



**Claudia Maria Garcia Medeiros de Oliveira**

## **O Substantivo-suporte**

### **Critérios Operacionais de Caracterização**

#### **Tese de Doutorado**

Tese apresentada ao Programa de Pós-graduação em Letras do Departamento de Letras da PUC-Rio como requisito parcial para obtenção Do título de Doutor em Letras

Orientador: Prof. Margarida Basilio

Rio de Janeiro  
Setembro de 2006

# **Livros Grátis**

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.



**Claudia Maria Garcia Medeiros de Oliveira**

## **O Substantivo-suporte**

### **Critérios Operacionais de Caracterização**

Tese apresentada ao Programa de Pós-graduação em Letras do Departamento de Letras do Centro de Teologia e Ciências Humanas da PUC-Rio como requisito parcial para obtenção Do título de Doutor em Letras. Aprovada pela Comissão Examinadora abaixo assinada.

**Prof. Margarida Basilio**

Orientador  
Departamento de Letras - PUC-Rio

**Prof. Violeta de San Tiago Dantas Barbosa Quental**

Departamento de Letras - PUC-Rio

**Prof. Helena Franco Martins**

Departamento de Letras - PUC-Rio

**Prof. Heronides Maurílio de Melo Moura**

Departamento de Língua e Literatura Vernáculas - UFSC

**Prof. Maria Carlota Amaral Paixão Rosa**

Departamento de Letras Vernáculas - UFRJ

**Prof. Paulo Fernando Carneiro de Andrade**

Coordenador Setorial do Centro de Teologia e Ciências  
Humanas - PUC-Rio

Rio de Janeiro, 1 de Setembro de 2006

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

### **Claudia Maria Garcia Medeiros de Oliveira**

Graduou-se em Computação no ITA (São José dos Campos, SP) em 1986. Obteve o título de Mestre em Sistemas e Computação em 1990, pelo Instituto Militar de Engenharia (Rio de Janeiro); e de PhD em Computação em 1995, pelo Imperial College (Londres, Reino Unido). Áreas de interesse: Estudos do Léxico, Lingüística Computacional e Inteligência Artificial.

#### Ficha Catalográfica

Oliveira, Claudia

O Substantivo-suporte: Critérios Operacionais de Caracterização / Claudia Maria Garcia Medeiros de Oliveira; orientador: Margarida Basilio. — 2006.

118 f: il. ; 30 cm

1. Tese (Doutorado em Letras) - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2006.

Inclui bibliografia

1. Letras – Teses. 2. Substantivo-suporte. 3. Substantivo. 4. Adjetivo denominal. 5. Classe de palavras. 6. Lexicografia de corpus. 7. Lingüística. I. Basilio, Margarida. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Letras. III. Título.

CDD: 400

## Agradecimentos

À professora Margarida Basilio, pela confiança e pelo respeito que sempre teve por minhas idéias e minha pesquisa. Considero um privilégio ter trabalhado com essa lingüista influente, extremamente produtiva e inovadora, que fez com que eu me apaixonasse pelos estudos do léxico.

À PUC-Rio pela bolsa de estudos concedida e pela infra-estrutura do campus que me ofereceu.

Ao Departamento de Letras, onde descobri a Lingüística como disciplina científica, tendo sido guiada por professores e pesquisadores admiráveis.

Aos amigos do Clic, Claudia Freitas, Milena Garrão, Luiz Alexandre Amaral, Cícero Santos, Violeta Quental e Maria Carmelita Dias, pelas discussões e idéias compartilhadas, e pelas parcerias na pesquisa em Lingüística Computacional, que contribuíram imensamente para esta tese.

Aos meus alunos, todos os que já tive, por perguntarem. Em especial, a Gilberto Silva por ter-me proposto um tema de pesquisa que fez com que eu buscasse o conhecimento lingüístico e que assim iniciasse esse novo caminho.

Mais uma vez, ao Departamento de Letras e sua infra-estrutura administrativa, em particular à Chiquinha, por suavizar as dificuldades e tornar quase imperceptível a burocracia inerente à gerência de uma unidade de ensino e pesquisa.

A Roger Troth, meu marido, pelo apoio, incentivo e por todas as pequenas coisas que são tudo. Não devo agradecer à Clara e à Raquel, minhas filhas, pois chegaram no meio do trabalho e atrapalharam muito, por terem se tornado o novo centro do meu mundo.

## Resumo

Oliveira, Claudia; Basilio, Margarida. **O Substantivo-suporte: Critérios Operacionais de Caracterização**. Rio de Janeiro, 2006. 118p. Tese de Doutorado — Departamento de Letras, Pontifícia Universidade Católica do Rio de Janeiro.

Este trabalho tem por objetivo prover um critério operacional para caracterizar substantivos em combinações de substantivo seguido de adjetivo, em que o substantivo se apresenta em situação análoga à dos chamados verbos leves ou verbos-suporte, largamente estudados em Lingüística e Processamento de Linguagem Natural nos últimos anos. O trabalho se situa na confluência entre estudos lingüísticos, lexicográficos e computacionais e pretende explorar a potencialidade da análise automática de corpora e instrumentos quantitativos em busca de uma maior objetividade na fundamentação de conceitos que norteiam a atividade de análise lingüística. O desenvolvimento da pesquisa alia a pesquisa em corpus ao dicionário tradicional para realizar o levantamento das principais propriedades das combinações *S-Adj* particularizado para o caso de ocorrência de adjetivos denominais. A partir das informações lexicográficas e contextuais demonstra-se a existência de um conjunto de substantivos que participam das construções estudadas de maneira semelhante aos verbos-suporte em combinação *V-SN*. Um método automático de reconhecimento dos substantivos-suporte em textos é elaborado com o objetivo de fornecer aos estudiosos um instrumento capaz de produzir evidências convincentes dada a insuficiência de julgamentos intuitivos para justificar a delimitação de expressões de aparente irregularidade.

## Palavras-chave

Substantivo-suporte; Substantivo; Adjetivo denominal; Classe de palavras; Lexicografia de corpus; Lingüística.

## Abstract

Oliveira, Claudia; Basilio, Margarida. **Support Nouns: Operational Criteria for Characterization**. Rio de Janeiro, 2006. 118p. PhD Thesis — Department of Letras, Pontifícia Universidade Católica do Rio de Janeiro.

The main goal of this work is to provide operational criteria for characterizing nouns in Noun - Adjective combinations, in which the noun occurs in an analogous way to so called light verbs or support verbs, widely studied in recent years in both Linguistics and Natural Language Processing. In the work, linguistic, lexicographic and computational studies converge in order to explore the potential for automatic analysis of corpora, whose aim is to provide quantitative tools and methods which would lead to a more objective way of establishing concepts which underlie linguistic analysis. The work unites corpus-based research with traditional lexicography in order to elicit the main properties of the N-Adj combinations occurring with denominal adjectives. The lexicographic and contextual data reveal the existence of a set of nouns that occur in the studied constructions in a way similar to light verbs in V-Noun phrasal combinations. An automatic method for recognizing support nouns in texts is developed, which will provide language specialists with an instrument capable of bringing solid evidence to add to intuitive judgments in the task of justifying the delimitation of expressions that are apparently irregular.

## Keywords

Support noun; Noun; Denominal adjective; Part of speech ; Corpus lexicography; Linguistics.

## Sumário

1	Introdução	<b>12</b>
1.1	Caracterização do problema	13
1.2	Objetivos	15
1.3	Posicionamento Metodológico	16
1.4	Organização do texto	17
2	Quadro Teórico	<b>19</b>
2.1	Introdução	19
2.2	Modelos lexicais	19
2.3	Polissemia	30
3	Classes de Palavras	<b>42</b>
3.1	Introdução	42
3.2	Propostas de Classificação	43
3.3	Considerações sobre a classe dos substantivos	46
3.4	Considerações sobre a classe dos adjetivos	52
3.5	Fronteiras entre classes de palavras	62
4	A Conceituação do Substantivo-suporte	<b>67</b>
4.1	Introdução	67
4.2	Lexicografia do Substantivo-suporte	67
4.3	A Função de Suporte: o Verbo e o Substantivo	83
4.4	Exemplário	93
5	O substantivo-suporte no corpus	<b>96</b>
5.1	Introdução	96
5.2	Composicionalidade semântica	96
5.3	O Método	98
5.4	Experimentos	104
6	Conclusões	<b>108</b>
6.1	Contribuições	108
6.2	Desdobramentos	109
6.3	Considerações Finais	110
	Referências Bibliográficas	<b>111</b>



## Lista de figuras

2.1	Rede de conexões para <i>cat</i> , <i>cats</i> , <i>rats</i> , <i>mats</i> e <i>caps</i>	23
2.2	Frame para QUESTION no FrameNet	28
2.3	Linhas contínuas unindo entre homonímia/monossemia e polilexia/vagueza	34
4.1	Entrada do dicionário Houaiss para COISA	73
4.2	Diagrama das referências cruzadas entre definições de substantivos-suporte	76
4.3	Entrada do dicionário Houaiss para FORMA	77
4.3	Entrada do dicionário Houaiss para FORMA (cont.)	78
4.4	Diagrama das referências cruzadas entre definições de substantivos-suporte (reiteração da figura 4.2)	90
5.1	Resumo do método de identificação de substantivos-suporte em contexto	104

## Lista de tabelas

1.1	Linguística baseada em corpus vs. linguística dirigida por corpus	17
3.1	Classes de palavras na tradição grega (esq.) e latina (dir.)	43
3.2	Motivações para as classes de palavras	44
3.3	Lista de substantivos genéricos do inglês	49
4.1	Lista de substantivos-suporte do Português	69
4.2	Número de adjetivos distintos em sintagmas <i>N Adj</i> com substantivos-suporte	72
4.3	Número de sentidos dicionarizados	73
4.4	Número de expressões multi-vocabulares dicionarizadas	74
4.5	Referências cruzadas em entradas do dicionário para substantivos-suporte	75
4.6	Acepções de <i>FORMA</i> , com exemplos do corpus	78
4.6	Acepções de <i>FORMA</i> , com exemplos do corpus (cont.)	79
4.7	Acepções de <i>FORMA</i> , com categorias de distinção	80
4.8	Resultados de concordância	82
4.9	Paráfrases para <i>S-Adj</i> com substantivos-suporte	92
4.9	Paráfrases para <i>S-Adj</i> com substantivos-suporte (cont.)	93
5.1	Expressões <i>S Adj</i> testadas	105
5.2	Resultados com as expressões <i>S Adj</i>	105
5.3	Resultados com <i>biológico</i>	106
5.4	Resultados com <i>cultural</i>	106

## Convenções Tipográficas

MAIÚSCULAS PEQUENAS

Para exemplos de palavras e expressões.

*Itálico*

Para exemplos de contextos.

**Negrito**

Para terminologia, quando for introduzida.

“Aspas duplas”

Para citações de outros autores.

‘Aspas simples’

Para exemplos de sentidos.

[Colchetes]

Para exemplos não retirados do corpus.

? Interrogação

Para estranheza semântica ou sintática.

\* Asterisco

Para impossibilidade semântica ou sintática.

*O físico Leo Szilard certa vez anunciou ao amigo Hans Bethe que estava pensando em escrever um diário: “Não pretendo publicar. Só vou registrar os fatos para a informação de Deus”. “Você não acha que Deus sabe dos fatos?”, Bethe perguntou. “Sim”, respondeu Szilard. “Ele sabe dos fatos mas não desta versão dos fatos”*

**Hans Christian von Baeyer**, *Taming the Atom* (1992).

# 1

## Introdução

Este trabalho tem por objetivo prover um critério operacional para caracterizar substantivos em combinações *S-Adj*, em que o substantivo se apresenta em situação análoga à dos chamados verbos leves ou verbos-suporte, largamente estudados em Lingüística e Processamento de Linguagem Natural nos últimos anos. O trabalho se situa na confluência entre estudos lingüísticos, lexicográficos e computacionais e pretende explorar a potencialidade da análise automática de corpora e instrumentos quantitativos em busca de uma maior objetividade na utilização e evidenciação de conceitos que norteiam a atividade de análise lingüística.

A utilização de análises lingüísticas baseadas em corpus vem ampliando as possibilidades de detecção de padrões construcionais das línguas. O apoio computacional disponível viabiliza o trabalho de pesquisa em corpora gigantescos, de grande cobertura e variabilidade textual. As principais características metodológicas da pesquisa baseada em corpus são (Biber, Conrad, & Reppen 1998):

- é uma pesquisa empírica, com base em padrões de uso, efetivamente produzidos, em textos reais;
- utiliza uma coleção de textos, o **corpus**, caracterizável por uma ou mais dimensões textuais (língua, gênero, registro, datas, etc);
- é extensivamente apoiada por computadores, de forma totalmente automática ou interativamente;
- utiliza mecanismos quantitativos e qualitativos de análise dos dados.

Do ponto de vista operacional, construções complexas, restritas por condições gramaticais de diversas naturezas, podem ser codificadas para que sejam identificadas em contexto, liberando o lingüista da tarefa enfadonha da busca manual de exemplos. O resultado da busca automática é mais consistente, pois se dá de maneira homogênea ao longo do tempo, o que é extremamente difícil para o lingüista individualmente e, ainda mais, em trabalhos cooperativos.

O léxico é um sistema dinâmico, em constante rearranjo, que não só armazena formas significantes, mas também fornece processos de produção de palavras e de expressões. O delineamento da classe de substantivos-suporte proposto no trabalho pressupõe um modelo de léxico que dê conta de construções regulares e semi-regulares, tais como as construções verbais com verbo-suporte, dentro de um quadro teórico que rejeita a separação entre a sintaxe e o léxico. Os modelos lexicais computacionais inspirados em teorias lingüísticas de orientação funcionalista, tais como (Fillmore 1976), oferecem um quadro adequado para a utilização dos resultados desta pesquisa em sistemas computacionais.

## 1.1

### Caracterização do problema

Na lingüística, o termo **substantivo vazio** é utilizado para fazer referência a substantivos que não denotam conceitos, ou substantivos com um conteúdo semântico mínimo, identificados, na análise sintática, com elementos sem expressão fonológica que atuam como núcleos de sintagmas nominais. De acordo com Panagiotidis (2003), no léxico de qualquer língua existe um número limitado de substantivos vazios, que formam uma classe fechada. Eles são considerados palavras gramaticais, não lexicais, com forte característica pronominal. Tendo em vista sua reduzida capacidade de denotação, substantivos vazios distinguem-se entre si por meio de seus traços fonológicos e morfo-semânticos, tais como o gênero. O substantivo vazio típico do inglês seria ONE, como em *big ONE*.

Na teoria lexical, tem havido muito pouco interesse no fenômeno, com algumas exceções tais como Schmid (2000), que apresenta um estudo sobre substantivos abstratos. De acordo com Schmid, há um subconjunto dos substantivos abstratos em inglês que atuam como conchas conceituais, que realizam seu potencial semântico em associação com outros itens lexicais. Palavras como THING, FACT, CASE, POINT, IDEA, REASON, PROBLEM e QUESTION, entre outros substantivos de alta frequência no inglês, são exemplos dessa classe.

Estudos lexicais aplicados ao processamento automático de textos ampliaram essa noção, incluindo palavras que podem ser apagadas ou ignoradas para o propósito de sumarização, classificação, agrupamento e outras operações computacionais sobre textos. Substantivos ocorrendo como quantificadores em expressões multi-vocabulares, tais como GROUP, BUNCH e LOTS, em *GROUP of students*, *BUNCH of students*, *LOTS of students*, são exemplos especificamente mencionados em (Muresan, Tzoukermann, & Klavans 2001) para o inglês. O caso geral, em que esse tipo de substantivo ocorre legitimamente como núcleo

de um sintagma nominal, não foi explorado.

No âmbito da Recuperação de Informações, a indexação automática do texto completo exige operações de compressão de texto tais como apagamento de palavras vazias, palavras que não contribuam para a precisão nem para a abrangência nas operações de busca. Tradicionalmente, palavras vazias são listas de palavras funcionais e verbos auxiliares. Substantivos-suporte não são palavras vazias, a julgar por esse critério usual.

O conceito de **densidade lexical** (Halliday 1985) subjaz essa decisão pela eliminação de palavras vazias na prática da Recuperação de Informações. Considerando os exemplos 1.1, o enunciado 2. seria mais “denso” que o 1. por expressar o mesmo significado proposicional de modo mais compacto.

### ex. 1.1

1. *É a coisa da ausência de interesse cultural.*
2. *É a ausência de interesse cultural.*

Apesar de adotar a tradicional divisão entre classes lexicais e gramaticais (cf. capítulo 3) Halliday admite o caráter fronteiro de certos substantivos e verbos. O autor também propõe que a frequência relativa de uma palavra pode indicar a quantidade de informação que ela traz para o enunciado.

Existe um conjunto não muito extenso de substantivos do português que se caracterizam por sua generalidade semântica. Os primeiros questionamentos sobre este tipo de palavra surgiram a partir de uma pesquisa de corpus que realizei em busca de sintagmas nominais *S-Adj*, onde *Adj* é um adjetivo denominal resultado de um processo  $[X]_S \rightarrow [[X]_S \text{ al}]_{Adj}$ . O resultado da pesquisa mostrou que, combinados aos mais variados adjetivos, sempre havia termos, como FATOR, PERSPECTIVA e ASPECTO, que são aparentemente intercambiáveis, apesar de terem significados literais bastante distintos.

$$\left. \begin{array}{l} \text{fator} \\ \text{perspectiva} \\ \text{aspecto} \end{array} \right\} \text{ ambiental} \quad \left. \begin{array}{l} \text{fator} \\ \text{perspectiva} \\ \text{aspecto} \end{array} \right\} \text{ racial} \quad \left. \begin{array}{l} \text{fator} \\ \text{perspectiva} \\ \text{aspecto} \end{array} \right\} \text{ fiscal}$$

Os dados mostram que a contribuição desses termos para o significado geral da expressão é bastante reduzida. As noções de significado e de contribuição de um item para o significado de uma expressão são, no mínimo, problemáticas, mas um apelo preliminar ao entendimento pré-teórico dessas noções parece confirmar que em ASPECTO AMBIENTAL a idéia mais proeminente é de ‘meio-ambiente’.

A identificação de tais substantivos, aqui denominados de **substantivos-suporte**, é muito importante dentro do contexto de interpretação automática

de textos e suas aplicações. Se a base do processamento é a palavra e o sintagma, então é fundamental que a computação do significado da expressão não seja prioritariamente baseada no substantivo, mas sim em seus complementos.

## 1.2

### Objetivos

O principal objetivo deste trabalho é a delimitação do substantivo-suporte por meio de suas características lexicográficas, funcionais e textuais.

A descrição lexicográfica do substantivo-suporte foi feita com propósitos especulativos e prospectivos. O conjunto inicial de substantivos-suporte foi obtido de uma forma empírica e assistemática e o dicionário foi muito útil em uma primeira tentativa de caracterizar esses substantivos como uma classe.

A proposta de enquadrar o substantivo-suporte dentro de um fenômeno mais amplo na linguagem, a função de suporte, é uma consequência da observação das ocorrências dos sintagmas nominais *S-Adj*, onde *S* pertence a um conjunto de substantivos de grande generalidade semântica. As características da construção remetem à descrição dos verbos leves apontadas por (Jespersen 1940) e (Poutsma 1926). Semelhante aos verbos-suporte, que realizam a função de suporte para o substantivo em sintagmas verbais, proponho que o substantivo-suporte realize a função de suporte nos sintagmas nominais. Dessa forma, proponho um visão lexical dos sintagmas *S-Adj*, enquadrando-os como expressões multi-vocabulares (EMV).

A questão da constituição do léxico vem à tona nessa proposta na definição da unidade lexical e na constatação de que a noção de palavra, da maneira como é delimitada tradicionalmente, não é adequada para definir o conjunto das unidades simbólicas básicas da língua. A perspectiva da análise da **construção** aproxima-se do modelo lexical da Gramática Funcional, onde a unidade é a oração, com a função de construir um modelo da experiência e das relações lógicas (significado ideacional), de realizar interação social (significado interpessoal) e conferir relevância ao contexto (significado textual) (Halliday 1994; Neves 2004). Mesmo admitindo, contrariamente a algumas vertentes do funcionalismo, que haja fronteiras possíveis entre a sintaxe e a morfologia, o trabalho é desenvolvido em torno de um modelo de léxico que prevê a formação de unidades maiores que a palavra morfológica, como explicitado em (Basilio 2005):

“[...] a cada passo adiante na investigação sobre unidades lexicais, mais os dados me forçam a concluir que o léxico, em seu papel de produzir e armazenar formas significativas a serviço dos macro-sistemas de significação e comunicação que são as línguas, utiliza



processos tanto morfológicos quanto sintáticos para a formação de suas unidades simbólicas básicas.”

Feito o enquadramento do substantivo-suporte como palavra de suporte, partiu-se para a sistematização da delimitação da construção **Substantivo-suporte–Adjetivo denominal**, tendo-se como base uma medição da composicionalidade semântica baseada em corpus. A valorização da ocorrência da construção no corpus marca uma posição similar à de J. R. Firth, que confere um papel primordial aos **eventos de fala (speech acts)** em sua Teoria Contextual do Significado, cf. (Newmeyer 1998). Em objeção à postulação do estruturalismo Saussureano de que instâncias da “parole” não passam de meras evidências para a estrutura da “langue”, Firth afirma que os eventos de fala são o objeto principal da lingüística, sendo elementos concretos, em contraste com a “langue” de Saussure que seria “um sistema de valores diferenciais, não de termos concretos e positivos” (Firth 1968), apud (Joseph, Love, & Taylor 2001). Metodologicamente, o interesse de Firth pelo fenômeno da colocação ficou registrado em sua frase “You shall know a word by the company it keeps”.

### 1.3

#### Posicionamento Metodológico

J. Sinclair, em uma aula sobre Lexicologia, Lexicografia e Lingüística Computacional, em Singapura, 1996, distinguiu duas posturas metodológicas frente à pesquisa em corpus (Ooi 1998), como mostra a tabela 1.1.

A distinção entre abordagens baseadas em corpus e dirigidas por corpus se assemelha ao contraste entre as abordagens **top-down** e **bottom-up** de resolução de problemas, utilizadas tradicionalmente nas disciplinas de Programação de Computadores e Engenharia de Software. No primeiro caso, o processo é analítico e os conceitos mais gerais da teoria do problema, suas abstrações de mais alto nível, são utilizadas para iniciar a análise. Os dados são utilizados em última instância, na confirmação, extensão ou rejeição da teoria. Por outro lado, a abordagem bottom-up inicia-se com os dados e, em processos de síntese, formulam a teoria que abstrai e generaliza a informação inerente aos dados.

Na prática da pesquisa lingüística, embora não na teoria, uma mistura das duas metodologias é invariavelmente necessária. No caso de uma pesquisa interdisciplinar, que busca meios lingüísticos de atingir objetivos computacionais, assim como prover meios computacionais para adicionar aos instrumentos de análise lingüística, a convergência das metodologias pode se acentuar. No entanto, minha pesquisa e seu encaminhamento tenderam mais à síntese. Em retrospectiva, foram longos períodos de observação do corpus, procurando

---

## Lingüística baseada em corpus    Lingüística dirigida por corpus

---

o corpus é utilizado para validar, verificar e melhorar observações lingüísticas que já tenham sido realizadas

o lingüista não questiona posições teóricas pre-estabelecidas ou categorias descritivas aceitas; sua posição com respeito à estrutura da língua já se estabilizou

o corpus é utilizado para ajudar a estender e melhorar a descrição lingüística

um exemplo de questão relevante: WHOM ainda é utilizado em inglês? como?

um corpus é de importância essencial no surgimento de novas idéias de como examinar os dados

o lingüista acredita que pode conciliar o tipo de evidências que emerge do corpus com as posições estabelecidas; ele deixa abertas as possibilidades de mudanças radicais na teoria para lidar com as evidências

a evidência do corpus é soberana portanto o lingüista minimiza os pressupostos sobre a natureza das categorias teóricas e descritivas

um exemplo de questão relevante: a distinção entre gramática e léxico é necessária?

---

Tabela 1.1: Lingüística baseada em corpus vs. lingüística dirigida por corpus

realizar a demarcação paulatina do fenômeno e, finalmente, formular uma proposta de caracterização da construção em foco.

Entretanto, não se pode negar que esta fase foi posterior a vários cursos de Teoria Lexical e a um estudo detalhado dos adjetivos denominais, em que se menciona a relevância da sua combinação com substantivos de semântica geral. Por outro lado, tanto na Teoria Lexical quanto na Lingüística Computacional, a questão da delimitação das unidades lexicais e dos verbos-suporte vem sendo permanentemente discutida. Neste sentido, poderíamos dizer que a pesquisa tende mais à pesquisa baseada em corpus.

### 1.4

#### Organização do texto

No capítulo 2, discuto os diferentes aspectos teóricos envolvidos no tratamento lexical e lexicográfico de unidades lexicais, assim como questões relacionadas ao significado dos itens lexicais, com especial ênfase na questão polissemia / vagueza, crucial no entendimento do conceito de substantivo-suporte.

No capítulo 3, abordo a questão das classes de palavras e apresento uma visão geral das classes envolvidas no trabalho: Substantivo e Adjetivo.

Os capítulos 4 e 5 são centrais na execução dos objetivos da tese. O capítulo 4 concentra-se na conceituação do substantivo-suporte, por meio de propriedades lexicográficas e textuais. A noção de substantivo-suporte é posicionada em paralelo à do verbo-suporte.

O capítulo 5 apresenta os resultados experimentais da análise de sintagmas *S-Adj* para fundamentar a proposição de um mecanismo objetivo de identificação em corpora da entidade substantivo-suporte.

O capítulo de conclusões discute os resultados alcançados e problemas remanescentes, e aponta para algumas linhas de continuação que pretendo seguir nesta área de investigação.

## 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**.

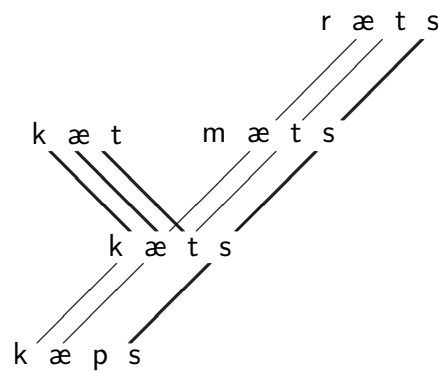


Figura 2.1: Rede de conexões para *cat*, *cats*, *rats*, *mats* e *caps*

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.

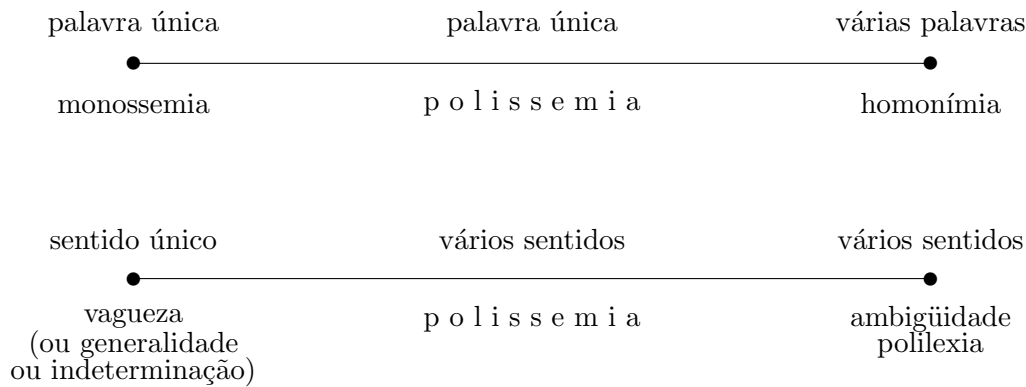


Figura 2.3: Linhas contínuas unindo entre homonímia/monosssemia e polillexia/vagueza

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.

## 3 Classes de Palavras

*As far as the laws of mathematics refer to reality, they are not certain, and as far as they are certain, they do not refer to reality. (Albert Einstein)*

### 3.1 Introdução

A categorização da palavra de acordo com traços que a posicionam dentro do sistema lingüístico é uma idéia presente na grande maioria dos modelos lexicais. Essa categorização é um elemento imprescindível na caracterização das construções em foco.

Neste capítulo são discutidos os principais aspectos da delimitação das chamadas **classes de palavras**. A seção 3.2 traz uma apresentação geral do tema, a respeito do qual os pontos de vista são teoricamente divergentes mas terminam gerando as propostas de classificação bastante semelhantes. Nas seções 3.3 e 3.4 as classes dos substantivos e dos adjetivos são analisadas em maior detalhe por serem mais centrais neste trabalho.

Ao analisar o substantivo, passo rapidamente pelos critérios de caracterização da classe e apresento em mais detalhe as diferentes sub-classificações que tangenciam de alguma forma a classe dos substantivos-suporte: os substantivos **genéricos** (Halliday & Hasan 1976) e as **conchas nominais** (Schmid 2000).

Já com relação ao adjetivo, discuto inicialmente a própria legitimidade da classe apresentando diversos critérios de caracterização e exemplificando. A distinção entre adjetivos predicadores e não predicadores e os processos morfológicos derivacionais que formam adjetivos são de particular importância pois meu foco é o **adjetivo relacional denominal**.

Na seção 3.5, discuto algumas relações importantes entre as classes de palavras relevantes nesta tese:

- o trânsito entre a classe dos substantivos e a dos adjetivos, muitas vezes fundidas na classe dos **nomes**;

- o trânsito entre a classe verbal e a nominal, por meio do processo de **nominalização**;
- as semelhanças entre verbos e adjetivos com respeito à predicação.

### 3.2

#### Propostas de Classificação

A separação das unidades lexicais em **classes de palavras** faz parte da visão aristotélica de linguagem. A vasta maioria dos modelos lexicais utilizaram essa idéia. A classificação das palavras em escaninhos rotulados, obedecendo a critérios que lhes atribuem propriedades em comum faz parte de uma tradição gramatical herdada do grego e do latim. É surpreendente, como mostra a tabela 3.1, como as categorizações clássicas perduram e podem ser encontradas nas mais modernas descrições gramaticais, (cf. Priscianus Charisius), apud (Jurafsky & Martin 2000; Biderman 1978).

<i>ὄνομα</i>	→	nome		<i>nomina</i>	→	nomes
<i>ἀντωνυμία</i>	→	pronome		<i>pronomina</i>	→	pronomes
<i>ῥήμα</i>	→	verbo		<i>verba</i>	→	verbos
<i>ἐπίρρημα</i>	→	advérbio		<i>adverbia</i>	→	advérbios
<i>μετοχή</i>	→	particípios		<i>participia</i>	→	particípios
<i>ῥύνδεσμος</i>	→	conjunções		<i>coniunctiones</i>	→	conjunções
<i>πρόθερις</i>	→	preposições		<i>praepositiones</i>	→	preposições
				<i>interiectiones</i>	→	interjeições
<i>ἄρθρον</i>	→	artigo				

Tabela 3.1: Classes de palavras na tradição grega (esq.) e latina (dir.)

Rosa (2000) discute extensivamente diversas propostas históricas de categorização das palavras e suas motivações, observando que esses esquemas se mantiveram com poucas alterações significativas (o adjetivo é uma delas, o que será discutido no seção 3.4). Entretanto, nota que:

“[...] a classificação das palavras deixou de basear-se em critérios semânticos e passou a ter por fundamentos critérios distribucionais, funcionais e sua categorização. A diferença de foco está, até certo ponto, refletida na nomenclatura: o uso da expressão **classe de palavra**, em lugar de **parte do discurso**, procura assinalar a ruptura com as noções que norteavam os estudos tradicionais.” (Rosa 2000, p. 99)

Utilizo o termo **classes de palavras** e, com apoio em (Câmara Jr 2000) e (Basilio 1999c), considero que é possível adotar três critérios na sua categorização: o critério semântico, o morfológico (pelas categorias flexionais

<b>aspecto semântico-sintático</b>	<b>aspecto ontológico</b>	<b>tradição gramatical</b>
sujeitos	entidades	substantivos
predicados	qualidades	adjetivos
predicadores	ações	verbos

Tabela 3.2: Motivações para as classes de palavras

que apresentam) e o sintático-funcional. Lyons (1977) apresenta, em dois paralelos, as motivações principais para as classes de palavras, como mostra a tabela 3.2

O critério semântico ou nocional é o motivador mais tradicional para a classificação das palavras. A grosso modo, os **substantivos** designam as pessoas, os objetos ou as situações, os **verbos** designam processo, os **adjetivos** designam as qualidades dos substantivos e os **advérbios** designam as qualidades dos verbos ou dos adjetivos (Dubois *et al.* 2001). As **preposições** e **conjunções** indicam relações lógicas entre outros elementos do discurso, os **artigos** determinam os substantivos e os **pronomes** os substituem.

Essa tradição da categorização nocional é duramente criticada por formalistas, que consideram a abordagem imprecisa e não preditiva, além de não se mostrar apropriada para expressar generalizações gramaticais que não se adequem aos limites das classes de palavras existentes. Emonds (1987) propõe um abandono irrestrito das classificações tradicionais em favor da abordagem formalista gerativa, como mostra a seguinte passagem.

“Overall, I think it must be concluded that this tightly arranged and interlocking system of bar notation categories is a tremendous advance over the *ad hoc*, loosely defined, overlapping, and generally unenlightening systems of traditional grammar. As such it is entirely reasonable and feasible that traditional classifications now give way to the more rational, cohesive, and essentially simple bar notation system discovered by generative grammar, at any educational level at which elements of syntax are taught.”

Por outro lado, Lyons (1977), mesmo acreditando na possibilidade de formulação de “procedimentos diagnósticos” de cunho sintático para o traçado preciso e definitivo de uma classificação vocabular, afirma que o interesse nesses procedimentos deve ser motivado pelo estabelecimento de propriedades semânticas das classes resultantes. Além disso, Lyons observa que:

“The fact that there appears to be a positive correlation in all languages between syntactically defined and semantically defined

expression-classes would tend to support the traditional view that there is a high degree of interdependence between the syntactic structure of sentence-nuclei and the semantic function of their constituent expressions. Despite what has been said at times by certain linguists there is no reason to doubt that the traditional view is, to this extent at least, well-founded.” (Lyons 1977, p. 438)

Apesar da procedência de uma das principais críticas às classificações tradicionais – a de que são inaplicáveis, em todos os seus detalhes, a línguas cujas estruturas gramaticais diferem significativamente das línguas indo-européias – existe uma quase unanimidade com respeito à distinção entre verbo e substantivo. Os verbos e os substantivos são classes tradicionalmente destacadas como as essenciais na gramática. Parecem ser classes indispensáveis na enunciação, possuindo realidade psicolinguística e tendo sido identificadas universalmente em extensos estudos trans-lingüísticos (Rosa 2000). Sapir, um conhecedor de numerosas e variadas línguas norte-americanas, reafirma a essencialidade dos verbos e nomes na seguinte passagem:

“There must be something to talk about and something must be said about this subject of discourse once it is selected. [...] The subject of discourse is a noun. As the most common subject of discourse is either a person or a thing, the noun clusters about concrete concepts of that order. As the thing predicated of a subject is generally an activity in the widest sense of the word, a passage from one moment of existence to another, the form which has been set aside for the business of predicating, in other words, the verb, clusters about concepts of activity. No language wholly fails to distinguish noun and verb [...]” (Sapir 1921, p.117)

Mesmo que o texto de Sapir não seja formalmente preciso (Lyons (1977) observa que deve haver uma distinção entre “nome” e “expressão nominal”, por exemplo), o autor consegue estabelecer o paralelo entre as classes verbal e nominal e os processos de predicação e referenciação, respectivamente.

Diversos autores estruturalistas, mais notadamente Bloomfield (1926), rejeitaram a universalidade das categorias dos nomes e dos verbos (ou de qualquer outra categoria); no entanto, contemporaneamente, o consenso a esse respeito tem sido restabelecido principalmente por estudos da psicologia cognitiva e da psicolinguística (Laudanna & Voghera 2002).

O conjunto das classes de palavras pode ser ainda subdividido para refletir determinadas propriedades das classes. Rosa (2000) discute a distinção entre

classes de palavras com **significado lexical** e com **significado gramatical**. Se por um lado é problemático admitir que algumas palavras isoladamente são portadoras de significado, e não outras, por outro lado, há variantes a essa oposição que terminam subdividindo as classes de palavras de maneira semelhante: palavra de conteúdo  $\times$  palavra de forma, palavra lexical  $\times$  palavra gramatical, palavra plena  $\times$  palavra vazia, contentivo  $\times$  functor, vocábulo forma  $\times$  vocábulo conectivo (Câmara Jr 2000). Um outro critério de subdivisão do conjunto das classes de palavras é quanto ao potencial de gerar vocabulário: as **classes abertas** e as **classes fechadas**.

A atribuição de classes de palavras é fundamental para muitas das aplicações do Processamento de Linguagem Natural, tais como *parsing* sintático e semântico, análise do discurso e processamento de fala. Para atender à necessidade computacional de lidar com classes menos ambíguas, ou seja, com menor interseção entre seus membros, os critérios utilizados para a categorização tendem a se basear em distribuição sintática e morfologia.

A precisão da classificação também é um requisito computacional importante, por isso recursos computacionais tais como bases de dados lexicais e corpora anotados com POS – do inglês *part of speech* – tendem a utilizar classes de maior granularidade. Por exemplo, o Penn Treebank (Marcus, Santorini, & Marcinkiewicz 1993) utiliza 45 classes, incluindo, junto com as clássicas, rótulos diferenciados para palavras estrangeiras, nomes próprios, símbolos fora do alfabeto, entre outros. Os verbos são também rotulados quanto à morfologia da ocorrência (forma básica, passado, gerúndio, particípio e terceira pessoa do singular), assim como os substantivos (singular e plural). Outros conjuntos chegam a ter centenas de classes.

Com o objetivo de estabelecer os aspectos fundamentais que participam da caracterização do substantivo e do adjetivo enquanto classes de palavras do português, apresento a seguir uma visão geral de cada aspecto dessas classes.

### 3.3

#### Considerações sobre a classe dos substantivos

##### 3.3.1

##### A caracterização do substantivo

Os substantivos formam uma classe primária, conforme discutido na seção 3.2, presentes em todos os sistemas de classificação de palavras, seja qual for o critério adotado. Além disso, os substantivos são classificados como palavra lexical, pertencentes a uma classe aberta.

Tradicionalmente, aos substantivos é atribuída primordialmente a função

de sujeito de uma proposição, ou seja, é a parte da oração à qual se atribui uma predicação. Secundariamente o substantivo pode exercer a função de complemento objeto de verbos (Biderman 1978). Mesmo quando pretende definir o substantivo como “a palavra com que designamos ou nomeamos os seres em geral”, Cunha & Cintra adicionam que:

“Do ponto de vista funcional, o substantivo é a palavra que serve, privativamente, de núcleo do sujeito, do objeto direto, do objeto indireto e do agente da passiva. Qualquer palavra de outra classe que desempenhe uma dessas funções equivalerá forçosamente a um substantivo.” (Cunha & Cintra 1985, p. 171)

De maneira mais formal, (Huddleston 2000) enumera três propriedades compartilhadas pelos elementos mais centrais da classe dos substantivos:

**Potencial Funcional I.** Funcionam como núcleos na estrutura de sintagmas nominais (SNs). Os SNs, por sua vez, realizam uma variedade de funções, dentre as quais destacam-se a de sujeito, complemento objeto ou predicativo na estrutura clausal, complemento em um sintagma preposicional;

**Potencial Funcional II.** Como núcleos de SNs, tem como termos dependentes membros de outras classes de palavras, principalmente determinantes e adjetivos.

**Flexão** Possuem as categorias flexionais de número (singular ou plural) e às vezes gênero (feminino ou masculino).

Não se encontram caracterizações que fujam muito desse esquema nas gramáticas tradicionais (Bechara 1980; Cunha 1972; Cunha & Cintra 1985; Rocha Lima 1998; Said Ali 2001), que se estendem, além disso, nos sistemas flexionais e outras características da morfologia do substantivo.

Uma distinção muito comum nas gramáticas tradicionais é feita entre os substantivos **concretos** e os **abstratos**. A subclasse dos substantivos abstratos será utilizada por (Schmid 2000) na caracterização das **conchas nominais**, na seção 3.3.3, por isso comento brevemente as definições propostas, porém ressalvo que são definições simples demais para serem utilizadas como critérios de delimitação.

Said Ali (2001) justifica a classe dos abstratos pela necessidade de nomear atributos de atributos, como beleza, tristeza, simpatia, etc.

“Os atributos, posto que sejam inerentes aos seres, são considerados muitas vezes como se existissem separados deles, como se fossem



outras entidades. Os substantivos que os representam chama-se abstratos; são concretos os nomes de referência direta aos seres.”  
(Said Ali 2001, p. 47)

Bechara (1980) introduz a noção de “existência independente”, para definir o substantivo concreto como o que “designa ser de existência independente: casa, mar, sol, automóvel, filho, mãe”; o abstrato designa ser de existência dependente”, tais como ações, estados e qualidades, “considerados fora dos seres, como se tivessem existência individual”.

### 3.3.2 Substantivos genéricos

Uma das características do substantivo-suporte é o grau de generalidade, ou falta de especificidade, com que é utilizado para denominar. Normalmente, denomina-se de substantivo genérico ao nome usado para fazer referência a um membro típico de um grupo. Por exemplo:

**ex. 3.1** *O ESTUDANTE de sétima série que estiver usando este manual para estudo poderá observar em seu próprio corpo as transformações progressivas provocadas pela atividade hormonal.*

apresenta um uso genérico de ESTUDANTE, enquanto que

**ex. 3.2** *Mas o golpe de 1964 e a situação política acabaram obrigando o jovem ESTUDANTE a trocar os livros, no segundo ano de faculdade, pelo microfone.*

apresenta um uso não genérico do mesmo substantivo.

Nesse sentido, qualquer substantivo pode ser utilizado no sentido genérico. Entretanto, a classe de substantivos genéricos, como definida por (Halliday & Hasan 1976) constitui-se em um pequeno conjunto de substantivos que possuem referência generalizada dentro de uma grande classe de substantivos, tais como “human noun”, “place noun” e “fact noun”. O substantivo genérico encontra-se no limite entre uma palavra lexical - classe aberta - e uma palavra gramatical - classe fechada. No exemplo a seguir, o sujeito UMA PACIENTE é reiterado pela superclasse A MULHER.

**ex. 3.3** *Uma PACIENTE em coma desde 1985 deu à luz a um bebê prematuro, de sete meses, com boas chances de sobrevivência . A MULHER, de 29 anos, foi estuprada e o hospital só descobriu que ela estava grávida quatro meses e meio depois, quando sua barriga começou a crescer e a idéia de ser obstrução digestiva foi afastada .*

Halliday e Hasan situam essa categoria de substantivos dentro do contexto de coesão textual, onde muitas vezes é necessário reiterar algo que já tenha sido mencionado previamente, como mostra o exemplo a seguir.

**ex. 3.4** *Romário sentiu a musculatura da coxa ao tentar participar do treino de ontem. O FATO levou os médicos do Flamengo a marcarem um exame de ressonância magnética para hoje, onde será detectado qual o real problema de Romário.*

Essa forma de reiteração envolve o uso de um nome de classe que inclua o que se pretende reiterar. Segundo Halliday & Hasan:

“The general principles behind this is simply that demonstratives, since (like other reference items) they identify semantically and not grammatically, when they are anaphoric require the explicit repetition of the noun, or some form of synonym, if they are to signal exact identity of specific reference; that is, to refer unambiguously to the presupposition at the identical level of particularization. A demonstrative without a following noun may refer to some more general class that includes the presupposed item.”  
(Halliday & Hasan 1976, p.64-65)

Uma lista de substantivos genéricos para o inglês pode ser vista na tabela 3.3.

Classe	Exemplos
humano	person, man, woman, child, boy, girl
não-humano animado	creature
inanimado concreto	thing, object
inanimado concreto contínuo	stuff
inanimado abstrato	business, affair, matter
ação	move
lugar	place
fato	question, idea

Tabela 3.3: Lista de substantivos genéricos do inglês

Outro tipo de substantivo genérico, encontrado em diversas línguas, são os **marcadores de lugar (place-holder words)**, um conjunto de palavras genéricas, utilizadas com propósitos pragmáticos e discursivos. Os usos mais comuns desse tipo de palavra são:

- substituir palavras esquecidas;

- substituir palavras tabu;
- fazer referência a algo desconhecido, atípico ou censurável.

Entre os marcadores mais comuns em português encontram-se: COISA, NEGÓCIO, TRECO, LANCE, TREM, SUJEITO, ELEMENTO, INDIVÍDUO, FULANO, SICRANO e BELTRANO. As gírias são ricas em marcadores, como por exemplo CARA e PARADA, e até mesmo expressões como NÃO SEI DAS QUANTAS.

Esse tipo de substantivo se parece com o substantivo-suporte com relação ao seu valor semântico vago, no entanto, distribucionalmente difere bastante. Os marcadores de posição possuem um caráter pronominal muito mais acentuado.

### 3.3.3

#### Conchas Nominais

Schmid (2000) apresenta uma classe de substantivos conceitualmente semelhantes ao substantivo-suporte que queremos caracterizar. São substantivos abstratos denominados **conchas nominais (shell-nouns)**.

Do ponto de vista gramatical, as conchas nominais podem ser encontradas nas seguintes construções:

1. Determinante + (Modificador) + Substantivo + cláusula-THAT ou cláusula-WH ou TO-infinitivo

**ex. 3.5** *The (deplorable) FACT that I have no money.*

Em português: Determinante + (Modificador) + Substantivo + (Modificador) + de + cláusula-QUE ou infinitivo

**ex. 3.6** *O método rítmico se baseia no FATO de que a ovulação ocorre 14 dias antes do dia de início da próxima menstruação.*

2. Determinante + (Modificador) + Substantivo + be + cláusula-THAT completiva ou cláusula-WH ou TO-infinitivo

**ex. 3.7** *The (big) PROBLEM was that I had no money.*

Em português: Determinante + (Modificador) + Substantivo + (Modificador) + ser + cláusula-QUE completiva ou infinitivo

**ex. 3.8** *O único PROBLEMA é que às vezes o pimpolho conta as anedotas mais calientes bem na presença dos sisudos amigos do poderoso papai.*

Considerando-se três classes nominais, os substantivos plenos, as conchas nominais e os pronomes com função anafórica, Schmid traça uma comparação entre o potencial de caracterização, o potencial de formação de conceito e o potencial de ligação dessas palavras.

Os substantivos plenos possuem potencial de caracterização detalhada, com denotação estável, capazes de denotar pessoas, objetos, animais, plantas, atividades, eventos, propriedades e circunstâncias. Os pronomes com função anafórica, tais como os pronomes retos, caracterizam os referentes com respeito a um número limitado de dimensões: gênero, número, pessoa e eventual humano versus não humano. As conchas nominais ficam em uma posição intermediária, pois o falante só consegue utilizá-la para caracterizar uma experiência de maneira muito geral e os detalhes da informação são providos pelo chamado **conteúdo da concha (shell content)**.

Com respeito ao potencial de formação de conceito, os substantivos plenos se relacionam estavelmente com o conceito, ou conceitos, relacionados ou não, que encapsulam. Pronomes, por outro lado, estão temporariamente ocupando o lugar de outras instâncias de conceitos. As conchas nominais novamente se colocam entre esses dois polos: exibem uma relação conceitual constante com um certo tipo de experiência, por serem substantivos abstratos, mas são temporários pois seus conteúdos mudam de acordo com o contexto lingüístico. Schmid considera que o conceito propiciado por uma concha nominal consiste de “uma parte simbólica estável e uma parte variável indexal”.

O potencial de ligação de um item lexical é tipicamente demonstrado pela função de pronomes anafóricos: eles indicam que dois elementos lingüísticos devem ser interpretados como dependentes. Os substantivos plenos não possuem essa capacidade, a não ser no sentido de estabelecer relações semânticas, no contexto de coesão lexical e substantivos genéricos (Halliday & Hasan 1976). As conchas nominais se aproximam bem mais dos pronomes anafóricos nesse aspecto.

Podemos identificar as conchas nominais com os substantivos-suporte com respeito ao potencial de caracterização e ao potencial de formação de conceito. O potencial de ligação descrito por Schmid é dependente do contexto gramatical em que o substantivo ocorre, e é justamente nesse ponto que as duas classificações de substantivos se separam. Estamos observando as construções *S + Adj*, bastante diversas das construções típicas de conchas nominais. Não há, em nosso caso, a função de ligação ou coesão textual.

A classe dos substantivos admite uma profusão de classificações, principalmente quanto a aspectos morfológicos, semânticos e pragmáticos. Os recortes que foram apresentados são baseados em diferentes composições desses

aspectos, motivados por uma visão funcionalista dessa classe essencial na descrição gramatical.

### 3.4

#### Considerações sobre a classe dos adjetivos

##### 3.4.1

##### A caracterização do adjetivo

Tradicionalmente nas línguas indo-européias, o adjetivo é distintivamente considerado como uma classe-de-palavras, ainda que as definições tradicionais para as classe-de-palavras sejam inadequadas para caracterizar os conjuntos que as representam. Segundo Quirk *et al.* (1978), mesmo assim

“[...] because of their general currency, it is convenient to continue to refer to adjectives [...] in English as word-classes but we must be aware that they do not constitute well-defined classes and, moreover, that neither class is homogeneous”. (Quirk *et al.* 1978, p.231)

Em linhas gerais, o adjetivo é conceituado como um modificador do substantivo, mesmo que, como reconhece Lyons (1977) “there are [...] some adjectives for which this statement is not valid; and there are other [...] adjectives for which its validity is questionable”. Algumas gramáticas do inglês, como (Nesfield 1907), consideram adjetivos todas as palavras que podem ocorrer como modificadores de substantivos, tais como artigos, numerais, pronomes demonstrativos, possessivos e indefinidos.

Nas línguas românicas, os adjetivos se aproximam mais dos substantivos do que no inglês, por exemplo, principalmente por seguirem o mesmo padrão flexional dos substantivos. Sendo assim, Said Ali chama de nome as “palavras com que se designam os seres e seus atributos”, subclassificando entre eles os adjetivos como denominadores de atributos. As gramáticas tradicionais do português, tais como (Nunes 1960) e (Said Ali 2001), abordam a classe dos nomes apenas sob o ponto de vista das categorias gramaticais que lhe são próprias, tais como número, gênero e gradação.

Câmara Jr (2000) justifica uma divisão das palavras em nomes, verbos e pronomes, pelo que chama de critério morfo-semântico, já que dentro destas grandes classes há uma certa uniformidade flexional e semântica. Entre os nomes encontram-se os adjetivos, que se diferenciam dos substantivos e dos advérbios por serem “termos determinantes de outro nome”. Neste ponto, Câmara Jr reconhece a dependência essencial do adjetivo com relação ao substantivo determinado, utilizando um critério funcional.

As tentativas de caracterização da classe dos adjetivos foram divididas abaixo entre as que utilizam um critério primariamente semântico e as que utilizam um critério primariamente sintático. Embora a interligação entre estas duas visões seja de grande interesse lingüístico, possivelmente por uma relação em que o aspecto semântico seja a causa do comportamento sintático, não me estendo nesta discussão, recomendando o texto (Lyons 1977) para o leitor interessado.

### O critério semântico

A classe dos adjetivos é abordada nas gramáticas tradicionais do português, com diferentes graus da profundidade. Rocha Lima (1998) caracteriza o adjetivo tão somente como “a palavra que restringe a significação ampla e geral do substantivo”.

Para Cunha & Cintra (1985), há dois tipos de adjetivos. O primeiro modifica o substantivo para caracterizar os seres, os objetos ou as noções por ele nomeadas, indicando-lhes uma qualidade, modo de ser, aspecto ou aparência ou um estado. O segundo tipo de adjetivo modifica o substantivo para estabelecer com este uma relação de tempo, espaço, matéria, finalidade, propriedade, procedência, etc. Denominados de adjetivos de relação, são de natureza classificatória, ou seja, precisam o conceito expresso pelo substantivo, restringindo-lhe a extensão do significado.

Bechara (1980) apresenta uma análise bem mais elaborada das funções do adjetivo, definido como a classe de lexema que se caracteriza por constituir a delimitação, isto é, caracterizar as possibilidades designativas do substantivo, orientando delimitativamente a referência a uma parte ou a um aspecto do denotado. A palavra-chave da definição de Bechara é delimitação, que apresenta três modalidades: a explicação, a especialização e a especificação.

Apesar da preocupação com distinções semânticas possivelmente importantes, a análise das funções dos adjetivos proposta por Bechara é de difícil emprego, já que os critérios utilizados são de extrema subjetividade e de formulação imprecisa.

### O critério sintático-funcional

No português, todos os adjetivos obedecem a concordância de gênero e número com o sintagma que complementa ou modifica:

#### ex. 3.9

1. *[As meninas são bonitas.]*
2. *[Os garçons são meio lerdos.]*

Além desta forte característica sintática, podem-se arrolar as seguintes propriedades, que, segundo (Huddleston 2000), os membros mais centrais da classe dos adjetivos possuem.

1. Os adjetivos ocorrem como núcleo em sintagmas que funcionam como complemento predicativo em uma oração: uso predicativo. Podem ser predicativos do sujeito, como em

**ex. 3.10 (Maria é bonita.)**

ou predicativos do objeto, como em

**ex. 3.11 (João considera Maria bonita.)**

A posição de aposto também é possível em alguns casos, como

**ex. 3.12**

(a) [*Bonita, a moça era muito assediada.*]

(b) [*Motivado, João era o meu melhor aluno.*]

Considero este tipo de ocorrência como predicativa, pois o adjetivo aposto vale semanticamente pela redução de uma oração em que o mesmo ocorre como predicativo, como pode ser observado em

**ex. 3.13**

(a) [*Por ser bonita, a moça era muito assediada.*]

(b) [*Quando estava motivado, João era o meu melhor aluno.*]

2. Os adjetivos ocorrem como núcleo em sintagmas que funcionam como modificadores de sintagmas nominais: uso atributivo ou adnominal, como em

**ex. 3.14**

(a) [*A menina bonita*]

(b) [*O garçom meio lerdo*]

A posição primária do adjetivo atributivo é pós-nominal, embora algumas vezes possa ocorrer em posição pré-nominal (A BONITA MENINA). Algumas vezes o pré-posicionamento do adjetivo acarreta alterações semânticas sensíveis, como mostra o contraste entre MEU VELHO AMIGO e MEU AMIGO VELHO.

No português existe ainda um tipo de ocorrência do adjetivo, minoritária é bem verdade, que não pode ser reduzida a 1 ou 2. É o caso de adjetivos em uma estrutura *Adj de Det N*, como em

**ex. 3.15**

(a) [*A idiota da menina*]

(b) [*O safado do meu cunhado*]

Nesta posição os adjetivos apresentam diversas peculiaridades, como por exemplo, não admite gradação e são em grande parte pejorativos.

3. Os adjetivos podem sofrer gradação, seja através de intensificadores, como advérbios de intensidade, seja através dos graus comparativo e superlativo, seja através dos graus diminutivo e aumentativo.

No português, existem algumas poucas formas sintéticas para o adjetivo comparativo, como MAIOR e MENOR para o adjetivo GRANDE. Em geral, o grau comparativo é obtido pelo uso dos modificadores MAIS, MENOS e TÃO, como em MAIS BONITA. O grau superlativo realiza-se, em alguns adjetivos, pela adição do morfema -íssimo, como em BELÍSSIMO, entre outros.

Nem todos os adjetivos ocorrem livremente nos contextos descritos em 1, 2 e 3. Por exemplo, as propriedades 1 e 3 não se aplicam a adjetivos denominais, como é discutido a seguir e, em maiores detalhes, em (Basilio, Oliveira, & Garrão 2003).

### 3.4.2

#### Subclasses de adjetivos

Raskin & Nirenburg (1995) apresentam uma revisão de diversas taxonomias para os adjetivos, elaboradas em diferentes épocas, tendo como foco diferentes línguas. As características que comparecem em todas as taxonomias, de uma maneira ou de outra, envolvem dimensão, propriedade física, cor, idade, valor, velocidade, etc. Estas características podem ser agrupadas de diversas maneiras, como por exemplo, relacionando-as segundo os sentidos perceptuais humanos.



Concordo com Raskin quando considera que a questão central que distingue os adjetivos entre si, de importância tanto sintática quanto semântica, é a diferença entre adjetivos predicadores e não predicadores. A denominação utilizada por Raskin é adjetivo **qualitativo** para o primeiro caso e adjetivo **relacional** para o segundo. Conforme (Basilio, Oliveira, & Garrão 2003), encontra-se majoritariamente nos trabalhos voltados para o português a denominação adjetivo **predicativo** (equivalente ao qualitativo) e **denotativo** (equivalente ao relacional).

As propostas de análise das funções do adjetivo denotativo que examinamos aqui têm como ponto em comum o fato de que, ora o adjetivo adiciona propriedades ao substantivo, ora restringe-lhe o referente. Esses comportamentos distintos podem ser observados nas propriedades sintáticas do sintagma nominal onde ocorre o adjetivo.

Perini (1978) parte da constatação de que alguns sintagmas nominais S Adj não podem ser transformados para orações relativas como em

esperam a bênção papal → \*esperam a bênção que é papal

Os adjetivos que compõem tais sintagmas, a que ele chama de *denominais*, com frequência podem ser transformados em sintagmas preposicionais como em

esperam a bênção papal → esperam a bênção do papa

Perini observa que uma análise morfológica dos adjetivos denominais revela a predominância da base substantiva, o que, entretanto, não determina as propriedades sintáticas do sintagma. Sua proposta é que a ocorrência do adjetivo seja interpretada como **referencial** ou **atributiva**. A distinção fica clara na comparação entre as seguintes ocorrências do adjetivo PRESIDENCIAL.

### ex. 3.16

1. [*A atitude PRESIDENCIAL foi desfavorável.*] (*atitude tomada pelo presidente*)
2. [*Ele vive assumindo atitudes PRESIDENCIAIS.*] (*atitude típica de um presidente*)

No primeiro caso a interpretação é referencial, pois o adjetivo é usado para fazer referência ao substantivo base; no segundo caso a interpretação é atributiva, pois o adjetivo atribui ao substantivo núcleo do sintagma as propriedades próprias da classe determinada pelo substantivo base do adjetivo.

Em (Basilio & Gamarski 1995), as autoras organizam o quadro das funções dos adjetivos nos sintagmas nominais, de modo a caracterizá-las gramaticalmente. A função de **preenchimento argumental** tem caráter sintático, a função **denotativa** tem caráter semântico lexical e a função **predicativa** tem caráter proposicional. Basilio & Gamarski observam que o adjetivo em função denotativa denota uma nova “classe natural” através da combinação de suas propriedades denotativas com as do substantivo que modifica. A motivação para a formação desse tipo de adjetivos é a potenciação das possibilidades denotativas da língua. Nesse artigo as autoras ressaltam a relevância da formação de adjetivos denominais denotativos dado o seu caráter fundamentalmente lexical, já que desempenham as funções de representar conceitos e fornecer elementos para a construção de enunciados.

Bastos (1980) examina os adjetivos denominais combinados a nominalizações de verbos, observando o papel de complemento do adjetivo. A primeira distinção apontada por Bastos se dá entre a interpretação classificativa – que classifica objetivamente a nominalização – a interpretação circunstancial – que determina o substantivo expressando circunstância – e a interpretação qualificativa – que atribui às nominalizações uma qualidade subjetiva. As interpretações classificativa e circunstancial são agrupadas como especificativas.

A partir dessa classificação, Bastos procura estabelecer critérios para um relacionamento argumental entre a nominalização e o adjetivo denominal. Para a autora os adjetivos podem ter uma função **subjetiva** se a base substantiva for o sujeito da base verbal da nominalização, como por exemplo em PENSAMENTO FREUDIANO. Por outro lado, o adjetivo pode ser **objetivo** se for o objeto, como em REFORMA TRIBUTÁRIA.

A proposta de (Lobato 1993) considera que a denotação é uma relação entre as propriedades semânticas abstratas de uma expressão – sua intensão – e o conjunto de referentes em potencial dessa mesma expressão – sua extensão. As diferenças entre os substantivos comuns e os adjetivos são, além das de natureza distribucional sintática, decorrentes do fato de que, embora ambas as categorias remetam a um conjunto de propriedades semânticas abstratas, somente nos substantivos a intensão pode levar à extensão.

Segundo Lobato, o adjetivo denotativo acrescenta propriedades semânticas às propriedades denotativas da expressão nominal a que ele se refere, daí a denotação do sintagma dá-se como uma conjunção de predicados. Já o adjetivo predicativo atribui propriedades semânticas ao referente denotado pela expressão nominal que ele modifica, resultando em uma leitura proposicional.

Segundo (Levi 1978), cujo propósito é evidenciar que adjetivos denomi-

nais são transformacionalmente derivados dos substantivos derivantes, existem características típicas de substantivo apresentadas pelo adjetivo não predicativo, sendo as essenciais:

- não gradatividade;
- não podem ser coordenados a adjetivos predicativos;
- quantificação por prefixos bi-, tri-, mono-, poli-, uni-, multi-, por exemplo, MULTI-DIMENSIONAL corresponderia ao sintagma VÁRIAS DIMENSÕES;
- impossibilidade de nominalização.

Muitos adjetivos morfologicamente denominais funcionam predicativamente, como pode ser visto no contraste entre “NOTA MUSICAL” e “CRIANÇA MUSICAL”. Esta duplicidade é acompanhada de um fenômeno de flutuação semântica reconhecido, mas pouco explicado na literatura.

Vilela & Silva (2003) propõem uma abordagem para o estudo dos adjetivos baseada na teoria da lingüística cognitiva, onde o foco são os princípios da sua própria categorização, ou seja, as propriedades prototípicas dos adjetivos. Assim, os autores buscam responder “Quais são as propriedades prototípicas dos adjetivos? Sob quês condições uma seqüência não é gramatical por causa da posição do adjetivo com relação ao substantivo?”. Segundo essa perspectiva, é possível estabelecer uma distinção entre dois grupos adjetivais de acordo com sua proximidade ao protótipo: (i) os adjetivos nucleares ou prototípicos; (ii) os adjetivos periféricos ou não-prototípicos. Segundo os autores, adjetivos nucleares:

- são morfologicamente simples;
- são unidimensionais (aceitam grau);
- são sincategoremáticos (dependem do substantivo para realizarem sua interpretação);
- pertencem à classe semântica das propriedades (a partir de uma frase *S Adj*, pode-se responder “Como é S? Adj.”);
- são atributivos ou predicativos;
- ocorrem em posição pós ou pré-nominal.

Neste sentido, Vilela & Silva posicionam o adjetivo predicativo entre os mais prototípicos e os denotativos entre os mais periféricos.

### 3.4.3

#### Morfologia do adjetivo

Deixando de lado a morfologia flexional dos adjetivos, que atua na formação de flexões de gênero e número, esta seção aborda a morfologia lexical, que dá conta da formação dos adjetivos derivados.

Existe um número considerável de processos de sufixação que resultam em adjetivos. Com base substantiva, o sufixo mais produtivo é o -al/-ar, principalmente por ser o mais neutro semanticamente, não apresentando restrições de uso quanto às bases. É adicionado a radicais latinos, primitivos ou derivados com a estrutura Xção, Xmento e Xncia, tais como EDUCACIONAL, PARLAMENTAR e EMERGENCIAL. Menos produtivo, o sufixo -ário adiciona-se a bases latinas, a partir de bases presas e formas livres ou derivadas com estrutura Xção, Xmento, Xncia e Xidade, tais como INFLACIONÁRIO e UNIVERSITÁRIO. O sufixo -ico carrega maior formalidade, sendo mais aplicado a radicais gregos e, com frequência, a bases presas, como em TECNOLÓGICO.

Entre os sufixos com conteúdo semântico encontram-se aqueles com o significado “provido de”. Os mais neutros deles são -oso e -ado. O primeiro adiciona-se a radicais latinos, primitivos ou derivados com a estrutura Xmento, Xncia e Xidade, como em MEDICAMENTOSO, SUBSTANCIOSO e CARIDOSO. O segundo adiciona-se a substantivos não derivados, muitas vezes em formações parassintéticas, como descrito mais à frente. O sufixo -ento tem caráter claramente pejorativo, como em SARNENTO e GRUDENTO. O sufixo -udo traz a idéia de ‘exageradamente provido de’, adicionando-se muitas vezes a partes do corpo como NARIGUDO e POPOZUDO.

Outro grupo de sufixos semanticamente não vazios é o de formadores de adjetivos pátrios ou denominadores de origem. Pertencem a este grupo os sufixos -ense, -ês e -ano, bastante produtivos a partir de nomes próprios de países, cidades, etc, tais como CEARENSE, JAPONÊS e BAIANO. Este último também pode indicar autoria, como em MACHADIANO e portadores de signos do zodíaco, como CAPRICORNIANO.

Entre os processos majoritários de formação de adjetivos a partir de verbos temos a formação V-do a partir de verbos transitivos, como CUMPRIDO, ou intransitivos, como NASCIDO. Em muitos casos as formações V-do são controversas, pois pode se tratar de flexões verbais do particípio ou conversões de verbos para adjetivos. Outras formações incluem a sufixação por -nte, -tivo, -tório e -dor, como ANDANTE, DISSERTATIVO, ACUSATÓRIO e REVELADOR, com variadas interpretações semânticas dependentes do verbo.

Com uma semântica mais estável, as formações com o sufixo -vel indicam a qualidade de afetado potencial pela ação verbal ou caracterizado pelo po-

tencial do verbo, como FLEXÍVEL, INFLAMÁVEL e VARIÁVEL. Alguns exemplos de base substantiva, como RAZOÁVEL, SAUDÁVEL e MISERÁVEL apresentam grande irregularidade semântica.

Finalmente, há a formação de adjetivos a partir de adjetivos e substantivos pelos sufixos dos graus aumentativo, diminutivo e superlativo. Muitas vezes, estes processos possuem caráter expressivo muito mais que denotativo, como em CANSADAÇO e APARTAMENTAÇO. Da mesma maneira, grande parte dos diminutivos podem indicar pejoratividade. Há também o sufixo -oso, que embora participe de um processo de formação de adjetivos denominais, também aparecem combinados a adjetivos, como em FEIOSO.

Processos de prefixação, por não envolverem mudança categorial, também geram adjetivos a partir de adjetivos. Por seu alto grau de produtividade destacam-se os prefixos negativos des- e in-; os prefixos de gradação como super-, ultra-, hiper-, sub-, bem-; os prefixos que indicam relações espaço-temporais, como pre-, pos-, inter-, trans-, super- e sub-.

O processo de derivação parassintética ocorre quando há adição simultânea de prefixo e sufixo a uma base, tanto verbal quanto nominal. Do ponto de vista morfológico, a verificação da parassíntese [a [b] c] ocorre quando não se identificam palavras correspondentes [a [b]] ou [b [c]]. Entre os processos mais produtivos podemos citar: as formações des-X-ado, com X substantivo, indicando desprovido de X (DESMIOLADO); a-X-ado, com X adjetivo ou substantivo, indicando aproximadamente X (ACINZENTADO, APRESSADO); en-X-ado, com X substantivo, indicando provido de X (ENDINHEIRADO); in-X-vel, com X verbo, indicando a negação de X-vel (IMPRATICÁVEL).

Um tipo curioso de formação de adjetivo é a conversão, que consiste na transposição de um item lexical de uma classe para outra sem alteração de sua forma fonológica. Alguns exemplos bastante recentes deste tipo de formação, como MODELO e PADRÃO mostram a vitalidade do processo.

#### 3.4.4

##### O adjetivo no enunciado

Adjetivos predicativos podem ser modificados, principalmente com respeito à gradação. Os advérbios de intensidade são seus modificadores naturais, como MUITO, MAIS, MENOS, TOTALMENTE, ABSOLUTAMENTE, etc.

Dada a sua natureza predicativa, adjetivos podem semanticamente corresponder a predicados n-argumentais. Isto se traduz sintaticamente em complementos, muito embora não se possa atribuir-lhes a mesma variedade e riqueza da complementação verbal. Os complementos mais simples são os sintagmas nominais, como em

**ex. 3.17**

1. *feliz da vida*
2. *irritados com José*

Orações podem fazer o papel de modificador e de complemento. Observando os seguintes enunciados:

**ex. 3.18** *João estava velho para ser presidente.*

**ex. 3.19** *João estava ansioso para ser presidente.*

percebe-se que em 3.18 temos semanticamente uma comparação entre a idade de João e uma idade esperada para o presidente, e neste caso considera-se a oração infinitiva como um modificador do adjetivo. Em 3.19 a oração infinitiva corresponde a um argumento na estrutura do predicado semântico de ANSIOSO, configurando-se, portanto, um caso de complementação.

Os adjetivos deverbais muitas vezes herdaram a estrutura argumental dos verbos derivantes, como discute Gamarski (1995), em especial para os adjetivos da forma Xnte. Isto pode ser notado nos exemplos discutidos em (Borba 1996):

**ex. 3.20**

1. *É estranhável (estranha-se) que os preços não baixem.*
2. *É conveniente (convém) que você visite os velhinhos.*

Este tópico merece uma discussão mais aprofundada, principalmente porque é pouquíssimo discutido nas gramáticas do português. Para uma apresentação mais abrangente deste aspecto dos adjetivos em inglês, remeto o leitor a (Huddleston 2000). Deve-se ter cuidado, no entanto, pois neste ponto há grandes diferenças entre as duas línguas. No inglês, por exemplo, encontram-se adjetivos que exigem complemento preposicionado, tal como *fond (of)*, o que em português não encontra paralelo.

### 3.5

#### Fronteiras entre classes de palavras

##### 3.5.1

##### Substantivos x adjetivos

Substantivos e adjetivos relacionam-se intimamente no português. Há processos de nominalização de adjetivos de dois tipos. Quando o adjetivo é usado como predicador de uma oração a nominalização aparece em função de anáfora, como por exemplo

**ex. 3.21**

1. [*É POSSÍVEL que chova.*]
2. [*A POSSIBILIDADE é mínima.*]

Por outro lado, a função semântica de denominação da qualidade expressa pelo adjetivo pode ocorrer de duas formas: a referência à qualidade geral, como em

**ex. 3.22** [*A SINCERIDADE anda escassa.*]

ou a referência à qualidade como propriedade de alguém, como em

**ex. 3.23** [*João é SINCERO. A SINCERIDADE de João é comovente.*]

Por outro lado, a formação do adjetivo a partir do substantivo tem como objetivo usar a semântica do substantivo para qualificar ou caracterizar. Os casos de formações morfológicamente marcadas foram apresentados na seção 3.4.3, no processo de sufixação. Cabe discutir os casos em que há ambigüidade na classificação de um item lexical por este se comportar conforme uma ou outra classe sem que haja alteração de sua forma fonológica.

Basilio (1995a) identifica a diferença entre conversão e a extensão de propriedades lexicais como fenômenos distintos de flutuação categorial no português. No primeiro caso, um item lexical assume totalmente as propriedades de uma outra classe, passando a constituir-se membro dessa outra. É o caso dos exemplos 3.24.

**ex. 3.24**

1. [*Seu instrumento MUSICAL favorito é o piano.*]
2. [*Fomos assistir a um MUSICAL espetacular.*]

No segundo caso, observa-se que o item lexical adquire determinadas propriedades da outra classe, mas não todas, o que se reflete nas restrições que se impõem ao seu uso. No exemplo a seguir, o uso do adjetivo POBRE como substantivo refere-se a todos os indivíduos caracterizados por ele, ou seja, tem caráter genérico, o que fica claro com a não alteração do sentido com o plural e pela estranheza do feminino.

**ex. 3.25**

1. [*O POBRE gosta de luxo.*]
2. [*Os POBRES gostam de luxo.*]
3. ?[*A POBRE gosta de luxo.*]

Há casos em que o substantivo ocorre em posição adjetiva, como em CUSTO BRASIL, e em abundantes denominações para cores, tais como AMARELO OVO, LARANJA CALIFÓRNIA e CINZA PRATA. Conforme destaca Basilio (1995b), nestas ocorrências os substantivos determinantes não se comportam morfologicamente como adjetivos, como pode ser observado em CUSTO BRASIL - \*CUSTOS BRASIS, não se caracterizando conversão. Existem substantivos que aparecem com frequência nesta função, em posição atributiva potencialmente com relação a um conjunto extenso de substantivos, tais como PROBLEMA (SITUAÇÃO PROBLEMA), PADRÃO (FONTE PADRÃO) e MODELO (ESTUDANTE MODELO).

Outro caso de fronteira entre substantivos e adjetivos é o conjunto dos agentivos deverbais e denominais: os designadores de profissões, agentes de ação habitual ou os instrumentais, como ENCANADOR, EMBROMADOR e REFRIGERADOR, respectivamente. Os agentivos são majoritariamente substantivos, mas boa parte deles pode ocorrer em função adjetiva, atribuindo agentividade ao substantivo especificado, por exemplo em EMPRESA ADMINISTRADORA. Talvez por não serem adjetivos plenos, não aparecem em função predicativa, portanto não gozam de uma série de propriedades dos adjetivos.

### 3.5.2

#### Verbos x substantivos

Embora em português, como ocorre em outras línguas românicas, persistam entre o substantivo e o verbo muitas das distinções formais e morfológicas do latim, em línguas como o inglês, por exemplo, muitas formas verbais são idênticas às nominais, tais como TO WALK e WALK.

As formas nominais do verbo - o infinitivo, o gerúndio e o particípio - são casos de fronteira entre o verbal e o nominal, por não exprimirem



nem tempo nem modo. O particípio compartilha diversas propriedades morfológicas, semânticas e sintáticas com os adjetivos; o gerúndio<sup>1</sup>, também desempenha funções adjetivas e adverbiais; o infinitivo exprime o processo verbal potencial, aproximando-se do substantivo (Cunha & Cintra 1985), como no exemplo 3.26, em que funciona como sujeito.

**ex. 3.26** CAMINHAR *era mais rápido que andar de carro*

Esse tipo de utilização nominal do infinitivo é discutido em (Biderman 1978), incluindo exemplos do francês e do romeno.

Para cumprir funções discursivas e funções semânticas de denominação e predicação, verbos e substantivos são adaptados morfológicamente em português, por meio de derivação. Deixando de lado os processos de formação de verbos a partir de substantivos, abordo aqui a direção inversa, **nominalização**, pela importância direta do resultado desse processo de formação de palavras nas construções com verbo-suporte.

A **nominalização** consiste em uma derivação cujo resultado é um substantivo. De acordo com (Basilio 1999c), a nominalização permite a referência a um processo verbal como a um tipo de evento, ação, estado, etc., independentemente de circunstâncias particulares.

Os processos de nominalização deverbiais mais comuns no português são:  $[X]_V \rightarrow [[X]_V\text{mento}]_N$ ,  $[X]_V \rightarrow [[X]_V\text{ção}]_N$ , designativos de ação;  $[X]_V \rightarrow [[X]_V\text{dor}]_N$ , designativo de agente, etc. Sem levar em consideração o sentido da derivação ( $V \rightarrow N$  ou  $N \rightarrow V$ ), mas simplesmente a relação verbo / nome, podemos adicionar à lista de processos a derivação regressiva.

A nominalização teve uma influência muito grande na promoção do léxico a uma posição de maior importância na gramática (cf. capítulo 2). Nas fases iniciais da teoria gerativa transformacional a morfologia derivacional foi abandonada, pois na teoria padrão os processos derivacionais gerais eram efetuados dentro das regras transformacionais sintáticas. Por exemplo, os enunciados em 3.27 seriam gerados da mesma estrutura profunda, sendo que na geração da segunda haveria a aplicação de uma regra fonológica de nominalização.

**ex. 3.27**

1. [*Os vizinhos terminaram de construir a casa.*]

<sup>1</sup>Há também o gerundivo, modalidade de gerúndio latino, com função de particípio passivo futuro; “exprime ação que está por se realizar ou que será realizada” (Houaiss 2001). Em português, há substantivos derivados de gerundivos latinos, como MEMORANDO e AGENDA, e neologismos, como FORMANDO, DIPLOMANDO, DOUTORANDO, etc. Entretanto, trata-se de um processo pouco produtivo de formação de palavras.

## 2. [Os vizinhos terminaram a construção da casa.]

Em (Basilio 1980) encontra-se uma análise dos prós e contras da hipótese transformacionista, resumida a seguir. Nessa hipótese, há duas virtudes que devem ser destacadas. Por um lado, o léxico fica reduzido pois muitas redundâncias seriam eliminadas. Por exemplo, dos pares <verbo/nominalização> apenas uma forma básica seria listada. Por outro lado, fica bem marcada a relação semântica entre as orações transformadas.

No entanto, alguns problemas sérios não são resolvidos transformacionalmente (recomendo (Gunzburger 1979) para uma análise bem exemplificada). Em (Chomsky 1970) o autor rejeita o tratamento transformacionista do léxico, propondo a Hipótese Lexicalista. Chomsky seleciona o problema da nominalização para argumentar que as relações verbo / nome nesses casos devem ser explicitadas no léxico. Mesmo nos casos idiossincráticos, uma vez que cada entrada lexical pode ter traços semânticos independentes, a hipótese lexicalista remedia o problema das extensões de sentido adquiridas pelas nominalizações.

A regularidade das relações palavra básica / nominalização é notável, como argumentado em (Gunzburger 1979), demonstrando a possibilidade de preverem-se as interpretações dos nominais a partir das características semânticas dos verbos correspondentes. Em (Basilio 1980), a autora é enfática nessa questão, tomando como básico o argumento de que o fenômeno da nominalização é uma associação paradigmática entre verbos e nomes, e não um mero processo de formação ou uma associação idiossincrática. As principais evidências dessa forte correspondência apresentadas pela autora são:

- grande parte dos verbos existentes no léxico do português apresentam uma contraparte nominal;
- os verbos sem um nome relacionado em geral podem ser caracterizados: verbos coloquiais, verbos auxiliares, verbos que sofreram *bloqueio*<sup>2</sup> por razões de diacronia, etc;
- a formação de verbos denominais é imprevisível, totalmente ao contrário da formação de nomes deverbais.

De um ponto de vista empírico, a importância do fenômeno da nominalização pode ser apreciada pelas evidências de que dentre os processos de

<sup>2</sup>Conforme definido por Mark Aronoff, o fenômeno do bloqueio ocorre quando existe no léxico uma forma que exerce a função que uma forma nominalizada exerceria.

formação de palavras, os mais produtivos são as nominalizações. Em (Sandmann 1989), o autor demonstra que novas formações de nomes ocorreram aproximadamente duas vezes mais que as de adjetivos e verbos juntas<sup>3</sup>.

Em (Beard 1998) encontra-se uma tipologia das derivações quanto à função que cumprem no contexto lingüístico. Beard distingue:

1. a derivação de traço (*featural derivation*), onde não há mudança de classe, mas sim de traços gramaticais, como gênero; possivelmente, em português esse tipo de derivação é considerado flexão;
2. a derivação funcional (*functional derivation*), onde pode haver mudança de classe, mas há principalmente mudança de caso, por exemplo, locativo (*-aria*), instrumental (*-dor*), origem (*-ense*), etc;
3. a transposição (*transposition*), onde há apenas mudança de classe para a adequação sintática;
4. a derivação expressiva (*expressive derivation*), onde não há mudança de classe; semanticamente, reflete a percepção subjetiva do falante.

Dentro dessa tipologia, há nominalizações funcionais e de transposição.

Em (Basilio 1999c) a autora propõe uma classificação mais refinada, que revela melhor o papel sintático / semântico da nominalização. Basilio observa que os processos de nominalização apresentam muitas vezes funções múltiplas e simultâneas. A principal delas seria a função sintática (tipo mais próximo à **transposição**). No entanto, à função sintática são agregadas as funções **textual** e da **estrutura do texto**, que lhe atribuem uma justificativa em termos de necessidades do discurso. Essa discussão das funções da nominalização reiteram a estreita ligação semântica entre verbos e suas nominalizações.

A relevância das classes de palavras na análise lingüística pode ser constatada no vasto conjunto de propostas de descrição da gramática que utilizam as classes como espinha dorsal do modelo descritivo. No âmbito deste trabalho, as categorias tradicionais se mostram adequadas como base do estudo. As características principais das categorias centrais para as construções com palavras de suporte foram expostas nesse capítulo.

<sup>3</sup>A metodologia da análise do cópuz por Sandmann bem como certas hipóteses de trabalho devem ser consideradas na apreciação dessa afirmação.

## 4

### A Conceituação do Substantivo-suporte

#### 4.1

##### Introdução

Esse capítulo propõe uma conceituação do substantivo-suporte de acordo com suas propriedades lexicográficas e textuais.

Na seção 4.2 relato a pesquisa de corpus inicial, que propiciou a seleção de uma lista fechada preliminar de substantivos-suporte do português. A seguir, o dicionário é utilizado no levantamento de características como: número de acepções, número de expressões multi-vocabulares dicionarizadas e circularidades nas definições. Utilizando o corpus, examino alguns exemplos de substantivos-suporte, analisando-os com respeito à polissemia e ao contexto de uso.

Na seção 4.3, a noção de substantivo-suporte é posicionada em paralelo à do verbo-suporte; busco, assim, demonstrar que o fenômeno de suporte é mais um dos processos de formação lexical que são compartilhados por verbos e substantivos.

#### 4.2

##### Lexicografia do Substantivo-suporte

*The covers of this book are too far apart.* (Ambrose Bierce, *The Devil's Dictionary*)

##### 4.2.1

##### O dicionário e a questão da multiplicidade de sentidos

O falante comum admite de bom grado que as palavras possuam múltiplos significados. Fora dessa hipótese, como explicar, sem recorrer a uma teoria lexicológica sofisticada, as possibilidades do exemplo 4.1?

##### ex. 4.1

1. *Raimundo sorriu, amarrotou a folha de PAPEL e lançou-a no chão.*
2. *O flúor desempenha um PAPEL importante na resistência dos dentes.*

3. O PAPEL subiu 20%, pulando dos R\$20 o lote de mil ações para R\$24...
4. Ele rasga uma cruz que havia sobre sua cama, joga santinhos no vaso sanitário e usa as vestimentas de primeira comunhão como PAPEL higiênico.

É para esse usuário que os dicionários separam entradas diferentes para cada significado percebido pelo lexicógrafo, procurando, de modo sistemático e consistente, elucidar as diferenças entre os significados por meio de definições e exemplos.

Por outro lado, é cada vez mais claro o que afirma Marcuschi:

“As expressões e seus sentidos, conteúdos, referentes, etc. não são dois lados da mesma moeda. Essa era uma idéia saussuriana que neste período pós-estruturalista está em crise. Não se trata de um problema de polissemia, vagueza, imprecisão etc., mas sim do efeito do **princípio de simbolização**, que resulta num sistema categorial fluido.” (Marcuschi 2004, p.269)

Apesar de todas as restrições aos significados prontos, ao modelo de léxico como conjunto de etiquetas que se atribuem a elementos de um mundo discretizado, à semântica imanente das palavras, ainda assim, é necessário operar com acepções dicionarizadas dos substantivos-suporte. Para a operacionalização deste trabalho, a categorização lexicográfica é de grande relevância por ser uma base de dados, um repositório do conhecimento tradicional e consagrado sobre o vocabulário da língua portuguesa.

Os substantivos-suporte constituem uma classe não muito extensa de substantivos abstratos do português que se caracterizam por sua generalidade semântica. As demais propriedades lexicográficas típicas dessa classe são estabelecidas no restante dessa seção.

#### 4.2.2

##### Frequência na língua

Frequência de ocorrência pode ser um fator de identificação de substantivos vazios. Muresan, Tzoukermann, & Klavans (2001) identificam 141 **substantivos-suporte** em um corpus do inglês, uma classe com muitas interseções com a dos substantivos-suporte, baseando-se apenas em contagem de frequência.

Com relação às conchas nominais, Schmid (2000) relata que, em um corpus de 225 milhões de palavras correntes do inglês britânico, as formas singulares de CASE, FACT, IDEA, NEWS, POINT, PROBLEM, REPORT e THING

estão entre os cem substantivos mais freqüentes. O substantivo *THING*, por exemplo, tem uma freqüência de 256 ocorrências por milhão, sendo que a vasta maioria dos itens lexicais do inglês possuem freqüência menor que 20 por milhão.

Em português, Marques (1995) descreve um estudo do léxico de alta freqüência numa parte do corpus do projeto NURC (Projeto de Estudo Conjunto e Coordenado da Norma Urbana Oral e Culta), um conjunto de textos orais, provenientes de entrevistas realizadas na cidade do Rio de Janeiro. Das 506.108 unidades textuais processadas, 370.777 (39% das unidades) eram da classe dos substantivos, correspondendo a ocorrências distintas de 9.539 lexemas (51,5% dos lexemas).

No corpus do NURC, os substantivos de alta freqüência (até 100 ocorrências) totalizam 488 lexemas. Marques considera que 75 deles são de “sentido geral”, ou seja, não dependente de contexto temático, e o restante é classificado por campo temático, como: ‘tempo’, ‘indivíduos’, ‘corpo humano’, ‘vestuário’, ‘casa’, ‘alimentação’, ‘família’, etc. Ao analisarmos os quadros de substantivos de sentido geral, encontramos 15 dos 32 substantivos-suporte preliminarmente selecionados (cf. tabela 4.1) no corpus do NILC (Aires & Aluisio 2001).

âmbito	cunho	matéria	perspectiva
área	dimensão	modo	plano
aspecto	elemento	natureza	ponto
base	esfera	nível	quadro
campo	fator	ordem	questão
caráter	forma	panorama	sentido
coisa	lado	papel	tipo
componente	maneira	parte	tom

Tabela 4.1: Lista de substantivos-suporte do Português

### 4.2.3

#### A combinação com o adjetivo denominativo

Dado que as unidades do processamento de linguagem natural são a palavra e o sintagma e que o substantivo é o núcleo do sintagma nominal, é comum esperar que o substantivo seja o centro do significado do sintagma. No caso de adjetivos predicativos, como em *céu aberto*, a idéia central está no substantivo; no caso de adjetivos relacionais, como em *lesão cerebral*, o substantivo e o adjetivo se fundem em uma única unidade denotativa.

Em contraste, em um sintagma contendo um substantivo-suporte, por exemplo em *ASPECTO comercial*, fica claro que comércio é a idéia central, ainda

que o substantivo núcleo seja ASPECTO. Assim, no contexto de processamento automático e interpretação de textos, a possibilidade de identificar esses substantivos é de grande importância. Estamos buscando os casos em que a ocorrência do substantivo não corresponda a um conceito específico, deixando de cumprir plenamente seu papel denominativo. Como exemplo, consideramos o substantivo ASPECTO na seguinte sentença do corpus.

**ex. 4.2** *Examinava o ASPECTO (constitucional) que lhe garantia o direito do voto.*

O apagamento do adjetivo entre parênteses enfraquece consideravelmente o conteúdo do enunciado, tornando o significado do sintagma amplo demais; o adjetivo especifica completamente o objeto. Esse é um exemplo de um substantivo-suporte, um substantivo cujo complemento é um adjetivo ou sintagma preposicional majoritariamente responsável pelo significado da expressão como um todo. Mesmo que sintaticamente o substantivo seja o núcleo do sintagma, semanticamente ele ocorre sem conteúdo específico.

Sendo assim, uma ótima pista para o reconhecimento de um substantivo-suporte está em seus complementos. Investiguei os tipos de adjetivos que ocorrem como complementos destes substantivos no corpus e observei que eles são em grande parte denominais, ainda que possam aparecer em função denotativo ou predicativo.

Basilio & Gamarski (1995) discutem os sintagmas nominais *S-Adj* do ponto de vista do adjetivo, observando como opera a função denotativa em combinação com substantivos “de cunho muito geral que necessitam de um preenchimento semântico a ser fornecido via de regra por um adjetivo denominal, substantivo ou sintagma preposicionado”. Exemplificando a partir do corpus do NURC, de discurso oral, as autoras selecionaram seis enunciados, quatro dos quais construídos com substantivos-suporte listados na tabela 4.1 (FATOR, PARTE, ASPECTO e COISA).

Uma análise inicial de exemplos coletados assistematicamente motiva os principais questionamentos desta seção. Por exemplo em:

**ex. 4.3** *... estudantes talentosos, atraídos para os campos cognitivistas, ditavam o TOM intelectual do momento...*

o substantivo TOM atua como um focalizador da essência da denotação do adjetivo INTELECTUAL. Em contraste, o substantivo ASPECTO

**ex. 4.4** *Os temas incluem o ASPECTO comercial da moda, as tendências internacionais...*

secciona o escopo do adjetivo denotativo COMERCIAL em sub-divisões de um todo. Em

**ex. 4.5** *Enfrentamos, portanto, um segundo dilema ... que é de NATUREZA ambiental.*

o substantivo NATUREZA pode ser apagado pois possui um papel apenas textual.

Algumas vezes o apagamento do substantivo-suporte não provoca mudanças significativas, como no seguinte exemplo.

**ex. 4.6** *[A decisão foi de cunho pessoal. → A decisão foi pessoal.]*

Em outras ocasiões, é necessário ajustar o enunciado, sintaticamente ou morfologicamente, como em:

**ex. 4.7** *[Agiu de FORMA cerebral. → Agiu cerebralmente.]*

**ex. 4.8** *[Atuava no ÂMBITO municipal. → Atuava no município.]*

A partir desses fatos pode-se concluir que um critério operacional promissor para a caracterização do substantivo-suporte é a extração de sintagmas nominais  $N Adj$ , onde  $Adj$  é um adjetivo denominal denotativo. A lista de substantivos resultantes pode ser analisada de acordo com sua intercambialidade, ou seja, sua co-ocorrência com o mesmo grupo de adjetivos, como no seguinte exemplo:

$$\left. \begin{array}{l} \text{fator} \\ \text{perspectiva} \\ \text{aspecto} \end{array} \right\} \text{racial} \quad \left. \begin{array}{l} \text{fator} \\ \text{perspectiva} \\ \text{aspecto} \end{array} \right\} \text{fiscal} \quad \left. \begin{array}{l} \text{fator} \\ \text{perspectiva} \\ \text{aspecto} \end{array} \right\} \text{ambiental}$$

Para dar partida na investigação dos substantivos-suporte e suas combinações, extraí de um dicionário do português brasileiro (Houaiss 2001) um grupo de 166 adjetivos denominais regulares, ou seja, adjetivos cujo significado é um produto motivado pela atuação do processo derivacional sobre o significado do substantivo básico. Não se pode garantir *a priori* se uma dada ocorrência de um adjetivo denominal será ou não denotativa, portanto não utilizei esse critério, que no entanto geraria maior precisão dos dados.

Usando o critério de intercambialidade, selecionei de um corpus uma lista preliminar de substantivos-suporte, colocados na tabela 4.1. A partir dessa lista, procurei inferir algumas propriedades lexicográficas importantes. Ainda que tenha encontrado um número reduzido de substantivos-suporte no corpus, classificá-los não é tarefa fácil, dado o seu alto grau de polissemia e a variedade



Substantivo	Adjs.	Substantivo	Adjs.
caráter	31	plano	17
aspecto	24	área	16
parte	23	questão	15
natureza	21	tipo	15
forma	20	campo	14
problema	20	papel	13
ordem	19	base	12
lado	17	sentido	12
nível	17	âmbito	10

Tabela 4.2: Número de adjetivos distintos em sintagmas *N Adj* com substantivos-suporte

de entradas que eles geram em dicionários convencionais. Além disso, muitas dessas entradas não correspondem em sentido ao substantivo-suporte.

Dos 166 adjetivos escolhidos, apenas 56 não ocorreram em associação com o substantivo-suporte. Esses adjetivos eram pouco freqüentes no corpus, como por exemplo DORSAL e CONDOMINIAL. A tabela 4.2 mostra os substantivos que se combinam com maior freqüência com diferentes adjetivos.

#### 4.2.4

##### O que diz o dicionário

Ao procurar a lista preliminar de substantivos-suporte no (Houaiss 2001), percebi que algumas peculiaridades se destacam. Em primeiro lugar, o número de acepções de um substantivo-suporte é, em geral, maior que o de um substantivo qualquer. Observa-se que os substantivos-suporte apresentam um elevado número de acepções, muitas delas marcadas como derivações semânticas, como metonímia, metáfora e analogia. A tabela 4.3 mostra os substantivos-suporte mais polissêmicos de nossa lista da tabela 4.1, de acordo com (Houaiss 2001).

Esta flutuação em significado mostra que o substantivo-suporte é semanticamente tão vago que parece exigir complementação. Por exemplo, o substantivo COISA aparece em (Houaiss 2001) com 21 sentidos distintos, como mostra a figura 4.1.

O substantivo COISA funciona como um coringa, cujo significado só pode ser compreendido no contexto de um enunciado. Na verdade, várias línguas Indo-Européias, como espanhol, alemão, inglês e francês possuem um substantivo-suporte que pode ser traduzido como COISA, que pode ser considerado como a manifestação ostensiva de um esquema cognitivo de substantivo e pode substituir quase todos os substantivos (Mihatsch 2003).

Substantivo	Sentidos	Substantivo	Sentidos
ponto	57	natureza	18
base	30	matéria	17
ordem	30	plano	17
forma	27	modo	16
quadro	27	sentido	14
coisa	21	área	13
tipo	20	caráter	13
tom	19	maneira	12
campo	18	parte	12

Tabela 4.3: Número de sentidos dicionarizados

**n substantivo feminino**

- 1 tudo quanto existe ou possa existir, de natureza corpórea ou incorpórea
- 2 qualquer ser inanimado
- 3 realidade, fato concreto, em relação ao que é abstrato ou assim considerado
- 4 algo que não se quer ou não se pode nomear
- 5 aquilo de que se está tratando ou falando
- 6 aquilo que se pensa; pensamento, idéia
- 7 relação, ligação, vínculo
- 8 interesse próprio, negócio, ocupação
- 9 ato, empreendimento, empresa
- 10 o que acontece; ocorrência, evento, caso
- 11 assunto, tema, matéria
- 12 negócio, transação
- 13 algo que provoque estímulo, que entusiasme; motivo, incentivo, compensação
- 14 o que não se sabe; mistério, enigma
- 15 mal-estar ou indisposição súbita; ataque, perda dos sentidos
- 16 Regionalismo: Paraíba. cigarro de maconha; baseado
- 17 órgão genital do homem ou da mulher
- 18 algo imprestável, velho ou maltratado; traste, troço, bagulho

**substantivo masculino**

- 19 o diabo

**coisas****n substantivo feminino plural**

- 20 bens, propriedades, valores
- 21 negócios, interesses, ocupações

Figura 4.1: Entrada do dicionário Houaiss para COISA

Subst.	EMVs	Exemplos
papel	113	p. almanaque, confiar ao p., passar p. com, ficar no p.
ponto	110	p. alto, p. cardeal, p. crítico, p. culminante, p. cego
campo	49	c. conceitual, c. de concentração, c. de força, c. visual
base	42	b. aérea, b. monetária, b. vetorial, à b. de, b. de dados
plano	38	p. geral, p. de saúde, p. inclinado, primeiro p.
fator	31	f. de correção, f. Rh, f. abiótico, f. de crescimento
ordem	30	o. civil, o. cronológica, o. de grandeza, o. do dia
forma	29	f. canônica, f. livre, de certa f., de f. alguma, de f. que
sentido	26	s. anti-horário, s. estrito, s. figurado, duplo s., fazer s.
parte	25	p. do discurso, p. ideal, p. íntimas, à p., a p. do leão
área	24	á. de livre comércio, á. de transferência, á. de serviço
modo	22	m. de ser, m. de ação, m. maior, de m. a, de todo m.
coisa	20	c. de, c. pública, cheio de c., não dizer c. com c.
nível	19	n. de energia, n. de vida, alto n., ao n. de, baixo n.
elemento	17	e. neutro, e. mórfico, e. de composição, estar no seu e.
questão	16	q. aberta, q. de ordem, q. de tempo, q. fechada
lado	15	l. a l., ao l. de, de l., de um l. para outro, pôr de l.
caráter	13	c. hereditário, c. tipográfico, a c., de c.
matéria	13	m. em questão, m. pré-estrelar, m. processual
quadro	11	q. clínico, q. mural, q. de horário, q. vivo
tipo	11	t. comum, t. de caixa, t. ideal, fazer t.

Tabela 4.4: Número de expressões multi-vocabulares dicionarizadas

Apesar da potencial riqueza de uma análise de COISA enquanto substantivo-suporte, constatei que o corpus, de viés jornalístico-científico, não produziu dados suficientes em termos de combinações *S-Adj*, com adjetivo denominal denotativo, por isso não prossegui na análise deste substantivo arquetípico (cf. capítulo 6).

Substantivos-suporte também aparecem em expressões cristalizadas em vários campos semânticos de especialidade. O substantivo BASE, por exemplo, possui 30 sentidos dicionarizados, 23 dos quais são em rubricas especializadas.

Outra característica lexicográfica marcante dos substantivos-suporte é o número de expressões multi-vocabulares dicionarizadas em que ocorrem. Devemos tratar dessa propriedade com muito cuidado, já que em muitas dessas expressões o sentido do substantivo não pode ser considerado de suporte. Por exemplo, se analisarmos a entrada de PAPEL, observamos que 100 das 113 expressões dicionarizadas são tipos especiais de papel, tais como PAPEL CELOFANE e PAPEL CREPOM, sob a rubrica 'indústria de papel'. Portanto não podem ser consideradas ocorrências do sentido de suporte de PAPEL. Já na entrada de PONTO, as 110 expressões listadas são de grande diversidade em sentido e campo semântico. Apesar dessas considerações, é justo afirmar que

Substantivo-suporte	Referências cruzadas
âmbito	esfera, campo
área	campo
aspecto	maneira, lado
base	parte, aspecto
campo	área, esfera, âmbito
caráter	cunho
cunho	caráter
dimensão	aspecto
elemento	parte
esfera	área
fator	elemento
forma	modo, maneira, tipo
lado	maneira, aspecto
maneira	modo, forma
modo	forma
natureza	caráter, tipo
panorama	matéria
parte	matéria, papel, área, lado
perspectiva	forma
plano	nível
ponto	parte, aspecto
questão	matéria, ponto
sentido	modo, aspecto
tom	modo, caráter

Tabela 4.5: Referências cruzadas em entradas do dicionário para substantivos-suporte

o substantivo-suporte tende a formar expressões cristalizadas, como ilustra a tabela 4.4.

Quando analisamos as entradas do sentido de suporte dos substantivos, encontramos uma grande quantidade de referências cruzadas, revelando uma rede muito interessante de relacionamentos, que pode ser utilizada para classificar os grupos de substantivos-suporte por função semântica.

Os dados da tabela 4.5 foram utilizados para construir um grafo de sentidos conectados da figura 4.2. As definições dos substantivos na primeira coluna fazem referência aos substantivos na segunda coluna. Por exemplo, na definição de ÂMBITO há referências a CAMPO e ESFERA.

### 1. Âmbito

n substantivo masculino

- (a) espaço que circunda, rodeia, envolve; periferia
- (b) espaço físico compreendido dentro de determinados limites; recinto, amplitude

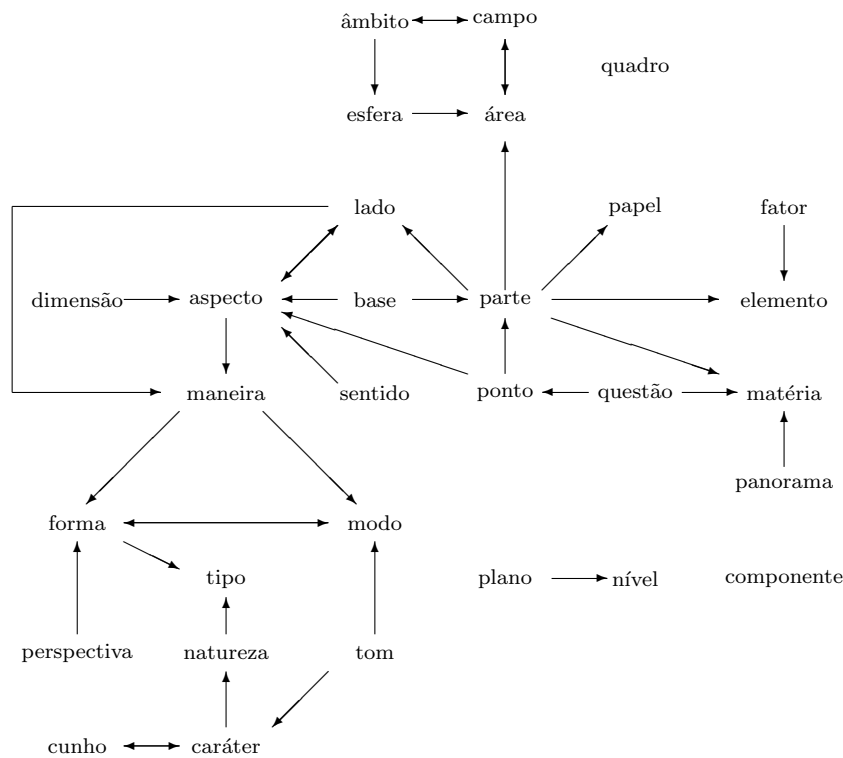


Figura 4.2: Diagrama das referências cruzadas entre definições de substantivos-suporte

Ex.: o â. do plenário

área que os antigos romanos eram obrigados a deixar em volta de suas casas

(c) Derivação: sentido figurado.

esfera de ação ou de pensamento; campo ou espaço em que ocorre ou se exerce alguma atividade

Ex.: era um assunto do â. da psicologia

(d) Derivação: sentido figurado.

núcleo central ou predominante dessa esfera

Ex.: aquela mulher não estava no â. de suas cogitações

(e) Rubrica: música.

intervalo entre a nota mais grave e a mais aguda de um trecho musical considerado

#### 4.2.5

##### A polissemia de FORMA

Na seção 2.3, a questão da ambigüidade/vagueza é discutida como gradação (cf. figura 2.3). Do ponto de vista do lexicógrafo, isso significa que não há testes diagnósticos decisivos para a separação dos sentidos. Kilgarriff

(1992) descreve uma idealização da prática lexicográfica, assumindo que para cada palavra o lexicógrafo:

1. coleta no corpus ocorrências da palavra;
2. divide as ocorrências em grupos, por afinidade dos significados da palavra nas ocorrências;
3. para cada grupo, racionaliza o que identifica a ocorrência com seu grupo;
4. codifica suas conclusões na linguagem extremamente restrita das definições de um dicionário.

Idealmente essa prática produziria acepções baseadas no que o autor chama de critério “SFIP - Sufficiently Frequent and Insufficiently Predictable”. Por um lado, o sentido deve ter como suporte um grupo de ocorrências significativo, desprezando usos fortuitos, como SANDUÍCHE DE PRESUNTO,

**n substantivo feminino**

- 1 configuração física característica dos seres e das coisas, como decorrência da estruturação das suas partes; formato, feitio, figura  
Ex.: a f. de uma mesa  
a f. humana
- 2 estado físico sob o qual se apresenta um corpo, uma substância etc.; estado  
Ex.: f. sólida, f. líquida, f. gelatinosa
- 3 a aparência física de um ser ou de uma coisa  
Ex.: f. agradável, f. exuberantes
- 4 um ser ou objeto indistinto, percebido imprecisamente  
Ex.: ao longe, avistamos uma f. vindo em nossa direção
- 5 a maneira como o músico, o artista plástico ou o escritor se expressa ou estrutura sua obra  
Ex.: um artista acadêmico só sabe trabalhar com f. consagradas  
o formalista é aquele que se preocupa mais com a f. do que com o conteúdo de sua obra
- 6 modo, jeito, maneira, método  
Ex.: isso é f. de se dirigir a uma pessoa?  
a f. em que se encontram os nossos serviços públicos  
encontrar uma f. de obter um metal mais puro
- 7 sistema, método  
Ex.: f. de governo
- 8 maneira particular em que uma categoria ou noção geral pode ocorrer; tipo, variedade  
Ex.: é uma f. nova do vírus  
descobriu nova f. de estelionato
- 9 estrutura coerente, segundo um padrão familiar  
aos poucos, o pensamento foi ganhando f.  
esta escultura possui uma f., já aquela é algo informe

Figura 4.3: Entrada do dicionário Houaiss para FORMA

10	condição ou aparência física ou mental; rigidez, saúde, elegância está voltando à antiga f. passam-se os anos, e ela não perde a f.
11	alinhamento, fila Ex.: os alunos estavam formados no pátio quando um saiu da f.
12	um dos diferentes modos de existência, ação ou manifestação de algo particular Ex.: entre as salamandras estão incluídas tanto f. aquáticas como terrestres

Figura 4.3: Entrada do dicionário Houaiss para FORMA (cont.)

no exemplo 2.5. Por outro lado, o uso totalmente previsível a partir de uma acepção já listada também não merece espaço no dicionário, como por exemplo a dupla possibilidade dos pares substantivo-adjetivo em usos característicos da flutuação categorial/extensão (Basilio 1995a) (cf. seção 3.5).

Nesta seção analiso a entrada de um substantivo-suporte FORMA no (Houaiss 2001). O objetivo é realizar uma espécie de engenharia reversa dessa entrada, procurando critérios para discernir as acepções que corresponderiam ao substantivo-suporte FORMA. A figura 4.3 reproduz parcialmente a entrada, formada, ao todo, por 27 acepções, sendo 15 em rubricas específicas. Considerando apenas as acepções não especializadas, busquei no corpus ocorrências de cada uma das 12 definições restantes. A tabela 4.6 apresenta, para cada definição, um exemplo de uso no corpus, quando este pôde ser atestado.

Definições	Exemplos do corpus
1 formato de seres e coisas	<i>Speranzoni está distribuindo em sua lotérica chaveiros com a FORMA de um pé de quatro dedos onde está escrito “O pé do João”.</i>
2 estado de corpo e substância	<i>As moléculas, na FORMA de gases, entram pelas narinas, atingem os receptores e desencadeiam um impulso elétrico que é levado até o bulbo olfativo espécie de central dos cheiros.</i>
3 aparência física de ser ou coisa	<i>Segundo ela, características visuais como cor da pele, textura e FORMA do cabelo e traços fisionômicos são resultado de “um processo de adaptação da espécie às regiões climáticas”.</i>
4 ser indistinto	—
5 maneira, de obra	<i>A categoria do autor leva à fetichização do estilo e da FORMA, e, finalmente, a uma trivialização do conteúdo já que o conteúdo de todo modo é sempre o mesmo, pois é um autor que conta sempre a mesma coisa.</i>

Tabela 4.6: Acepções de FORMA, com exemplos do corpus

Definições	Exemplos do corpus
6 modo	<i>Em seguida ele traça de FORMA irrefutável a história pertinente das ciências cognitivas constituintes - filosofia, psicologia, inteligência artificial, lingüística, e as disciplinas fronteiriças da antropologia e da neurociência - demonstrando seu desenvolvimento paralelo e convergente.</i>
7 sistema	<i>A Hungria é o país que mais acredita nessa FORMA de casamento.</i>
8 maneira, de categoria ou noção	<i>Estes seres tem uma FORMA de vida muito primitiva e se utilizam do próprio corpo para absorver nutrientes que estão ao seu redor.</i>
9 estrutura coerente	<i>a) De repente, um cientista percebeu aquela bola gigantesca tomando FORMA nos confins do sistema solar e, com alguns cálculos, descobriu que ela iria causar um estrago sideral em Júpiter.</i> <i>b) A escritura, no seu momento genético, é sempre plural; ela se dá como feixe de possibilidades e a grandeza do resultado final está menos em escolher alternativa do que em dar FORMA orgânica à multiplicidade.</i>
10 condição física	<i>Já quem não está com o corpo muito em FORMA para correr 90 minutos, pode experimentar o futebol soçate ou o salão.</i>
11 alinhamento, fila	—
12 modos de existência	<i>No final dos anos 1940, na época do Simpósio Hixon, estava ficando patente que nem a FORMA fisiológica nem a psicológica do behaviorismo eram viáveis.</i>

Tabela 4.6: Acepções de FORMA, com exemplos do corpus (cont.)

Não faz parte deste estudo apresentar uma visão crítica do dicionário, mas cabe comentar o quanto a divisão das acepções carece de uma justificativa mais convincente. Muitas vezes os exemplos não parecem enquadrar-se em nenhuma delas, a não ser por um esforço de transformação generalizadora ou especializadora do uso exemplificado. No entanto, na maioria das vezes o oposto ocorre: os exemplos se encaixam em várias acepções simultaneamente, levantando questionamentos sobre a validade das acepções arroladas.

Com respeito à concordância, a palavra FORMA no corpus ocorreu em um conjunto de 13.963 contextos, dos quais 8.000 foram selecionados randomicamente. Subconjuntos desse conjunto inicial foram obtidos pela eliminação das locuções DE CERTA F., DE F. A, DE F. ALGUMA, DE ALGUMA F., DE F. QUE, entre outras, gerando um conjunto de contextos analisável manualmente.

O exame dos contextos, à procura de critérios para a distinção entre as acepções em que FORMA é substantivo-suporte, motivou algumas categorizações, organizadas na tabela 4.7. A primeira é com respeito à concretude do



Definições	concreto	tipo de propriedade*	informação suplementar exigida*
1 formato de seres e coisas	–	de ‘coisa’	n/a
2 estado de corpo e substância	–	de ‘coisa’	n/a
3 aparência, de ser ou coisa	–	de ‘coisa’	n/a
4 ser indistinto	+	n/a	n/a
5 maneira, de obra	–	de ‘noção’	não
• 6 modo	–	de ‘processo’	PREP V Inf
• 7 sistema	–	de ‘noção’	(ADJ) PREP N
• 8 maneira, de noção	–	de ‘noção’	(ADJ) PREP N
9 estrutura coerente	–	de ‘coisa’	n/a
		de ‘noção’	não
10 condição física	–	de ‘coisa’	n/a
11 alinhamento, fila	+	n/a	n/a
• 12 modos de existência	–	de ‘noção’	(ADJ) PREP N

• substantivo-suporte  
\* n/a: não avaliado

Tabela 4.7: Acepções de FORMA, com categorias de distinção

sentido. As acepções 1, 2, 3, 5, 6, 7, 8, 9, 10 e 12 nomeiam ‘propriedades’, sendo portanto abstratas; as acepções 4 (ser, objeto) e 11 (fila) nomeiam ‘coisas’, sendo portanto concretas. Coincidentemente, as acepções concretas não puderam ser atestadas no corpus utilizado, indicando que são usos pouco frequentes. Entre as acepções abstratas, separei as que nomeiam propriedades de ‘coisas’ (1, 2, 3, 9, 10), de ‘noções’ (5, 7, 8, 9, 12) e de ‘processos’ (6, 7).

A última categorização obedece a um critério sintático, de adjacência da palavra a outras palavras. A consideração é se o uso de FORMA vem necessariamente acompanhado de informação suplementar à direita (Sinclair 2001), posição preferencial para especificadores adjetivos, e de que classe de palavras ou tipo de sintagma cumpre essa suplementação. Sem a preocupação de definir exaustivamente os padrões de suplementação, a tabela 4.7 contém apenas os mais usuais. Por exemplo, a acepção ‘sistema’ (7) exige um complemento preposicionado, como em FORMA DE CASAMENTO.

A exigência de informação suplementar é uma indicação de vagueza. A análise revelou que as acepções 5 e 9 não exigem suplementação, ao passo que 6, 7, 8 e 12 exigem. Este grupo de acepções seria considerado o grupo dos substantivos-suporte.

#### 4.2.6

### O substantivo-suporte no corpus

#### A descrição do corpus

Como já dito previamente, este estudo do substantivo-suporte originou-se de uma pesquisa em corpus cujo foco era o sintagma *S-Adj*, onde *Adj* é um adjetivo denominal resultante de um processo de derivação.

Utilizei um corpus de português do Brasil, compilado pelo NILC (Núcleo Interinstitucional de Linguística Computacional), descrito em (Aires & Aluisio 2001). As consultas, realizadas por meio de uma linguagem especializada de especificação de padrões lingüísticos, resultam em um conjunto de segmentos textuais que, de maneira geral, correspondem a parágrafos dos textos da coleção. O corpus contém cerca de 37 milhões de palavras em textos em prosa, divididos em textos corrigidos, textos não corrigidos e textos semicorrigidos. Os textos classificados como corrigidos, totalizando 33.081.000 palavras, são aqueles publicados para grande número de leitores (livros, jornais, revistas, etc.), que são, portanto, supostamente corrigidos por especialistas em revisão de textos. Há cerca de 4.490 textos de diversos gêneros: livros (de literatura brasileira; didáticos - biologia, química, física, história, geografia; enciclopédias; temáticos - arte, ciências, etc.); revistas; constituição brasileira e textos jurídicos; jornais. Os textos não corrigidos, totalizando 736.000 palavras, são textos autênticos, escritos por pessoas de nível médio de escolaridade (segundo grau) e universitários. Há 2.430 textos que incluem redações, monografias e textos de publicidade, por exemplo. Finalmente, os textos semicorrigidos, em número de 340, com aproximadamente 1.390.000 palavras, são textos publicados para um pequeno número de leitores, ou não publicados, que são corrigidos, mas, geralmente, não por especialistas em revisão de textos. Estão entre estes contratos, relatórios, dissertações acadêmicas, etc.

#### Uma análise dos dados

A análise detalhada de todos os substantivos-suporte em sintagmas *N Adj* é uma necessidade computacional óbvia, dados os fatos que vimos até aqui. Devo ressaltar que esses itens lexicais devem ser observados individualmente já que se trata de uma classe provavelmente fechada, de substantivos muito peculiares.

Para demonstrar a viabilidade, segurança e produtividade da estratégia utilizada, apresento aqui a análise de um subgrupo dos substantivos-suporte com função de essência (cf. seção 4.3.2): TOM, CARÁTER, CUNHO e NATUREZA. Os seguintes padrões foram utilizados para concordância no cor-

Padrão	Total ocorr.	Ocorr. distintas	% S é de suporte
CUNHO <i>Adj</i>	89	49	100%
NATUREZA <i>Adj</i>	230	176	93%
CARÁTER <i>Adj</i>	730	427	97%
TOM <i>Adj</i>	478	113	72%

Tabela 4.8: Resultados de concordância

pus: “*cunho*” [*pos*=“*ADJ*”], “*natureza*” [*pos*=“*ADJ*”], “*caráter*” [*pos*=“*ADJ*”] e “*tom*” [*pos*=“*ADJ*”], onde *pos* abrevia **classe de palavras**<sup>1</sup>. A tabela 4.8 resume os resultados.

O substantivo CUNHO apareceu com sentido de suporte em 100% dos casos, no contexto *N Adj*; nessas ocorrências 55% dos adjetivos eram denominais.

O substantivo NATUREZA apareceu como substantivo-suporte em 93% dos casos, 56% dos quais foram seguidos de adjetivos denominais. Nos casos em que o sentido não foi considerado de suporte (7% – 17 ocorrências), encontramos 4 adjetivos de origem: NATUREZA AMAZÔNICA, NATUREZA CALIFORNIANA, NATUREZA BRASILEIRA e NATUREZA AMERICANA; dos 13 restantes apenas 4 eram seguidos de adjetivos denominais.

O substantivo CARÁTER apareceu como substantivo de suporte em 97% dos casos, apenas 36% dos quais eram seguidos de adjetivos denominais. Entre os sentidos plenos, 3 eram em línguas de especialidade (biologia); dos restantes apenas 3 eram seguidos de adjetivos denominais.

O substantivo TOM não apresentou a maioria esmagadora de ocorrências vazias – 72% dos casos analisados – por aparecer frequentemente (37 em 113) no sentido de ‘tom de voz’, que não é de suporte. Nesses casos, TOM ocorre como um complemento de um verbo *dicendi* ou outra palavra dessa classe semântica, tais como ‘discurso’ or ‘saudação’. Como exemplos, apresentamos:

#### ex. 4.9

1. *disse em TOM malicioso*
2. *murmurou em TOM queixoso*
3. *falou em TOM grave*

Outros casos do sentido pleno do substantivo TOM inclui tons de cor, tais como TOM *vermelho*, TOM *laranja* e tons musicais, tais como TOM *menor*. Nos

<sup>1</sup>POS é a abreviatura de **part of speech**, um rótulo consagrado mesmo em bases de dados do português.

casos de suporte, o adjetivo que segue o substantivo é denominal em 46% da ocorrência; nos casos plenos, 12% são adjetivos denominais.

Finalmente, observei a intercambialidade entre CUNHO e NATUREZA, com respeito aos seus sentidos de suporte. Todas as ocorrências de sintagmas nominais com CUNHO foram encontrados também com NATUREZA, como os pares

CUNHO acadêmico - NATUREZA acadêmica  
 CUNHO social - NATUREZA social  
 CUNHO jornalístico - NATUREZA jornalística

### 4.3

#### A Função de Suporte: o Verbo e o Substantivo

*Gravitation is not responsible for people falling in love.* (Albert Einstein)

##### 4.3.1

#### O Verbo-suporte

O uso de certos verbos na formação de uma unidade predicadora em conjunto com um complemento nominal tem sido estudado em diversas línguas, sob diversas visões metodológicas e terminologias. Na língua portuguesa, em geral, tem-se um verbo de alta frequência, como atesta (Garrão 2006), cujas ocorrências se dão, em uma grande proporção, na situação de suporte. A denominação atribuída a esses verbos, verbos-suporte ou verbos leves, parece estar associada a essa função, por isso parece contraditória a referência ao sentido pleno do verbo-suporte. A rigor, teríamos que fazer referência ao “verbo V em função de suporte”, mas o uso corrente do termo é “o verbo-suporte V”.

Já que nesse capítulo o dicionário tem um papel tão importante, a primeira conceituação para o verbo-suporte que registro nessa seção é retirada do (Houaiss 2001), associada ao verbete DAR:

- 1) a) em algumas acepções, dar funciona como verbo pleno, com seu próprio significado (p.ex., dar um documento a um funcionário = passá-lo às suas mãos); enquanto em inúmeras outras, faz de verbo-suporte, constituindo, com o substantivo (que na gramática tradicional é seu objeto direto), um todo semântico (p.ex., dar um abraço = abraçar); a.1) neste segundo caso, a função do verbo pendula entre a de um elemento de semântica quase vazia e aquela de um verbo não exatamente pleno, mas ainda portador de certo valor

semântico maior ou menor, conforme o caso; o estabelecimento de seu sentido depende dos substantivos que com ele ocorrem na posição de objeto, tornando o número de acepções enorme; a.2) quando dar faz de verbo-suporte, o chamado objeto direto não funciona como argumento, tendo, na verdade, a natureza de um predicado, orientando o evento e classificando ou identificando o referente; a.3) por sua importância, diversas acepções de dar, usado como verbo-suporte, estão registradas no corpo deste verbete; diversas outras devem ser procuradas pelo substantivo que faz parte do objeto direto, como de hábito no restante deste dicionário;

Na definição acima aparecem dois dos elementos que mais participam da conceituação do verbo-suporte na literatura. Em primeiro lugar é colocada a questão do conteúdo semântico do verbo, na oposição verbo pleno × verbo-suporte, chamando atenção para a semântica vazia, ou esvaziada, do verbo-suporte. Em segundo lugar, é destacado o aspecto da predicação com verbo-suporte, em que o objeto direto forma com o verbo um todo semântico, funcionando como um predicado.

Ambos os aspectos são abordados em (Neves 1999), que trata das construções com verbo-suporte com atenção voltada para a lexicografia, tendo em vista a delimitação de unidades lexicais (ver também (Basilio 1999b; Basilio 1999a; Garrão & Dias 2001)). A autora traça uma linha contínua que une construções livres a expressões cristalizadas. As livres são constituídas de verbos plenos complementados por sintagmas nominais, que são completamente livres (CONSOLIDAR A ESTRADA, FINDER PROPOSTAS), onde os dois elementos exercem papéis independentes na estrutura argumental; as expressões cristalizadas incorporam um significado unitário, onde “nem mesmo parece ser possível postular um sintagma nominal em posição de objeto” (Neves 1999), como DAR *um pulo* e TOMAR *partido*. Entre estes dois extremos de construção, colocam-se as construções constituídas com verbos-suporte, verbos que sofrem um certo esvaziamento do sentido lexical, porém contribuem semanticamente para o significado total da construção (DAR UM RISO, TER CONFIANÇA).

Neves explica que algumas construções com verbo-suporte se situam mais próximas de construções livres, outras mais próximas de expressões cristalizadas. Elas são compostas por um verbo com determinada natureza semântica básica, que funciona como instrumento morfológico e sintático na construção do predicado, e por um sintagma nominal que entra em composição com o verbo para configurar o sentido do todo, bem como para determinar os papéis temáticos da predicação.

Em (Neves 1996), a autora posiciona o estudo das construções com verbo-suporte como parte integrante da investigação das predicções da língua, já que, em  $V_{sup} + SN$ , o  $SN$  não pode ser considerado objeto do verbo, mas uma espécie de “predicante”. Em (Neves 2000), encontra-se a definição do *verbo-suporte* como “ verbos de significado esvaziado que formam com seu complemento (objeto direto), um significado global, geralmente correspondente ao que tem um outro verbo da língua”. A autora também propõe tipos semânticos que caracterizariam o uso do verbo-suporte: verbo de ação (DAR *um beijinho*), processo (TOMAR *conhecimento*), e estado (TER *noção*). Ainda em (Neves 1996), a autora aponta as funções das construções com verbo-suporte nos enunciados, exemplificando, em relação à construção com o verbo pleno correspondente ao SN, casos de necessidade de maior versatilidade sintática, maior adequação comunicativa, maior precisão semântica e variação na configuração textual.

Deixando de lado a questão do conteúdo semântico e o valor pragmático do verbo-suporte, Ranchhod (1983) segue Zelig Harris, ao analisar a construção com verbo-suporte como uma **transformação gramatical** de uma sentença com verbo pleno, em que este é nominalizado, como por exemplo:

He studies eclipses. → He makes studies of eclipses.

Sendo assim, o verbo-suporte carrega a flexão do elemento predicador para marcas de tempo, aspecto, pessoa e número. Variantes aspectuais do verbo-suporte seriam realizadas algumas vezes por substituição de um verbo por outro na posição de suporte, como por exemplo:

ter esperança → acalentar esperança  
 → alimentar esperança  
 → nutrir esperança

Os critérios de identificação de uma construção com verbo-suporte apontados por Ranchhod para o verbo ESTAR baseiam-se na possibilidade de delimitação dos **nomes predicativos**, ou seja, substantivos abstratos que denotem processo ou estado. Dentro de uma perspectiva semelhante, a da léxico-gramática, (Baptista 2000) observa que a propriedade mais geral das construções com verbo-suporte é a existência de uma relação particular entre o sujeito e o substantivo predicativo.

As nominalizações são os casos mais claros de substantivos predicativos, por poderem ser morfológicamente reconhecidas. De uma maneira geral, tais critérios esbarram em dificuldades notórias, como a distinção abstrato × concreto e estado inerente × não-inerente.

O **verbo de ligação** também aparece em análises semelhantes, como (Mejlachowicz 2003), que enfatizam o esvaziamento semântico do verbo, que faria apenas o papel de elo de ligação entre o sujeito e o elemento predicativo, de valor adjetivo.

A delimitação dos **verbos gerais** em (Allerton 1984) exemplifica o quanto o fenômeno verbal pode ser estendido aos substantivos, como elaboro na seção 4.3.2. Essa denominação é utilizada por Allerton para descrever uma classe de verbos muito próxima à dos verbos-suporte, com status semi-lexical e complementados por nominalizações deverbais. Além disso, ele observa que os verbos gerais formam uma classe fechada, ou seja, não se trata apenas de uma função de suporte que pode ser desempenhada eventualmente por qualquer verbo.

Outro termo designador da classe dos verbos-suporte é **verbo leve**, introduzido por Jespersen (1940). O termo *verbo leve* parece ser mais utilizado nas gramáticas do inglês, no entanto também é corrente em estudos semântico-lexicais do português (Scher 2003; Viotti 2003).

Butt (2003) discute os verbos leves de uma perspectiva inter-lingüística, citando trabalhos para o Japonês (construções Nome + Verbo), Romance (construções Verbo + Verbo), Hindi (construções Nome + Verbo) e Urdu (construções Verbo + Verbo). Butt enquadra os verbos leves dentro de uma teoria mais ampla de **predicados complexos**. As características mais marcantes desses verbos, que aparecem nas diversas construções em diferentes línguas, são relacionadas à vagueza do verbo qualificado como leve e à existência de uma forma verbal plena idêntica.

Sinclair coloca os verbos-suporte como parte de um fenômeno mais geral das línguas: a deslexicalização:

The meaning of words chosen together is different from their independent meanings. They are at least partly delexicalized. This is the necessary correlate of co-selection [...] there is a strong tendency to delexicalize in the normal phraseology of modern English. (Sinclair 1994, p. 22)

Sinclair exemplifica a deslexicalização com uma classificação de adjetivos entre os **seletivos**, que selecionam um conjunto menor a partir de um conjunto maior, e os **enfocadores** (“focusing”), que se fundem ao substantivo por serem de uma certa forma redundantes, repetindo parte do significado do substantivo, como por exemplo em *proximidade FÍSICA, experimento CIENTÍFICO e público GERAL*.

Scher (2003) observa que a pesquisa sobre as construções com verbos leves vem se concentrando em três questões principais:

- a determinação de suas propriedades lexicais (Basilio, Dias, & Martins 1994), sintáticas (Ranchhod 1990), (Baptista 2000) e semânticas (Neves 1999);
- a determinação da real natureza - lexical (Basilio, Dias, & Martins 1994) ou sintática (Gross 1984) - dos complexos;
- a investigação da relação entre estas propriedades.

Observações sobre as construções com verbo leve em inglês, já indicadas em (Poutsma 1926), apud (Scher 2003), apontam três características importantes que apresentam as construções com esses verbos:

1. o verbo leve é semanticamente vago;
2. o complemento do verbo leve tem como núcleo um substantivo que carrega uma parte da predicação, em geral morfologicamente deverbais;
3. a construção admite uma variação, ou paráfrase, em que o verbo base do núcleo deverbais do SN é o verbo principal (*deu um grito* → *gritou*).

### 4.3.2

#### O substantivo-suporte

Tomando como base a seção anterior, o objetivo desta seção é apresentar os argumentos para o enquadramento do substantivo-suporte no mesmo fenômeno lingüístico do verbo-suporte. Apresento a seguinte ordem de argumentação:

1. o substantivo-suporte é semanticamente vago;
2. o complemento do substantivo-suporte, um adjetivo denominal, tem como núcleo um nome, que veicula a parte mais relevante da denominação;
3. a construção admite paráfrase, quer em termos da substituição do substantivo-suporte, mantendo-se o adjetivo, quer em termos da substituição da construção *S-Adj* por uma expressão cujo núcleo é o substantivo base do adjetivo.

Os exemplos comentados estão listados no exemplário, na seção 4.4. Ressalvo que os exemplos não são exaustivos com relação aos detalhes dos contextos sintáticos de cada substantivo-suporte analisado. Seria necessária uma análise em detalhe de cada um deles para que fossem especificadas as possibilidades de ocorrência plena ou em função de suporte. No entanto, há um exemplo para cada substantivo-suporte da tabela 4.1.



**O substantivo-suporte é semanticamente vago.** Identificando como vagueza o esvaziamento lexical, observado por vários autores citados, pode-se demonstrar o quanto a ausência de especificação do substantivo-suporte torna o enunciado problemático. Observando os exemplos 4.10, em 2. o apagamento do adjetivo denominal POLÍTICA torna o enunciado anômalo pois ESFERA como substantivo-suporte é demasiadamente vago; em contraste, em 4. o apagamento do adjetivo não provoca o mesmo efeito.

**ex. 4.10** *Exemplo 11.*

1. *Na ESFERA política, os manda-chuvas aprenderam a tratar os conflitos intestinos com maior flexibilidade e eficácia.*
2. *? Na ESFERA, os manda-chuvas aprenderam a tratar os conflitos intestinos com maior flexibilidade e eficácia.*
3. *No PARTIDO político, os manda-chuvas aprenderam a tratar os conflitos intestinos com maior flexibilidade e eficácia.*
4. *No PARTIDO, os manda-chuvas aprenderam a tratar os conflitos intestinos com maior flexibilidade e eficácia.*

Com relação ao exemplário da seção 4.4, a impossibilidade de interpretação (não extravagante), pode ser verificada nos exemplos:

1, 3, 4, 5, 6, 8, 9, 11, 12, 13, 14, 15.b) 16, 17, 18, 19, 20, 23, 24, 25, 28, 29, 30, 31.b)

No caso do exemplo 8, CUNHO não ocorre no corpus fora de construção com substantivo-suporte. Já nos exemplos 5 e 6, os substantivos grifados assumiriam um de seus sentidos plenos (CAMPO = ‘fora da cidade’ e CARÁTER = ‘firmeza moral’):

**ex. 4.11** *Exemplos 5 e 6.*

1. *No CAMPO, implicava na obediência aos cânones do bom gosto. formas universais baseadas em outra idéia universal: a razão, elo de ligação entre todos os homens.*
2. *A monarquia egípcia tinha CARÁTER: o faraó era considerado deus, filho de Osíris.*

Os exemplos 2, 7 e 10 só seriam interpretáveis no caso de uma anáfora, em que ÁREA / COMPONENTE / ELEMENTO funcionariam como elementos de coesão textual:

**ex. 4.12** *Exemplo 2.*

1. *Com relação ao retorno à sociedade Sotovia diz que ele ocorre “não apenas no atendimento específico à população, acontece na diferenciação dos profissionais da ÁREA, sejam os graduados por aqui, ou os formados por outras faculdades que aqui se especializam ou atualizam”.*

Nos exemplos com MANEIRA (15.a e b) apresento duas possibilidades: na primeira, a MANEIRA *dicotômica* o apagamento não impossibilita a interpretabilidade, ao contrário de de MANEIRA *hierárquica*. Os exemplos 31.a e b (TOM) apresentam o mesmo comportamento.

**ex. 4.13** *Exemplo 15.*

1. a) *Postulados dessa natureza deixam bastante evidente a MANEIRA com que se costuma, novamente em certos círculos intelectuais, tratar as questões da cultura e da tecnologia.*
2. b) ? *A origem da religião está registrada nos escritos destes povos: na poesia épica suméria, os deuses eram representados de MANEIRA.*

Nos exemplos restantes, 15.a), 21, 22, 26, 27, 31.a), não houve impossibilidade de interpretação, apesar de haver evidência de vagueza, dada a flutuação de sentido percebida com relação ao substantivo. A situação nesses exemplos é semelhante aos casos de verbos transitivos, como por exemplo BEBER, que na ausência de complementação assumem um sentido intransitivo peculiar.

**O complemento do substantivo-suporte tem como núcleo um adjetivo**, que carrega uma parte da denominação, em geral morfológica ou sintaticamente denominal. Este ponto está bastante evidente, pois todos os casos listados exemplificam a afirmativa. Por outro lado, a ferramenta de concordância de que disponho não permite uma consulta de cunho morfológico, logo não poderia quantificar os casos em que algum dos substantivo-suporte estivesse complementado por um adjetivo morfológica e sintaticamente simples. Vale a pena mencionar os casos de adjetivos não denominais que, por razões históricas, já não permitem uma análise morfológica sincrônica, tais como: DOMÉSTICO, LÍRICO, METABÓLICO, DIALÉTICO, entre outros.

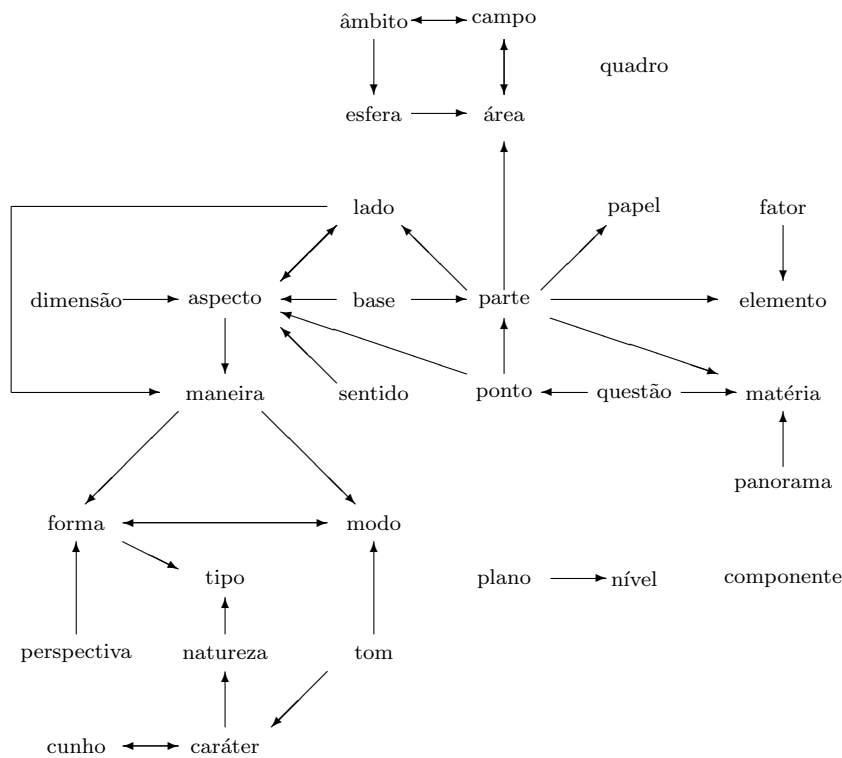


Figura 4.4: Diagrama das referências cruzadas entre definições de substantivos-suporte (reiteração da figura 4.2)

**A construção admite paráfrase,** quer em termos da (a) substituição do substantivo-suporte, mantendo-se o adjetivo, quer em termos da (b) substituição da construção *S-Adj* por uma expressão cujo núcleo é o substantivo base do adjetivo.

O tipo (a) de paráfrase é equivalente à intercambialidade entre substantivos-suporte. A evidenciação desta propriedade motiva uma classificação em que os membros das classes se equivalham, ou desempenham uma função muito semelhante, dentro das construções. Re-analisando o grafo da figura 4.2, reiterado na figura 4.4, percebe-se o quanto ele parece induzir uma classificação, por sua configuração agrupada. Os quatro grupos de substantivos-suporte resultantes da interpretação da figura podem ser descritos de acordo com as seguintes funções<sup>2</sup>:

**1. função de delimitação,** que especifica o objeto do enunciado numa estrutura abstrata de campos de informação, onde são classificados ÂMBITO, CAMPO, ESFERA e ÁREA; essa função ficou concentrada na região superior do grafo;

**2. função de sub-divisão,** que determina os níveis ou partes dentro de um

<sup>2</sup>Alguns substantivos-suporte da tabela 4.1 não aparecem no grafo porque suas definições em (Houaiss 2001) não se relacionam por referência cruzada com as definições dos demais

todo, onde são classificados LADO, ASPECTO, PARTE, PONTO, FATOR ELEMENTO, MATÉRIA, QUESTÃO e NÍVEL;

**3. função de enfoque**, que especifica o modo como o objeto é abordado, onde são classificados SENTIDO, MODO, MANEIRA, FORMA, DIMENSÃO, PERSPECTIVA e PLANO;

**4. função de essência**, que define o objeto do enunciado total ou basicamente, onde são classificados CUNHO, CARÁTER, NATUREZA, BASE, TIPO e TOM. Essa função ficou concentrada na região inferior do grafo.

Observando o grupo correspondente à função de delimitação, extraí dos exemplos 1 (ÂMBITO), 2 (ÁREA), 5 (CAMPO) e 11 (ESFERA) os SNs em que ocorrem os substantivos-suporte, obtendo o seguinte esquema de intercambialidade:

$$\begin{array}{cc}
 \text{âmbito} \left\{ \begin{array}{l} \textit{doméstico} \\ \textit{odontológico} \\ \textit{artístico} \\ \textit{político} \end{array} \right. & \text{área} \left\{ \begin{array}{l} \textit{doméstica} \\ \textit{odontológica} \\ \textit{artística} \\ \textit{política} \end{array} \right. \\
 \\
 \text{campo} \left\{ \begin{array}{l} \textit{doméstico} \\ \textit{odontológico} \\ \textit{artístico} \\ \textit{político} \end{array} \right. & \text{esfera} \left\{ \begin{array}{l} \textit{doméstica} \\ \textit{odontológica} \\ \textit{artística} \\ \textit{política} \end{array} \right.
 \end{array}$$

Essa classificação é suficientemente motivada pelos dados mas ainda não foi testada exaustivamente no corpus. No capítulo 6, de conclusões, sugiro a ampliação da análise dentro de uma perspectiva cognitivista.

O tipo (b) de paráfrase, substituição da construção *S-Adj* por uma expressão cujo núcleo é o substantivo base do adjetivo, é analisado, para cada exemplo da seção 4.4, na tabela 4.9.

De uma maneira geral, a tabela demonstra a grande regularidade das paráfrases, ainda mais quando contabilizadas por grupos de função.

1. o grupo da função de delimitação se apresenta totalmente regular;
2. o grupo da função de sub-divisão, o mais numeroso, apresenta dois exemplos problemáticos: FATOR e PONTO;
3. o grupo da função de enfoque apresentam maior irregularidade nas paráfrases (4 / 4); ou seja, as construções com substantivo pleno apre-

sentam um grau de estranheza maior, no entanto não apresentam agramaticalidade, como pode ser observado em:

- ... *tanto da PERSPECTIVA neurológica como da ecológica.*
- ... *tanto na neurologia como na ecologia*

4. o grupo da função de essência fica bem parafraseado somente com o apagamento do substantivo-suporte, de maneira regular.

As análises apresentadas confirmam de modo objetivo a hipótese de que o substantivo-suporte tem um comportamento completamente análogo ao verbo-suporte, indicando que a função de suporte não se restringe aos verbos.

Ex.	Função	Paráfrase	exata (+) ou aproximada (-)
1	1	ÂMBITO doméstico → nos lares	+
2	1	ÁREA odontológica → odontologia	+
5	1	CAMPO artístico → nas artes	+
11	1	ESFERA política → na política	+
21	1	PANORAMA político → na política	+
3	2	ASPECTOS metabólicos e morfológicos → o metabolismo e a morfologia	+
7	2	COMPONENTE política → a política	+
10	2	ELEMENTO dialético → a dialética	+
12	2	FATOR biológico → a biologia	-
14	2	LADO harmônico → a harmonia	+
16	2	MATÉRIA filológica → em filologia	+
19	2	NÍVEL sintático → a sintaxe	+
23	2	PARTE tática → na tática	+
26	2	PONTO metodológico → ?	?
27	2	QUADRO macroeconômico → na macroeconomia	+
28	2	QUESTÃO mercadológica → pelo mercado	+
9	3	DIMENSÃO diacrônica → a diacronia	+
13	3	FORMA diplomática → com diplomacia	+
15	3	a) MANEIRA dicotômica → a dicotomia b) MANEIRA hierárquica → com hierarquia	+
17	3	MODO lógico → com lógica	+
22	3	PAPEL biológico → a biologia	-
24	3	PERSPECTIVA neurológica → na neurologia	-
25	3	PLANO fenomenológico → da fenomenologia	-
29	3	SENTIDO matemático → na matemática	-

Tabela 4.9: Paráfrases para *S-Adj* com substantivos-suporte

Ex.	Função	Paráfrase	exata (+) ou aproximada (-)
4	4	BASE filosófica → da filosofia filosófica (somente o apagamento)	- +
6	4	CARÁTER teocrático → teocracia teocrático(somente o apagamento)	- +
8	4	CUNHO ideológico → de ideologia	+
18	4	NATUREZA mecânica → da mecânica mecânica (somente o apagamento)	- +
20	4	ORDEM semântica → de semântica semântica (somente o apagamento)	- +
30	4	TIPO mongólico → da Mongólia mongólica (somente o apagamento)	- +
31	4	a) TOM homérico → ? b) TOM profético → profecia profético (somente o apagamento)	? - +

Tabela 4.9: Paráfrases para *S-Adj* com substantivos-suporte (cont.)

#### 4.4 Exemplário

Os exemplos abaixo foram retirados de um sub-corpus que extraí do corpus do NILC, contendo adjetivos em ÍCO. A lista contém um exemplo para cada substantivo-suporte, sendo que os números 15 MANEIRA e 31 TOM exibem dois exemplos cada um.

1. Apesar do sucesso das modinhas no ÂMBITO doméstico, o cenário musical durante o Império foi dominado pelo canto lírico.
2. Com relação ao retorno à sociedade Sotovia diz que ele ocorre “não apenas no atendimento específico à população, acontece na diferenciação dos profissionais da ÁREA odontológica, sejam os graduados por aqui, ou os formados por outras faculdades que aqui se especializam ou atualizam”.
3. Os ASPECTOS metabólicos e morfológicos são mantidos no novo organismo, graças ao núcleo.
4. Também não se pode prever quais cientistas, ou que ciência, concordarão com uma formulação de BASE filosófica do novo campo.
5. No CAMPO artístico, implicava na obediência aos cânones do bom gosto. formas universais baseadas em outra idéia universal: a razão, elo de ligação entre todos os homens.
6. A monarquia egípcia tinha CARÁTER teocrático: o faraó era considerado deus, filho de Osíris.
7. Sob o prisma da estratégia de desenvolvimento sustentável a adotar-se para o Haiti, é óbvio que tanto o BID como o Banco Mundial estariam melhor qualificados que a ONU para gerir o processo, mas no caso haitiano a COMPONENTE política é tão importante e sensível que melhor seria confiá-la a um organismo também político.
8. As relações diplomáticas Jamaica-Estados Unidos só melhoram já na década de 80, quando o governo socialista é substituído por um governo de CUNHO ideológico mais direitista, que

marca o início de uma era mais conservadora na história jamaicana.

9. Parece-me que o texto caminha na seguinte direção: de um lado quer fazer uma análise sincrônica, de caso, e, de outro, deseja projetar a DIMENSÃO diacrônica, histórica.

10. Hill observa outro ponto: “O ELEMENTO dialético no pensamento científico foi grandioso nas descobertas da filosofia mecânica, um reconhecimento do irracional (no sentido do inexplicável mecanicamente) ficou perdido quando ele triunfou, e está sendo dolorosamente recuperado em nosso próprio século.”

11. Na ESFERA política, os manda-chuvas aprenderam a tratar os conflitos intestinos com maior flexibilidade e eficácia.

12. A questão se resumiria em saber se foi mesmo um FATOR biológico ou um acontecimento se situando num plano puramente espiritual.

13. Os treinadores, do mesmo jeito, tratam-se de uma FORMA diplomática, mas não levam uma conversa por mais de 10 minutos.

14. Engraçado é que eles se interessam pelo seu trabalho inicial, que é mais rítmico, e têm menos interesse pelo LADO harmônico.

15.a) Postulados dessa natureza deixam bastante evidente a MANEIRA dicotômica com que se costuma, novamente em certos círculos intelectuais, tratar as questões da cultura e da tecnologia.

15.b) A origem da religião está registrada nos escritos destes povos: na poesia épica suméria, os deuses eram representados de MANEIRA hierárquica.

16. Se lhes faltava competência em MATÉRIA filológica, sobrava-lhes certamente percepção em MATÉRIA artística e musical.

17. Cohen (1981) dá três golpes na noção de que a maioria dos indivíduos não se comporta de MODO lógico ou racional.

18. Nós já vimos que as ondas sonoras são de NATUREZA mecânica e por isso não podem se propagar no vácuo.

19. Chomsky de fato falou sobre como o NÍVEL sintático interage com a fonologia e a semântica; mas durante anos concentrou suas energias analíticas no nível da sintaxe (ver, porém, Chomsky e Halle 1968) .

20. A diferença está na imensa dificuldade de se formalizar e traduzir em linguagem de máquina considerações de ORDEM semântica.

21. No PANORAMA político, era possível notar a disposição governamental de implantar um projeto de liberação controlada do regime, apesar de dificuldades de diferentes naturezas.

22. Caracterizaremos, de forma resumida, o PAPEL biológico básico das principais organelas celulares.

23. Na PARTE tática, a grande preocupação palmeirense é com o jogo aéreo do Grêmio, baseado principalmente no centroavante Jardel.

24. E talvez um estudo do modo imagístico, menos conhecido, pudesse ajudar a esclarecer alguns dos estágios posteriores da percepção visual e fazê-lo de uma maneira diferente tanto da PERSPECTIVA neurológica como da ecológica.

25. Acho que minha abordagem se dá dentro de um PLANO fenomenológico, por meio de modos de percepção e não por meio de temas.

26. Isso suscita um importante PONTO metodológico acerca do papel das hipóteses na teorização: esse papel é, entre outros, o de especificar a amplitude das supostas aplicações de uma teoria.

27. Apesar da euforia nas Bolsas de Valores e do apoio dos empresários ao Plano FHC, o QUADRO macroeconômico traçado pelos economistas para o primeiro semestre deste ano não é cor-de-rosa.
28. Por uma QUESTÃO mercadológica, o português não está incluído, mas a versão 4.0 é o primeiro produto da Lotus que vai apresentar uma versão em português apenas 30 dias depois da versão em inglês.
29. No SENTIDO matemático da teoria dos conjuntos, onde dois termos são distos equivalentes segundo uma certa relação de ordem, a relação de equivalência.
30. A população japonesa, do TIPO mongólico (da raça amarela), concentra-se sobretudo nas áreas litorâneas, principalmente nas planícies de Tóquio, de Nagóia e de Osaka.
- 31.a) Até o TOM homérico em que são escritas, concorre para essa monotonia.
- 31.b) Seus dois representantes mais evidentes apesar de sua ligação com o surrealismo nunca ter sido direta eram Saint-John Perse e René Char, que faziam uma poesia de TOM profético, oracular.



## 5

### O substantivo-suporte no corpus

*It doesn't matter how beautiful your theory is, it doesn't matter how smart you are. If it doesn't agree with experiment, it's wrong. (Richard P. Feynman)*

#### 5.1

##### Introdução

A Lexicografia Computacional iniciou-se como um conjunto de métodos automáticos de utilização de versões digitais de dicionários convencionais como fonte de informações para sistemas de PLN (Ooi 1998). Com a ampliação dos recursos disponíveis - evolução das tecnologias e a disponibilidade de vastas coleções de textos - e o aumento das exigências sobre os recursos lexicais utilizados por sistemas de PLN cada vez mais sofisticados, a Lexicografia Computacional passou a contribuir com questões relevantes da Lexicologia e da Lexicografia Lingüística.

Dentro desse contexto, o trabalho experimental de delimitação das expressões com substantivo-suporte consiste na análise de contextos para caracterizar os elementos do significado que se organizam dentro da expressão considerada, e como se comporta semanticamente essa expressão dentro de contextos maiores.

Na seção 5.2 consideramos a primeira parte do problema, ou seja o processo de composição de informações semânticas provenientes de duas palavras em uma expressão.

Na seção 5.3 consideramos a segunda: a expressão em contexto.

#### 5.2

##### Composicionalidade semântica

O termo *composicionalidade* foi utilizado por Katz & Fodor (1963) para descrever um aspecto da competência semântica do falante de uma língua. Para os autores, composicionalidade é um modelo de interpretação semântica e produção baseado em um cálculo: cada átomo ou unidade lingüística, como a palavra, contribui para o significado global do enunciado, sistematicamente.

A principal idéia é a existência de um mecanismo pelo qual o significado de uma cadeia lingüística é decodificada a partir do que o falante sabe sobre cada unidade componente e sobre a organização dessas unidades. Cada vez que uma estrutura ou enunciado ocorre, seu significado é composicionalmente calculado. Segundo Searle, a idéia de composicionalidade de um enunciado está associada ao seu significado literal:

“The literal meaning of a sentence is entirely determined by the meanings of its component words (or morphemes) and the syntactical rules according to which these elements are combined”. (Searle 1979, p. 117)

Assim, o autor garante a existência de pelo menos uma interpretação do enunciado por meio desse cálculo. O otimismo com relação à possibilidade de seleção desse conjunto de pares ⟨palavra, significado literal⟩, para cada enunciado, vem sendo abandonado desde então.

Cruse (1986) apresenta a semântica de uma sentença como o resultado da combinação de sentidos mais simples, ou seja, contribuições semânticas de constituintes cada vez mais simples. Assim, o autor caracteriza um **constituente semântico** (p.25) como um par ⟨forma, significado⟩. Cruse pressupõe a associação estável entre um significado e um constituinte sentencial que se repete a cada ocorrência do constituinte. Fica patente a noção de composicionalidade subjacente ao conceito. Nesse contexto, a não-composicionalidade de constituintes semânticos atípicos, referidos por Cruse como “peripheral types of semantic constituents” (p. 29) seria devida à “unicidade colocacional”, que explicaria a discrepância do cálculo composicional.

Fillmore (1979), por outro lado, é particularmente crítico ao modelo de cálculo, a que chama de “a segunda idealização” da Lingüística, em referência ao construto do falante ideal do modelo chomskyiano.

Ao argumentar que o substantivo, ou o adjetivo, é o centro do significado em um enunciado, a noção de composicionalidade é utilizada implicitamente. Será necessário então adotar o modelo do cálculo para postular a existência das expressões com substantivo-suporte? Utilizo (Cruse 2004) para me posicionar com relação a esta questão. O autor enuncia o **princípio da composicionalidade** como:

“The meaning of a grammatically complex form is a compositional function of the meaning of its grammatical constituents.” (Cruse 2004, p.65)

Sendo esta a versão “forte” do princípio, Cruse separa três proposições subsumidas:

1. “The meaning of a complex expression is completely *determined* by the meaning of its constituents.”
2. “The meaning of a complex expression is completely *predictable* by general rules from the meaning of its constituents.”
3. “Every grammatical constituent has a meaning which contributes to the meaning of the whole” (*Ibid.*)

Das três, a terceira é a menos ambiciosa, exigindo apenas uma participação sistemática dos constituintes na interpretação do todo da expressão. Essa é a única possibilidade de compromisso entre a metodologia apresentada a seguir e o conceito tradicional de composicionalidade.

### 5.3 O Método

A identificação das ocorrências de substantivos-suporte no corpus pode ser vista como uma instância de Desambiguação do Sentido Lexical. A questão central é como o falante distingue o sentido de uma palavra ambígua ou vaga quando se depara com um enunciado onde ela ocorre. Trata-se de um problema muito relevante em Linguística Computacional visto que o computador, a princípio, não conta com esta capacidade de discernir, uma função essencial do sistema lingüístico humano, de extrema eficiência. Uma formulação simplificadora do problema seria separar em agrupamentos as diversas ocorrências de uma palavra ambígua, de acordo com seu significado na ocorrência.

O método desenvolvido para este trabalho tem como motivação a questão colocada por Kilgarriff (1997) quando discute o que é o sentido de uma palavra. A partir de um modelo computacional em que o sentido da palavra (*word sense*) ou a unidade lexical não são unidades básicas, mas sim as ocorrências de uma palavra em contexto, operacionalizadas como extratos de um corpus, ele elabora:

“To know what a word sense  $s_1$  is, is to know which uses of the word are part of  $s_1$  and which are not, probably because they are part of  $s_i$ , where  $i \neq 1$ . If we are to know what word senses are, we need operational criteria for distinguishing them” (Kilgarriff 1997, p.97)

Nosso método é, portanto, um critério operacional para distinguir substantivos-suporte e suas ocorrências dentro de um corpus.

Em um trabalho que analisa as expressões verbais da forma (Verbo + Sintagma Nominal), Garrão (Garrão 2006) discute esta questão ao propor que, na construção de léxicos eletrônicos, a delimitação das unidades lexicais complexas, as chamadas Expressões Multi Vocabulares (EMVs), deva ser compreendida com base em uma análise de corpus. A abordagem experimental utilizada em (Garrão 2006) e introduzida em (Garrão *et al.* 2006) é bem próxima à empregada aqui.

### 5.3.1

#### O Modelo do Espaço Vetorial

Schütze & Pederson (1995) introduziram o Modelo do Espaço Vetorial em uma abordagem bastante original para o problema de Desambiguação de Sentidos Lexicais. A idéia é utilizar vetores de grandes dimensões na representação do contexto de cada ocorrência de uma palavra em foco ( $w$ ), e agrupar esses contextos por uma medida de similaridade entre os vetores. Os agrupamentos obtidos seriam avaliados de acordo com sua coesão: os mais coesos corresponderiam a sentidos diferentes da palavra  $w$ . Para identificar o sentido de uma nova ocorrência de  $w$ , essa ocorrência é representada por um vetor que será comparado aos centróides dos agrupamentos previamente construídos. Os sentidos eleitos serão os representados pelos agrupamentos mais próximos ao vetor da nova ocorrência.

O Modelo do Espaço Vetorial, descrito com grande clareza em (Manning & Schütze 1999), é o mais utilizado entre os modelos de recuperação de documentos, devido à sua simplicidade conceitual e a clareza da metáfora da proximidade espacial entre documentos, representados como vetores de palavras. A representação de um documento nesse modelo é feita considerando cada palavra como uma dimensão no espaço vetorial; o tamanho do vocabulário contido na coleção de documentos determina a dimensão total do espaço. A proximidade entre documentos é medida pelo ângulo entre os respectivos vetores.

Como exemplo, vamos considerar as seguintes expressões como os dois únicos textos de uma coleção:

Doc1: a dog and a cat

Doc2: and a frog and a cat

O espaço vetorial correspondente seria composto por cinco dimensões

$(a, dog, and, cat, frog)$

e conteriam dois vetores. Considerando a frequência bruta da palavra como o tamanho do vetor na dimensão dessa palavra, os vetores seriam:

$$(2, 1, 1, 1, 0)$$

$$(2, 0, 2, 1, 1)$$

Qualquer que seja a língua em questão, é natural que nem todas as palavras que ocorrem no documento sejam importantes na avaliação de seu tópico. Em sistemas manuais de indexação de documentos, os **termos** são determinados manualmente por um especialista, que selecionam palavras ou expressões que descrevam o conteúdo do documento, tendo em vista as palavras-chaves que o usuário do sistema irá fornecer para encontrar o documento indexado. Já em um sistema automático, principalmente os que lidam com documentos de domínios diversos, o primeiro passo para a seleção dos termos para a indexação é a eliminação das palavras gramaticais, normalmente listadas em um conjunto denominado **stopwords**. Muitas vezes o conjunto de stopwords inclui também palavras de altíssima frequência, já que isso indicaria baixo poder discriminatório. Não utilizei esse tipo de stopword pois certamente teria que incluir alguns substantivos-suporte.

O segundo passo no tratamento das palavras que serão indexadas é **pesagem dos termos**. Em Recuperação de Informações, esta pesagem é uma maneira de capturar a noção de **poder discriminatório**, promovendo com pesos altos as palavras de maior poder. Dessa maneira, existem três fatores principais na pesagem dos termos: **frequência do termo** (número de ocorrências de um termo em um documento), **frequência documental** (número de documentos onde ocorre o termo) e **coeficiente de normalização** em função do tamanho do documento. Esses fatores são combinados para definir o esquema de pesagem dos termos.

A frequência do termo quantifica a participação do termo no documento, portanto conta positivamente na pesagem. A frequência documental, sendo alta, tende a indicar que o termo não é muito informativo, pois palavras de maior especificidade tendem a se concentrar em pequenos conjuntos de documentos, enquanto que palavras de significado mais geral tendem a ocorrer em toda a coleção. Portanto o fator frequência documental conta negativamente. Finalmente, o coeficiente de normalização é utilizado para compensar o fato de que documentos maiores tendem a conter um maior número de termos distintos, o que faz com que sejam mais facilmente recuperados que os menores. Dentre os muitos esquemas de pesagem avaliados na literatura, o esquema que utilizei, que multiplica a frequência do termo pelo inverso da frequência documental, junto com a normalização do tamanho do documento, apresenta um

desempenho equiparável aos melhores (Salton & Buckley 1988).

O cálculo das similaridades entre dois vetores de documentos pode ser feito de várias maneiras, utilizando medidas como a distância de Euclides, coeficiente de Jaccard, coeficiente de Dice e a medida do cosseno, para citar as mais usadas. Essa última baseia-se no ângulo entre dois vetores de documentos (Baeza-Yates & Neto 1999). O cosseno entre os vetores  $\vec{u}$  e  $\vec{v}$  é dado pelo coeficiente entre o produto escalar dos vetores por seus módulos:

$$\cos(u, v) = \frac{\sum(u_i \times v_i)}{\sqrt{\sum u_i^2 \times \sum v_i^2}}$$

Por exemplo, a medida do cosseno de similaridade entre “a dog and a cat”(2,1,1,1,0), e “a frog and a cat”(2,0,2,1,1) é dada por:

$$\frac{2 \times 2 + 1 \times 0 + 1 \times 2 + 1 \times 1 + 0 \times 1}{\sqrt{(2^2 + 1^2 + 1^2 + 1^2 + 0) \times (2^2 + 0 + 2^2 + 1^2 + 1^2)}} = 0.84$$

A matriz de similaridades dos  $n$  documentos é uma matriz simétrica  $n \times n$ , onde cada célula  $m_{ij}$  contém a medida de similaridade entre os documentos  $d_i$  e  $d_j$ .

### 5.3.2

#### Explorando o contexto

Nessa parte experimental do trabalho, definimos **contexto** de uma dada palavra ou expressão como o parágrafo do corpus onde ocorre. Visto que o corpus foi construído a partir de resultados de busca na Internet<sup>1</sup>, ao invés de coletar uma só sentença ou um número arbitrário de sentenças, decidi coletar parágrafos inteiros, de tamanhos variados.

Para cada expressão  $S + Adj$  a ser examinada, coletei três conjuntos de contextos:

**conj1** - contextos contendo  $S Adj$

**conj2** - contextos contendo  $S$  mas não  $S Adj$

**conj3** - contextos contendo  $Adj$  mas não  $S Adj$

Os seguintes contextos são exemplos dos conjuntos coletados para a análise da expressão NATUREZA CULTURAL.

#### ex. 5.1

<sup>1</sup>O corpus descrito na seção 4.2.6 mostrou-se insuficiente para prover o número de exemplos necessários para o nosso método. Busquei coletar os parágrafos na Internet de maneira sistemática, utilizando o Google<sup>TM</sup>, processando as páginas de resultados de 3 em 3, para cobrir em extensão as páginas fornecidas pelo buscador.

**conj1** - *Os Estados-Partes estimularão a adoção de medidas que facilitem o trânsito de agentes culturais, vinculados à execução dos projetos de NATUREZA CULTURAL.*

**conj2** - *A modéstia deveria ser cultivada da mesma maneira pelos dois sexos [...] mas há homens que vão continuar repetindo que a mulher deve ter mais modéstia do que o homem [...] eles são incapazes de apreciar o prazer que provoca a virtude da renúncia [ no entanto: ] As mulheres, tanto quanto o homem, deveriam ter as paixões e os desejos próprios à sua NATUREZA; a bestialidade só aparece quando a razão não os controla, mas a obrigação de controlar paixões e desejos é o dever de toda a humanidade e não o dever de um sexo só.*

**conj3** - *Com seu orçamento invariavelmente contingenciado, o Ministério da Cultura bem que gostaria, mas tem optado por não mexer na lei de incentivo por temer um colapso na produção CULTURAL brasileira que nestes dez anos se tornou absolutamente dependente deste tipo de patrocínio - que, na verdade, não é real, porque está baseado em isenção fiscal. Esta é a justificativa apresentada, por exemplo, para que o decreto que traz as primeiras mudanças na Lei Rouanet na gestão de Gilberto Gil esteja juntando pó na Casa Civil.*

### 5.3.3

#### Medindo a composicionalidade

No modelo do espaço vetorial, a composicionalidade de uma expressão  $[X Y]$  pode ser heurísticamente medida usando as similaridades entre os contextos de  $[X Y]$ ,  $X$  e  $Y$ . Assume-se que, se  $[X Y]$  é não composicional, os contextos de  $[X Y]$  são muito distantes dos de  $X$  e de  $Y$ . Zhai (1997) chama essa medida de “Context Similarity”, e a utiliza na delimitação de expressões multi-vocabulares (ou “átomos lexicais”).

Em (Garrão *et al.* 2006), um método semelhante é empregado para verificar quando a expressão  $V SN$  é não composicional e, de uma perspectiva lexicográfica, propor que sejam armazenadas em um léxico computacional. Os exemplos 5.2 e 5.3 mostram contextos daquele trabalho: em 5.2, USAR CAMISINHA é considerado composicional e em 5.3 TOMAR PARTIDO é considerado não composicional. Garrão *et al.* também consideram alguns casos limítrofes, onde o método é inconclusivo.

#### ex. 5.2

USAR CAMISINHA

*Se estão com Aids, eles informam o freguês ou companheiro e pedem para USAR CAMISINHA.*

*Você não ia ser louca de ter vida sexual sem USAR CAMISINHA, não é verdade?*

*É consenso entre os jovens de ambas nacionalidades a necessidade de se USAR CAMISINHA durante a transa.*

#### CAMISINHA

*Nesse sentido, pode ser inconveniente para muitas mulheres chamar a atenção para a CAMISINHA e estabelecer negociações sobre sexo seguro antes que tenham intimidade com seu parceiro.*

*O pai de André, o médico epidemiologista Glacus de Souza Brito, 36, um dia levou para casa uma CAMISINHA, desenrolou-a e mostrou ao filho.*

*Como psicólogo, ele acha que, para os meninos menores, a escolha da CAMISINHA como objeto de coleção tem a ver com o tamanho do pênis.*

#### ex. 5.3

#### TOMAR PARTIDO

*Mas é fácil, como dizia Swift, reconhecer o gênio: todos os imbecis da época TOMAR PARTIDO contra ele.*

*Era uma coisa complicada, que minha mãe não aceitava de vez em quando explodia, a gente via as consequências, tinha que TOMAR PARTIDO, isso ao longo de anos.*

*Os sérvios acusam a ONU e a Otan de terem TOMAR PARTIDO dos muçulmanos e croatas na guerra.*

#### PARTIDO

*A busca de coligações será um trabalho feito pela direção do PARTIDO.*

*As pressões virão de líderes antiquercistas do PARTIDO que ainda não conseguiram articular uma ação conjunta contra ele.*

*Petistas tentam mobilizar militância do PARTIDO para reverter a queda contínua nas pesquisas de opinião.*

Essa medida é adequada para os testes de identificação do meu objetivo, considerando-se que o padrão de composicionalidade esperado em uma expressão com substantivo-suporte *S Adj* será de tal modo que *S* contribui muito pouco para o significado, em comparação com a contribuição de *Adj*, uma



espécie de composicionalidade desequilibrada. Nos resultados das experiências, as evidências para essa hipótese são dadas por altos níveis de similaridade entre os contextos da expressão *S Adj* e os contextos da palavra que mais contribui na expressão. Portanto,

Similaridade entre os contextos de *S Adj* (**conj1**) e *Adj* (**conj3**)  
é maior que  
Similaridade entre os contextos de *S Adj* (**conj1**) e *S* (**conj2**).

Um resumo algorítmico do método proposto para computar o caráter de suporte do substantivo é apresentado na figura 5.1.

- 1 Selecionar  $n$  contextos com *S Adj*, *S* sem *Adj* e *Adj* sem *S*;
- 2 Remover as palavras *S* e *Adj* dos contextos
- 3 Indexar os contextos; e
- 4 Calcular a matriz de similaridades entre os três conjuntos de contextos.

Figura 5.1: Resumo do método de identificação de substantivos-suporte em contexto

O passo 3 foi incluído para que as palavras que constituem o padrão segundo o qual o contexto foi selecionado não sejam um fator de influência na similaridade entre contextos, especialmente porque os contextos são pequenos textos que poderiam ficar distorcidos pela presença de termos frequentes.

## 5.4 Experimentos

Um bom critério operacional para a caracterização do substantivo-suporte consiste na extração de expressões nominais *S Adj*, onde *Adj* é um adjetivo denominal. A lista resultante de substantivos pode ser analisada para determinar intercambialidade, isto é, co-ocorrência com o mesmo grupo de adjetivos mantendo o significado da expressão relativamente estável.

Utilizando o critério de intercambialidade, foi selecionada desse corpus uma lista preliminar de substantivos-suporte. Tomando como base essa lista, uma seleção de combinações *S Adj* foram identificadas, para servirem como dados empíricos para validar a hipótese subjacente ao método descrito na seção 5.3.3. A tabela 5.1 mostra as expressões selecionadas.

As tabelas 5.2, 5.3 e 5.4 mostram os resultados desses testes, multiplicados por um fator de 100.000 para que os valores sejam melhores visualizados.

<i>S Adj</i>	<i>S</i>	<i>Adj</i>
conj1	conj2	conj3
fator biológico	fator	biológico
papel biológico	papel	biológico
natureza cultural	natureza	cultural
natureza econômica	natureza	econômica
cunho cultural	cunho	cultural

Tabela 5.1: Expressões *S Adj* testadas

<i>S Adj</i>	<i>S</i>	<i>Adj</i>
fator biológico	19.27	43.42
papel biológico	8.38	48.86
natureza cultural	19.90	84.24
natureza econômica	13.68	99.13
cunho cultural	14.98	56.47

Tabela 5.2: Resultados com as expressões *S Adj*

A tabela 5.2 pode ser interpretada da seguinte forma. Tomando como exemplo a expressão na primeira linha, FATOR BIOLÓGICO, a similaridade entre os contextos de FATOR BIOLÓGICO e os contextos do substantivo FATOR é 19.27; a similaridade entre os contextos de FATOR BIOLÓGICO e os contextos do adjetivo BIOLÓGICO é 43.42. É importante observar que os valores de similaridade em cada linha são relativos somente aos contextos envolvidos na sintagma *S Adj*, portanto, somente os valores na mesma linha podem ser comparados. Cada grupo de três conjuntos de contextos (*S Adj*, *S* e *Adj*) foi indexado separadamente. Na prática, para cada conjunto, tem-se uma coleção independente, e valores de peso das palavras não se correlacionam de uma coleção para outra.

Em cada linha da tabela 5.2, a similaridade entre *S Adj* e *Adj* é mais do que o dobro da similaridade entre *S Adj*, *S*. Assim, pode se concluir que o contexto em que o adjetivo aparece e os contextos nos quais o sintagma nominal aparece são muito mais próximos daqueles contendo o substantivo e a sintagma nominal, dando suporte à hipótese inicial.

De modo a promover ainda mais a hipótese de substantivos-suporte, foi explorada a idéia de que, se o substantivo nas expressões nominais *SAdj* for substantivo-suporte, satisfazendo assim o critério de intercambiabilidade, as similaridades entre expressões nominais  $S_1Adj$  e  $S_2Adj$ , com  $S_1 \neq S_2$ , devem ser altas. Portanto, foi realizado o seguinte teste: depois de selecionar um adjetivo denominal e dois substantivos-suporte, foram comparados os contextos das expressões resultantes.

Na tabela 5.3, os substantivos são FATOR e PAPEL e o adjetivo é BIOLÓGICO; na tabela 5.4, os substantivos são CUNHO e NATUREZA e o adjetivo

fator biológico	
vs biológico	21.94
papel biológico	
vs biológico	25.44
fator biológico	
vs papel biológico	192.30

Tabela 5.3: Resultados com *biológico*

cunho cultural	
vs cultural	27.06
natureza cultural	
vs cultural	39.60
cunho cultural	
vs natureza cultural	56.12

Tabela 5.4: Resultados com *cultural*

é CULTURAL. Os resultados demonstram que as expressões  $S_1Adj$  e  $S_2Adj$  são consideravelmente mais próximas do que  $S_1Adj$  e  $Adj$  ou  $S_2Adj$  e  $Adj$ , sustentando ainda mais a hipótese inicial.

## Discussão

Muitas vezes, uma só palavra ou expressão é o suficiente para que o falante localize o discurso dentro de um assunto específico. Por exemplo, analisando os parágrafos do corpus que contêm o padrão  $S$ -DIDÁTICO, encontram-se 329 exemplos, a vasta maioria a respeito de ‘educação’ e ‘ensino’. A noção de **campo semântico**, grupo de palavras que co-ocorrem em um certo tipo de contexto, identificando uma relação de tópico (Stubbs 2002), é adequada para descrever o vocabulário encontrado no extrato do corpus, em que se destacam as frequências de algumas palavras como LIVRO, MATERIAL, RECURSO, TEXTO e AULA. Poucos parágrafos desse conjunto não se enquadravam nesse tópico, entre eles uma exceção digna de nota:

**ex. 5.4** *As coisas, os acidentes de percurso, os loucos atores do pastelão DIDÁTICO: Collor, Alves dos Santos, PC, as CPIs, tudo.*

Essas expectativas com relação às palavras e seus contextos são obviamente individuais, de cada falante, resultantes de sua vivência lingüística e seu conhecimento de mundo. No entanto, um corpus suficientemente representativo da experiência lingüística e cultural de um grupo é um instrumento fidedigno de medição de expectativa.

Os adjetivos que compõem as expressões apresentadas nos estudos de casos foram escolhidos por gerarem esse tipo de expectativa quanto ao campo semântico em que se inserem. Busquei o contraste entre o substantivo-suporte, vago, e o adjetivo, específico. Pelas características discutidas na seção 3.4.2, os adjetivos denotativos se apresentam muitas vezes com esse poder discriminatório quanto ao tema do contexto em que ocorrem. Assim, os resultados tão categóricos que obtive têm relação com essas escolhas.

## 6 Conclusões

*To forget one's purpose is the commonest form of stupidity.* (Friedrich Nietzsche)

### 6.1 Contribuições

Este trabalho foi desenvolvido a partir de um quadro teórico em que o léxico é formado, não só por itens listados, mas também por mecanismos que permitem e determinam sua expansão; não só por palavras mas também por construções.

O estudo foi desenvolvido de uma maneira essencialmente experimental, regido pelos dados provenientes do corpus. O próprio recorte lingüístico selecionado resultou da observação dos sintagmas nominais extraídos e tabulados de forma que a intercambialidade dos substantivos-suporte se tornou evidente e o contorno dessa classe começou a ser definido.

Utilizando o dicionário como fonte de conhecimento lexical e o corpus como fonte de dados lingüísticos, tracei um perfil lexicográfico dos substantivos-suporte, evidenciando principalmente seu caráter vago.

A partir de um estudo dos verbos-suporte, que manteve em primeiro plano a questão das classes de palavras envolvidas nas construções  $V_{sup}-SN$ , a noção de substantivo-suporte foi posicionada em paralelo à do verbo-suporte, demonstrando que o fenômeno de suporte é mais um dos processos de formação lexical que são compartilhados por verbos e substantivos.

Como um fecho para o objetivo principal da tese, o de prover um critério operacional para caracterizar substantivos em combinações  $S-Adj$ , o trabalho experimental de delimitação das expressões com substantivo-suporte foi realizado, utilizando a análise de contextos discursivos de modo a capturar uma certa noção de composicionalidade semântica quantificável a partir do corpus. Confirmando que no padrão de composicionalidade esperado em uma expressão com substantivo-suporte  $S-Adj$ , a contribuição de  $S$  é muito pequena em comparação com a contribuição de  $Adj$ , os resultados das experiências

mostraram os altos níveis de similaridade entre os contextos da expressão *S-Adj* e os contextos da palavra que mais contribui na expressão, o *Adj*.

## 6.2

### Desdobramentos

O trabalho de pesquisa desenvolvido nesse formato, um projeto de doutoramento, termina arbitrariamente no fim de um prazo, provavelmente quando se adquire finalmente amadurecimento no assunto e um entendimento razoável do objeto para construir modelos e tirar conclusões interessantes. É nesse ponto que se encontra essa pesquisa.

A lista de possíveis desdobramentos deste trabalho é extensa e ainda em expansão. Enumero aqui as extensões mais imediatas. Em primeiro lugar, o trabalho ficou restrito às expressões *S-Adj*, no entanto há outros padrões que devem ser examinados, principalmente o  $S_1$ -de- $S_2$ , pela semelhança semântica da expressão adjetiva de- $S_2$  com o próprio adjetivo e pela alta frequência de  $S_1$ -de- $S_2$  com  $S_1$  substantivo-suporte.

O substantivo-suporte COISA foi abandonado por escassez de dados. Por ser considerado o mais vago dos substantivos, considero que seu estudo é de grande interesse, principalmente na lexicologia de orientação cognitivista.

As funções do substantivo-suporte merecem ser melhor analisadas, em termos de escalas prototípicas, de tal modo que as funções seriam preferencialmente exercidas por certos itens, mas alguns itens poderiam exercer mais de uma função. Por exemplo, PLANO poderia funcionar como ‘sub-divisão’ ou como ‘ênfase’; NÍVEL, como ‘delimitação’, ‘sub-divisão’ ou ‘ênfase’; TOM seria basicamente ESSÊNCIA, mas poderia funcionar como ‘ênfase’; e assim por diante. A análise de corpus pode fornecer uma contraparte quantitativa para a questão central/periférico.

De grande relevância para o Léxico Computacional, uma proposta de representação computacional dos substantivos-suporte faz parte da seqüência natural de desdobramentos desta tese. Verbos-suporte já são representados nos sofisticados léxicos existentes (cf. (Fillmore, Wooters, & Baker 2001)) e o estudo desses esquemas de representação poderiam ser adaptados para abranger também os substantivos.

O método quantitativo de aferição de composicionalidade é extremamente sensível ao grau de dispersão dos tópicos que constam dos conjuntos analisados. Como está definido, o diagnóstico de substantivo-suporte depende do adjetivo e sua especificidade semântica. Seria importante desenvolver para o corpus um parâmetro de difusão do contexto, para que as coleções de contextos *S-Adj* possam ser comparáveis quanto aos campos semânticos dos vo-

cabulários.

### 6.3

#### Considerações Finais

A pesquisa em Lexicografia Computacional baseada em corpus depende muito da quantidade e da qualidade dos dados disponíveis. A vasta maioria das investigações sobre a língua inglesa encontradas na literatura utilizam dicionários digitais, cujas informações são acessáveis por meio de programas, e não apenas por interface com o usuário. Avaliações quantitativas, tais como a distribuição das entradas de acordo com classe de palavras e número de acepções também podem ser obtidas bem como, em certos casos, as distribuições de frequência no corpus de exemplos e contexto sintático da acepção nos exemplos (cf. (Sinclair 2001)).

Essa área está apenas engatinhando no que tange ao português, portanto muito está por fazer. Os corpora disponíveis academicamente não são robustos o bastante; por isso, muitas vezes a argumentação estatística ressent-se um pouco da falta de exemplos, da baixa ocorrência de determinadas expressões. Por outro lado, os dicionários tradicionais, de que dispomos para consulta manual apenas, oferecem exemplos ora literários ora artificiais, tornando difíceis as análises sincrônicas dos sentidos dicionarizados.

Espero que as contribuições desta tese se mostrem relevantes na área, não só como fontes de informação sobre a classe dos substantivos-suporte, mas também como exemplo do caminho percorrido no desenvolvimento de um trabalho essencialmente interdisciplinar.

**Referências Bibliográficas**

- AIRES, R., E ALUISIO, S. Criação de um corpus com 1.000.000 de palavras etiquetado morfossintaticamente. Relatório Técnico NILC-TR-01-8, NILC, Campinas, SP. 2001. 4.2.2, 4.2.6
- ALLERTON, D. J. Levels of co-occurrence restriction. *Lingua* 63:17–40. 1984. 4.3.1
- BAEZA-YATES, R., E NETO, B. R. *Modern Information Retrieval*. Addison Wesley. 1999. 5.3.1
- BAPTISTA, J. 2000. *Sintaxe dos predicados nominais construídos com o verbo-suporte SER DE*. Tese de doutorado, Universidade do Algarve, Faro. 2000. 4.3.1
- BASILIO, M., E GAMARSKI, L. Adjetivos denominais no português falado. In CASTILHO, A., ed., *Gramática do Português Falado*, volume V. Brasil: Editora Unicamp. 1995. 3.4.2, 4.2.3
- BASILIO, M.; DIAS, M. C.; E MARTINS, H. Expressões dar+sn: um estudo de representação lexical. In *Anais do Encontro da ASSEL*. 1994. 4.3.1
- BASILIO, M.; OLIVEIRA, C.; E GARRÃO, M. A não-delimitação da unidades lexicais. In HENRIQUES., ed., *Linguagem Conhecimento e Aplicação: estudos de língua e lingüística*. Editora Europa. 2003. 3.4.1, 3.4.2
- BASILIO, M. *Estruturas Lexicais do Português: uma abordagem gerativa*. Coleção Perspectivas Lingüísticas. Editora Vozes. 1980. 3.5.2
- BASILIO, M. Flutuação categorial de base adjetiva no português falado. In CASTILHO, A., ed., *Gramática do Português Falado*, volume V. Brasil: Editora Unicamp. 1995a. 3.5.1, 4.2.5
- BASILIO, M. O fator semântico na flutuação substantivo/adjetivo em português. In HEYE, J., ed., *Flores verbais*. Rio de Janeiro: Editora 34. 177–192. 1995b. 3.5.1
- BASILIO, M. Padrões de configuração estrutural de unidades lexicais. In *Para sempre em mim: Homenagem a Ângela Vaz Leão*. Belo Horizonte: PUC-Minas. 205–212. 1999a. 4.3.1
- BASILIO, M. Questões clássicas e recentes na delimitação de unidades lexicais. *Palavra* 5. 1999b. 4.3.1
- BASILIO, M. *Teoria Lexical*. Editora Ática, Brasil. 1999c. 3.2, 3.5.2, 3.5.2



- BASILIO, M. 2005. Para além das fronteiras morfológicas: a formação de unidades lexicais complexas no português do Brasil. Conferência no Centro de Linguística Geral e Aplicada, Universidade de Coimbra. 2005. 1.2, 2.2
- BASTOS, L. C. 1980. Interpretação de adjetivos denominais. Dissertação de mestrado, Pontifícia Universidade Católica do Rio de Janeiro. 1980. 3.4.2
- BEARD, R. Derivation. In SPENCER, A., E ZWICKY, A. M., eds., *The Handbook of Morphology*. Blackwell Publishers. 1998. 3.5.2
- BECHARA, E. *Moderna Gramática Portuguesa*. São Paulo: Cia. Editora Nacional. 1980. 3.3.1, 3.4.1
- BIBER, D.; CONRAD, S.; E REPPEN, R. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge, UK: Cambridge University Press. 1998. 1
- BIDERMAN, M. T. *Teoria Lingüística (lingüística quantitativa e computacional)*. Rio de Janeiro: LTC. 1978. 3.2, 3.3.1, 3.5.2
- BLOOMFIELD, L. A set of postulates for the science of language. *Language* 2:153–164. 1926. 2.2, 3.2
- BLOOMFIELD, L. *Language*. University of Chicago Press. 1933. 2.2
- BORBA, F. *Uma Gramática de Valências para o Português*. Rio de Janeiro: Editora Ática. 1996. 3.4.4
- BUTT, M. The light verb jungle. *Harvard Working Papers in Linguistics* 9. Papers from the Harvard/Dudley House Light Verb Workshop. 2003. 4.3.1
- BYBEE, J. Morphology as lexical organization. In HAMMOND, M., E NOONAN, M., eds., *Theoretical morphology*. Academic Press. 119–141. 1988. 2.2
- CÂMARA JR, J. M. *Estrutura da Língua Portuguesa*. Editora Vozes. 2000. 3.2, 3.2, 3.4.1
- CHOMSKY, N. *Aspects of the Theory of Syntax*. The MIT Press. 1965. 2.2
- CHOMSKY, N. Remarks on nominalizations. In JACOBS, R. A., E ROSENBAUM, P. S., eds., *Readings in English Transformational Grammar*. Ginn and Company. 1970. 2.2, 3.5.2
- CRUSE, D. A. *Lexical Semantics*. Cambridge, UK: Cambridge University Press. 1986. 2.3.1, 5.2
- CRUSE, D. A. Polysemy and related phenomena from a cognitive linguistic viewpoint. In SAINT-DIZIER, P., E VIEGAS, E., eds., *Computational Lexical Semantics*. EUA: Cambridge University Press. 33–49. 1995. 2.3.1

- CRUSE, D. A. Aspects of the micro-structure of word meanings. In RAVIN, Y., E LEACOCK, C., eds., *Polysemy: Theoretical and Computational Approaches*. Oxford, UK: Oxford University Press. 30–51. 2000. 2.3.3, 2.3.4
- CRUSE, D. A. *Meaning in Language: An Introduction to Semantics and Pragmatics*. UK: Oxford University Press. 2004. 2.3.3, 2.3.4, 5.2
- CUNHA, C., E CINTRA, L. *Nova Gramática do Português Contemporâneo*. Rio de Janeiro: Nova Fronteira. 1985. 3.3.1, 3.4.1, 3.5.2
- CUNHA, C. *Gramática do Português Contemporâneo*. Belo Horizonte: Editora Bernardo Álvares. 1972. 3.3.1
- DI SCIULLO, A. M., E WILLIAMS, E. *On the Definition of Word*. The MIT Press. 1987. 2.2
- DUBOIS, J.; GIACOMO, M.; GUESPIN, L.; MARCELLESI, C.; MARCELLESI, J.-B.; E MEVEL, J.-P. *Dicionário de Lingüística*. Cultrix, 8ª edição. 2001. 3.2
- DUNBAR, G. Towards a cognitive analysis of polysemy, ambiguity and vagueness. *Cognitive Linguistics* 12(1):1–14. 2001. 2.3.3, 2.3.4
- EMONDS, J. Parts of speech in generative grammar. *Linguistic Analysis* 17(1–2):3–42. 1987. 3.2
- FELLBAUM, C., ed. *WordNet An Electronic Lexical Database*. Cambridge, MA: MIT Press. 1998. 2.2.1
- FILLMORE, C.; WOOTERS, C.; E BAKER, C. F. Building a large lexical databank which provides deep semantics. In *Proceedings of the Pacific Asian Conference on Language, Information and Computation*. 2001. 2.2.1, 6.2
- FILLMORE, C. Frame semantics and the nature of language. *Annals of the New York Academy of Sciences: Conference on the Origin and Development of language and Speech* 280:20–32. 1976. 1, 2.2.1
- FILLMORE, C. Innocence: a second idealization for linguistics. *Berkeley Linguistics Society*, 5 63–76. 1979. 5.2
- FIRTH, J. R. *Selected Papers of J. R. Firth 1952-59*. London: Longman. 1968. 1.2
- GAMARSKI, L. Efeitos da morfologia sobre a estrutura argumental: adjetivos deverbais em –nte. In CASTILHO, A., ed., *Gramática do Português Falado*, volume V. Brasil: Editora Unicamp. 1995. 3.4.4
- GARRÃO, M., E DIAS, M. C. P. Um estudo de expressões cristalizadas do tipo V+SN e sua inclusão em um tradutor automático bilíngüe (português/ inglês). *Cadernos de Tradução* 2(8):165–182. 2001. 4.3.1

- GARRÃO, M.; OLIVEIRA, C.; DE FREITAS, M. C.; E DIAS, M. C. Corpus-based compositionality. In VIEIRA, R.; QUARESMA, P.; NUNES, M. D. G. V.; MAMEDE, N.; OLIVEIRA, C.; E DIAS, M. C., eds., *Computational Processing of the Portuguese Language, PROPOR 2006*, volume 3960 of *Lecture Notes in Computer Science*. Rio de Janeiro: Springer. 2006. 5.3, 5.3.3
- GARRÃO, M. 2006. *O Córpus não mente jamais: sobre a identificação e uso de combinações multivocabulares do tipo verbo mais sintagma nominal*. Tese de doutorado, Pontifícia Universidade Católica do Rio de Janeiro, Brasil. 2006. 4.3.1, 5.3
- GAZDAR, G.; KLEIN, E.; PULLUM, G.; E SAG, I. *Generalized Phrase Structure Grammar*. Cambridge, EUA: Harvard University Press. 1985. 2.2
- GEERAERTS, D. Vagueness's puzzles, polysemy's vagaries. *Cognitive Linguistics* 4(3):223–272. 1993. 2.3, 2.3.1, 2.3.3, 2.3.4, 2.3.4
- GODDARD, C. Polysemy: A problem of definition. In RAVIN, Y., E LEACOCK, C., eds., *Polysemy: Theoretical and Computational Approaches*. Oxford, UK: Oxford University Press. 129–151. 2000. 2.3.3
- GROSS, M. Lexicon-grammar and the syntactic analysis of French. In *Proceedings of the COLING*, 275–282. 1984. 4.3.1
- GUNZBURGER, M. L. G. 1979. *Previsibilidade Semântica em Nominais Correspondentes a Verbos Intransitivos*. Tese de doutorado, Pontifícia Universidade Católica do Rio de Janeiro, Brasil. 1979. 3.5.2
- HALLE, M. Prolegomena to a theory of word formation. *Linguistic Inquiry* 4(1):3–16. 1973. 2.2
- HALLIDAY, M. A. K., E HASAN, R. *Cohesion in English*. Longman. 1976. 3.1, 3.3.2, 3.3.2, 3.3.3
- HALLIDAY, M. A. K. *Spoken and Written Language*. Oxford: Oxford University Press. 1985. 1.1, 1.1
- HALLIDAY, M. A. K. *An Introduction to Functional Grammar*. London: Edward Arnold, 2ª edição. 1994. 1.2
- HOUAISS, A. *Dicionário Eletrônico Houaiss da Língua Portuguesa*. Editora Objetiva. 2001. 1, 4.2.3, 4.2.4, 4.2.4, 4.2.5, 4.3.1, 2
- HUDDLESTON, R. *Introduction to the Grammar of English*. Cambridge University Press, 8ª edição. 2000. 3.3.1, 3.4.1, 3.4.4
- HUDSON, R. A. The linguistic foundations for lexical research and dictionary design. *International Journal of Lexicography* 1(4):287–312. 1988. 2.2.1

- JACKENDOFF, R. Morphological and semantic regularities in the lexicon. *Language* 51(3):639–671. 1975. 2.2, 2.2
- JACKENDOFF, R. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford University Press. 2002. 2.2, 2.2, 2.3.1
- JANSSEN, T., E REDEKER, G., eds. *Cognitive Linguistics: Foundations, Scope, and Methodology*. Berlim: Mouton de Gruyter. 1999. 2.3.1
- JESPERSEN, O. *A Modern English Grammar on Historical Principles*. London: George Allen & Unwin. 1940. 1.2, 4.3.1
- JOSEPH, J. E.; LOVE, N.; E TAYLOR, T. J. Firth on language and context. In *Landmarks in Linguistic Thought II*. Routledge. 57–71. 2001. 1.2
- JURAFSKY, D., E MARTIN, J. *Speech and Language Processing*. EUA: Prentice Hall. 2000. 3.2
- KATZ, J., E FODOR, J. The structure of a semantic theory. *Language* 170–210. 1963. 5.2
- KATZ, J. *Semantic Theory*. Nova Iorque: Harper & Row. 1972. 2.3.1
- KILGARRIFF, A. 1992. *Polysemy*. Tese de doutorado, University of Sussex, UK. 1992. 2.3.3, 4.2.5
- KILGARRIFF, A. “I don’t believe in word senses”. *Computers and the Humanities* 31:91–113. 1997. 2.3.4, 5.3
- LAUDANNA, A., E VOGHERA, M. Nouns and verbs as grammatical classes in the lexicon. *Rivista di Linguistica* 14(1):9–26. 2002. 3.2
- LEVI, J. *The Syntax and Semantics of Complex Nominals*. Academic Press. 1978. 3.4.2
- LOBATO, L. Adjetivo: Tipologia e interpretação semântica. *Boletim da ABRALIN* 14. 1993. 3.4.2
- LYONS, J. *Semantics*. Cambridge University Press. 1977. 2.3.2, 3.2, 3.2, 3.4.1
- LYONS, J. *Linguistic Semantics*. Cambridge University Press. 1995. 2.3
- MANNING, C., E SCHÜTZE, H. *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press. 1999. 5.3.1
- MARCUS, M. P.; SANTORINI, B.; E MARCINKIEWICZ, M. A. Building a large annotated corpus of English: the Penn Treebank. *Computational Linguistics* 19(2):313–330. 1993. 3.2
- MARCUSCHI, L. A. O léxico: Lista, rede ou cognição social. In NIGRI, L.; FOLTRAN, M. J.; E OLIVEIRA, R. P., eds., *Sentido e Significação: em torno da obra de Rodolfo Ilari*. São Paulo: Editora Contexto. 263–284. 2004. 4.2.1

- MARQUES, M. H. D. Léxico de alta frequência na língua portuguesa. In HEYE, J., ed., *Flores verbais*. Rio de Janeiro: Editora 34. 247–282. 1995. 4.2.2
- MEJLACHOWICZ, S. 2003. Uma análise semântico-aspectual dos verbos de ligação. Dissertação de mestrado, Pontifícia Universidade Católica do Rio de Janeiro. 2003. 4.3.1
- MIHATSCH, W. Nouns are things - new evidence for a grammatical metaphor. In *8th International Linguistics Conference*. 2003. 4.2.4
- MOURA, H. *Significação e contexto: uma introdução a questões semânticas e de pragmática*. Florianópolis: Insular. 1999. 1
- MOURA, H. Apresentação: Polissemia e indeterminação semântica. *DELTA* 18(Número Especial):IX–XVI. 2002. 2.3.3
- MURESAN, S.; TZOUKERMANN, E.; E KLAVANS, J. Combining linguistic and machine learning techniques for email summarization. In *Proceedings of CoNLL-2001: The Fifth Workshop on Computational Language Learning*. 2001. 1.1, 4.2.2
- NESFIELD, J. C. *Aids to the Study and Composition of English*. Macmillan and Co. 1907. 3.4.1
- NEVES, M. H. M. Estudo das construções com verbo-suporte em português. In CA KOCH, I. V., ed., *Gramática do Português Falado*, volume VI: Desenvolvimentos. Brasil: Editora Unicamp. 1996. 4.3.1
- NEVES, M. H. M. A delimitação das unidades lexicais: o caso nas construções com verbo-suporte. *Palavra* 5. 1999. 4.3.1
- NEVES, M. H. M. *Gramática de Usos do Português*. São Paulo, Brasil: Editora Unesp. 2000. 4.3.1
- NEVES, M. H. M. *A Gramática Funcional*. São Paulo, Brasil: Martins Fontes. 2004. 1.2
- NEWMAYER, F. J. *Language Form and Language Function*. Cambridge, MA: The MIT Press. 1998. 1.2, 2.2
- NUNES, J. J. *Compêndio da Gramática Histórica Portuguesa (Fonética e Morfologia)*. Lisboa: Livraria Clássica Editora, 6ª edição. 1960. 3.4.1
- OOI, V. B. Y. *Computer Corpus Lexicography*. Edinburgo: Edinburgh University Press. 1998. 1.3, 5.1
- PANAGIOTIDIS, P. Empty nouns. *Natural Language and Linguistic Theory* 21(2):381–432. 2003. 1.1

- PERINI, M. Para a análise dos adjetivos denominais. In *Anais do II ENL*. Rio de Janeiro: PUC-Rio. 1978. 3.4.2
- POLLARD, C., E SAG, I. *Head-driven Phrase Structure Grammar*. Chicago, EUA: Chicago University Press. 1994. 2.2
- POUTSMA, H. *A Grammar of the Late Modern English: part II*. Groningen: P. Noordhof. 1926. 1.2, 4.3.1
- PUSTEJOVSKY, J., E BOGURAEV, B. Introduction: Lexical semantics in context. In PUSTEJOVSKY, J., E BOGURAEV, B., eds., *Lexical Semantics: the Problem of Polysemy*. Oxford University Press. 1996. 2.3.1
- PUSTEJOVSKY, J. Computational lexicons. In WILSON, R. A., E KEIL, F. C., eds., *The MIT Encyclopedia of the Cognitive Sciences*. MIT. 1999. 2.2.1
- QUIRK, R.; GREENBAUM, S.; LEECH, G.; E SVARTVIK, J. *A Grammar of Contemporary English*. Longman Group Limited, 7ª edição. 1978. 3.4.1
- RANCHHOD, E. On the support verbs ser and estar in portuguese. *Linguisticae Investigationes* VII(2):317–353. 1983. 4.3.1
- RANCHHOD, E. *Sintaxe dos Predicados Nominais com Estar*. Lisboa: Instituto Nacional de Investigação Científica, Centro de Lingüística da Universidade de Lisboa. 1990. 4.3.1
- RASKIN, V., E NIRENBURG, S. Lexical semantics of adjectives: A microtheory of adjectival meaning. Technical Report MCCS-95-288, Computing Research Laboratory, New Mexico State University. 1995. 3.4.2
- RAVIN, Y., E LEACOCK, C. Polysemy: An overview. In RAVIN, Y., E LEACOCK, C., eds., *Polysemy: Theoretical and Computational Approaches*. Oxford, UK: Oxford University Press. 1–29. 2000. 2.3.2, 2.3.3, 2.3.4
- ROCHA LIMA, C. H. *Gramática Normativa da Língua Portuguesa*. José Olympio Editora, 36ª edição. 1998. 3.3.1, 3.4.1
- ROSA, M. C. *Introdução à Morfologia*. São Paulo: Contexto. 2000. 3.2, 3.2
- SAID ALI, M. *Gramática Histórica da Língua Portuguesa*. Editora UNB. 2001. 3.3.1, 3.4.1
- SALTON, G., E BUCKLEY, C. Term-weighting approaches in automatic text retrieval. *Information Processing and Management* 24(5):513–523. 1988. 5.3.1
- SANDMANN, A. J. *Formação de Palavras no Português Brasileiro Contemporâneo*. Editora UFPR, 2ª edição. 1989. 3.5.2

- SCHER, A. P. Quais são as propriedades lexicais de uma construção com verbo leve? In MÜLLER, A. L.; NEGRÃO, E. V.; E FOLTRAN, M. J., eds., *Semântica Formal*. São Paulo: Contexto. 205–219. 2003. 4.3.1
- SCHMID, H.-J. *English Abstract Nouns As Conceptual Shells: From Corpus to Cognition*. Mouton de Gruyter. 2000. 1.1, 3.1, 3.3.1, 3.3.3, 4.2.2
- SCHÜTZE, H., E PEDERSON, J. Information retrieval based on word senses. In *Proceedings of the ACM Special Interest Group on Information Retrieval*. 1995. 5.3.1
- SCHÜTZE, H. Automatic word sense discrimination. *Computational Linguistics* 24(1):97–124. 1998. 2.3.3
- SEARLE, J. Literal meaning. In *Expression and Meaning: Studies in the Theory of Speech Acts*. Cambridge University Press. 117–136. 1979. 5.2
- SINCLAIR, J. Trust the text. In COULTHARD, M., ed., *Advances in Written Text Analysis*. Londres: Routledge. 12–25. 1994. 4.3.1
- SINCLAIR, J., ed. *COBUILD English Dictionary for Advanced Learners*. Harper-Collins Publishers, 3a.ª edição. 2001. 2.3, 4.2.5, 6.3
- STEEDMAN, M. Categorical grammar. *Lingua* 90:221–258. 1993. 2.2
- STUBBS, M. *Words and Phrases: Corpus Studies of Lexical Semantics*. Blackwell Publishing. 2002. 5.4
- TUGGY, D. Ambiguity, polysemy and vagueness. *Cognitive Linguistics* 4(3):273–290. 1993. 2.3.1
- VILELA, M., E SILVA, F. The position of the adjective in portuguese: centre and periphery of the adjective class. In *Conference Proceedings: Language, Culture and Cognition, An International Conference on Cognitive Linguistics*. Catholic University of Portugal, Braga. 2003. 3.4.2
- VIOTTI, E. A composicionalidade nas sentenças com o verbo ter. In MÜLLER, A. L.; NEGRÃO, E. V.; E FOLTRAN, M. J., eds., *Semântica Formal*. São Paulo: Contexto. 221–241. 2003. 4.3.1
- WEINREICH, U. Webster's third: a critique of its semantics. *Journal of American Linguistics* 30:405–409. 1964. 2.3.1
- ZHAI, C. Exploiting context to identify lexical atoms - a statistical view of linguistic context. In *Proceedings of the International and Interdisciplinary Conference on Modelling and Using Context*, 119–129. 1997. 5.3.3

# Livros Grátis

( <http://www.livrosgratis.com.br> )

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)



[Baixar livros de Literatura](#)  
[Baixar livros de Literatura de Cordel](#)  
[Baixar livros de Literatura Infantil](#)  
[Baixar livros de Matemática](#)  
[Baixar livros de Medicina](#)  
[Baixar livros de Medicina Veterinária](#)  
[Baixar livros de Meio Ambiente](#)  
[Baixar livros de Meteorologia](#)  
[Baixar Monografias e TCC](#)  
[Baixar livros Multidisciplinar](#)  
[Baixar livros de Música](#)  
[Baixar livros de Psicologia](#)  
[Baixar livros de Química](#)  
[Baixar livros de Saúde Coletiva](#)  
[Baixar livros de Serviço Social](#)  
[Baixar livros de Sociologia](#)  
[Baixar livros de Teologia](#)  
[Baixar livros de Trabalho](#)  
[Baixar livros de Turismo](#)