3. *Ian Robinson, 37, e Julia Dias, 29, atuam na ÁREA comercial e cultural do Consulado da Austrália em São Paulo.*

**ex. 2.16**

1. *Na revenda de veículos da Mesbla da avenida do Estado, a ÁREA sob a marquise foi cercada com grades.*
2. *“Até porque os escândalos na ÁREA da saúde têm mostrado os fantasmas”, disse.*
3. *? [a ÁREA sob a marquise e a da saúde estão sob inspeção judicial.]*

**Critério lógico:** pressupõe que uma afirmação não pode ser falsa e verdadeira ao mesmo tempo para o mesmo referente. No exemplo 2.17, a leitura ambígua entre o concreto – ‘região’ – e o abstrato – ‘setor de atividade’ – permite que a afirmação possa ser falsa para uma leitura e verdadeira para a outra.

**ex. 2.17** *A Associação Comercial de Campo Grande realiza de amanhã a 21 deste mês um seminário para discutir e levantar subsídios para o projeto de revitalização da ÁREA comercial da cidade.*

Uma outra maneira de utilização do critério lógico é a construção de uma afirmação aparentemente contraditória, que só será semanticamente consistente quando a predicação ou a referência forem ambíguas, como ilustra o exemplo 2.18.

**ex. 2.18**

1. *Real Madrid, o Barcelona e La Coruña, com um PONTO a menos, jogam em seus estádios.*
2. *Mas o PONTO central não está aí.*
3. *[O PONTO não é o PONTO.]*

**Critério definicional:** pressupõe que, se a palavra é polissêmica, não há como formular um conjunto de condições necessárias e suficientes que cubram todos os conceitos expressados pela palavra, uma definição que contenha o significado inerente à palavra, independente do contexto em que se insere. Esse é o critério mais clássico de identificação de polissemia, exibindo a tendência à generalização comum no processo de categorização humano.

Geeraerts (1993) critica os testes de ambigüidade observando que, de maneira geral, os julgamentos metacognitivos seriam tendenciosos pois estariam sujeitos a diversos tipos de influências externas. Além disso, os testes seriam inconsistentes. Por um lado, existem palavras capazes de passar pelo critério lingüístico em certas ocasiões mas não pelo definicional. Como exemplo, o autor analisa a palavra JORNAL em exemplos como 2.19. Admitindo a aceitabilidade do exemplo, conclui-se que o elemento anafórico SUA e o antecedente JORNAL tenham a mesma leitura. Tal fato parece ser conflitante com o uso do verbo DECIDIR, mais afim à leitura de JORNAL como ‘grupo de pessoas’, e do complemento TIPOGRAFIA, mais afim à leitura ‘publicação/produto’. Seria muito improvável acomodar ambos os sentidos em uma só definição.

**ex. 2.19** *[O JORNAL decidiu mudar sua tipografia.]*

Ou seja, os testes não concordam entre si. Por outro lado, o mesmo teste pode ser manipulado para gerar resultados diferente, o que pode ser observado no exemplo 2.20: o primeiro enunciado causa estranheza, enquanto o segundo parece superar o problema através de uma reescrita simples.

**ex. 2.20**

1. *? [A TESE é instigante e amarelada pelo tempo.]*
2. *[A TESE ainda é instigante embora esteja amarelada pelo tempo.]*

As críticas aos critérios parecem dirigidas muito mais à metodologia de aplicação dos testes do que à sua utilidade informativa. Como argumenta Cruse, os testes colocam os sintomas da polissemia em evidência. Tendo em vista que a ambigüidade se manifesta em um espectro contínuo, não há como esperar que haja testes infalíveis e definitivos que decidam os limites claros entre o ambíguo e o não ambíguo. Em (Cruse 2000) os testes são utilizados cumulativamente, assumindo que quanto mais critérios são satisfeitos maior o grau de separação entre os sentidos. Em (Cruse 2004), tendo definido os conceitos de discreção e antagonismo, Cruse restabelece os testes de ambigüidade, ressaltando que na verdade eles identificam leituras discretas da palavra.

A discussão sobre como se relacionam polissemia e vagueza, e conseqüentemente, sobre o papel do contexto na interpretação semântica das unidades lexicais, é de grande importância no estudo das expressões com palavras de suporte, e será retomada no capítulo 4. Por um lado, essas palavras que ocorrem na função de suporte - os verbos e os substantivos - são polissêmicas por admitirem, no mínimo, o sentido pleno e o sentido de suporte (cf. exemplos 2.18 2.16). Por outro lado, o sentido de suporte é vago pois exige complementação - seja do substantivo, seja do adjetivo - e participa pouco do significado da expressão.