

UNIVERSIDADE DO VALE DO RIO DOS SINOS  
UNIDADE ACADÊMICA DE PESQUISA E PÓS-GRADUAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM LINGUÍSTICA APLICADA  
NÍVEL MESTRADO

Lílian Figueiró Teixeira

A SEMÂNTICA DOS COMPOSTOS NOMINAIS – UM ESTUDO DE  
*CORPUS* PARALELO INGLÊS/PORTUGUÊS

São Leopoldo

2009

Lílian Figueiró Teixeira

A SEMÂNTICA DOS COMPOSTOS NOMINAIS – UM ESTUDO DE  
*CORPUS* PARALELO INGLÊS/PORTUGUÊS

Dissertação apresentada como requisito parcial para a obtenção do título de Mestre pelo Programa de Pós-Graduação em Linguística Aplicada da Universidade do Vale do Rio dos Sinos.

Orientadora: Profa. Dra. Rove Luiza de O. Chishman

São Leopoldo

2009

Ficha catalográfica

T266s Teixeira, Lílian Figueiró  
**A semântica dos compostos nominais : um estudo de corpus paralelo inglês/português / por Lílian Figueiró Teixeira. – 2009.**  
209 f. : il. ; 30cm.

Dissertação (mestrado) — Universidade do Vale do Rio dos Sinos, Programa de Pós-Graduação em Linguística Aplicada, 2009.

“Orientação: Prof<sup>a</sup>. Dr<sup>a</sup>. Rove Luiza de O. Chishman, Ciências humanas”.

1. Compostos nominais. 2. Corpus - Linguística. 3. Corpus paralelo. 4. Semântica lexical I. Título.

CDU 801.28

Catálogo na Fonte:  
Bibliotecária Vanessa Borges Nunes - CRB 10/1556

Lílian Figueiró Teixeira

A semântica dos compostos nominais – um estudo de *corpus* paralelo inglês/português

Dissertação de Mestrado apresentada como requisito parcial para a obtenção do título de Mestre pelo Programa de Pós-Graduação em Linguística Aplicada da Universidade do Vale do Rio dos Sinos.

Aprovado em 10 de março de 2009.

BANCA EXAMINADORA

---

Profa. Dra. Aline Villavicencio - UFRGS

---

Profa. Dra. Maria da Graça Krieger – UNISINOS

---

Orientadora – Profa. Dra. Rove Luiza de Oliveira Chishman – UNISINOS

Ao meu pai querido.

## **Agradecimentos**

Mesmo que o hífen esteja fora de moda, vou fazer alguns rápidos agradecimentos ainda utilizando este recurso. Espero poder agradecer pessoalmente cada uma das pessoas mencionadas aqui: a-professora-Rove-que-sempre-diz-as-coisas-que-precisam-ser-ditas, a-professora-Ana-que-é-a-coordenadora-mais-elegante-da-Unisinos, a-mãe-que-tem-muita-paciência-comigo, o-Ju-que-continua-sendo-o-melhor-namorado-do-mundo, a-Juliana-amiga-e-anjo-particular, a-tia-Renata-confiando-sempre-em-mim, o-tio-Miranda-que-me-ensinou-a-ser-mais-independente, o-colega-Mário-que-me-livrou-de-alguns-apuros-achando-soluções-imediatas-para-problemas-persistentes, o-colega-Lucas-que-criou-os-programas-certos-para-que-este-trabalho-se-tornasse-possível, o-Diarmuid-respondendo-sempre-aos-e-mails-de-última-hora.

A todos os colegas e professores: muito obrigada.

*Because the world is round it turns me on  
Because the world is round...aaaaahhhhhh*

*Because the wind is high it blows my mind  
Because the wind is high.....aaaaaaaahhhh*

*Love is old, love is new  
Love is all, love is you*

*Because the sky is blue, it makes me cry  
Because the sky is blue.....aaaaaaaahhhh*

*Aaaaahhhhhhhhhh....*

The Beatles

## RESUMO

Os compostos nominais são construções produtivas em diversos idiomas, ou seja, novas combinações são facilmente criadas em contextos de uso da língua. No entanto, esse fenômeno linguístico é idiossincrático, fato que torna o seu estudo um desafio para a linguística e para as investigações que se ocupam do Processamento da Linguagem Natural. Neste trabalho, é feita uma investigação sobre a forma como os elementos constituintes dos compostos nominais em inglês formados por dois substantivos (compostos NN) se relacionam semanticamente e quais as características dos seus correspondentes de tradução em língua portuguesa encontrados em dez edições da revista *National Geographic*. O objetivo desta investigação é identificar as relações mais frequentes no *corpus* a fim de que se possa propor uma tipologia que expresse a composicionalidade semântica dessas construções. Para alcançar esse fim, o trabalho está dividido em três etapas. A primeira etapa consiste em apresentar os pressupostos teóricos adotados no trabalho. Primeiramente, são discutidos alguns conceitos sobre a semântica dos compostos nominais, como produtividade, transparência semântica, nuclearidade, lexicalização e nominalização. Em seguida, são apresentadas duas teorias que são utilizadas para a elaboração de uma tipologia da composicionalidade semântica dos compostos NN: a Teoria do Léxico Gerativo (PUSTEJOVSKY, 1995) e a teoria de *templates* ou *frames* (RYDER, 1994, FILLMORE, 2006). Também é apresentado um panorama das possibilidades de estudos linguístico-computacionais dos compostos nominais; entre os estudos possíveis a partir dessa perspectiva estão as expressões multivocabulares, a anotação semântica, o aprimoramento de gramáticas e léxicos, a tradução automática e a multilinguagem. Na segunda etapa, são apresentados alguns recursos metodológicos da área da linguística de *corpus* - como coleta e alinhamento de *corpus* paralelo - e descreve-se como esses recursos foram utilizados no processamento do *corpus* deste trabalho. Além disso, são feitas algumas considerações sobre as estruturas utilizadas para traduzir estes compostos para o português. Para a análise, um *corpus* paralelo foi compilado e as ocorrências de compostos NN foram analisadas. Entre as ferramentas utilizadas para que se chegasse a uma lista de compostos pertinentes a este trabalho, destacam-se: itemizador, etiquetador, extrator, alinhador e concordanciador. De um total de 4.693 possíveis compostos, 200 compostos confirmados foram selecionados e analisados. A terceira etapa do trabalho consiste na análise e na discussão dos dados. Entre os resultados gerais, destaca-se que 195 compostos NN apresentam um núcleo evidente, 199 são transparentes semanticamente e produtivos e 199 são endocêntricos, em que o compostos como um todo representa uma subclasse do seu núcleo. As relações semânticas mais frequentes entre os dois elementos dos compostos presentes no *corpus* são: telicidade, agentividade, meronímia, localização, posse e hiponímia. A consulta à base de dados lexical do *FrameNet* contribuiu para a confirmação das relações semânticas, reforçando a adequação da tipologia utilizada. Quanto aos correspondentes de tradução no *corpus*, não foi possível relacionar uma única possibilidade de tradução para cada relação semântica. Porém, percebeu-se que as diferenças culturais são questões que influenciam diretamente nas escolhas de tradução. Este trabalho sugere uma análise semântica dos compostos nominais em inglês utilizados no contexto de uma revista, apresentando também uma reflexão sobre os seus correspondentes de tradução.

Palavras-chave: compostos nominais, linguística de *corpus*, *corpus* paralelo, *frames* semânticos, semântica lexical.



## ABSTRACT

Noun compounds are productive constructions in many languages. However, they are idiosyncratic, fact that makes the study of this linguistic phenomenon a challenge for the linguistics and for the Natural Language researches. The purpose of this paper is to study the semantics of the noun compounds formed by two nouns (NN compounds). It is also intended to identify the translation equivalents in Portuguese found in ten editions of the National Geographic Magazine. The final product is a proposal of typology which expresses the compositionality of the NN compounds according to the data found in the corpus. This paper has three distinctive parts, where the following subjects are introduced: the theoretical bases for this paper; the methodological resources from Corpus Linguistics that were adopted; the analysis and discussion about the data. Concepts about the semantics of nominal compounds as productivity, semantic transparency, headness, lexicalization and nominalization are commented. Two theories were used for the elaboration of a typology that concerns about the semantic compositionality of the NN compounds: the Generative Lexicon Theory (PUSTEJOVSKY, 1995) and the templates or frames theory (RYDER, 1994, FILLMORE, 2006). Some linguistic computational studies about the noun compounds are also analysed and their main subjects include: multiword expressions, semantic annotation, improvement of grammars and lexicon, machine translation and multilingualism. As data for the analysis, a corpus was collected and aligned. Besides, some considerations about the structures used for translating these compounds to Portuguese were made. The parallel corpus was analysed through the use of a concordancer (WordSmith Tools, SCOTT, 2008). The other computational resources that were very useful for the purposes of this study were: itemizer, POS tagger, extractor and aligner. Starting from a list of 4,693 candidates to compounds, 200 were confirmed and analysed. Some of the results of the study are that 195 NN compounds have a head, 199 are semantically transparent and productive and 199 are endocentric. The most frequent relations between the two elements from the compounds found in the corpus are: telicity, agentivity, meronymy, localization, possession and hyponymy. A great contribution for confirming the semantic relations was from the FrameNet Database, which reinforced the adaptation of the typology applied. In relation to the translation correspondents in the corpus, it was not possible to match to only one translation for each semantic relation. This paper suggests a semantic analysis of the noun compounds in English, which are found in a magazine context, and also provides a reflexion about their translation correspondents, which are most of the time influenced by cultural differences.

Keywords: noun compounds, corpus linguistics, parallel corpus, semantic frames, lexical semantics.

## LISTA DE FIGURAS

<b>FIGURA 1</b> – <i>Frame</i> Localização_por_evento da base de dados do <i>FrameNet</i> .....	53
<b>FIGURA 2</b> – <i>Corpus</i> itemizado.....	88
<b>FIGURA 3</b> – <i>Corpus</i> alinhado.....	91
<b>FIGURA 4</b> – Exemplo de árvore de decisão (SCHMID, 1994, p. 46).....	94
<b>FIGURA 5</b> – <i>Corpus</i> etiquetado.....	94
<b>FIGURA 6</b> – Resultados do extrator.....	96
<b>FIGURA 7</b> – Parte da concordância de <i>people</i> .....	101

## LISTA DE TABELAS

<b>TABELA 1</b> – Principais conceitos apresentados na subseção 2.1.1.....	26
<b>TABELA 2</b> – Resumo das relações selecionadas a partir de trabalhos estruturalistas.....	33
<b>TABELA 3</b> – Resumo dos principais conceitos apresentados por Ryder (1994).....	42
<b>TABELA 4</b> – Esquemas sugeridos por Ryder (1994).....	49
<b>TABELA 5</b> – Relações semânticas de Ó Séaghdha (2007).....	64
<b>TABELA 6</b> – Principais etiquetas de Ryder (1994).....	78
<b>TABELA 7</b> – Dados dos <i>corpora</i> .....	87
<b>TABELA 8</b> – Alinhamentos do <i>corpus</i> paralelo – total: 9.514.....	90
<b>TABELA 9</b> – Frequências de <i>core words</i> de compostos no <i>corpus National Geographic</i> .....	97
<b>TABELA 10</b> – Candidatos a compostos.....	99
<b>TABELA 11</b> – Filtros para a seleção dos compostos.....	101
<b>TABELA 12</b> – Compostos do <i>corpus National Geographic</i> selecionados para a análise.....	104
<b>TABELA 13</b> – Resultados gerais da análise inicial dos 200 compostos selecionados.....	106
<b>TABELA 14</b> – Sufixos deverbais.....	110
<b>TABELA 15</b> – Chave de análise dos compostos.....	114
<b>TABELA 16</b> – Relações encontradas no <i>corpus</i> .....	116
<b>TABELA 17</b> – Dados gerais das frequências de padrões de tradução.....	126
<b>TABELA 18</b> – Ocorrências da relação SERVE PARA.....	128
<b>TABELA 19</b> – Ocorrências da relação VEM DE, É FEITO A PARTIR DE.....	130
<b>TABELA 20</b> – Ocorrências da relação É FEITO DE.....	130
<b>TABELA 21</b> – Ocorrências da relação POSSUI.....	131
<b>TABELA 22</b> – Ocorrências da relação CONTÉM.....	132
<b>TABELA 23</b> – Ocorrências da relação É LOCALIZADO EM.....	133
<b>TABELA 24</b> – Ocorrências da relação OCORRE EM.....	134
<b>TABELA 25</b> – Ocorrências da relação TEM.....	134

## SUMÁRIO

<b>1 INTRODUÇÃO.....</b>	<b>14</b>
<b>2 REVISÃO DE LITERATURA.....</b>	<b>17</b>
2.1 PERSPECTIVA LINGUÍSTICA.....	17
<b>2.1.1 Compostos nominais: conceitos pertinentes e principais características..</b>	<b>18</b>
2.1.1.1 Produtividade.....	20
2.1.1.2 Transparência semântica.....	22
2.1.1.3 Nuclearidade.....	23
2.1.1.4 Compostos endocêntricos e exocêntricos.....	24
2.1.1.5 Lexicalização.....	24
2.1.1.6 Nominalização.....	25
<b>2.1.2 Estudos clássicos sobre os compostos.....</b>	<b>26</b>
<b>2.1.3 Linguística Cognitiva: Ryder (1994) e Fillmore (2006).....</b>	<b>36</b>
<b>2.1.4 Teoria do Léxico Gerativo.....</b>	<b>54</b>
2.2 PERSPECTIVA LINGUÍSTICO-COMPUTACIONAL.....	59
<b>2.2.1 Expressões multivocabulares.....</b>	<b>60</b>
<b>2.2.2 Anotação semântica.....</b>	<b>62</b>
<b>2.2.3 Aprimoramento de gramáticas e de léxicos.....</b>	<b>66</b>
<b>2.2.4 Tradução automática.....</b>	<b>70</b>
<b>2.2.5 Multilinguagem.....</b>	<b>73</b>
<b>3 METODOLOGIA.....</b>	<b>80</b>
3.1 LINGUÍSTICA DE <i>CORPUS</i> .....	80
3.2 ESCOLHA DO <i>CORPUS</i> .....	85
3.3 COLETA E ALINHAMENTO DO <i>CORPUS</i> .....	86
3.4 EXTRAÇÃO E BUSCA NO <i>CORPUS</i> PARALELO.....	92
3.5 RESULTADOS DA EXTRAÇÃO.....	96
<b>4 ANÁLISE E DISCUSSÃO DOS DADOS.....</b>	<b>103</b>
4.1 OS COMPOSTOS NOMINAIS EM LÍNGUA INGLESA.....	103
<b>4.1.1 Nuclearidade.....</b>	<b>106</b>
<b>4.1.2 Compostos endocêntricos.....</b>	<b>108</b>
<b>4.1.3 Transparência semântica e produtividade.....</b>	<b>109</b>
<b>4.1.4 Nominalização.....</b>	<b>110</b>

4.2 PADRÕES SEMÂNTICOS.....	112
<b>4.2.1 Padrões de pessoas</b> .....	117
<b>4.2.2 Padrões de animais</b> .....	117
<b>4.2.3 Padrões de plantas</b> .....	118
<b>4.2.4 Padrões de artefatos</b> .....	119
<b>4.2.5 Padrões de substâncias</b> .....	120
<b>4.2.6 Padrões de eventos</b> .....	120
4.3 TRADUÇÃO DOS COMPOSTOS PARA A LÍNGUA PORTUGUESA.....	124
<b>4.3.1 Sumário da análise do <i>corpus</i> paralelo</b> .....	135
<b>5 CONSIDERAÇÕES FINAIS</b> .....	137
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	140
ANEXO A – Traduções livres feitas pela autora deste trabalho.....	144
ANEXO B – Grupos de compostos do <i>corpus National Geographic</i> .....	153
ANEXO C – Análise rasa dos 200 compostos.....	181
ANEXO D – Relações semânticas dos compostos analisados.....	186
ANEXO E – Equivalentes de tradução do <i>corpus National Geographic</i> .....	194
ANEXO F – <i>Corpus</i> Paralelo – CD-ROM	

## 1 INTRODUÇÃO

O estudo dos compostos nominais é um tema de interesse para diversas áreas, seja no âmbito da ciência linguística, incluindo morfologia, sintaxe, lexicologia e semântica, como no âmbito da ciência da computação, com ênfase nas investigações sobre o Processamento da Linguagem Natural (PLN). Aos linguistas interessa a reflexão sobre a forma como estas construções compostas são constituídas e as infinitas possibilidades de combinação. Aos estudiosos da área do PLN, por sua vez, o grande desafio é fazer que os sistemas reconheçam que tais construções constituem uma unidade.

Tomando como ponto de partida este contexto, escolheu-se como tema desta dissertação os compostos nominais em inglês formados por dois substantivos. Ainda que tal temática possa ser explorada sob diferentes vieses, todos igualmente interessantes e complexos, elegeu-se a semântica como foco deste estudo. Como justificativa para a escolha deste recorte teórico, citam-se os seguintes aspectos: (i) diferentemente de outras línguas, como o português, em que a relação entre dois substantivos é expressa através de alguma outra palavra, como uma preposição (*pano de prato*), em inglês basta antepor um substantivo a outro para formar um composto (*bus stop*), sem utilizar nenhuma outra palavra que possa dar pistas da relação entre estes dois elementos, ou seja, a relação fica implícita; (ii) além disso, as construções em inglês formadas por dois substantivos são extremamente numerosas e compostos novos deste tipo podem ser facilmente criados.

O propósito deste estudo é analisar a semântica dos compostos nominais (doravante compostos NN<sup>1</sup>); em outras palavras, interessa aqui compreender como os elementos dos compostos em língua inglesa se relacionam semanticamente. Para tal, segue-se os princípios da Linguística de *Corpus*, priorizando dados reais da língua, e parte-se para a construção de um *corpus* paralelo. A revista *National Geographic*, em sua versão eletrônica, com suas edições em inglês e português, foi o ponto de partida. Acredita-se que a possibilidade de aproximar, através dos dados de um *corpus* paralelo, os compostos em inglês de suas traduções em português permite que se avance em direção aos estudos bilíngues, haja vista que pode-se cotejar como as relações semânticas expressas em um língua se expressam em outra.

Tendo em vista este objetivo geral e as justificativas apontadas acima, as seguintes questões de pesquisa são formuladas:

---

<sup>1</sup> Essa denominação vem do inglês, em que os compostos formados por dois substantivos são chamados de *Noun-Noun Compounds*; por isso, utiliza-se de forma abreviada compostos NN.

- (i) Quais as relações semânticas mais recorrentes entre os elementos de compostos NN, conforme estudos já realizados?
- (ii) Considerando-se a diversidade de perspectivas teóricas para tratar da semântica dos compostos NN, que tipologia é descritiva e explanatoriamente adequada para dar conta da composicionalidade semântica destas construções?
- (iii) De que forma as ferramentas relacionadas ao estudo de *corpus* podem ser utilizadas para uma análise semântica dos compostos nominais a partir de um *corpus* paralelo formado por textos em língua inglesa e a sua tradução em português?
- (iv) Quais as relações semânticas mais frequentes encontradas nos compostos nominais do *corpus* paralelo analisado?

Para responder estas questões, a dissertação foi organizada em três capítulos, além dos capítulos de Introdução e Considerações Finais.

No capítulo 2, é apresentada uma revisão teórica que parte de alguns conceitos gerais relacionados à composição. É também propósito desse capítulo revisitar alguns estudos clássicos sobre a semântica dos compostos, referências estas que ainda são consideradas em estudos atuais, principalmente na área da informática. Por fim, o capítulo 2 traz as duas teorias que servem de pilares para este estudo: a teoria do Léxico Gerativo de Pustejovsky (1995) e a representação por modelos (*templates*) linguísticos de Ryder (1994). A escolha por estas teorias se justifica pelas seguintes razões: (i) através da teoria do Léxico Gerativo é possível descrever a semântica dos compostos levando em conta a semântica de cada um dos elementos, ou seja, é possível compreender o sentido do composto, identificando como um substantivo completa o sentido do outro, e vice-versa; (ii) os *templates* linguísticos de Ryder não constituem um grupo fechado de relações entre os substantivos, mas sugerem que, a partir de um contexto específico, seja possível relacionar um papel, uma interpretação para cada elemento. Esta relação que Ryder faz entre os compostos e um contexto situacional permitiu a utilização dos dados da base lexical *FrameNet* neste estudo, uma base construída a partir da teoria de *frames* de Fillmore (2006).

No capítulo 3, a metodologia, que é norteada por conceitos da área da Linguística de *Corpus*, é apresentada. Um *corpus* paralelo (formado pelas reportagens da revista *National Geographic*) é construído e desta fonte de dados os compostos a serem analisados são extraídos. Para a formação deste *corpus* e para a busca pelos dados analisados, várias

ferramentas computacionais que auxiliam os estudos linguísticos são apresentadas, como: itemizador, etiquetador, alinhador, extrator e concordanceador.

O capítulo 4, por fim, tomando como base os compostos extraídos do *corpus* de estudo, traz a análise e discussão dos dados.

Como resultado deste estudo, espera-se, em um primeiro momento, ter aprofundado a reflexão sobre a semântica dos compostos NN. Buscou-se, através da análise das traduções, avançar nos estudos bilíngues, à medida em que o foco passa a ser também a comparação entre as duas línguas. Em termos aplicados, espera-se contribuir para áreas específicas da Linguística Computacional, em especial para o desenvolvimento de programas que se valem de anotação semântica ou de aprendizado de máquina. A construção de base de dados multilíngues ou sistemas de memória de tradução também são aplicações que podem se beneficiar de estudos desta natureza.



## 2 REVISÃO DA LITERATURA

O estudo dos compostos nominais tem recebido a atenção de pesquisadores em diversas áreas, seja na morfologia, na terminologia, na sintaxe ou até mesmo em áreas mais relacionadas diretamente ao PLN, como anotação de *corpus*, criação de léxicos computacionais e tradução automática. Neste trabalho, o foco é o estudo da semântica dos compostos NN. Este capítulo oferece uma perspectiva geral, sem trazer maiores detalhes sobre os estudos que não sejam semânticos, já que não é esse o propósito deste trabalho. Esta revisão da literatura apóia-se, também, em estudos que assumem uma perspectiva computacional no tratamento da semântica dos compostos. O compromisso com esses dois vieses justifica a estrutura deste capítulo.

A seção 2.1 traz uma caracterização da semântica dos compostos NN a partir do enfoque linguístico. Chama a atenção, nessa parte, a diversidade de tratamentos, valendo-se de conceitos muitas vezes divergentes entre si. Na seção 2.2, é dada ênfase aos estudos computacionais que contemplam os compostos NN.

O estudo dos compostos justifica-se por se tratar de um fenômeno extremamente produtivo na língua inglesa. Embora facilmente se possa criar uma expressão composta, a sua criação não obedece a regras fixas pré-estabelecidas. Como há certa imprevisibilidade quanto ao uso e à interpretação dos compostos, não há um consenso quanto à melhor forma de se tratar esse fenômeno linguístico. Por serem produtivos, uma das soluções seria identificar regularidades e estabelecer algumas regras quanto ao seu uso. Mas como nem sempre é possível analisar os compostos por meio de regras, o melhor a se fazer seria criar uma lista, um léxico com os compostos da língua. Essa tarefa parece ser apenas parcialmente possível, já que novos compostos podem ser facilmente criados.

### 2.1 Perspectiva linguística

Os compostos nominais podem ser definidos sob diferentes aspectos, sejam eles lexicais, sintáticos, morfológicos ou semânticos. Mesmo que o objetivo deste trabalho seja o estudo da semântica dos compostos, o primeiro desafio é definir o que constitui um composto nominal na língua inglesa. Na subseção 2.1.1, serão apresentados alguns conceitos quanto à composição, tornando possível, assim, a compreensão de como os compostos nominais se diferenciam de outros tipos de compostos. Feitas essas considerações iniciais, na subseção 2.1.2, será proposta uma reflexão sobre as propriedades semânticas dos compostos NN. Para

tanto, serão abordados alguns estudos clássicos presentes na revisão teórica apresentada em Downing (1977) e Ryder (1994).

Dois diferentes estudos sobre a semântica dos compostos são apresentados nas subseções 2.1.3 e 2.1.4, as propostas de Ryder (1994) e Pustejovsky (1995). A partir dessas duas teorias e da constatação de alguns aspectos relevantes de estudos semânticos anteriores, propõe-se uma descrição combinada dos compostos analisados nesta pesquisa, que será apresentada no capítulo de análise dos dados.

### 2.1.1 Compostos nominais: conceitos pertinentes e principais características

Nesta subseção, são apresentados alguns conceitos pertinentes ao estudo dos compostos, sendo estes: produtividade, transparência semântica, nuclearidade, composição exocêntrica e endocêntrica, lexicalização e nominalização. Também são expostas diferentes perspectivas quanto à definição de compostos nominais e algumas de suas principais características, tanto em inglês como em português. Busca-se, ao final desta revisão, identificar quais definições e especificidades melhor se prestam à descrição dos dados empíricos deste estudo.

Há diversas definições para os compostos nominais em língua inglesa e cada autor sugere diferentes critérios para identificar um grupo de palavras como composto. Ryder (1994), que realizou um estudo na área do cognitivismo, sugere inicialmente que se considerem todas as combinações de dois substantivos como compostos. A partir desse grupo de possíveis compostos, aplicam-se certos critérios com o objetivo de identificar os “verdadeiros compostos”. A grande dificuldade dos linguistas é estabelecer esses critérios. Uma das principais preocupações dos trabalhos que se ocupam da definição de composição é diferenciar os compostos de sintagmas nominais. Levi (1978 *apud* RYDER<sup>2</sup>, 1994), que fez uma análise dos compostos nominais sugerindo paráfrases que explicam a relação entre os elementos constituintes a partir da identificação de um predicado que os une, decidiu não fazer essa distinção por não encontrar critérios suficientemente eficientes. Cada pesquisador sugere um teste diferente para identificar um composto, em que a sintaxe ou a semântica de determinada construção é considerada.

Quanto às considerações morfológicas, pesquisadores como Marchand (1969 *apud* RYDER, 1994) e Lees (1970 *apud* RYDER, 1994) sugerem o teste de padrões de pronúncia,

---

<sup>2</sup> Foi utilizado o trabalho de Ryder (1994) porque não foi possível encontrar todos os originais citados por ela, principalmente por terem sido publicados há certo tempo.

em que o primeiro elemento de um composto formado por dois substantivos ou por um adjetivo seguido de um substantivo é a palavra que recebe a pronúncia mais forte (*strong stress*). Para os autores, *black bird*<sup>3</sup> e *blackbird* têm dois significados diferentes. Se a pronúncia for mais forte em *bird*, tem-se um sintagma nominal que pode ser traduzido por “pássaro preto”. Já quando a pronúncia é mais forte em *black*, pode-se interpretar como um composto que se refere a um tipo de pássaro, o melro. No entanto, há muita variação quanto à pronúncia, havendo inclusive variações dialetais. Para Bauer (1978 *apud* RYDER, 1994), o mesmo falante pode pronunciar o mesmo composto de forma diferente em momentos diferentes. Há inclusive casos de incoerência semântica, se apenas o critério de pronúncia for adotado para a identificação dos compostos. Um exemplo trazido por Levi (1978 *apud* RYDER, 1994) é o fato de que *apple cake* seria considerado um composto e *apple pie* não o seria, pois, no primeiro, *apple* é a palavra de pronúncia mais forte, e no segundo, *pie* é mais forte.

Levando-se em conta as considerações sintáticas, pode-se dizer que os compostos comportam-se como palavras, e que, dessa forma, não é possível inserir uma outra palavra entre os dois elementos. Outra característica sintática dos compostos é que não é possível modificar parte da palavra sem modificar o composto todo. Critérios como esses são adotados por Adams (1973 *apud* RYDER, 1994) e Bloomfield (1933 *apud* RYDER, 1994). Ao se levar em conta a sintaxe, alguns compostos que não passariam no teste de pronúncia seriam aprovados, como aqueles em que o modificador indica de qual material o objeto é feito. Por exemplo, em *glass door* não poderíamos incluir *beautiful* no meio: *\*glass beautiful door*. Também não seria possível incluir um adjetivo antes do composto querendo apenas modificar o primeiro elemento, como em *hard metal floor*. Assim, não há como dizer que o metal do qual a porta é feita é duro, mas a porta em si não o é.

Levi (1975 *apud* DOWNING, 1977) prioriza questões semânticas na identificação de um composto, adotando um critério quanto ao aspecto permanente ou habitual. Se for possível estabelecer alguma relação habitual entre os dois substantivos, tem-se um composto como em *water bug*, um inseto que vive na água. Ao afirmar que o composto deve ter um aspecto permanente, Levi (1978 *apud* RYDER, 1994) não consideraria *moth hole* um composto, porque a mariposa pode já ter abandonado o buraco. No entanto, o aspecto permanente, segundo Ryder (1994), é o fato de que a mariposa criou o buraco. De alguma forma, tanto os

---

<sup>3</sup> No anexo A, encontram-se traduções livres, feitas pela autora deste trabalho, dos compostos nominais exemplificados neste capítulo.

critérios sintáticos como os semânticos apresentados até aqui parecem trazer resultados coerentes.

Outra conclusão interessante de Ryder (1994) quanto aos compostos é o fato de que, ao se escolher um determinado modificador para um núcleo, apenas parte do significado do todo é informada. Por exemplo, em *glass door*, só é possível saber que essa porta é feita de vidro. Não há informação alguma quanto ao seu tamanho ou para onde se chega ao atravessá-la. Assim, pode-se dizer que a relação presente entre os elementos de um composto fornece apenas parte do significado da palavra.

Neste trabalho, adota-se o seguinte critério por ser basicamente semântico e também o mais utilizado em trabalhos da computação: a unidade referencial, ou seja, tem-se duas palavras mas apenas um referente. Em caso de dúvida, podem ainda ser aplicados critérios sintáticos, como a impossibilidade de inserção de outra palavra entre os dois elementos e o fato de que não é possível modificar apenas parte da palavra sem modificar o todo. Esses critérios parecem interessantes, pois reforçam a ideia de unidade referencial. Neste trabalho, pretende-se também verificar até que ponto a permanência da relação entre os elementos de um composto pode facilitar a identificação de compostos no *corpus* utilizado.

Feitas essas considerações sobre as dificuldades de se definir um composto, serão apresentados a seguir alguns pontos centrais para a compreensão da semântica dos compostos.

#### 2.1.1.1 Produtividade

Para Ryder (1994), a composição é um processo derivacional muito produtivo em língua inglesa. No caso dos compostos NN, ocorre a combinação de dois morfemas livres para formar um substantivo. A partir dessa afirmação, poder-se-ia concluir que, a todo instante, novos compostos surgem e a sua criação tem como base regras presentes em compostos já conhecidos pelos falantes. No entanto, caracterizar as regras gerais presentes nas formas já conhecidas e nas novas não é uma tarefa fácil. Mesmo que se considere que a língua é um sistema governado por regras, não há como negar que ainda se podem identificar casos que fogem a qualquer padrão. Quando eles ocorrem, são chamados de exceções, e, se há necessidade de se listarem esses elementos imprevisíveis da língua, cria-se um léxico. Segundo tal ótica, a gramática, constituída de regras em sua maioria sintáticas, e o léxico, visto como uma lista de palavras, são tratados separadamente. No caso dos compostos nominais, tema deste trabalho, por serem produtivos, imaginar-se-ia que eles obedecem a

regras e por isso seriam previstos pela gramática. Porém, a formação e a interpretação destes são idiossincráticas, ou seja, imprevisíveis.

A questão, segundo Ryder (1994), é que há graus de produtividade, variando entre a plena e a limitada. A produtividade plena se dá quando é possível prever as combinações de morfemas quanto às propriedades semânticas ou sintáticas. Já nos casos em que a produtividade é limitada, as combinações são imprevisíveis, bem como o significado do substantivo criado. Os compostos formados por animais em que o primeiro substantivo refere-se a um lugar são plenamente produtivos, pois o composto refere-se a animais e seus *habitats* correspondentes, tais como: *polar bear*, *sea cow*, *sea horse*, *water bug*. Porém um composto como *blackmail* possui produtividade limitada, pois não há uma regra que descreva este padrão e que dê origem a outros compostos. No entanto, uma forma mais atual pôde ser criada por analogia, *whitemail*.

Devido basicamente ao que Ryder (1994) chama de *semantic drift* (movimento semântico), usos que eram produtivos antigamente hoje em dia não são mais, pois o composto originalmente pode ter um significado e, sincronicamente, um de seus elementos adquire um novo significado. O padrão que motivou a formação do composto *Angora cat* era indicar a origem do animal, sendo que *Angora* se refere a um local. Atualmente, quando alguém se refere a um gato angorá, a informação original não é mais recuperada, pois se associa o modificador de gato a uma característica do animal, o fato de ter pelos grandes. A partir desse novo significado, novos compostos são criados, como *Angora rabbit*. O que Ryder (1994) tenta deixar claro é que os compostos não podem ser divididos em produtivos e não-produtivos, mas se deve sugerir que alguns compostos são mais produtivos do que outros.

Os compostos são produtivos porque novas combinações de palavras em contextos variados podem ser criadas, mas dispõem de características idiossincráticas, já que não é possível estabelecer padrões quanto à sua produção. Esse é o principal problema relacionado aos compostos nominais, o que justifica não apenas o interesse de muitos pesquisadores por essa temática, como também a falta de consenso no que tange ao seu tratamento teórico. É difícil criar regras, estabelecer padrões, pois o fenômeno é idiossincrático, mas também não é possível criar um léxico com todos os compostos, sendo que novas formas podem ser criadas com frequência. No entanto, ao se identificar graus de produtividade ou de semi-produtividade em algumas construções, é possível propor algum tratamento para os compostos.

### 2.1.1.2 Transparência semântica

Segundo Vale (2001), se uma expressão apresenta transparência semântica, o seu significado é obtido a partir da soma do significado dos seus componentes. Este mesmo conceito foi apresentado anteriormente por Bolinger (1975, p. 411), que descreve transparência semântica como a “qualidade de expressões que falantes conseguem facilmente ‘decifrar’”. Há também uma forte ligação da transparência semântica com a produtividade. Smarsaro (2004) traz como exemplo de expressão transparente *dente de ouro*, pois é possível interpretá-la como um objeto que é feito com determinado material. Essa mesma expressão também é extremamente produtiva, pois existem outras construções semelhantes em que o objeto e o material do qual este é feito são relacionados, como: *dente de amálgama*, *dente de porcelana*, *dente de resina*, entre outros.

Quando uma palavra é produtiva e apresenta transparência semântica, Smarsaro (2004) a considera composicional, ou seja, o seu significado pode ser deduzido do significado dos seus componentes. A relação entre produtividade e composicionalidade se justifica, pois os elementos que formam uma expressão composicional são livres e podem dar origem a outras construções, sendo desta forma também produtivos. Quando isso não é possível, dizemos que a palavra não é composicional, como em *lua-de-mel*. Nesse caso, não se trata de uma lua, nem de algo feito de mel; não há nenhuma transparência entre os elementos que a compõem. Com esse tipo de expressão sem transparência semântica, também chamado de opaco semanticamente, a produtividade é baixa. Não há outras expressões com o mesmo tipo de relação, como *lua de sal*, *lua de água*, *lua de óleo* ou *lua de açúcar*.

Tanto os compostos transparentes semanticamente quanto os opacos serão considerados para a análise neste trabalho. No entanto, imagina-se que os transparentes se prestem mais a uma abordagem co-composicional<sup>4</sup>, em que o sentido de cada elemento contribui para o significado do composto. Por ser mais frequente a criação de novos compostos a partir de outros considerados transparentes, eles são extremamente produtivos, e espera-se, por meio de uma análise de *corpus*, encontrar usos de compostos transparentes cuja proporção em relação aos opacos possa ser verificada.

Jespersen (1909 *apud* RYDER, 1994) considera compostos somente as expressões que não são transparentes semanticamente, como *blackmail* e *honey moon*. Uma construção

---

<sup>4</sup> A noção de co-composição adotada aqui é a de Pustejovsky (1991), para quem o significado de uma sentença é determinado não só pela aplicação do verbo sobre o argumento, mas também pela aplicação da função do argumento sobre o verbo.

transparente como *glass door*, uma porta que é feita de vidro, para o autor, não seria um composto. Como neste trabalho o tema são os compostos NN em geral, busca-se alguma forma de interpretar tanto as construções transparentes quanto as opacas, ambas consideradas compostas, seguindo Copestake (2003, ver subseção 2.2.3) e Ryder (1994), para quem as combinações de dois substantivos separados por um espaço podem constituir compostos.

### 2.1.1.3 Nuclearidade

Os compostos em língua inglesa são geralmente formados por modificador seguido de núcleo. Segundo McDonald (1995), o núcleo morfológico de um composto pode ser identificado por meio de critérios sintáticos ou semânticos. Neste trabalho, considera-se o núcleo semântico dos compostos NN, porém é importante diferenciá-lo, em linhas gerais, do núcleo sintático.

O núcleo sintático de um composto é a palavra que determina o composto como um todo, e a sua posição, na língua inglesa, é em geral a do elemento da direita. Em *polar bear*, o núcleo sintático é *bear*, pois, embora tenhamos um composto formado por um adjetivo e um substantivo, o composto como um todo é um substantivo. Como neste trabalho só serão analisados os compostos formados por dois substantivos, a principal consideração sintática que se aplica a este estudo é que o núcleo do composto é o elemento da direita.

Em língua portuguesa, os equivalentes de traduções dos compostos NN da língua inglesa são formados por um núcleo sucedido por um modificador; ou seja, em português, assim como em outras línguas latinas, tal como o italiano, o núcleo é a palavra da esquerda. Até mesmo em inglês, há casos que fogem à regra. Compostos como *attorney general* e *mother-in-law* servem de exemplo: o núcleo é o substantivo da esquerda, o que se evidencia pela forma no plural dessas palavras – *attorneys general* e *mothers-in-law*.

Para Jespersen (1924 *apud* MCDONALD, 1995), o núcleo semântico é a palavra mais importante, que é definida ou modificada por outra. Assim, em *apple cake*, o núcleo semântico é *cake*, pois a expressão refere-se a um bolo que é feito de maçã, sendo *apple* o seu modificador. Na língua inglesa, o núcleo semântico nem sempre será o elemento da direita. No entanto, também espera-se identificar casos em que o composto NN não possui núcleo, como em *poet-painter*, em que o referente é as duas coisas ao mesmo tempo ou em compostos com sentido pejorativo, tal qual *bonehead*. Quando não é possível identificar um núcleo, considera-se a referência; ou seja, se os dois elementos constituem uma unidade, um único referente, tem-se um composto.

#### 2.1.1.4 Compostos endocêntricos e exocêntricos

Quando um composto é considerado o hipônimo de seu núcleo, tem-se um endocêntrico. Para Cruse (1986, p. 88), hiponímia é “a relação lexical que corresponde à inclusão de uma classe em outra”<sup>5</sup>, sendo também chamada de relação “é um”. Veja-se o composto *desktop computer*. Como é possível dizer que *desktop computer* é um tipo de computador, esse composto pode ser considerado endocêntrico.

Em alguns casos de compostos nominais não é possível estabelecer essa relação de classe e subclasse. Se a relação “é um” ou “é um tipo de” não pode ser identificada, a expressão é exocêntrica, com em: *bird brain*. A relação “é um” não se aplica nesse caso, pois não podemos afirmar que um *bird brain* é um tipo de *brain*; se chamamos alguém dessa forma, referimo-nos figurativamente a um tipo de pessoa, cujo tamanho do cérebro está sendo comparado com o de um pássaro.

#### 2.1.1.5 Lexicalização

O conceito de lexicalização está relacionado à transparência semântica dos constituintes de um composto. Para Sandmann (1997), são lexicalizados os compostos que não são transparentes semanticamente. *Pé-de-galinha* seria um composto lexicalizado para o autor, exatamente porque não se refere a um pé de uma galinha. Neste trabalho, adota-se esse conceito de lexicalização, pois autores da área da informática, como Copestake (2003), também consideram lexicalizados os compostos opacos. É importante, para um trabalho interdisciplinar, adotar conceitos utilizados nas duas áreas, sendo aqui a informática e a linguística. A solução encontrada por Copestake (2003) para os compostos lexicalizados é listá-los em um léxico.

Para Downing (1977), quanto mais lexicalizada for uma expressão, mais facilmente ela poderá ser interpretada isoladamente, fora do seu contexto de uso. Também é importante considerar que, dentro do grupo dos compostos lexicalizados, há diferentes graus de lexicalização, que estão relacionados às mudanças de significado da relação entre os dois membros do composto.

Há uma espécie de lexicalização parcial em *pano de prato*, por exemplo, pois o *pano* não serve somente para secar pratos, secando quaisquer tipos de louças. Já em *pé-de-galinha*,

---

<sup>5</sup> Todas as citações em língua estrangeira foram traduzidas livremente pela autora deste trabalho.



teríamos um composto lexicalizado, pois este não é transparente, não se tratando de pé algum, mas de rugas próximas aos olhos. A transparência semântica de um composto não é constante, pois, quando é criado, ele pode ser transparente, mas com o tempo ele passa a ser aceito pela comunidade como um nome convencional. Se a relação de sentido entre os substantivos não pode ser recuperada, o composto se torna arbitrário, apresentando um nível intermediário de lexicalização.

Neste trabalho, foram analisados tanto os compostos lexicalizados quanto os não-lexicalizados. No entanto, não se procurou levar em consideração aspectos diacrônicos da semântica dos elementos de um composto, buscando-se analisá-los apenas sincronicamente. Ainda que seja possível identificar graus de lexicalização ou de transparência nos compostos, tentou-se analisá-los a partir do seu uso no *corpus*, sem se levar em consideração como surgiu a expressão ou se o seu significado sofreu alguma alteração ao longo do tempo.

#### 2.1.1.6 Nominalização

Para Lieber (2004), que realizou um estudo sobre a formação dos significados de expressões complexas, os compostos são divididos em dois grandes grupos conforme a classe gramatical do segundo elemento. Os compostos nominais são aqueles cujo N2 não é derivado de verbos, enquanto os compostos sintéticos (ou deverbais) apresentam um N2 deverbal, como em *truck driver, load tolerance, city employee* etc.

Quanto à composição, Lieber (2004) afirma que se trata do processo mais produtivo de formação de palavras em inglês, e os compostos nominais formados por substantivos estão entre os mais produtivos. Essa afirmação fortalece a justificativa para este estudo, pois o foco aqui são os compostos nominais formados por substantivos. Além disso, a composição nominal é “um meio altamente produtivo de formação de palavras para criar substantivos concretos” (LIEBER, 2004, p. 41). Talvez isso justifique o fato de que boa parte dos estudos semânticos dos compostos nominais seja sobre objetos concretos, conforme será apresentado na próxima subseção.

Apesar de o escopo deste trabalho se limitar aos compostos nominais, quando se faz uma busca por compostos formados por dois substantivos em um *corpus* de língua inglesa, depara-se (conforme dados de TEIXEIRA e CHISHMAN, 2008) com expressões em que o segundo elemento é deverbal. Casos como *flood losses, foundation investigation, horseback gathering* e *drainage improvement*, em que o N2 é deverbal, estão presentes nos dados deste

estudo, pois, sendo deverbais ou não, esses itens sofreram uma nominalização e são usados como substantivos no contexto específico de estudo.

Por nominalização, Basilio (2004) entende o conjunto de processos que forma substantivos a partir de adjetivos e principalmente a partir de verbos. Assim, neste estudo, pretende-se analisar também os compostos deverbais encontrados no presente *corpus*, pois os seus elementos sofreram uma nominalização e têm a função de substantivos, requerendo alguma explicação semântica.

Na tabela 1, a seguir, são resumidos alguns dos conceitos apresentados nesta subseção:

**TABELA 1:** Principais conceitos apresentados na subseção 2.1.1

<b>Conceitos</b>	<b>Definições</b>
Composto nominal	Dois substantivos que constituem uma unidade e têm como referência um único ser ou objeto.
Produtividade	Possibilidade de se criarem novas combinações de palavras em contextos variados a partir de um padrão estabelecido.
Transparência semântica	O significado de um composto pode ser derivado do significado de seus elementos.
Nuclearidade semântica	A palavra mais importante, que é definida ou modificada por outra, é o núcleo semântico.
Compostos endocêntricos	O composto é o hipônimo (relação “é um”) de seu núcleo.
Compostos exocêntricos	A relação “é um” não pode ser identificada entre o composto e o seu núcleo.
Lexicalização	São lexicalizados os compostos que não são transparentes semanticamente.
Nominalização	Formação de substantivos a partir de adjetivos e verbos.

### 2.1.2 Estudos clássicos sobre os compostos

Antes da apresentação das teorias que serviram de pilares para este estudo, será apresentada, nesta subseção, uma reflexão sobre alguns estudos que abordam a semântica dos compostos. Essa visão panorâmica serviu de base para a escolha das perspectivas adotadas neste trabalho.

As maiores dificuldades encontradas em estudos linguísticos sobre os compostos são definir composto nominal e estabelecer as suas principais categorias. Os autores concordam quanto a algumas características dos compostos, como os padrões sintáticos, no entanto, a área mais problemática é a semântica, especialmente quando se busca determinar o significado dos compostos NN.

Ryder (1994) faz um levantamento dos estudos sobre a semântica dos compostos realizados até então. Mesmo que os estudos apresentados pela autora sejam mais antigos, alguns ainda servem de referência para estudos realizados na atualidade. Além disso, o propósito desta revisão de literatura é identificar as relações semânticas comuns em diferentes trabalhos sobre os compostos NN.

Examinando-se as propostas de análise semântica dos compostos, percebe-se que algumas categorias são semelhantes ou recorrentes. A categoria *local*, por exemplo, está presente nas descrições de Jespersen (1909), Marchand (1969), Warren (1978) e Adams (1973) (*apud* RYDER, 1994). Um exemplo seria *garden-party*, em que o N1 indica onde o N2 ocorre. Para Adams (1973 *apud* RYDER, 1994), o local pode ser tanto o N1, como em *field mouse* e *pocket handkerchief*, como o N2, como em *biscuit factory* e *law court*. Já Warren (1978 *apud* RYDER, 1994) inclui tanto as relações de tempo como as de espaço na categoria localização. O diferencial do trabalho de Warren (1978 *apud* RYDER, 1994) é que, ao listar as relações semânticas possíveis entre os elementos dos compostos, ela também se preocupa em expor as principais características dos substantivos que podem preencher as posições de N1 e N2. A localização pode se referir a um lugar concreto ou abstrato, como em *sea horse* e *baseball personality*. Já o outro substantivo pode ser um objeto, uma entidade animada ou uma entidade abstrata, exemplificados respectivamente em *hotel desk*, *jail birds* e *school dance*.

Em Hatcher (1960 *apud* RYDER, 1994), não há uma categoria para o lugar, mas, por meio dos exemplos, é possível perceber que substantivos que indicam lugares são explicados por meio da relação parte/todo, como em: *sea gull*, *field mouse*, *pond lily*, *water cress*. Um rato do campo é considerado parte do campo, pois no campo há ratos. Um lírio do lago é considerado parte do lago, e assim por diante.

Nos compostos *eveningsong* e *nightclub*, o N1 indica o tempo, categoria que aparece em dois trabalhos, Marchand (1969) e Warren (1978) (*apud* RYDER, 1994). Conforme já mencionado anteriormente, a última autora inclui tanto o tempo como o local em um conceito mais geral, a localização, em que a relação tempo-objeto pode ser de três tipos:

- a) Tempo seguido de entidade animada: *weekend guests*, *afternoon clerk*, *spring squad*, *summer people*;
- b) Tempo seguido de entidade inanimada e concreta: *Sunday paper*, *afternoon sun*, *summer home*, *day school*;

- c) Tempo seguido de entidade abstrata: *fall colors, 1948 season, evening hours, winter air, night sky, summer job.*

Além das relações de tempo e lugar encontradas em mais de uma sugestão de análise, outras categorias são recorrentes e importantes para este trabalho, como função ou finalidade, relação parte/todo e comparação.

A relação de função é apresentada nos trabalhos de Jespersen (1909), Marchand (1969), Warren (1978) e Adams (1973) (*apud* RYDER, 1994). Pode-se dizer que o N1 indica para que serve o N2. Como exemplo, temos os seguintes compostos: *flagstaff, beehive, keyhole, birdcage, wineglass, cigar-case, fuel oil, gear wheel.* Para Warren (1978 *apud* RYDER, 1994), tanto o tempo como o lugar podem expressar a finalidade de algo, diferentemente dos casos de localização. *Nightdress* é um vestido que é usado à noite, esta é a sua função, enquanto *weekend guests* são os convidados que se hospedam em uma determinada residência durante o fim de semana, sendo que *weekend* não é a sua função, apenas a sua localização temporal. Diferenciar se o N2 é um nome abstrato ou concreto, animado ou inanimado pode auxiliar na identificação da relação que este apresenta com o N1. Casos do tempo como um indicador de função, segundo Warren (1978 *apud* RYDER, 1994), só ocorrem antes de objetos. Já o tempo como localização pode ocorrer antes de objetos inanimados, nomes abstratos e entidades animadas. O mesmo ocorre quanto ao local, conforme a autora.

A relação de parte e todo está presente nos cinco trabalhos estudados por Ryder (1994), sendo que duas direções são possíveis: N1 é uma parte de N2, como em *stone fruit*, e N2 é uma parte de N1, como em *broomstick*. Jespersen (1909), Marchand (1969) e Adams (1973) (*apud* RYDER, 1994) também criaram mais uma categoria, material ou “N2 é feito de N1”, em que também se poderia dizer que o material está de alguma forma contido em um dos elementos. No entanto, há uma diferença entre a relação de parte e todo e a relação de material: em *feather-bed*, as penas fazem parte da cama; se tirarmos as penas, ainda temos uma cama; mas em *gold ring*, o ouro é o material do qual o anel todo é feito; sem o ouro não temos um anel. Novamente encontra-se uma categoria que inclui tanto tempo como lugar, porém diferentemente da localização, em que o tempo ou o lugar ocupam a posição N1; no caso da relação parte-todo, eles aparecem no N2. Exemplos deste tipo seriam *flower garden* e *summertime*, em que as flores fazem parte do jardim e o verão é uma época do tempo.

A última relação que merece destaque entre os trabalhos apresentados por Ryder (1994) é a comparação, em que os dois elementos do composto apresentam alguma

característica em comum. Exemplos de Marchand (1969 *apud* RYDER, 1994) são *bell-flower* e *goldfish*, em que temos uma flor que lembra um sino pelo seu formato e um peixe que lembra ouro pela sua cor. Adams (1973 *apud* RYDER, 1994) também apresenta exemplos desse tipo, mas Warren (1978 *apud* RYDER, 1994) vai mais além incluindo nesse grupo os copulativos e os aditivos. Nos copulativos, como em *man servant*, pode-se dizer que o homem é um criado, e nos aditivos, como em *secretary-treasurer* e *poet-painter*, o referente constitui os dois elementos ao mesmo tempo em um mesmo nível. Compreende-se que a categoria comparação precisa ser desmembrada de alguma forma e sugere-se a seguinte: comparação em que alguma característica de N1 é presente em N2 (*bell-flower*); compostos em que é possível identificar uma relação de hiponímia (classe e subclasse ou classe e instância), também chamados de copulativos (*man servant*); compostos aditivos em que não é possível encontrar um núcleo para o referente (*poet-painter*).

Conforme já mencionado anteriormente, o panorama de trabalhos mais antigos apresentado nesta subseção serviu de ponto de partida para esta investigação. Mesmo que este estudo tenha como base duas teorias específicas (RYDER, 1994; PUSTEJOVSKY, 1995), foi importante analisar trabalhos já feitos para reconhecer quais as relações mais produtivas ou recorrentes. Dessa forma, foi possível traçar alguns aspectos teóricos presentes nos dados deste estudo, aprimorando assim o estudo semântico dos compostos em uso corrente na língua.

Considerando-se o que foi exposto até aqui, percebe-se a importância de algumas relações e parece necessário que uma teoria de análise semântica dos compostos deva dar conta desses casos de alguma forma. As relações a serem destacadas são as seguintes:

- a) tempo;
- b) lugar;
- c) parte-todo;
- d) material;
- e) função;
- f) comparação;
- g) classe-subclasse;
- h) adição.

Mesmo que tempo e lugar possam aparecer em outras relações como parte ou função, o seu uso como localização precisa de alguma forma ser diferenciado e a sua principal

característica é ocupar o N1. Da mesma forma, o material que é indicado no N1 pode ser considerado parte do todo, mas, como geralmente o objeto é completamente feito de um determinado material e não há como separar a parte desse todo, há uma relação diferente de alguma forma. A comparação propriamente dita seria a de que N2 possui alguma característica de N1. Os outros tipos de semelhança entre elementos em uma estrutura composta parecem ficar mais bem acomodados em categorias separadas, para não haver sobreposição, sendo estas as relações de classe-subclasse e adição.

A partir dessas conclusões iniciais, chega-se a conclusão de que uma das teorias estudadas neste trabalho, a do Léxico-Gerativo (PUSTEJOVSKY, 1995), parece poder dar conta de muitas dessas relações, como parte-todo, função e classe-subclasse. Essa teoria será apresentada com maiores detalhes na seção 2.1.4, mas pode-se adiantar que uma outra relação, apresentada somente por Jespersen (1909 *apud* RYDER, 1994), é prevista nessa teoria; trata-se do caso em que o N1 é um instrumento ou uma ferramenta que dá origem ao N2, como em *gunshot* e *sabre-cut*.

Muitos autores acreditam que um tratamento exaustivo dos compostos não é possível. Com o objetivo de abranger os diversos tipos de compostos, alguns autores tornam as categorias mais abstratas, o que dificulta a escolha da categoria para relacionar o composto. Quando poucas categorias são propostas, elas são tão gerais que não é possível enquadrar um composto em uma delas sem gerar dúvidas. No entanto, quando o autor sugere várias categorias, o problema é identificar um limite entre elas, de forma que não haja sobreposições. Um exemplo deste último caso é o estudo de Warren (1978 *apud* RYDER, 1994), em que a indicação de lugar aparece em diferentes categorias.

Ainda em relação aos estudos apresentados nesta subseção, parece interessante destacar duas questões: a referência dos compostos e os compostos exocêntricos. Hatcher (1960 *apud* RYDER, 1994) não se dedicou ao estudo da referência, mas, para a autora, as suas categorias (fonte, parte/todo e propósito) podem receber subcategorias de acordo com aspectos relacionados à referência do composto. Por exemplo, se o composto se refere a uma pessoa, um animal ou um objeto, que tipo de relações podem ser previstas entre o referente e o seu modificador? A autora não chega a responder a essas questões. Assim, a hipótese de que cada referente possa ter relações particulares com o modificador pode ser confirmada a partir de uma análise de uso da língua, como no caso do presente trabalho, que utiliza os dados de um *corpus*. O estudo de Ryder (1994, apresentado na subseção 2.1.3) estabelece algumas categorias que consideram se o referente é ser humano, animal ou local e sugere algumas funções possíveis do seu modificador.

Entre os autores citados nesta subseção, apenas Marchand (1969) e Warren (1978) (*apud* RYDER, 1994) preocupam-se em caracterizar de alguma forma os compostos exocêntricos, ou seja, aqueles em que a relação de hiponímia (classe e subclasse) não é identificada. Marchand (1969 *apud* RYDER, 1994) não esclarece se há alguma classificação que dê conta de todos os tipos de compostos exocêntricos, mas afirma que esses compostos denotam seres vivos, como pessoas, animais ou plantas, além de algumas outras coisas. Quando o composto exocêntrico é relacionado a pessoas, geralmente o seu uso é pejorativo, como em *bonehead*, *blockhead*, *doughface*, *humpback*, *lionheart*, *potbelly*, *rubberneck*, *egghead*. Nos exocêntricos, segundo Warren (1978 *apud* RYDER, 1994), nem sempre é possível identificar um núcleo, fato que coloca em xeque a estrutura clássica dos compostos: modificador e núcleo.

Um outro estudo que merece destaque e é apresentado em Ryder (1994) é o de Levi (1978). Mesmo sendo um trabalho mais antigo, há muitas referências a esse estudo em trabalhos da área da informática, como em Ó Séaghdha (2007). Os estudos apresentados anteriormente são estruturalistas, ao passo que o de Levi é gerativista, fazendo parte de uma corrente gerativista da década de 1970. Provavelmente esses estudos ainda sirvam de referência para trabalhos atuais por não terem surgido propostas de estudos linguísticos mais completas sobre os compostos que forneçam classificações que possam ser utilizadas por estudos computacionais.

Levi (1978 *apud* RYDER, 1994) propõe que por trás de cada composto há uma estrutura subjacente formada por predicados apagados. Ela sugere nove predicados apagados recuperáveis que podem fazer parte da forma subjacente de um composto NN: CAUSE, HAVE, MAKE, USE, BE, IN, FOR, FROM e ABOUT (causar, ter, fazer, usar, ser, em, para, de e sobre). Fornecer uma lista finita com os predicados possíveis facilita o trabalho do analisador que busca interpretar as relações semânticas dos compostos presentes em um *corpus*, uma das atividades propostas neste estudo. No entanto, é preciso conferir se esses predicados são abrangentes, ou seja, se eles dão conta dos variados tipos de compostos; além disso, também é importante ver se, ao se tentar encaixar um composto em uma dessas categorias, não há sobreposição, permitindo mais de uma interpretação a um mesmo composto. Neste trabalho, a fim de esclarecer essas dúvidas quanto aos predicados de Levi (1978 *apud* RYDER, 1994), analisou-se cada um dos predicados.

No estudo de Levi (1978 *apud* RYDER, 1994), a relação entre os elementos de um composto é explicada por meio da sintaxe de uma possível oração relativa que teria dado origem ao composto; ou seja, a autora procura explicitar se o N1 exerceria o papel de sujeito

ou objeto de uma oração relativa. O primeiro predicado sugerido por ela é o CAUSE, em que o N1 pode ser o sujeito (*drug death, birth pains*) ou o objeto direto (*tear gas, disease germ*).

Quanto ao predicado HAVE, para Levi (1978 *apud* RYDER, 1994), o N1 pode ser o sujeito (*reptilian scales, lemon peel*) ou o objeto direto (*picture book, apple cake*). A autora apresenta dois tipos de uso do predicado MAKE, mencionando os casos em que o N1 é o sujeito (*daisy chains*) ou o objeto direto (*honeybee, songbird*) de uma oração relativa.

Nos exemplos do predicado USE de Levi (1978 *apud* RYDER, 1994), o N1 é o objeto direto (*voice vote, steam iron*). Já no composto *soldier ant*, a autora identifica o predicado BE e o N1 funcionando como objeto direto; assim sendo, poderíamos construir a oração “a formiga é um soldado”. O predicado IN da autora indica a localização de uma forma geral, em que o N1 pode ser o tempo (*morning prayer*) ou o local (*field mouse*) ou indicar uma localização metafórica (*marital sex*). Compostos como *horse doctor, arms budget* e *nose drops* são interpretados pela autora por meio do predicado FOR, sugerindo as paráfrases *doctor for horses, budget for arms* e *drops for the nose*. Entre os predicados de Brekle (1970 *apud* RYDER, 1994), o mais próximo do significado de FOR seria a relação final. Um bom exemplo é *baby powder*, que poderia ser parafraseado como *talco para bebê*.

O penúltimo predicado de Levi (1978 *apud* RYDER, 1994) é FROM (*olive oil, candlelight, battle fatigue*), que indica uma relação de direção, de onde algo veio (*oil from the olive, light from the candle, fatigue from battle*). O predicado mais neutro entre os sugeridos pela autora é ABOUT, cujos exemplos são *tax law* e *price war*, cujas paráfrases correspondentes são *law about tax* e *war about price*.

Grande parte das categorias de Levi (1978 *apud* RYDER, 1994) parece pertinente, pois é possível identificar essas relações em um *corpus*. Porém, o predicado ABOUT parece representar uma relação pouco específica, tornando-se interessante avaliar se, a partir dos dados encontrados no *corpus* de estudo, é necessário especificar uma relação geral como essa. Outra questão que chama a atenção na tipologia da autora é o fato de ela usar tanto verbos como preposições, escolha que pode trazer algumas inconsistências durante a classificação. Ao se utilizar a categoria IN é possível se referir a algo que faz parte de um determinado lugar, assim o predicado HAVE também poderia ser utilizado.

Ao se tentar explicar a ambiguidade de um composto por meio da sua estrutura subjacente, um composto com diferentes interpretações possíveis terá mais de uma estrutura. Mesmo propondo-se estruturas subjacentes diferentes para os diferentes significados e referentes, não há como saber de qual estrutura subjacente o composto é derivado, uma vez que o predicado foi apagado durante a transformação da estrutura em composto. Ou seja, a



informação com o significado entre os elementos que foi apagada não pode ser recuperada. Mesmo que um composto possa gerar diferentes interpretações, Levi (1978 *apud* RYDER, 1994) afirma que geralmente uma interpretação é mais frequente em uma determinada época ou comunidade de fala. A primeira interpretação para *horse doctor* é de um “médico para cavalos”, mas outras interpretações menos frequentes também são possíveis, como “médico que é um cavalo”, “médico que possui cavalos” ou “médico que usa cavalos”. Embora uma interpretação seja a mais provável, a autora não considera isso no seu trabalho, pois de qualquer forma a ambiguidade está presente na competência do falante e ela acredita que uma descrição linguística deve abranger todos os possíveis sentidos de um nominal complexo.

Além de não haver limites claros entre as estruturas, a escolha de uma estrutura subjacente é arbitrária. Não há como afirmar que a relação entre os membros de um composto é expressa em uma oração relativa. Um exemplo claro disso é o de Gleitman e Gleitman (1970 *apud* DOWNING, 1977) quanto ao composto *garbage man*, pois nem todo homem que recolhe o lixo é um lixeiro, somente aquele que profissionalmente e regularmente remove o lixo pode ser um lixeiro. Outra justificativa de Downing (1977) para rejeitar que as orações relativas possam expressar o sentido geral entre os dois substantivos é que nem todas as orações relativas necessariamente foram reduzidas a compostos. Mesmo existindo o composto *horse doctor*, é possível encontrar a 423 ocorrências da expressão *doctor for horses* no site de busca da *Google*<sup>6</sup>.

Na tabela 2, a seguir, são apresentadas as relações semânticas que serão consideradas neste trabalho, por serem reforçadas por diversos autores e também porque, por meio dessa seleção, busca-se uma descrição mais geral que possa evitar casos de sobreposição de categorias.

**TABELA 2:** Resumo das relações selecionadas a partir de trabalhos estruturalistas

<b>Relação</b>	<b>Descrição</b>	<b>Exemplos</b>
Tempo	Indica localização e o tempo ocupa o N1.	<i>weekend guests, Sunday paper.</i>
Lugar	Indica localização e o lugar ocupa o N1.	<i>garden-party, pocket handkerchief, school dance.</i>
Parte-todo	N1 é uma parte de N2 ou N2 é uma parte de N1. Se o todo for tempo ou lugar, eles ocuparão o N2.	<i>stone fruit, broomstick, feather-bed, flower garden, golf-season.</i>
Material	N2 é feito de N1.	<i>gold ring.</i>
Função	N1 indica para que serve o N2.	<i>flagstaff, beehive, keyhole,</i>

<sup>6</sup>Quando realizada a busca por *horse doctor*, obteve-se 60.100 resultados. As buscas foram feitas com as expressões entre aspas.

	Tempo e lugar podem expressar a finalidade de algo. O tempo e o local como indicadores de função, segundo Warren, só ocorrem antes de objetos.	<i>birdcage, wineglass, cigar-case, fuel oil, gear wheel, nightdress.</i>
Comparação	N2 possui alguma característica de N1.	<i>bell-flower e goldfish.</i>
Classe-subclasse	N1 é uma instância de N2.	<i>football game.</i>
Adição	Não é possível encontrar um núcleo para o referente.	<i>secretary-treasurer e poet-painter.</i>
Origem	N1 é um instrumento ou ferramenta que dá origem ao N2.	<i>gunshot e sabre-cut.</i>

Comparando-se a tabela 2 com as categorias sugeridas por Levi (1978 *apud* RYDER, 1994), é possível estabelecer alguma relação entre os dois tipos de estudo. O predicado HAVE pode ser interpretado, em grande parte dos casos, como um exemplo de meronímia (relação parte-todo), com exceção dos casos em que o verbo indica posse.

A relação com verbo MAKE pode ser dividida em dois casos específicos, dependendo da sua paráfrase: quando o significado da preposição que acompanha o verbo cria subcategorias, tais quais *made by* e *made of*; quando o uso de uma determinada preposição não é relevante, como em *honey bee*, que pode ser parafraseado por “abelha que faz mel”. Somente com as paráfrases formadas por *made of* é possível estabelecer alguma correspondência com uma das relações do quadro 2, a de material.

O predicado BE, conforme apresentado em Levi (1978 *apud* RYDER, 1994), assemelha-se às relações de comparação, classe-subclasse e adição, pois é possível parafrasear os compostos da tabela 2 utilizando este verbo: o peixe é dourado, o jogo é de futebol e o poeta é um pintor. Já IN indica tempo ou lugar e FOR expressa a função do N2. Os casos em que FROM expressa a origem de N2 não parecem corresponder à relação de origem encontrada nos outros trabalhos, pois nestes é um instrumento ou ferramenta que dá origem ao N2.

Não foi possível identificar algum caso na tabela 2 que seja correspondente ao uso dos predicados MAKE, USE, ABOUT e CAUSE de Levi (1978 *apud* RYDER, 1994). Isso pode indicar que essas relações merecem maior estudo.

Além dos trabalhos apresentados até aqui, há mais um que merece destaque, o estudo com abordagem psicolinguística de Downing (1977). Um diferencial desse trabalho é que houve uma preocupação em especificar características sobre o referente do composto. Downing (1977) associa algumas categorias semânticas do substantivo núcleo com

interpretações preferidas da relação entre modificador e núcleo. As categorias do núcleo e os papéis do modificador identificados pela autora são:

- a) ser humano – identidade ocupacional, sexual e racial: *women officers* – N2 é do gênero N1;
- b) animais – aparência, habitat: *giraffe bird* – N2 se parece com N1;
- c) plantas – aparência, habitat: *trumpet plant* – N2 se parece com N1;
- d) objeto natural – composição, localização espaço-temporal, origem, aparência: *cow hair* – N2 tem origem em N1;
- e) objeto sintático – propósito: *banana fork* – N2 é usado para cortar N1.

Nessas categorias, observa-se, na primeira coluna, a natureza do núcleo e, à sua direita, as principais funções do seu modificador. Mesmo sem propor um modelo que dê conta das suas descobertas, Downing (1977) faz considerações interessantes em seu trabalho quanto a objetos naturais e sintáticos e sobre o aspecto permanente e habitual. A autora compara os objetos naturais com os sintáticos, em que os primeiros são classificados a partir de suas características intrínsecas e os últimos quanto a sua função, já que um objeto é criado tendo-se uma utilização em mente para ele. Essa hipótese parece interessante de ser testada em um *corpus* com dados reais da língua.

Downing (1977) revela que a relação entre os membros de um composto geralmente apresenta um caráter permanente e habitual. Comparando um composto com a sua paráfrase em forma de oração relativa, Gleitman e Gleitman (1970 *apud* DOWNING, 1977) sugerem que o uso do composto implica que a relação é de natureza habitual, enquanto na oração relativa isso não ocorre. Como exemplo, eles comentam sobre o composto *owl-house*, que não pode ser interpretado como uma casa sobre a qual uma coruja cai ou sobrevoa. Há uma tendência em interpretar o composto a partir de uma unidade integral, genérica ou necessária. No entanto, há casos em que a relação no composto é fortuita ou temporária. No caso de *bus money*, dentro do contexto *I can't find my bus money*, comentado por Zimmer (1975 *apud* DOWNING, 1977), não há outra interpretação além de que se trata do dinheiro reservado para pagar a tarifa do ônibus. Não há uma relação habitual necessariamente, em que se acreditaria que alguém sempre coloca o dinheiro no bolso com esse propósito. Mesmo que os dois tipos de compostos sejam possíveis, Downing (1977) identificou que os compostos novos baseados em relações habituais ou genéricas são muito mais frequentes.

### 2.1.3 Linguística Cognitiva: Ryder (1994) e Fillmore (2006)<sup>7</sup>

Nesta subseção, é apresentado o estudo de Ryder (1994), que parte dos princípios da gramática cognitiva e da teoria dos esquemas para abordar a semântica dos compostos NN. Como Ryder (1994) organiza as relações semânticas dos compostos a partir de conceitos muito semelhantes aos conceitos da teoria de *frames* semânticos, alguns aspectos da base de dados lexical *FrameNet*<sup>8</sup> (BAKER *et al*, 1998) também serão apresentados. O *FrameNet* é organizado a partir da teoria de *Frames* de Fillmore (2006) e contém informações semânticas que podem ser utilizadas para uma análise co-composicional dos compostos nominais. Assim, nesta subseção, parte-se do trabalho de Ryder (1994) e logo após é feita a relação com o estudo de Fillmore (2006).

Ryder (1994) optou por essa base teórica, a gramática cognitiva de Langacker (1987), por perceber que estudos anteriores (apresentados na seção 2.1.2 deste trabalho) não dão conta de todos os compostos existentes ou que venham a ser formados pelos falantes de uma língua. Ryder (1994) acredita que categorizar os compostos por meio de regras e incluir as exceções em um léxico não é a melhor forma de se analisarem os compostos. O que ela defende é a existência de padrões com diferentes graus de produtividade, indo do menos produtivo ao extremamente produtivo.

Outra questão apontada por Ryder (1994) como problemática está relacionada à função da aplicação das regras. Se os mesmos padrões são utilizados para descrever formas já conhecidas ou para gerar formas novas, corre-se o risco de se encontrarem dificuldades. Alguns padrões que possam ter sido produtivos há alguns anos podem atualmente não ser mais. *Scarecrow* pode ter sido um tipo de construção produtiva antigamente, mas nos tempos atuais não se formam novos compostos a partir de uma relação semelhante. Há também casos em que até mesmo formações sincrônicas são criadas por analogia a formas que sofreram alguma modificação no seu significado.

Por essas razões, estabelecer regras gerais de formação de compostos é uma tarefa muito complexa. Ryder (1994) escolheu o modelo de gramática cognitiva criado por Langacker (1987) para o estudo dos compostos, pois nesse modelo não há a divisão em regras e exceções, têm-se padrões que são mais ou menos produtivos.

---

<sup>7</sup> A referência utilizada neste trabalho foi originalmente publicada em 1982 em *Linguistics in the Morning Calm*, Linguistics Society of Korea (ed.), p. 111-137, Seul: Hanshin Publishing Company.

<sup>8</sup> Disponível em: <<http://framenet.icsi.berkeley.edu/>>

O grande diferencial dessa gramática é que ela é baseada no uso. Assim, é a partir de frases realmente utilizadas pelos falantes que diferentes estruturas são extraídas, ou seja, estabelecem-se regras. Nessa perspectiva, não se objetiva estabelecer um grupo de regras gerais, mas o que se busca são casos especiais, incluindo assim formas totalmente idiossincráticas e padrões com graus intermediários de generalização.

Um dos conceitos mais importantes nessa teoria é o de valência gramatical, em que os compostos são formados por meio do compartilhamento de alguns elementos dos seus componentes. Nas relações de valência, a estrutura composta herda a maioria das características de um dos componentes da estrutura, ou seja, do núcleo, que, na teoria de Langacker (1987), é chamado de determinante do perfil (*profile determinant*). No caso dos compostos em língua inglesa, o determinante geralmente é o segundo substantivo, como em *night school*, cujo determinante de perfil ou núcleo é *school*.

Langacker (1987) faz uso de esquemas em sua gramática, os mesmos que Ryder (1994) decide aplicar em seu estudo. Por esquema, compreende-se uma estrutura abstrata do conhecimento que resume o que se sabe sobre uma variedade de casos e representa as relações entre as partes componentes. Essas partes são as variáveis. Por exemplo, entre as variáveis para o esquema  *festa*, tem-se o objetivo da festa e o que os convidados devem trazer. Quando se especifica uma das variáveis dizendo-se que é uma festa de aniversário, logo a segunda variável é preenchida, instanciada; sabe-se que se deve trazer um presente para o aniversariante.

Os esquemas apresentados por Ryder (1994) podem ser de três tipos: (i) esquemas de eventos ou *scripts*, (ii) esquemas de entidades e (iii) esquemas de características. Os *scripts* são sequências de eventos em um contexto particular, constituídos por variáveis e os requisitos de o que pode preencher essas variáveis. Nesses esquemas de eventos encontram-se as situações rotineiras pré-determinadas e que não sofrem alterações. Além das variáveis, que podem ser preenchidas ou não, cada *script* apresenta papéis diferentes e cenários e tempos esperados.

Em um substantivo, os seus possíveis referentes são vistos em uma variedade de papéis em que um ou mais esquemas de eventos são possíveis. Ao pensarmos na palavra *médico*, por exemplo, o seu referente pode estar conectado a diversos esquemas de eventos que diferem quanto à centralidade desses eventos. Segundo Ryder (1994), o esquema de evento central depende da experiência da pessoa e para a palavra *médico*, poderia ser *paciente que visita o médico*, *médico que faz uma operação cirúrgica* ou *médico que realiza pesquisa*. Há outros esquemas também possíveis, que são, no entanto, menos centrais quanto à

conexão com médico, como nos casos de *médico que dirige um mercedes-benz*, *médico que possui uma casa cara* e *médico que joga golf*. Por um médico ser um ser humano e também uma criatura viva, há vários outros esquemas menos centrais que são conectados a ele.

A partir de um grupo de esquemas de eventos, é possível abstrair subesquemas que são, de certa forma, independentes do esquema de eventos. No esquema *médico que faz uma operação cirúrgica*, podemos identificar algumas características possíveis no subesquema *médico*, como ser adulto, estar vestido de branco, ser inteligente e casado. Esses subesquemas são chamados de esquemas de entidades, que possuem características que são mais fortemente conectadas ao esquema de entidade médico do que outras. A principal diferença entre os dois tipos é que um esquema de eventos é construído a partir de um único ponto de vista, por exemplo, o ponto de vista do paciente que vai até o consultório médico, mas um esquema de entidade inclui informação de diferentes pontos de vista, a saber, todos aqueles que o falante conhece. De forma resumida, pode-se dizer que os esquemas de eventos referem-se às atividades nas quais o médico está envolvido e os esquemas de entidades trazem as propriedades da entidade, neste caso do médico.

Também é possível “abstrair dos esquemas de eventos e entidades as ações, relações ou qualidades que são comuns a todos eles” (RYDER, 1994, p. 69), tais como tamanho, cor, textura, proximidade ou ação que são frequentes nos esquemas de que fazem parte. Essas informações constituem os esquemas de características e são básicas, pois é fácil e quase automático para um ser humano identificá-las e abstrai-las. Uma característica como *redondo* faz parte de esquemas como *bola* e a fruta *laranja*.

Na teoria de Langacker (1987), a valência tem um papel importante na composição. Para fazer parte de uma mesma expressão, os elementos devem possuir subestruturas em comum, que o autor chama de correspondências. Em um composto NN, é necessário estabelecer uma correspondência de identidade entre um esquema e cada uma das estruturas. No composto *bread knife*, pode-se identificar um esquema central de cortar alguma coisa; nesse caso, as variáveis para o instrumento e para o que é cortado são preenchidas respectivamente por faca e pão. As características importantes que possibilitaram a criação deste composto são as constatações de que a faca corta e o pão pode ser cortado.

Há dois tipos de valência que podem ser encontrados nos compostos NN: (i) quando a combinação de palavras evoca um único esquema; (ii) quando diversos esquemas podem ser evocados em relação a um único item lexical. O primeiro pode ser exemplificado em *bread knife*, em que é possível identificar o esquema, em que a faca corta o pão. No segundo caso, há mais de um esquema de evento, como em *man*, que possui diversos esquemas conectados a

ele que são igualmente centrais, podendo ser a profissão, os gostos, a nacionalidade, etc. Mesmo assim, é possível estabelecer quais os esquemas possíveis relacionados a *man*. Além de estabelecer esquemas a partir de itens lexicais específicos, um mesmo esquema pode ser ativado a partir de diversas palavras, como palavras que se referem a algum tipo de recipiente. Podem-se utilizar as palavras *box*, *glass*, *pot*, *pan*, etc. Nesse caso, cada uma dessas palavras remete ao esquema “X contém Y”.

Outra questão referente à valência dos substantivos em um composto nominal é que os seus elementos possuem uma certa autonomia. Em um composto como *street fights*, é possível imaginar vários lugares onde uma briga pode acontecer, assim como se podem citar diversas atividades que podem ocorrer na rua. Além disso, é possível se referir a uma rua sem considerar quais atividades ocorrem nela. O inverso também é possível; em alguns momentos não interessa dizer onde a briga ocorre ou ocorreu.

A seguir, será apresentada a forma como Ryder (1994) utilizou estas duas teorias, a Gramática Cognitiva (LANGACKER, 1987) e a Teoria de Esquemas (RUMELHART e ORTONY, 1977 *apud* RYDER, 1994), para o estudo dos compostos nominais. O primeiro conceito importante, já mencionado anteriormente, é o de um determinante de perfil (*profile determinant*). Para Langacker (1987), o determinante é o núcleo do substantivo composto, que é acompanhado por um modificador. No caso dos compostos exocêntricos, não há um determinante de perfil.

Em relação aos compostos NN, Ryder (1994) afirma que eles não são prototípicos, ou seja, nem sempre é possível identificar um único determinante de perfil. Há casos em que o referente pode ser o primeiro ou o segundo substantivo. Ryder (1994) subdivide esse grupo de compostos em cinco categorias, considerando-se um composto formado por XY<sup>9</sup>:

- a) o referente é ao mesmo tempo um X e um Y: *student athlete*;
- b) o referente pode ser um X ou um Y em momentos e contextos diferentes: *hunter-gatherer*;
- c) o referente funciona como um X, mas tem a aparência de Y: *houseboat*;
- d) o referente possui a forma de X, e é um Y metafórico: *spoon people*;
- e) o referente possui algumas características de um X e algumas características de um Y: *spiderwoman*.

---

<sup>9</sup> Quanto aos exemplos de Ryder (1994) mencionados nesta seção, optou-se por manter sua representação original para os compostos NN, em que X refere-se ao primeiro substantivo e Y, ao segundo.

Quando um substantivo é usado como modificador de outro, de um núcleo, as duas palavras compartilham elementos conforme a sua valência. Assim, a palavra *cage* pode se referir a diversos tipos de construções usadas para trancar algum animal. No entanto, se o objetivo for especificar essa construção, referindo-se ao tipo de jaula, por exemplo, uma jaula que serviria para prender girafas, basta adicionar o modificador, formando *giraffe-cage*. Com esse composto, é possível identificar um esquema de evento sobre gaiolas que contêm girafas. Por meio desse esquema de evento é possível saber o tamanho e as principais características dessa gaiola.

Utilizar uma estrutura de modificador e núcleo, formando um composto nominal, tem como principal função categorizar as coisas. O modificador expressa características de dois tipos: funcionais ou descritivas. As características funcionais constituem as informações sobre o que o referente faz ou o que pode ser feito com ele. Duas características funcionais de *cavalo*, em um exemplo de Ryder (1994), são as seguintes: eles podem ser montados por alguém e eles servem para puxar carroças. Por meio de características descritivas, obtêm-se alguns detalhes sobre o referente, tais como tamanho, formato, cor, textura etc. É por meio de uma única característica descritiva, como formato, que duas coisas são diferenciadas, como, por exemplo, um triângulo e um círculo.

Uma das formas de se interpretar a característica funcional de um composto NN é tentar identificar um esquema de evento em que os dois elementos participam. É o caso de *baby-spoon*. Ao se imaginar que uma colher participa do esquema de eventos em mexer a comida e conduzi-la até a boca, pode-se interpretar o composto como um tipo de colher que participa de um esquema em que a comida é conduzida à boca de um bebê, pois esse esquema de eventos é compartilhado pelos dois elementos.

Quando o modificador traz uma característica descritiva, os dois elementos não participam do mesmo esquema. Eles tão pouco compartilham um esquema de evento. Em um *tiger cowrie*, o referente recebeu esse modificador devido à característica de que os dois são listrados.

Para expressar os esquemas possíveis em compostos NN, Ryder (1994) utiliza *templates*, que não deixam de ser esquemas que incluem características fonológicas e semânticas dos componentes e da estrutura composta. Se um *template* específico dá origem a novos compostos, ele é chamado de base de analogia.

As formas fixas e complexas, incluindo os compostos NN, que são frequentemente usadas pelos falantes são chamadas de “expressões convencionais” por Langacker (1987).



Uma expressão convencional como *sea legs* pode servir de base analógica para novos compostos, como *ice legs*.

Os *templates* linguísticos formados a partir de grupos de compostos que possuem um elemento em comum, ou seja, que constituem famílias de compostos, servem de bases de analogia. A palavra que se repete várias construções compostas é chamada de nódulo ou *core word* por Ryder (1994) e pode ocupar tanto o lugar do núcleo quanto o do modificador. Exemplos: *sea lion*, *seaman*, *sea cow*, *seaweed* ou *boathouse*, *warehouse*, *tree house*, *firehouse*. A partir de uma expressão convencional em que um padrão de uso é identificado, uma palavra nódulo deu e pode dar origem a novos compostos que seguirão o mesmo *template* linguístico.

Para Ryder (1994), um esquema como “X + *box* = uma caixa que contém/armazena X” é motivado pragmaticamente, já que um esquema de *box* sempre inclui uma variável central a ser ocupada por algo que se encontra dentro da caixa. Quando uma *core word* sempre possui o mesmo significado, independentemente da palavra que a acompanhar, Ryder (1994) afirma que há uma credibilidade de sugestão (*cue reliability*) absoluta. A noção de *cue reliability*, de Bates e MacWhinney’s (1987 *apud* RYDER, 1994), é uma sugestão que conduz à conclusão correta. Quando a sugestão sempre leva a uma conclusão correta, a credibilidade é alta.

Apesar de haver várias palavras com a mesma *core word* e que podem ser interpretadas a partir de um único *template* linguístico, pode haver algum tipo de variação. Uma *core word* como *man* apresenta uma *cue reliability* relativa, pois há uma variedade de significados para o composto, conforme a palavra com a qual ela se combine. Com a *core word man* que geralmente ocorre no segundo substantivo de um composto, é possível identificar diferentes padrões do primeiro substantivo:

- a) localização natural (*mountain man*), que indica onde o homem vive ou trabalha;
- b) localização construída pelo homem (*dairyman*), sendo o lugar onde ele trabalha;
- c) um instrumento (*plowman*) que é usado pelo homem para trabalhar;
- d) um animal (*cowman*) com o qual o homem trabalha ou de quem ele cuida;
- e) uma substância básica intrinsecamente sem forma (*snowman*) que apresenta-se em formato humano;
- f) um objeto produzido (*milkman*) que o homem faz, vende ou transporta.

Através da presença da *core word* não é possível identificar o significado do composto, mas, ao combinar a palavra com outras palavras com classes semânticas específicas, é possível reduzir as possíveis interpretações a resultados mais confiáveis.

Outro tipo de variação pode ocorrer se, a partir de expressões estabelecidas, não se consigam relacionar de alguma forma as diferentes palavras que acompanham a *core word*. Havendo uma credibilidade extremamente baixa, é necessário se ter um *template* linguístico diferente para cada composto. Ryder (1994) exemplifica esse tipo de credibilidade por meio de compostos com *board*: *fiberboard*, *corkboard*, *billboard*, *breadboard*, *chalkboard*, *checkerboard*, *cupboard*, *surfboard*, *headboard*. Além de a relação semântica entre os elementos ser de diversos tipos, o referente para *board* também varia bastante.

Embora a relação de um composto possa ser interpretada com uma certa previsibilidade através de uso de *core words*, em alguns casos, o significado de um dos elementos precisa ser ajustado, ocorrendo o que Langacker (1987) chama de acomodação. Em um esquema como *Cão bebendo vinho*, a acomodação ocorre por meio da alteração do significado de *cão*, mas não é uma mudança muito drástica. Cães não bebem vinho, mas teria-se que imaginar um cão específico que tivesse adquirido este hábito. Também é possível acomodar uma constante do *script* *Cliente comendo em um restaurante*, se nesse restaurante o cliente não precisa pagar pela sua refeição.

A tabela 3, a seguir, resume os principais conceitos apresentados nesta subseção:

**TABELA 3:** Resumo dos principais conceitos apresentados por Ryder (1994)

<b>Conceito</b>	<b>Definição</b>
Valência gramatical	Compartilhamento de alguns elementos dos componentes de uma expressão composta. Nas relações de valência, a estrutura composta herda a maioria das características de um dos componentes da estrutura, ou seja, do núcleo.
Esquema	Estrutura abstrata do conhecimento que resume o que se sabe sobre variedade de casos e representa as relações entre as variáveis. Um esquema refere-se a uma situação que envolve variáveis. Por exemplo, o esquema <i>jogo</i> evoca os elementos que fazem parte deste evento, como local, jogadores, tempo, etc.
<i>Profile determinant</i>	Constitui o núcleo do substantivo composto, sendo geralmente o N2 em língua inglesa.
<i>Templates</i>	Esquemas que incluem características fonológicas e semânticas dos componentes e da estrutura composta, sendo representado através de fórmulas. Uma das possibilidades de evocar o esquema <i>jogo</i> pode ser através do <i>template</i> linguístico <i>esporte + game</i> . Os compostos <i>soccer game</i> , <i>football game</i> são exemplares deste <i>template</i> e o N1 preenche

	uma das variáveis do esquema, especificando o tipo de competição.
Base de analogia	<i>Template</i> específico que dá origem a novos compostos.
<i>Core word</i> (núdulo)	Palavra que se repete em construções compostas e pode ocupar tanto o lugar do núcleo quanto o do modificador.
Expressões convencionais	Formas fixas e complexas, incluindo os compostos NN, que são frequentemente usadas pelos falantes.
<i>Cue reliability</i>	Sugestão que conduz a conclusão correta sobre o significado do composto NN. Quando uma <i>core word</i> sempre possui o mesmo significado, independentemente da palavra que a acompanhar, a credibilidade de sugestão é alta.
Acomodação	O significado de um dos elementos do composto precisa ser ajustado para ser integrado com o outro substantivo, formando uma expressão complexa coerente.

Partindo dos conceitos apresentados na tabela 3, Ryder (1994) analisou 1.600 compostos extraídos dos livros *American Heritage* e *American Heritage Word Frequency Book* (1971). Por meio dessa análise, ela conseguiu identificar cinco *templates* linguísticos frequentes e confiáveis, em que um dos elementos refere-se a: localização, recipiente, ser humano, parte do corpo e animal. Por confiáveis, conforme já foi apresentado anteriormente, entende-se que substantivos com certos tipos de significado quando aparecem juntos formando compostos produzirão um tipo previsto de significado no composto. Na maioria dos casos estudados por Ryder (1994), o significado do composto representa um esquema central para um dos substantivos ou para ambos. Por exemplo, em *wine bottle*, sabe-se que o vinho geralmente é mantido em garrafas e que garrafa participa de um esquema que possui uma variável a ser preenchida com o que é mantido na garrafa. Assim, não causa surpresa que o composto seja interpretado como “uma garrafa que contém vinho”. Os *templates* linguísticos, também chamados de padrões pela autora, apenas diferem pelo fato de que alguns são mais frequentes do que outros.

O padrão mais frequente encontrado nos dados de Ryder (1994) é:

- Localização  $Y + X = X$  localizado em  $Y$ <sup>10</sup>. Exemplos: *camp stool, altarpiece, contrywoman, seaman, schoolroom, housemaid, dairymaid, bushman, townsfolk, yard boy, spaceman, stableboy, seahorse, field mouse, mountain laurel, beach plum, earthworm, prairie dog, housefly, tree bear, tree ant, rock badger, meadowlark, riverboat, garageman*.

<sup>10</sup> Quando refere-se aos *templates*, Ryder (1994) utiliza Y para referir-se ao primeiro substantivo do composto NN e X, para o segundo. O sinal de igual relaciona o padrão com uma paráfrase equivalente.

O reverso desse padrão também é possível:

- Y + Localização X = X no qual Y é tipicamente encontrado. Exemplos: *apple orchard, cranberry bog, hen house, birdhouse, cow shed, pigpen, fairyland, barbershop, bedroom, icehouse, shipyard, boathouse, woodshed, grassland, cornfield, hayloft, rose garden, grocery store, gas station.*

A autora inclui na etiqueta “localização” outras relações semânticas como parte/todo e instrumento. Ela define localização como “referente imóvel e inanimado com espaço interior, ou se é um objeto plano, com expansão interna” (RYDER, 1994, p. 98). Uma localização pode ser natural (cavernas, oceanos, campos, lagos, rios) ou artificial (casas, salas, corredores, barracões, fábricas) e, para que seja considerada uma localização de um dos elementos em um composto, deve-se levar em consideração o tamanho do outro elemento em relação à localização. Uma árvore pode ser a localização de um esquilo, mas não a de um elefante típico.

Outras interpretações podem fazer com que esse padrão se sobreponha a outras relações, como parte/todo e instrumento. Dependendo do ponto de vista assumido, o motor de um carro (*car engine*) pode ser considerado uma parte do carro ou um motor que é localizado dentro do carro. Da mesma forma, um banco de acampamento (*campstool*) pode ser considerado um banco que é geralmente encontrado dentro da área de um acampamento, ou um componente da coleção de coisas que fazem parte de um acampamento, ou um instrumento usado na atividade de acampar. Para Langacker (1987), as várias interpretações possíveis não são um problema, já que isso é esperado de alguma forma. Ryder (1994) decidiu incluir os três casos em uma única categoria porque todos os diferentes sentidos são possíveis e não há como determinar uma única interpretação.

Para esses dois padrões para localização, Ryder (1994) sugere subpadrões:

- a) se o item localizado é um animal ou uma planta, geralmente ele mora ou cresce na localização: *mountain laurel, seaweed*;
- b) se ele é um humano e a localização não é uma casa, o local é onde ele mora ou possivelmente trabalha: *mountain man, countrywoman*;
- c) se é uma casa, é onde o humano trabalha: *houseboy*.

O último subpadrão não poderia indicar que alguém mora em uma casa, pois casa é o valor padrão para o local onde os humanos vivem, e informar isso seria redundante.

Nos dados de Ryder (1994), o segundo padrão comum é:

- Y + Recipiente X = X que tipicamente contém Y. Exemplos: *suitcase, teapot, ice bag, saucepan, wastebasket, bookcase, mailbox, wine glass, matchbox, dishpan, matchbook, coal scuttle, dustbin*.

O padrão recipiente apresenta as mesmas características de localização, porém ele é móvel, ou seja, é geralmente menor do que localização e na maioria das vezes é artificial. O reverso desse padrão – Y Recipiente + X – não apresenta uma interpretação homogênea. Isso ocorre porque os recipientes são artificiais e foram criados com um determinado propósito, podendo conter diversos itens. Entretanto informar que algo pode ser contido em um recipiente não é uma informação relevante. Assim, duas interpretações são possíveis: “X habitualmente contido em Y” (*bag lunch, box wine*) e “X semelhante a Y em formato” (*boxcar, box stall, box office, box bed*).

Substantivos que se referem a seres humanos também apresentam *templates* altamente frequentes:

- animal doméstico + humano = um humano que cria, cuida ou treina animal doméstico. Exemplos: *horseman, cattleman, poultryman, sheepman, goatboy, goosegirl, horsecowman*;
- veículo/maquinaria + humano = um humano que opera veículo/maquinaria. Exemplos: *boatman, cabman, trainman, ferryman, busman, bargeman, coachman, taximan, trolleyman, tugboat man*;
- instrumento/ferramenta + humano = um humano que trabalha usando instrumento/ferramenta. Exemplos: *ploughboy, brakeman, cameraman, drillman, oarsman, penman, radioman, tillerman*;
- arma + humano = um humano que usa a arma, geralmente como um assassino ou caçador profissional. Exemplos: *gunman, spearman, bowman, rifleman, axman, crossbowman, swordsman, hatchet man, pike man*.

Há um *template* mais geral que geralmente se sobrepõe aos outros ou os substitui:

- produto + humano = um humano que faz, vende, entrega, transporta ou processa um produto como uma profissão. Exemplos: *mailman, milkmaid, garbage man, iceman, basket woman, butterman, cinderman, dry goods man, dustman, fishwife, fruit woman, groceryman, hardwareman, meatman, oysterman, pastry man, scrapman, teaman, egg woman, flower girl, newspaperwoman, newspaperboy*.

Por produto, compreende-se “qualquer coisa cuja produção ou desenvolvimento é influenciado por pessoas” (RYDER, 1994, p. 101). Leite, por exemplo, seria um tipo de produto. Dessa forma, como uma mesma pessoa pode executar mais de uma atividade em relação a um mesmo item, como produzir, vender ou processar, esses *templates* se relacionam.

Também é possível que um dos substantivos indique um *hobby* ou a atividade favorita de alguém. No entanto, Ryder (1994) encontrou apenas dois compostos estabelecidos com esse sentido, *sportsman* e *yachtsman*.

Os dois últimos padrões são menos frequentes, mas apresentam certa regularidade. Um deles se refere à parte do corpo:

- parte do corpo + roupa/jóias = roupa/jóias vestidas na parte do corpo. Exemplos: *headband, necktie, earmuffs, breastplate, neckerchief, wristwatch, eyeglasses, earring, necklace, body stocking, neckband, wristband, hairpiece, knee plate, chin strap, neckpiece, neckcloth*. Esse grupo é bem pequeno, já que a maioria das roupas já possui um valor padrão bem forte para onde elas serão usadas. Assim, um composto como *waist-belt* seria redundante. Geralmente o segundo substantivo não é uma peça de roupa, mas algo que originalmente é usado em outra parte do corpo, tal como lenços (usados na cabeça), regatos (usados nas mãos), relógios (guardados nos bolsos), anéis (usados nos dedos) e meias (usadas nos pés);
- parte do corpo + roupa = roupa que se estende até a parte do corpo. Exemplos: *waistcoat, ankle socks, knee socks, knee pants, elbow cuffs, hip boots, knee boots*. Esse grupo também é pequeno, já que nem todos os itens de roupas podem cobrir diferentes extensões do corpo;
- parte do corpo + algo que não é roupa = algo que não é roupa é operado por/usado na parte do corpo. Exemplos: *foot pedal, foot brake, handcart, hand soap, toothbrush, hairbrush, face cream, neck brace, face towel, facecloth, hand brake, hand drill*,

*handsled, handkerchief, handbag*. Esse grupo é limitado, pois muitas ferramentas possuem uma parte do corpo como padrão usada para operá-las e porque muitos itens usados no corpo não podem ser limitados a somente uma parte, como perfume.

Há um outro *template* que começa com determinadas partes do corpo, *head, foot e tail*, em que essa parte se refere a algo que está no topo, abaixo ou atrás de um todo sem vida. Exemplos: *headboard, foothills, tailgate e tail pipe*.

O quinto padrão é com os compostos formados por algum animal:

- animal + animal carnívoro = animal carnívoro que come/caça animal. Exemplos: *bee fly, bee moth, bee louse, bird dog, boarhound, buckhound, foxhound, wolfhound, elkhound, buffalo wolf, fox terrier, buffalo fly, mouse hawk*.

Quando o segundo substantivo é um cachorro, a idéia de que ele caça junto com um caçador humano é incluída. Há um outro padrão que concorre com esse, mas é um pouco diferente. Ele indica que há uma relação de semelhança entre os animais:

- animal<sub>1</sub> + animal<sub>2</sub> = animal<sub>2</sub> tem semelhança com animal<sub>1</sub>. Exemplos: *tiger shark, kangaroo rat, mule deer, elephant seal, catbird, dogfish, spider monkey, mole rat, spider wasp*.

Ao analisar os seus dados, Ryder (1994) dividiu a localização em três: *Location In*, *Location On* e *Location By*. Essa divisão se tornou necessária por dois motivos relacionados ao objetivo da sua pesquisa: a possibilidade de encontrar ou criar um esquema semântico sem que a acomodação seja necessária para a sua interpretação depende do tipo de relação de localização usada; a relação de localização mais comum é *Location In*, mas com os compostos formados por *sea* e *beach* é diferente. Os compostos com a *core word beach* apresentam as relações *Location In* e *Location By*. Já nos compostos com *sea*, foram encontradas *Location In* e *Location On*.

Ryder (1994) realizou experimentos que levam em consideração como diferentes sujeitos compreendem o significado de compostos NN. O principal objetivo é testar previsões quanto às relações entre *templates* linguísticos e esquemas de informação semântica. Para isso, ela criou compostos por meio de uma seleção aleatória de alguns substantivos referentes a itens concretos.

Ao criar os compostos, Ryder (1994) utilizou os *templates* identificados no seu *corpus* de estudo. Além dos casos citados anteriormente, outros *templates* menos frequentes também serviram de base para os seus experimentos. Um deles é o seguinte:

- cachorro + produto = produto usado por/no cachorro. Exemplos: *dog collar, dog chow, dog brush, dog biscuit, dog dish, doghouse, dog bed*.

A partir desse *template*, pode-se chegar a um mais geral:

- animal doméstico + produto = produto usado por/no animal doméstico. Exemplos: *cat collar, birdcage, hamster wheel, horse trough, cow pasture, mouse trap, rat poison*.

O *template* relacionado com partes do corpo também recebeu maior especificação:

- parte do corpo + produto = produto usado em conexão com parte do corpo. Exemplos: *elbow brace, face cream, toothbrush, footstool, ankle supporter*.

Há ainda um *template* menos específico relacionado com o produto:

- X + Produto = Produto usado em X. Neste caso específico, X refere-se ao primeiro substantivo. Exemplos: *baby powder, floor vacuum, laundry detergent, house paint, pet shampoo*.

Ryder (1994) identificou também outro *template*, que relaciona uma planta com o que ela produz:

- produto + planta = planta que produz o produto. Exemplos: *apple tree, cranberry bush, tomato plant*.

Em alguns casos, houve uma acomodação do núcleo como uma parte do padrão do *template*, como em:



- substância que pode ser moldada + entidade com forma = a entidade feita da substância que pode ser moldada. Exemplos: *clay pigeon*, *rubber duck*, *plastic gun*.

De uma forma geral, o que Ryder (1994) concluiu foi que alguns tipos de composição são mais previsíveis, pois seguem, de alguma forma, uma regra. A previsibilidade a que ela se refere envolve o conhecimento de mundo que o falante possui. O modelo sugerido por ela considera que os protótipos fazem parte da gramática. Assim, os *templates* linguísticos servem de base para a criação de novos compostos. Como esse modelo considera como o usuário da língua interpreta o composto, é a partir do seu conhecimento de mundo sobre os dois elementos de um composto que ele identifica um esquema comum aos dois para interpretá-lo. Quando não consegue achar uma situação em que os dois substantivos possam se relacionar, o falante usa outras estratégias, como a acomodação.

Para esquematizar as categorias dos compostos, Ryder (1994) sugere 48 etiquetas, representadas por meio de um verbo que relaciona os dois substantivos do composto. Na tabela 4, a seguir, cada uma dessas relações é apresentada com o seu esquema correspondente. A primeira etiqueta, BRING (trazer), representa o esquema em que o N2 entrega, traz ou distribui o N1.

**TABELA 4:** Esquemas sugeridos por Ryder (1994)

Etiqueta	Esquema	Etiqueta	Esquema
BRING	N2 traz/entrega/distribui N1	LIKSO	Semelhança em som
CON	N2 contém N1	LIKTE	Semelhança em textura
CONIN	N2 é contido em N1	LOCIN	N2 é localizado dentro de N1
EAT	N2 come/bebe/fuma N1	LOCBY	N2 é localizado na proximidade imediata de N1
EATBY	N2 é consumido por N1 (comido/bebido/fumado)	LOCON	N2 é localizado na superfície de N1
EQUB	Algo que é N2 e N1 ao mesmo tempo ou uma mistura de N2 e N1.	LOCOF	N2 é a localização de N1
EQUF	N2 que funciona como um N1	LOCUN	N2 é localizado abaixo de N1
EQUM	Um N2 metafórico que é um N1	MADBY	N2 é feito por N1
FIX	N2 arruma/conserta N1	MADOF	N2 é feito de N1
FOND	N2 gosta de N1	MAK	N2 faz N1
GOTO	N2 estende-se até o N1	MEM	N2 é um membro de N1
HADBY	N2 é possuído por N1	MSC	a relação entre N2 e N1 não se encaixa em nenhuma outra categoria
HAS	N2 possui N1 como uma parte/componente	OPER	N2 é operado por N1
HASPT	N2 tem N1 como uma parte/componente	PTOF	N2 é uma parte de N1
HUNT	N2 caça/coleciona (animais vivos) N1	RIDE	N2 conduz/dirige N1
LIK	N2 é semelhante a N1	SELL	N2 vende/entrega N1
LIKAC	Semelhança em ações	SHOOT	N2 opera N1 (arma)
LIKCO	Semelhança em padrões visuais/de cor	TEND	N2 cuida/cria/treina N1

LIKFU	Semelhança em função	TENDBY	N2 é cuidado por N1
LIKPO	Semelhança em posição	USE	N2 usa N1 como uma ferramenta
LIKPR	Semelhança em personalidade ou habilidades mentais	USEBY	N2 é usado por N1 como uma ferramenta
LIKPT	Semelhança em que uma parte de N2 é semelhante a uma parte de N1 ou ao N1	USEON	N2 é usado em N1
LIKSH	Semelhança em forma	WEAR	N2 é usado em ou por N1
LIKSI	Semelhança em tamanho	WORKW	N2 trabalha com N1 (objeto inanimado)

Ao expor as categorias conforme mostrado na tabela 4, Ryder (1994) não traz as características dos substantivos que compõem o composto. No caso do segundo esquema, em que N2 contém N1, podemos imaginar que o N2 seja um recipiente e que o N1 é algo que pode estar contido em um recipiente, mas essa informação não está evidenciada. Há também um grande número de relações que se repetem. Não há necessidade de criar duas etiquetas diferentes para a relação CONTÉM, como N2 contém N1 e N1 é contido em N2. Basta apenas uma relação semântica, cuja direção, se o sujeito é o N1 ou o N2, é explicitada.

Os *templates* de Ryder (1994) também podem ser chamados de *frames*, por isso foi feita a relação com o *FrameNet* neste trabalho. Fillmore (2006, p. 373) descreve a semântica de *frames* como “um programa de pesquisa em semântica empírica e um modelo descritivo para apresentar os resultados de tal pesquisa”. A semântica de *frames* é empírica, pois a língua é descrita a partir da experiência, do seu uso. A descrição, assim como no trabalho de Ryder (1994), é feita a partir da valência semântica, ou seja, a descrição do papel semântico e dos seus argumentos.

Segundo Fillmore (2006), um *frame* é um sistema de conceitos relacionados de tal modo que, para entendê-lo, é necessário entender a estrutura toda na qual ele se encaixa. Por exemplo, ao pensarmos em uma compra, vários elementos de uma situação como essa são automaticamente ativados, como preço, comprador, produto, vendedor, entre outros. Todos esses elementos fazem parte de um conceito maior que engloba todos esses participantes, que é o *frame* transação comercial. Pode-se dizer assim que os *frames* envolvem conhecimento geral, compartilhado por toda uma comunidade de fala.

O evento em que alguém compra algo evoca o *frame* transação comercial, porém um ponto de vista é privilegiado, o do comprador. Isso é um aspecto recorrente nos *frames*, o que permite diferentes perspectivas. De qualquer forma, todos os outros elementos que fazem parte do *frame* são trazidos em questão quando se cita apenas um dos elementos do *frame*.

Resumindo esses conceitos, podemos dizer que os eventos são organizados por meio de diversas situações nas quais eles podem ocorrer, constituindo diferentes domínios

semânticos, como comunicação, tempo, espaço, corpo, emoção, entre outros. Esses domínios são descritos a partir de *frames* por Fillmore (2006), ou seja, a partir de situações bem específicas que envolvem diversos elementos participantes.

Os elementos que fazem parte dos *frames* são semelhantes aos papéis temáticos (agente, ator, paciente etc.), porém são mais específicos de acordo com o *frame* evocado, como comprador, vendedor, coisa comprada, preço etc. Uma palavra específica pode evocar um *frame*, como, por exemplo, o verbo “comprar”. A partir desse elemento lexical, chega-se ao *frame* transação comercial com todos os seus elementos participantes envolvidos.

O projeto *FrameNet* desenvolvido pelo *International Computer Science Institute (ICSI)*, de Berkeley, tem como propósito a construção de uma base de dados organizada a partir da teoria de *frames* de Fillmore (2006). As principais características desse projeto são o uso de evidência empírica, de *corpus*, para a elaboração de generalizações sintáticas e semânticas; a utilização da semântica de *frames* para a representação das valências de palavras.

A base de dados criada pelo projeto está disponível on-line e os dados são organizados por meio de três componentes, conforme dados de Baker, Fillmore e Lowe (1998):

- a) o léxico, com uma lista de itens que são relacionados aos elementos *frames* e a frases de exemplo extraídas de um *corpus*;
- b) a base de dados em si, que traz descrições sobre a estrutura de cada *frame*, informando os seus elementos *frame* (participantes) e alguns itens lexicais relacionados;
- c) frases anotadas semanticamente extraídas de um *corpus* e que exemplificam as propriedades dos itens lexicais.

Para exemplificar como esse recurso pode ser utilizado, apresenta-se brevemente como é possível chegar a um *frame* e quais as informações encontradas na base de dados. Ao se fazer uma busca pela palavra *field*, chega-se ao *frame Locale\_by\_Event* (localização por evento), que é descrito como “um local definido em termos de um evento que ocorreu ou ocorrerá lá”. Entre os seus elementos *frames*, tem-se: evento definidor, local, partes constituintes, descritor, nome e localização relativa. Cada elemento aparece com uma cor diferente que é utilizada na anotação dos exemplos, conforme mostra a figura 1.

Na primeira frase de exemplo da figura 1, *Woodstock, NY was the SITE of the most important rock concert of 1969* (Woodstock, Nova Iorque foi o local do concerto mais importante de 1969), *Woodstock* é o local, e *the most important rock concert of 1969* é o

definidor do evento. Mesmo que o projeto não objetive identificar relações semânticas entre elementos de estruturas complexas, como os compostos, pode-se utilizar essa base de dados para caracterizar os substantivos que fazem parte de um composto. Outro aspecto que torna possível a relação desse recurso com o trabalho de Ryder (1994) é que a relação entre as unidades lexicais no *FrameNet* é feita por meio de verbos, assim como as etiquetas sugeridas por Ryder (1994). Um composto como *baseball field* pode ser descrito como um local onde um evento específico ocorre, o jogo de *baseball*.

A consulta aos dados do *FrameNet* contribuiu para uma maior compreensão sobre a relação semântica dos compostos. A partir da busca feita por *field*, apresentada na figura 1, é possível saber quais os seus argumentos, ou seja, quais as características dos itens lexicais que podem se combinar com essa palavra formando um composto.

## Locale\_by\_event

**Definition:** A place defined in terms of an event which occurred (or will occur) there.

Woodstock, NY was the SITE of the most important rock concert of 1969.

Our meeting SITE was chosen for its solitude.

In the second example, the event FE is labeled with two labels, because the two words do not form a syntactic constituent; the first would be GF Genitive PT Possessive, and the second, GF Dependent and PT NP.

**FEs: Core:**

Defining\_event [DefEvt] The event for which the locale is known or is named  
The driver died at the SCENE of the accident.

Locale [Lcl] A stable bounded area. It is typically the denotation of the nouns  
Semantic Type Location in Locale\_by\_event, which will be annotated on the target word.

We need to find a suitable VENUE for such a large wedding.  
Tours was the SITE of the defeat of the Turks by the French.

**Non-Core:**

Constituent\_parts [Cnst] Salient parts that make up a Locale.  
We need a meeting SITE with more computers.

Descriptor [des] The Descriptor is a temporary condition of the Locale.  
The next day, the governor visited the still-smoldering SITE of the fire.

Name [Name] The name of the locale.  
The Bonneville SITE has been a favorite of drivers for many years.

Relative\_location [re] A place that a Locale is located with respect to.

**Semantic Type**

Locative\_relation

The SITE of their first meeting, west of the Mississippi, has never been disclosed.

**Lexical Units**

*battlefield.n, field.n, scene.n, site.n, theater\_((of\_war)).n, venue.n*

Figura 1: *Frame* Localização\_por\_evento da base de dados do *FrameNet*.

#### 2.1.4 Teoria do Léxico Gerativo

A Teoria do Léxico Gerativo, de Pustejovsky (1995), propõe níveis de representação para as entradas semânticas lexicais. Para o estudo dos nominais, podem-se utilizar os elementos da estrutura *qualia*, que são os seguintes: formal, constitutivo, télico e agentivo.

O diferencial dessa teoria é o fato de que ela foi desenvolvida com o objetivo de servir aos interesses da computação, propondo uma forma lógica de representação do conhecimento com fins de auxiliar na tarefa de PLN. Segundo Chishman (2002), essa abordagem considera a geratividade do léxico em vez de tentar descrevê-lo de forma enumerativa, como ocorre em um método mais tradicional. O léxico, para Pustejovsky (1995), não é considerado um conjunto estático de palavras, já que a língua é usada de forma criativa, gerando novos significados para as palavras de acordo com os seus contextos. No entanto, segundo o autor, é possível capturar várias regularidades linguísticas, formando um léxico reduzido com as possíveis relações lógicas entre os sentidos de cada item lexical.

No artigo *The Generative Lexicon*, Pustejovsky (1991) apresenta alguns princípios básicos da sua teoria, que são resumidos a seguir:

- o estudo da semântica lexical só é possível se parte-se da apreciação da estrutura sintática de uma língua. O significado da palavra depende da estrutura na qual ela está inserida;
- os significados das palavras refletem as estruturas conceituais do sistema, ou seja, a semântica da língua natural deve ser a imagem dos princípios conceituais não-linguísticos;
- a semântica da boa formação é necessária para caracterizar uma teoria do significado, assim sendo, os fatores pragmáticos e o discurso devem ser tratados separadamente das contribuições semânticas dos itens lexicais na composição;
- uma representação como a teoria dos papéis temáticos não é suficiente para fornecer uma interpretação semântica de uma sentença. O autor sugere o método da decomposição lexical, que constitui na divisão mínima dos itens lexicais em formas estruturadas e não mais em conjuntos de características;
- o léxico não é formado apenas por verbos, embora se tenham realizado diversos trabalhos acerca da semântica verbal e muitas questões ainda estejam em aberto.

Como essa teoria aborda diversas questões linguísticas, alguns pontos mais relevantes para a aplicação deste estudo serão comentados ao longo deste trabalho: a representação em níveis, a estrutura *qualia* e a co-composição.

Os quatro níveis de representação semântica são a estrutura argumental, a estrutura de eventos, a estrutura *qualia* e a estrutura de herança. Uma estrutura de argumento especifica o número e o tipo de argumentos para uma palavra, tomada como predicado. É por meio desse mapeamento que as relações com as expressões sintáticas são estabelecidas. A estrutura de eventos, por meio de uma análise subeventual, classifica os verbos em estados, processos e transições. Um verbo como *construir* possui subeventos, em que no momento da construção, tem-se um processo e quando a construção está concluída, tem-se um estado. Enquanto a estrutura *qualia* é uma representação dos aspectos essenciais do significado de uma palavra, a estrutura de herança relaciona a palavra com outros conceitos do léxico, contribuindo para a sua organização.

Conforme já mencionado anteriormente, o foco deste trabalho é na estrutura *qualia*, pois esse nível é voltado para a semântica dos nominais, principal interesse deste trabalho. A estrutura *qualia* é composta por quatro papéis, que constituem os aspectos do significado de uma palavra, quais sejam:

- papel Constitutivo: a relação entre um objeto e suas partes. Exemplo: mão como parte do corpo;
- papel Formal: é o que distingue o objeto dentro de um domínio maior, trazendo os seus atributos físicos. Exemplo: um dos significados de jornal tem como papel formal o seu hiperônimo, ou seja, um objeto físico;
- papel Télico: propósito ou função do objeto. Exemplo: o papel télico de refrigerante é beber, ou seja, é para isso que ele serve;
- papel Agentivo: fatores envolvidos no surgimento ou na criação de um objeto. Exemplo: um bolo só surgiu porque ele foi assado, então o verbo assar é o seu papel agentivo.

Ao se referir a um objeto, pode-se também explicar como este surgiu, para que serve, ou seja, indicar os papéis que compõem a estrutura *qualia*. Uma palavra como *livro* pode ser explicada por meio do seu papel télico, sua função, ou seja, o livro serve para ler. No entanto, se mudarmos o ponto de vista e pensarmos no seu papel agentivo, como ele surgiu, pode-se

afirmar que, para o livro existir, alguém precisou escrevê-lo. O papel constitutivo também se aplicaria para esse exemplo, pois o livro possui capa, páginas, sumário etc. Já quanto ao seu papel formal, pode-se dizer que o livro é um tipo de objeto físico. Decompondo-se as palavras dessa forma, a descrição lógica torna-se mais detalhada, possibilitando a relação entre os itens lexicais de tal forma que possam ser criados novos significados de acordo com as combinações lexicais, com o seu uso. Isso se dá basicamente por meio do mecanismo gerativo chamado de co-composição. Para Pustejovsky (1991), o significado de uma sentença é determinado não só a partir aplicação do verbo sobre o argumento, mas também pela aplicação da função do argumento sobre o verbo. Nos compostos também é possível aplicar esse tipo de interpretação bidirecional. Em um composto como *beer drinker*, podemos interpretar que alguém bebe cerveja, mas também é importante ressaltar que a palavra *beer* possui características que permitem que ela ocupe essa posição. Cerveja é um líquido que serve de bebida para as pessoas. Da mesma forma, alguém só pode beber algo que seja líquido.

Busa e Johnston (1999) propõem uma interpretação para os compostos nominais nas línguas inglesa e italiana a partir da estrutura *qualia*. Descreve-se aqui como os autores utilizaram cada um dos papéis para representar a semântica dos compostos. O aspecto formal remete à relação “é um”, em que o hiperônimo do composto será o seu núcleo. Vejamos a expressão: *bread knife*. Como o hiperônimo de *knife* é *artifact\_tool*, esse será o mesmo para *bread knife*.

Por modificação télica, entende-se o propósito de algo. Em *bread knife*, *bread* modifica telicamente *knife*, pois essa expressão pode ser definida como “uma faca usada para cortar pão”.

A modificação agentiva pode ser identificada em *bullet wound*. O elemento *bullet* especifica como o ferimento foi feito, exercendo o papel agentivo, já que o nome relaciona a origem do objeto, como ele surgiu. Um *bullet wound* é um ferimento que surgiu através do ato de disparar uma arma.

A relação “parte de” é expressa no papel constitutivo, em que o modificador especifica uma parte ou subparte do núcleo, como em *glass door*. Uma *porta de vidro* é uma porta feita de vidro, em que *vidro* expressa o material do qual esse objeto é feito. Para os autores (BUSA; JOHNSTON, 1999), o composto é interpretado como um hipônimo do seu núcleo; assim, pode-se dizer que uma porta de vidro é uma porta. Para interpretarmos o composto *porta de vidro*, basta representar a estrutura *qualia* de porta, preenchendo o papel constitutivo com vidro.



Para Pustejovsky, Bergler e Anick (1993), a partir da análise de padrões de ocorrência em um *corpus*, é possível prever sistematicamente a semântica lexical de um substantivo estabelecendo grupos de estruturas colocacionais<sup>11</sup>, que são estruturas *qualia* genéricas que capturam não somente a relação semântica entre os tipos de argumentos de uma relação, mas também, por meio de um *corpus*, as relações colocacionais que concretizam esses papéis. Assim, além de identificar padrões de ocorrência como substância + recipiente, é possível obter uma lista com estas combinações que poderá incluir compostos como: *ice bag*, *tea pot* e *wine bottle*. Tomemos como exemplos as seguintes frases apresentadas pelos autores:

- 1) *John painted the door.*
- 2) *John walked through the door.*
- 3) *John is scrubbing the fireplace.*
- 4) *The smoke filled the fireplace.*

A partir da observação dessas frases, é possível identificar padrões de usos para os verbos que ocorrem antes de *door* e *fireplace*. Os verbos *paint* e *scrub* são ações praticadas em objetos físicos, já *walk through* e *fill* são processos que ocorrem em lugares. Com a identificação desses padrões pode-se interpretar que, nos exemplos 1 e 3, o complemento do verbo é um objeto físico e, em 2 e 4, *door* refere-se à entrada e *fireplace* ao buraco da lareira.

Acredita-se que considerar os padrões de uso dos elementos de um composto pode contribuir para a compreensão da sua semântica, fazendo com que a tarefa não se concentre apenas no núcleo ou no modificador do composto, mas nos dois elementos.

No trabalho de Pustejovsky, Bergler e Anick (1993), o *corpus* de estudo é da área técnica, da informática. Para eles, nesse tipo de texto, os compostos geralmente expressam uma relação de subclasse em que o primeiro elemento é instância do segundo e o segundo elemento é a classe a que pertence o primeiro, como em *Unix operating system* e *C language*.

Os autores sugerem que, para a interpretação dos compostos, a evidência colocacional deve ser usada na tarefa de classificação das relações semânticas. Por meio de estudos de similaridade, a partir das ocorrências, é possível listar as possíveis classes de uma instância. O procedimento para que se possa chegar a relações semânticas mais profundas automaticamente, através de colocações, é feito da seguinte maneira:

---

<sup>11</sup> Colocação refere-se à recorrência de duas palavras juntas.

- (i) dada uma palavra P, todos os substantivos (N) núcleo de SNs que ocorrem logo após P são extraídos da base de dados. Obtém-se uma lista de possíveis candidatos à classe da instância. Com a busca pela palavra *backup*, chega-se às palavras *operation*, *device*, *disk*, *path message*;
- (ii) é gerada uma lista de verbos em que P é objeto direto e o valor de *Mutual Information (MI)*<sup>12</sup> é calculado. O mesmo é feito com cada N obtido na tarefa anterior. Essa etapa se justifica, pois se espera que tanto a instância como a classe co-ocorram com os mesmos verbos. Entre os verbos que ocorrem com a palavra *tape*, foram encontrados os seguintes: *unload*, *position*, *mount*, *initialize*, entre outros;
- (iii) a taxa de similaridade entre P e cada um dos Ns é calculada. O substantivo com a maior pontuação de similaridade é geralmente a classe. O substantivo com o valor de MI mais alto, no caso de *backup*, é *operation*. Desta forma, *backup* pode ser considerado um tipo de *operation*.

Reconhecer a classe a que pertence um composto é importante para sistemas de interpretação automática de compostos que são baseados em taxonomia. A partir de um *corpus* de treinamento em que se identificam padrões de uso de determinadas classes, é possível que um sistema reconheça e interprete compostos novos, não presentes no léxico de treinamento.

Cálculos como esse podem facilitar o trabalho de representação da estrutura semântica dos compostos, aprimorando a descrição da estrutura *qualia*. Ao pesquisar, nas ocorrências de um *corpus*, que verbos ocorrem com um determinado composto, podemos obter o papel télico. A partir de uma lista de candidatos, calcula-se a *MI*, obtendo como resultado os verbos que são mais fortemente associados ao composto ou a classe. Se houver mais de um verbo com um alto valor de *MI*, podem-se incluir os mais frequentes na estrutura *qualia*.

Copestake (2003) sugere que alguns compostos não-deverbais podem ser descritos a partir de papéis da estrutura *qualia*, sendo eles os compostos télicos e a categoria *Made-of*, que corresponderia ao papel constitutivo. A necessidade de uma categoria *Made-of* se justifica pela ambiguidade da preposição *of*. Ao representar o composto *polystyrene box* pela paráfrase *box of polystyrene*, há grande chance de ela ser interpretada com *box containing polystyrene*; dessa forma, ao se propor uma espécie de paráfrase verbal com a construção *made of*, a relação entre os dois substantivos ficaria mais clara.

---

<sup>12</sup> *Mutual Information*, segundo Church e Hanks (1990), é um cálculo a partir da probabilidade de duas palavras ocorrerem juntas.

Os compostos télicos identificados por Copestake (2003) são todos nominais, mas há algum tipo de relação verbal que poderia ser recuperada por meio de uma paráfrase verbal. Um composto como *relaxation class* poderia ser analisado como *class that teaches relaxation*; dessa forma, o verbo *teach* preencheria o papel télico. Muitos télicos podem ser parafrazeados pela preposição *for*, mas uma construção desse tipo não parece ser muito informativa.

O problema com os télicos é que geralmente não é possível recuperar um único verbo, embora o trabalho de Lapata e Lascarides (2003) sugira que algum tipo de extração verbal é possível. Para os autores, por meio da análise do contexto, é possível identificar algumas dependências sintáticas, entre elas o fato de que os compostos são precedidos ou sucedidos por verbos. O mesmo verbo que é utilizado com o composto pode ocorrer também com o núcleo do composto, e imagina-se que poderia ser utilizado em uma paráfrase verbal do composto. A partir de um composto como *hen house*, pode-se fazer uma busca em um *corpus* pelos verbos mais frequentes que ocorrem com o núcleo *house*. Por meio de uma busca simples no *British National Corpus*, foi possível identificar os seguintes verbos ocorrendo com a palavra *house*: retornar, ver, conter, vender, alugar, viver, ir, partir, encontrar, incendiar, parecer, etc. Se se tivesse acesso a um *corpus* maior e se pudessem elaborar cálculos de frequência, comparando-se não somente os verbos que ocorrem com o núcleo, mas também os que ocorrem com o modificador, poder-se-ia chegar a uma conclusão quanto ao melhor verbo que expressa a telicidade entre os elementos deste composto. Porém, a partir dos resultados obtidos, poder-se-ia intuitivamente descrever a *hen house* por meio do verbo viver, pois um galinheiro se trata de um lugar onde as galinhas vivem, ou também poder-se-ia afirmar que a função do galinheiro é manter as galinhas dentro dele.

A partir dos pressupostos teóricos expostos neste trabalho, propõe-se uma adaptação da abordagem semântica de Pustejovsky (1995), da teoria de *frames* (FILLMORE, 2006) e dos *templates* linguísticos (RYDER, 1994) para uma análise dos compostos nos dois idiomas em questão, em português e em inglês. Não parece possível utilizar uma única proposta de classificação para os compostos; dessa forma, busca-se, por meio de uma abordagem mista, a melhor representação dos compostos identificados neste estudo.

## 2.2 Perspectiva linguístico-computacional

Nesta seção, são apresentados trabalhos na área da computação que se ocupam também do estudo da semântica dos compostos. Os compostos nominais fazem parte de um

grupo de expressões chamadas de multivocabulares. Há diversos estudos que se ocupam em achar soluções para o processamento de uma expressão formada por mais de uma palavra. Um exemplo de como essas construções podem ser tratadas é exposto na subseção 2.2.1.

Nas subseções seguintes, são apresentadas outras aplicações da semântica dos compostos em estudos computacionais, como anotação semântica, aprimoramento de gramáticas e léxicos, tradução automática e multilinguagem.

### 2.2.1 Expressões multivocabulares

Um dos principais desafios para o PLN são as expressões multivocabulares, que são definidas como “interpretações idiossincráticas que ultrapassam os limites das palavras (ou espaços)” (Sag *et al.*, 2002, p. 2). Têm-se buscado soluções para que um sistema automático de processamento da língua possa identificar esses conjuntos de palavras como uma unidade semântica. Objetiva-se também que o significado dessas expressões possa ser identificado automaticamente ou semiautomaticamente. Os compostos nominais, foco desta pesquisa, são um tipo de expressão multivocabular. Assim, o objetivo desta subseção é apresentar de forma geral a problemática das expressões multivocabulares e situar os compostos nominais nesse contexto.

Sag *et al.* (2002) apresentam as principais características das expressões multivocabulares (doravante EMV) e as dividem em sintagmas lexicalizados (*lexicalized phrases*) e sintagmas institucionais (*institutionalized phrases*), conforme terminologia adaptada de Bauer (1983 *apud* SAG ET AL., 2002). Para se ter uma idéia da importância desta questão, recorra-se a Jackendoff (1997), que observa que as EMV são encontradas em mesma quantidade que as unidades formadas por uma só palavra em um léxico de um falante nativo. Dessa forma, para que um sistema possa analisar adequadamente a linguagem produzida por seres humanos, precisa-se encontrar um tratamento computacional adequado para essas expressões.

Os sintagmas lexicalizados apresentam sintaxe ou semântica parcialmente idiossincrática, podendo conter palavras que não ocorrem isoladamente. São subdivididos em expressões fixas, expressões semifixas e expressões flexíveis sintaticamente. Já os sintagmas institucionalizados são sintaticamente ou semanticamente composicionais, mas ocorrem com uma alta frequência em um contexto específico.

Os métodos mais utilizados para se lidar com as EMV são a análise linguística composicional e abordagem de palavras-com-espacos (*words-with-spaces*). Nenhuma das

duas soluções traz resultados totalmente satisfatórios, primeiramente porque nem todas as EMV são composicionais e pelo fato de que as palavras-com-espacos não aceitam nenhum tipo de variação morfológica ou interna. Mais adiante esses métodos são apresentados em relação aos compostos nominais.

As expressões fixas, conforme sua própria denominação, não aceitam nenhum tipo de variação ou modificação e não são composicionais. Entre estas temos: *by and large*, *in short*. Como elas sempre ocorrem nesse formato, uma representação por palavras-com-espacos é suficiente. Neste tipo de representação, sugere-se unidades formadas por mais de uma palavra em um léxico, unindo-as com espacos.

As expressões semifixas possuem níveis de composicionalidade e podem sofrer alguma variação lexical, incluindo nessa classe os compostos nominais. Os compostos *car park*, *attorney general* e *part of speech* são unidades que não sofrem variações sintáticas, mas flexionam em número. Não é possível incluir alguma outra palavra entre *car* e *park*, mas pode-se utilizar essa forma no plural, *car parks*. No caso dos compostos em que o núcleo é o elemento da direita, eles podem ser representados por palavras-com-espacos acrescentadas pela marca do plural, o “s”: [*car park*]/s. Se o núcleo do composto for o elemento da esquerda, essa solução não é possível, pois *attorney general* e *part of speech* não pluralizam no último elemento. Uma solução bastante trabalhosa seria listar as formas no singular e no plural de cada composto. Tratar os compostos nominais por meio de uma abordagem totalmente composicional, em que o significado do todo é obtido através da soma do significado das partes, segundo Sag *et al.* (2002), não é possível, pois eles possuem diferentes graus de composicionalidade e, se os elementos forem tratados separadamente, não se chega ao seu significado. Os autores sugerem que cada composto seja considerado uma unidade e incluído em uma lista, de forma que o elemento que pode ser flexionado seja indicado de alguma forma. No caso de *part of speech*, a informação de que *part* pode ser flexionada deve aparecer no léxico.

Quanto às expressões flexíveis sintaticamente, elas podem sofrer variações na ordem das palavras e incluem construções com verbos e partículas, expressões idiomáticas decomposicionais e verbos leves.

Entre os sintagmas institucionalizados ou convencionalizados, é possível identificar colocações e construções de NN, tal como *traffic light*. Essas construções são composicionais, mas a frequência de uso desses dois elementos juntos é muito mais alta do que qualquer outra lexicalização alternativa do mesmo conceito. Quando se quer referir ao semáforo, usa-se com

mais frequência o compostos que já é convencionalizado, em vez de uma oração relativa que explique no que este objeto constitui.

Um dos aspectos mais relevantes quanto aos estudos sobre as EMV é que há muito ainda a ser feito com o objetivo de representá-las da melhor forma possível. Quanto aos compostos nominais, estes são também um tópico de interesse para os estudos computacionais relacionados com tradução automática, recuperação de informação, anotação semântica, entre outros. Estudos sobre a semântica, como o realizado nesta dissertação, podem servir de contribuição para estas aplicações computacionais.

### 2.2.2 Anotação semântica

A partir de um texto em formato digital é possível realizar diversas tarefas relacionadas ao PLN. Podem-se fazer buscas por uma determinada palavra ou estrutura, como, por exemplo, as buscas em *sites* como o *Google* e o *Altavista*. No entanto, algumas informações linguísticas podem ser incluídas em textos, ou em um conjunto de textos, um *corpus*. Entre essas informações está a indicação da classe morfológica de cada palavra e a função sintática de um grupo de palavras. Na frase “Um grupo roubou um carro...”<sup>13</sup>, podemos informar que “grupo” é um substantivo e que “Um grupo” é o sujeito da oração. Para incluir esse tipo de informação, utilizam-se etiquetas como marcadores. Essas etiquetas são convenções ou siglas adotadas para a anotação de informação linguística, podendo ser de ordem morfológica ou sintática. Assim, pode-se estabelecer que a letra N (do inglês, *noun*) refere-se a substantivos e que Subj (do inglês, *subject*), a sujeito.

Atualmente há ferramentas que incluem essas informações automaticamente e com bons resultados. Dessa forma, ao referir-se à anotação morfossintática, compreende-se a inclusão de informações linguísticas feita por um programa em um texto ou *corpus* determinado. Tendo-se um *corpus* anotado ou etiquetado, é possível fazer buscas mais específicas e não apenas pelos itens lexicais. Se o objetivo for realizar um estudo sobre os verbos que são seguidos de gerúndio, as suas ocorrências podem ser extraídas por meio de uma busca pelas etiquetas correspondentes a verbos e a verbos no gerúndio, criando-se o que se chama de fórmula ou padrão.

As aplicações de um *corpus* anotado são as mais variadas possíveis. Há estudos na área de sumarização automática, recuperação de informação e tradução automática que o

---

<sup>13</sup> Frase do Corpus NILC/São Carlos, versão 8.0. Disponível em: <<http://www.linguateca.pt/ACDC/>>.

utilizam. Um dos grandes desafios tem sido a inclusão de anotação semântica, tais como papéis temáticos, relações semânticas ou taxonomias específicas, pois com isso os resultados das ferramentas computacionais podem ser muito mais precisos. Tem-se buscado a realização de anotação semântica de forma automática. Para isso geralmente realiza-se uma anotação manual de parte do *corpus*. Com base nesse *subcorpus* anotado, criam-se estatísticas que têm como base não só as etiquetas semânticas, como também as morfossintáticas, com o objetivo de “treinar” uma determinada ferramenta com estas informações para que ela possa aprender a etiquetar o restante do *corpus* automaticamente. Essa parte menor do *corpus* que é etiquetada manualmente é chamada de *corpus* de treinamento e os resultados da anotação semântica automática ainda não são tão satisfatórios como os de anotação morfológica ou sintática. Por isso, este tem sido o tema de muitos trabalhos na área da Linguística Computacional, tais como Ó Séaghdha (2007), Baldwin *et al.* (2003) e Girju *et al.* (2007).

Apresenta-se a seguir parte do trabalho de Ó Séaghdha (2007), que tem como objetivo utilizar um *corpus* de treinamento e a partir de métodos probabilísticos poder etiquetar automaticamente as relações semânticas entre os elementos de compostos de um *corpus*. Diferentemente de Ó Séaghdha (2007), não se propõe aqui a realizar um experimento desse tipo; o foco deste trabalho é a análise semântica manual dos compostos NN em si. De qualquer forma, é interessante apresentar e avaliar a tipologia adotada pelo autor nos seus experimentos. Assim, pode-se sugerir neste trabalho uma tipologia que vá ao encontro das reais necessidades dos profissionais que realizam tarefas ou desenvolvem ferramentas sobre o processamento da língua.

Entre os problemas com algumas tipologias dos compostos, Ó Séaghdha (2007) destaca a dificuldade tanto de humanos como de máquinas para identificar a sua ocorrência em texto. Quanto maior for o conjunto de relações semânticas, mais esparsas elas serão. Assim, é possível que algumas relações ocorram poucas vezes em um *corpus* inteiro.

Considerando essas questões, Ó Séaghdha (2007) afirma que a anotação semântica dos compostos deve ser baseada nos seguintes princípios teóricos:

- a) as categorias devem dar conta da maioria dos compostos possível;
- b) os limites entre as categorias devem ser claros e elas devem descrever conceitos coerentes;
- c) a distribuição das classes não pode ser esparsa e deve ser imparcial;
- d) os conceitos devem generalizar para outros fenômenos linguísticos;
- e) a anotação deve ser simples;

f) a informação fornecida pelas categorias deve ser útil.

O ponto inicial da classificação sugerida por Ó Séaghdha (2007) são as nove categorias de Levi (ver subsecção 2.1.1). No entanto, para incluir o conceito de eventos e participantes, o autor altera algumas dessas categorias. Por exemplo, a relação FOR é substituída por duas outras, INST (instrumento) e AGENT (agente). A seguir, apresenta-se uma tabela com as relações sugeridas por Ó Séaghdha (2007):

**TABELA 5:** Relações semânticas de Ó Séaghdha (2007)

Relação	Tradução	Exemplo <sup>14</sup>
BE	SER	<i>steel knife</i>
HAVE	TER	<i>street name</i>
IN	EM	<i>forest hut</i>
INST	INSTRUMENTO	<i>rice cooker</i>
ACTOR	ATOR	<i>honey bee</i>
ABOUT	SOBRE	<i>fairy tale</i>
REL	RELAÇÃO GERAL	<i>camera gear</i>
LEX	LEXICALIZADO	<i>home secretary</i>
UNKNOWN	DESCONHECIDO	<i>similarity crystal</i>
MISTAG	ERRO DE ETIQUETAÇÃO	<i>blazing fire</i>
NONCOMP	NÃO-COMPOSTO	<i>[real tennis] club</i>

O agente ou o ator e o instrumento podem ser ou não animados e conscientes, mas sempre serão participantes de algum evento. Por REL, compreendem-se as relações não específicas, em que se percebe que os dois elementos do composto são relacionados de alguma forma, mas não é possível explicar essa relação. Os lexicalizados são os semanticamente opacos e as etiquetas MISTAG e NONCOMP estão relacionadas com sequências de elementos que não constituem um composto NN.

Analisando-se essas etiquetas, é possível estabelecer uma relação com outras classificações sugeridas. A relação BE pode expressar três tipos de relações<sup>15</sup>: copulativa (*woman driver*), material (*plastic box*) ou de comparação (*angler fish*).

A etiqueta HAVE lembra o papel constitutivo de Pustejovsky (1995), já que é possível associá-la à relação parte e todo. No entanto, Ó Séaghdha (2008) identificou mais quatro possíveis interpretações: posse (*customer account*), condição física ou estado mental (*cat*

<sup>14</sup> Exemplos de Ó Séaghdha (2007, p. 74).

<sup>15</sup> Dados baseados em *Annotation Guidelines for Compound Noun Semantics*, material não publicado de Ó Séaghdha (2008), disponível em: <<http://www.cl.cam.ac.uk/~do242/guidelines.pdf>>.



*instinct*), uma propriedade (*water volume*) e um conjunto de entidades (*stamp collection*). A relação IN está relacionada com a localização no espaço e no tempo. O ACTOR é um participante animado e consciente, enquanto que o INST é um participante que não possui essas características. Quanto aos compostos REL e UNKNOWN, a principal diferença é que nos primeiros é possível compreender os compostos, embora não se consiga explicar a relação entre os elementos; já a categoria UNKNOWN traz compostos em que a sua interpretação não é possível, mesmo com o auxílio do contexto. Os compostos que recebem a etiqueta LEX por Ó Séaghdha (2007) são aqueles que não se enquadram em nenhuma das outras relações sugeridas e não são produtivos, ou seja, o seu significado restringe-se a uma combinação única de dois elementos. Os lexicalizados não são posicionais. A relação entre os substantivos descrita através da etiqueta ABOUT é utilizada em algum tipo de descrição. Os compostos *fairy tale*, *science lesson*, *crime investigation* podem ser parafraseados com ABOUT, *tale about a fairy*, *lesson about science*, *investigation about a crime*, e de certa forma retratam o composto como um hipônimo do seu núcleo. Uma aula de ciências não deixa de ser um tipo de aula.

No seu estudo, Ó Séaghdha (2007, p. 75) extraiu as sequências de dois substantivos do *British National Corpus* por meio da seguinte heurística: “todas as sequências de dois substantivos simples foram extraídas com exceção daquelas adjacentes a outro substantivo e aquelas que contêm caracteres não alfabéticos”. Neste trabalho, adota-se também uma heurística semelhante para a extração dos candidatos a compostos. A partir dos resultados da extração, Ó Séaghdha (2007) obteve uma cobertura de 92% dos compostos válidos, conforme sua classificação manual das seis primeiras relações mais específicas apresentadas na tabela 5. Decidiu-se deixar de fora as relações que não pudessem receber uma interpretação mais específica.

O principal diferencial do trabalho de Ó Séaghdha (2007) foi o fato de que, durante a anotação manual do *corpus* de treinamento, os anotadores humanos tiveram acesso às frases em que os compostos foram utilizados e não simplesmente ao composto de forma isolada. Com uma interpretação em contexto o anotador se aproxima do uso real da língua.

Para a tarefa de anotação semântica automática, foram selecionados 2 mil compostos, dos quais 500 foram anotados manualmente e serviram de *corpus* de treinamento. O restante do *corpus* foi anotado automaticamente por meio do método chamado de *Support Vector Machine*<sup>16</sup>. A média de precisão da anotação semântica automática das seis relações a partir

---

<sup>16</sup> Para maiores informações sobre este método, ver Vapnik (1998).

da ocorrência do modificador e do núcleo de um composto foi de 56,76%. Para ampliar esse resultado, Ó Séaghdha (2007) utilizou como referência uma base de dados lexical, a *WordNet*<sup>17</sup>. A motivação para utilizar esse recurso está relacionada ao fato de que os dados do *WordNet* estão organizados a partir de relações semânticas. Dessa forma, é possível fazer uma busca pelos hipônimos de uma palavra. Como por meio dos hipônimos chega-se a alguns compostos nominais, acreditou-se que incluir essas informações, representadas através de vetores, pudesse aumentar a precisão do método de anotação. Por exemplo, entre os hipônimos de *knife*, tem-se *paper knife*, *bread knife*, *butcher knife*, *pocket knife*, entre outros. De fato, com a utilização dos dados da *WordNet*, Ó Séaghdha (2007) conseguiu aumentar a média de precisão do seu sistema para 58,35%.

Esse trabalho relatado serve de exemplo de como um estudo sobre as relações semânticas dos compostos NN pode ser utilizado para a anotação de *corpus*.

### 2.2.3 Aprimoramento de gramáticas e de léxicos

Para a área da Linguística Computacional, uma gramática é considerada um conjunto de regras sobre uma língua, ou seja, generalizações que restringem as possibilidades de uso de determinadas categorias ou elementos sintáticos.

Essas regras são representadas por meio de expressões lógicas de forma que um programa computacional possa analisar ou produzir textos automaticamente. Por exemplo, para que um sistema identifique em uma frase um sintagma nominal, é necessário incluir de alguma forma, nesse programa, quais os elementos que podem fazer parte de um sintagma nominal. Se incluirmos algumas categorias gramaticais, as classes fechadas são facilmente identificadas por um programa, pois basta listar os seus elementos, como os artigos definidos e indefinidos, por exemplo. No entanto, as classes abertas não são listadas facilmente. Para que um programa identifique o sintagma nominal de uma frase, além de saber quais as classes ou seqüências de elementos podem fazer parte do SN, é necessário associar a palavra *um* com o artigo indefinido, ou a palavra *menino* com substantivo. Para isso, é necessário criar um léxico, que é constituído por uma lista de palavras com alguma informação gramatical. O léxico pode ser construído manualmente ou semiautomaticamente por meio de recursos computacionais como dicionários eletrônicos ou *corpora*.

---

<sup>17</sup> Disponível em: <<http://wordnet.princeton.edu/>>.

A gramática e o léxico são utilizados em conjunto e há trabalhos que se ocupam somente da criação e do aprimoramento desses recursos. Alimentar sistemas como esses com regras quanto às expressões multivocabulares, em especial quanto aos compostos, ou incluí-las em um léxico ainda representa um desafio. Um trabalho que tem essa preocupação é o de Copestake (2003). A pesquisadora utiliza uma abordagem composicional para os compostos e a sua integração em uma gramática da língua inglesa a partir de dados extraídos de *corpus*. Nesta subseção, apresenta-se a sugestão de análise semântica de Copestake (2003).

Com o objetivo de estabelecer dados sobre os compostos NN que integrarão uma gramática, Copestake (2003) propõe soluções diferentes para cada tipo de composto. Ela parte do princípio de que um composto NN é formado por dois substantivos separados por um espaço e sem outro substantivo antes ou depois. São identificados os principais tipos: deverbais, relacionais e lexicalizados. Nos compostos deverbais, como *birth spacing* e *rose fertiliser*, em que o N2 vem de um verbo, o N1 será um dos argumentos desse verbo, podendo ser o sujeito ou o objeto. Os relacionais são aqueles substantivos que exigem um complemento, geralmente sintagmas que iniciam com a preposição *of*. *Jazz fan* é relacional, pois alguém é fã de algo e o composto pode ser parafraseado em *fan of jazz*. Quanto aos lexicalizados tais como *home secretary* e *monkey business*, a autora sugere que eles sejam tratados separadamente, por meio da criação ou implementação de uma base de dados ou léxico.

Outros compostos descritos por Copestake (2003) são os hiperonímicos, os que podem ser expressos por meio da relação feito-de, os preposicionais, os télicos não-deverbais e os não-parafraseáveis. Por compostos hiperonímicos compreendem-se os endocêntricos (ver subseção 2.1.1), em que o composto é um hipônimo ou subclasse do N2, tais como *tuna fish* e *oak tree*. Para diferenciar a relação de material da relação parte-todo, estabeleceu-se a categoria dos compostos feitos-de, em que um dos substantivos denota o material do qual o composto é feito: *steel sword*. Quando é possível parafrasear um composto acrescentando ao substantivo um sintagma preposicional, Copestake (2003) os chama de compostos preposicionais. Um exemplo desse caso seria *airshow accident*, que pode ser parafraseado por *accident at (an) airshow*. Os compostos cuja relação é parte-todo são considerados relacionais, pois a paráfrase de *car engine* pode ser *engine of a car*. Os compostos télicos são baseados na estrutura *qualia* de Pustejovsky (1995, ver subseção 2.1.5), em que o papel télico indica a função de algo. Embora esses compostos não sejam deverbais, é possível utilizar um verbo para parafraseá-los. Uma *relaxation class* poderia ser explicada por meio do verbo

*teach, a class that teaches relaxation*. Os compostos não-parafraseáveis são os que dependem do contexto para serem entendidos, como no exemplo de Downing (1977), *apple-juice seat*.

No seu trabalho, já que Copestake (2003) busca integrar os seus dados sobre os compostos a uma gramática geral da língua, os dados analisados também foram extraídos de um *corpus* geral, o *British National Corpus*, que contém 100 milhões de palavras. Por *corpus* geral entende-se um conjunto de textos de gêneros e assuntos variados. A partir dessa análise, a autora concluiu que os compostos hiperonímicos não são frequentes em *corpora* gerais, ocorrendo mais em *corpora* especializados, com textos de áreas específicas.

A metodologia adotada para um experimento em que o formalismo utilizado por Copestake (2003) para representar a semântica dos compostos é testado segue descrita resumidamente nesta subseção. Inicialmente coleta-se uma amostra aleatória de possíveis compostos NN de acordo com a sua frequência de ocorrência no *corpus*. São considerados três grupos de frequência e, para cada grupo, 20 compostos são coletados: frequência maior do que 500, frequência igual a cinco e os casos de *hapaxes*, compostos que ocorreram apenas uma vez. As formas no singular e no plural de um mesmo composto são agrupadas para a contagem. Logo após os compostos são conferidos manualmente e as sequências que não são compostos são descartadas.

Com esses dados, criam-se paráfrases com preposições para os compostos. Esse tratamento é defendido porque, em especial, a tradução automática tiraria proveito desse recurso. Algumas línguas como o português, por exemplo, utilizam uma preposição para relacionar dois substantivos, diferentemente do inglês, que faz essa relação de forma direta. Os compostos que são parafraseados por meio de uma preposição aceitam essa estrutura por serem composicionais, fato que denota a compatibilidade da abordagem com a gramática a ser integrada, que também é composicional. Há casos especiais que requerem um outro tratamento, conforme visto anteriormente. A seguir, é listado o tratamento dado a cada tipo de composto por Copestake (2003):

- a) os compostos lexicalizados ou em que o núcleo não é o N2 são listados em um léxico;
- b) os compostos deverbais e relativos são relacionados com os seus argumentos e por isso são chamados de nominais predicativos;
- c) os compostos feitos-de, preposicionais e télicos são parafraseados;
- d) aqueles compostos que não podem ser parafraseados, pois a sua interpretação depende do contexto, da pragmática, recebem a indicação de que há uma relação entre os dois

substantivos, que não é especificada. Somente com uma análise posterior do seu contexto é que essa relação pode ser explicada.

Conforme já mencionado anteriormente, foi utilizado um formalismo para representar a semântica desses compostos. O formalismo escolhido foi a Semântica de Recursividade Mínima (*Minimal Recursion Semantics, MRS*) de Copestake *et al.* (1999), pois este é compatível com a gramática a ser aprimorada, a saber, a *English Resource Grammar*<sup>18</sup> (ERG). Na ERG, os dois substantivos de um composto NN são relacionados por meio de uma relação não-especificada. Para informar que relação é esta, Copestake (2003) altera o seu modelo criando uma MRS robusta. Assim, a paráfrase de *airshow accident* é representada da seguinte forma:

$$airshow(x) \hat{U} at(y, x) \hat{U} accident(y)$$

Nessa representação, o N1 é o elemento x, o N2 é o elemento y, e a preposição une o y ao x, pois temos *accident at airshow*. O símbolo  $\wedge$  refere-se à conjunção e. Para os compostos predicativos como *jazz fan*, a autora sugere:

$$jazz(x) \hat{U} fan(y, x)$$

A relação entre os dois substantivos é expressa no elemento *fan*, pois este é o y e o seu complemento é o x, jazz.

Considerando as categorias de compostos sugeridos por Copestake (2003), é possível identificar alguns padrões de composição de acordo com a relação entre os elementos. Na relação feito-de, o N1 será uma substância e o N2, um objeto físico, tal como *cardboard box*. Outra informação sobre a relação feito-de, segundo Copestake e Lascarides (1997), é que ela não ocorre em compostos sobre líquidos ou artefatos. Já os compostos télicos terão um N1 que pode ser qualquer coisa e o N2 é um artefato. Exemplos: *ice-cream container* e *linen chest*.

Além dos casos apresentados nesta subseção, Copestake e Lascarides (1997) destacam um tipo específico de composto NN, os genitivos. Compostos como *blacksmith's hammer* e *driver's cab* são genitivos, e esse caso é marcado por meio do acréscimo da apóstrofe, do

---

<sup>18</sup> Disponível em <<http://lingo.stanford.edu/>>.

possessivo. Este é um tipo de composto possível quando o N1 for um ser humano. Copestake (2003) não explicita como um composto possessivo seria representado na gramática, mas imagina-se que ele possa ser parafraseado com a preposição *of*.

Avaliando o trabalho de Copestake (2003), percebe-se que ela utilizou um formalismo adequado ao seu objetivo, a implementação da gramática. Por meio de uma linha, ou expressão, é possível explicitar a relação semântica entre os elementos de um composto. O caso mais problemático, segundo a autora, é estabelecer uma categoria específica para os compostos télicos, pois não é possível identificar um único verbo que relacione os dois substantivos. Para resolver esse problema, Lapata e Lascarides (2003) sugerem uma interpretação automática para a metonímia lógica. Nesse tipo de metonímia o argumento sintático de uma construção é diferente do seu argumento lógico. Por exemplo, em *finish the cigarette*, o argumento do verbo “terminar” é cigarro, mas dizer que alguém terminou o cigarro implica que alguém terminou de fumar o cigarro. Para identificar o verbo que fornece a interpretação metonímica da construção, os autores utilizaram um *corpus* grande e anotado parcialmente. Por meio de cálculos probabilísticos, foi possível obter uma lista ordenada de verbos a partir dos significados possíveis preferidos.

Quanto às categorias sugeridas por Copestake (2003), é forte a influência da estrutura *qualia* de Pustejovsky (1995). Além dos compostos télicos, explicitamente considerados a partir da estrutura *qualia*, é possível identificar outros papéis, como o formal nos compostos hiperonímicos e o constitutivo nos compostos feitos-de. Destaca-se, aqui, que a relação *made-of* refere-se apenas ao material e que a relação de *parte/todo* seria expressa por meio de paráfrases preposicionais. Houve aqui uma preocupação em distinguir quando um objeto possui partes distintas e quando este é feito totalmente de uma determinada substância. Vale lembrar que essa preocupação em diferenciar esses dois casos também está presente nos primeiros estudos sobre a semântica dos compostos (ver subseção 2.1.2).

#### 2.2.4 Tradução automática

Neste trabalho, a semântica dos compostos nominais é analisada tendo-se em vista algum padrão de tradução do inglês para a língua portuguesa. Busca-se contrastar os usos dos compostos nas duas línguas, estabelecendo-se algumas generalizações quanto às relações semânticas entre os elementos de um composto e a sua estrutura.

Para realizar esta tarefa, que é apresentada no capítulo de análise deste trabalho, parte-se de um conjunto de textos composto por originais e traduções manuais, disponibilizados em

formato eletrônico. Salienta-se o fato de esses textos serem traduzidos por humanos, pois há também textos traduzidos por máquina, assunto que será apresentado brevemente a seguir.

Como parte deste estudo concentra-se na tradução dos compostos do inglês para o português, é importante comentar alguns aspectos gerais quanto ao tema. Ao observar um texto traduzido, percebe-se que o tradutor humano possui uma certa liberdade, pois ele pode reorganizar as frases, excluir informações que não sejam importantes ou até mesmo incluir informações necessárias para a compreensão do leitor, principalmente se forem culturas diferentes. Quanto aos compostos nominais formados por dois substantivos, foco deste trabalho, eles podem ser traduzidos de diversas formas. Geralmente encontram-se, nos textos traduzidos, compostos do tipo N de N, mas também se pode encontrar um adjetivo e um substantivo, ou o tradutor pode simplesmente traduzir o composto com um único substantivo ou excluí-lo da tradução quando o referente já tiver sido mencionado e puder ser retomado pelo leitor.

Segundo Costa (2005 *apud* AZEREDO, 2007), o tradutor é limitado pelo texto original, já que ele precisa reescrevê-lo com a preocupação de manter o mesmo tom e conteúdo, mesmo que não concorde com as ideias expressadas. Uma das questões mais importantes para os estudos de tradução é a noção de equivalência, que, conforme Costa (2005 *apud* AZEREDO, 2007), sugere uma visão preconceituosa quanto ao texto traduzido, pois leva à ideia de que o texto original é superior ao traduzido. Julgar a qualidade de uma tradução requer uma análise minuciosa, e deve-se levar em consideração o objetivo dessa tradução.

Azeredo (2007) afirma que o texto traduzido é dependente do original, constituindo um gênero textual que nunca será uma equivalência perfeita do original. No entanto, mesmo que não haja uma equivalência total, um determinado referente representado em uma língua por um nominal composto precisa ser traduzido de alguma forma, e é nessa representação que este trabalho se concentra.

Embora este trabalho não tenha um compromisso direto com a tarefa de tradução automática, espera-se que as conclusões obtidas com este estudo possam contribuir para tarefas de PLN, em especial, para a tradução realizada por máquinas. Dessa forma, será apresentada nesta seção uma breve descrição de alguns aspectos relacionados à tradução automática (*machine translation*).

Os estudos de tradução automática, também chamada de tradução mecânica, iniciaram na década de 1950, quando se esperava ingenuamente que fosse possível obter um texto traduzido de qualidade sem nenhuma intervenção humana. Hoje em dia, sabe-se que um

sistema de tradução tem suas limitações, não sendo possível a sua realização sem a intervenção de humanos. Dessa forma, fala-se em sistemas de tradução auxiliada por humanos (*human-aided machine translation*) e em tradução humana auxiliada por máquinas (*machine-aided human translation*).

Segundo Martins *et al.* (2004), os sistemas atuais não buscam uma equivalência da língua-fonte na língua-alvo, mas outras estratégias são adotadas. Entre elas, temos: (i) a redefinição da língua-fonte, que é reduzida a um subconjunto normalizado da variedade real; (ii) a descrição e o treinamento de uma ferramenta a partir de uma sublíngua, uma linguagem especializada; (iii) a noção de tradução crua (*raw translation*), que constitui a saída do sistema e o ponto de partida do trabalho do tradutor humano; (iv) a tradução automática como um mecanismo de triagem, formando uma tradução grosseira (*rough translation*) que requer uma revisão humana.

Entre as abordagens utilizadas pela tradução automática, Dorr *et al.* (1999 *apud* MARTINS *et al.*, 2004) apontam três: tradução baseada exclusivamente em conhecimento linguístico, tendo como base dicionários e gramáticas; tradução baseada em conhecimento, incluindo dicionários, gramáticas, enciclopédias e bases de conhecimento; e tradução baseada em exemplos, considerando as ocorrências em dicionários, gramáticas e *corpora*. Os resultados do presente trabalho, com certeza, poderão contribuir para os sistemas que adotam a terceira abordagem, pois há também a preocupação com análises de dados reais da língua, sem a preocupação em partir de regras oferecidas por uma gramática tradicional ou por definições presentes em dicionários.

Segundo Martins *et al.* (2004), são duas as principais técnicas: a tradução direta e a tradução indireta. Na tradução direta, a tradução é feita palavra-por-palavra, buscando-se em uma base de dados, como um dicionário bilíngue, um item lexical correspondente na língua-alvo. Esse tipo de técnica não produz resultados muito bons, em especial, para casos como os compostos nominais, pois a expressão não seria considerada uma unidade e cada palavra seria traduzida separadamente.

Na tradução indireta, busca-se uma representação intermediária entre a língua-fonte e a língua-alvo, representação esta que pode ser uma interface específica baseada em transferência ou uma interlíngua, uma língua artificial independente tanto da língua-fonte quanto da língua-alvo. A tradução por transferência envolve a criação de uma metalíngua que constitui, na maioria das vezes, em uma tabela de correspondências sintático-semânticas entre as duas línguas. Ao buscar, de alguma forma, estabelecer como se dá a tradução dos



compostos do inglês para o português, imagina-se que os resultados da análise proposta neste trabalho poderiam contribuir para uma técnica baseada em transferência.

A utilização de interlíngua para a tradução parece ser a solução mais adequada para o processamento automático, pois, além de ser livre de casos de ambiguidade, uma interlíngua se presta como ponto de processamento intermediário entre quaisquer línguas, e não de uma língua-fonte e uma língua-alvo específicas como no caso de uma metalíngua. Para o tratamento dos compostos, a representação da estrutura *qualia* poderia servir como um tipo de interlíngua, conforme McDonald (1995). Embora seja mais adequada para sistemas multilíngues, a aplicação de uma abordagem interlingual ainda não é efetiva, pois ela deveria conter todas as línguas existentes, algo como uma gramática universal que fosse subjacente a todas as línguas naturais.

No que diz respeito aos sistemas de tradução para a língua portuguesa, a situação é muito desanimadora. Há poucos sistemas e os que existem são baseados em transferência basicamente sintática e não produzem resultados satisfatórios ao processarem construções e léxico mais complexos, pois geralmente a tradução é feita palavra-por-palavra. Com as conclusões obtidas neste estudo, espera-se contribuir para o aprimoramento desse quadro dos tradutores automáticos para a língua portuguesa.

### 2.2.5 Multilinguagem

“Multilinguagem é uma característica de tarefas que envolvem o uso de mais de uma linguagem natural” (KAY, 1996, p. 229). A principal tarefa multilíngue é a tradução, podendo ser feita de uma única língua-fonte para uma língua-alvo, ou possuir diversas línguas-alvo. Outras tarefas que envolvem mais de uma língua são identificação automática de língua, processamento de fala multilíngue e recuperação de informação multilíngue. Esta subseção se justifica, pois há diversos estudos sobre a semântica dos compostos nominais que se ocupam em comparar como esse fenômeno ocorre em diferentes línguas.

Um programa de tradução automática precisa inicialmente identificar a língua na qual o texto está escrito, quando essa informação não é oferecida. A identificação pode ser feita a partir de textos escritos ou de fala. Para um sistema, identificar a língua escrita é mais simples, desde que já se tenham incluído informações prévias sobre as línguas em questão. A maior dificuldade é desenvolver um sistema que possa identificar a língua de fala em pouco tempo. Por exemplo, estrangeiros que utilizam serviços telefônicos precisam ser encaminhados para

operadores que falam a sua língua, e saber em que idioma eles falam de forma rápida tem sido a preocupação de algumas empresas.

A identificação da língua é só um exemplo de tarefa multilíngue e da importância de se produzir material que traga soluções para um mundo globalizado, em que saber apenas a sua língua materna não é suficiente. Como neste trabalho a preocupação é com a tradução bilíngue, do inglês para o português, serão apresentados, por fim, alguns trabalhos que se ocupam da semântica dos compostos a partir da comparação do inglês com outras línguas.

Johnston e Busa (1999) propõem uma análise composicional da semântica dos compostos nominais em inglês e italiano, que tem como base a Teoria do Léxico Gerativo (Pustejovsky, 1995, ver subseção 2.1.5). Para os autores, a interpretação do composto é composicional e sua forma de expressar esta relação é representando o substantivo núcleo sendo modificado por outro substantivo. É na estrutura *qualia* do substantivo núcleo que um dos componentes semânticos é especificado. Por exemplo, no composto *bread knife*, parte-se do núcleo *knife* e um dos seus argumentos, sobre o que é cortado, será preenchido com *bread*.

Comparando os compostos em inglês e italiano, Johnston e Busa (1999) conseguiram identificar alguns padrões relacionados com a relação semântica entre os elementos, que é representada por um dos papéis da estrutura *qualia*. A seguir, observa-se uma lista com alguns correspondentes nas duas línguas (JOHNSTON; BUSA, 1999, p. 169):

a. <i>bread knife</i> <i>coltello da pane</i>	b. <i>wine glass</i> <i>bicchiere da vino</i>	c. <i>bullet hole</i> <i>foro di pallottola</i>
d. <i>lemon juice</i> <i>succo di limone</i>	e. <i>glass door</i> <i>porta a vetri</i>	f. <i>silicon breast</i> <i>seni al silicone</i>

Se em inglês os compostos NN são formados por um modificador seguido do núcleo, em italiano, temos o núcleo, uma preposição e o modificador. Nos compostos *bread knife* e *wine glass* é possível identificar o papel télico, pois o modificador traz informação sobre a função ou o propósito do núcleo. Uma faca de pão serve para cortar pão, enquanto o copo de vinho é usado para conter vinho. Quando a relação entre os elementos do composto pode ser expressa pelo papel télico, a preposição que une os dois em italiano é *da*.

Quando o modificador se refere à origem do núcleo, como ele surgiu, a preposição usada em italiano é *di*. Em *bullet hole* e *lemon juice*, o modificador pode ser relacionado ao papel agentivo, pois o buraco só surgiu a partir da bala e o suco só passou a existir após alguém ter espremido alguns limões.

Nos dois últimos exemplos, *glass door* e *silicon breast*, o modificador refere-se à constituição do núcleo, ou seja, o papel constitutivo. Isso ocorre porque é possível afirmar que o peito é feito de silicone e que a porta é feita de vidro. Nesse caso, a preposição em italiano é *a*.

Partindo-se dessas conclusões, em que os autores relacionam um papel da estrutura *qualia* com uma preposição específica em italiano, é possível identificar padrões de tradução relacionados com a semântica dos compostos. Como no presente estudo a preocupação também é com a comparação entre duas línguas, o inglês e o português, parte-se da hipótese de que seja possível identificar algum padrão desse tipo, pois em português os compostos possuem uma estrutura semelhante ao italiano, podendo ou não ser ligados por uma preposição.

Os compostos NN do inglês não possuem a mesma estrutura em italiano, pois, conforme já exposto nesta subseção, o correspondente é formado pelos seguintes elementos em ordem: núcleo, preposição e modificador. Assim, Johnston e Busa (1999) precisaram estabelecer alguns critérios para diferenciar os compostos de um substantivo seguido de um sintagma preposicional. Para ser um composto, não é possível aceitar algo entre o núcleo e o modificador. A expressão *coltello sul frigo* (*knife on the fridge*) não constitui um composto, pois é possível inserir um adjetivo entre os substantivos, *coltello tagliente sul frigo* (*sharp knife on the fridge*). Já no composto *coltello da pane* (*bread knife*) não é possível incluir um adjetivo, a não ser que ele modifique o composto como um todo. \**Coltello tagliente da pane* (*knife sharp for bread*) não é aceitável, mas *coltello da pane tagliente* (*sharp bread knife*) é possível, pois o adjetivo modifica o composto todo e não somente um dos elementos. Esses critérios são os mesmos utilizados para identificar um composto em língua inglesa (ver subseção 2.1.1).

Neste trabalho, conforme já anunciado anteriormente, um dos objetivos que se pretende alcançar é a proposta de uma descrição semântica que dê conta dos compostos nominais. Haja vista o compromisso com os compostos NN do inglês e seus correspondentes em português, espera-se contribuir com estudos linguístico-computacionais voltados para as questões multilíngues.

Para concluir este capítulo, em que foram apresentadas diversas perspectivas para tratar da semântica dos compostos NN, é interessante chamar a atenção para a opção por não se apoiar em uma única perspectiva teórica para explicar e descrever a semântica dos compostos NN. Retomando-se o propósito deste capítulo, apresenta-se a seguir como cada um dos conceitos apresentados pode contribuir para este estudo:

- a) **produtividade:** quanto mais produtivo é um composto, mais facilmente se identificam regras ou *templates* (generalizações sobre o seu uso e a sua interpretação). Esse aspecto está relacionado com a solução dada para o composto. Quando tem-se um composto produtivo, é possível identificar um certo padrão, *template*. Se ele for totalmente idiossincrático, a única solução é listá-lo em um léxico. A partir de famílias de compostos, com *core words*, pode-se verificar até que ponto é possível identificar grupos semânticos de compostos;
- b) **compostos endocêntricos e exocêntricos:** nos endocêntricos, que constituem a maioria dos casos, o composto é um hipônimo do núcleo, já nos exocêntricos não é possível identificar um núcleo. Os compostos endocêntricos, para Pustejovsky *et al.* (1993), são altamente frequentes em *corpora* especializados e estabelecem uma relação mais geral entre o composto e o seu núcleo. Os exocêntricos também podem ser altamente produtivos e possuir diferentes graus de lexicalização. Um composto endocêntrico não apresenta uma relação específica entre os dois substantivos do composto, mas traz uma relação mais geral entre o composto e o seu núcleo. Os exocêntricos, dependendo do grau de lexicalização, podem ser interpretados utilizando-se a metonímia;
- c) **compostos lexicalizados e composicionais:** nos composicionais, o significado do todo pode ser obtido a partir da soma dos seus elementos, como em *rat poison*, um veneno usado para matar ratos. Com os lexicalizados, isso nem sempre é possível, pois uma *honey moon* não é uma lua feita de mel, mas refere-se ao período inicial do casamento. Somente com os compostos composicionais é possível identificar uma relação entre os elementos. Os lexicalizados são listados e os composicionais analisados;
- d) **compostos nominais e nominalizações:** os compostos nominais são formados por substantivos, já com as nominalizações um dos elementos é deverbal. Se o composto possuir um elemento deverbal, pode-se partir desse elemento para se chegar à sua interpretação. Em Ó Séaghdha (2007), os dois recebem as mesmas etiquetas, posicionamento a ser seguido neste trabalho, mas o verbo da nominalização facilita a sua identificação;
- e) **nucleariedade:** nem sempre é possível identificar um núcleo do composto, mas se há um núcleo, são as suas características que são herdadas pelo composto. A direção da

interpretação parte do núcleo. Identificar o núcleo é o primeiro passo para a análise do composto.

Além dos conceitos apresentados nos parágrafos anteriores, neste trabalho, serão utilizadas algumas noções semânticas mais específicas e que são recorrentes em diversos trabalhos. Os dados a seguir, organizados a partir da estrutura *qualia* de Pustejovsky (1995) repondem a primeira questão de pesquisa apresentada na introdução deste trabalho, *Quais as relações semânticas mais recorrentes entre os elementos de compostos NN, conforme estudos já realizados?*:

Considerando-se a diversidade de perspectivas teóricas para tratar da semântica dos compostos NN, que tipologia é descritiva e explanatoriamente adequada para dar conta da composicionalidade semântica dessas construções

- 1) papel télico – o N1 indica a função/ para que serve o N2: *keyhole, gear wheel, flagstaff, beehive, birdcage, wineglass, cigar-case, fuel oil*;
- 2) papel agentivo – subdividido em:
  - a. O N1 é um instrumento que dá origem ao N2: *gunshot e sabre-cut*;
  - b. Um dos substantivos é a causa do outro: *drug death, tear gas*;
  - c. A origem do composto é o N1: *goat milk, olive oil, candlelight*;
  - d. O N1 pode ser a fonte de energia para o N2: *voice vote*.
- 3) papel constitutivo (a parte pode ocupar tanto o lugar do N1 como o N2) – subdividido em:
  - a. Material do qual o todo é feito: *gold ring, stone wall*;
  - b. Todo e parte integrante: *broomstick, feather-bed*;
  - c. Todo e característica abstrata: *room temperature*;
  - d. Recipiente e conteúdo: *toy box, tea pot*.
- 4) localização (local e tempo – N1): indica onde algo ocorre ou está/ quando algo ocorre: *city street, sea horse, school friends, weekend guests, Sunday paper*;
- 5) relação de posse e possuído, em que N1 possui N2 ou N2 possui N1: *family car*;
- 6) comparação – N2 possui alguma característica em comum com N1: *bell-flower, goldfish*;
- 7) copulativo – classe e subclasse – um dos substantivos (N1 ou N2) é hipônimo do outro: *man servant, woman driver*;

- 8) aditivo – não tem núcleo, o referente é as duas coisas ao mesmo tempo e as duas coisas estão no mesmo nível: *poet-painter*;
- 9) relação mais geral – endocêntrico – quando é possível identificar uma relação entre os dois, mas não é possível explicá-la: *tax law, price war*.

As categorias semânticas empregadas pelo projeto *FrameNet* são utilizadas, neste trabalho, para a conferência das categorias dos substantivos e para a identificação de verbos que possam unir os dois, estabelecendo uma relação. Quanto aos *templates* de Ryder (1994), verifica-se se são recorrentes no *corpus* paralelo. O interessante dos *templates* é que se sugere uma classificação tanto para o núcleo quanto para o modificador. Isso colabora para uma interpretação co-composicional. Quanto às 48 classes apresentadas por Ryder (1994), estas também foram reformuladas (ou adaptadas), pois há muita repetição. Por exemplo, ela sugere duas categorias para o verbo conter, N2 contém N1 e N2 é contido em N1. Poder-se-ia utilizar apenas uma categoria, conter, e estabelecer a direção, se o N1 contém o N2, ou se o N2 contém o N1.

A seguir, é apresentada uma lista com as principais etiquetas sugeridas por Ryder (1994) com a inclusão de informações quanto às características dos elementos que podem ocupar o papel de N1 ou N2:

**TABELA 6:** Principais etiquetas de Ryder (1994)

Relação	Significado	N1 + N2	Exemplo
BRING	traz, entrega	produto + humano	<i>milkman</i>
CON/CONIN	contém, é contido em	parte integrante + todo todo + parte integrante	<i>stone fruit</i> <i>lemon peel</i>
EAT, HUNT	come, caça	animal + animal carnívoro	<i>bird dog</i>
EQUB	equivalência, adição	profissão + profissão local + local, etc.	<i>poet-painter</i>
EQUF, EQUM, LIK	equivalência, semelhante a, comparação	animal + animal flor + flor, etc.	<i>spider monkey</i>
HAD/HAS	pertence, é pertencido por	possuidor + possuído possuído + possuidor	<i>family car</i> <i>yachtsman</i>
LOC	localização	local + pessoas, artefatos, animais, plantas, frutas	<i>country boy</i> <i>forest hut</i>
MAD	origem de algo, é feito	instrumento + efetuado <sup>19</sup>	<i>gunshot</i>
TEND	cuida, treina	animal doméstico + humano	<i>horseman</i>
USE	usa como instrumento	instrumento/ferramenta + humano	<i>axman</i>

<sup>19</sup> Por efetuado, segundo definição de BORBA (1996), compreende-se algo que passou a existir.

SHOOT	opera (arma)	arma + humano	<i>swordsman</i>
OPER	opera	veículo/maquinaria + humano	<i>truck driver</i>
WEAR	é usado/vestido	parte do corpo + produto	<i>waist-belt</i>

Uma representação que considera estas etiquetas e as nove relações sugeridas anteriormente fornece a resposta para a segunda pergunta de pesquisa deste trabalho: *considerando-se a diversidade de perspectivas teóricas para tratar da semântica dos compostos NN, que tipologia é descritiva e explanatoriamente adequada para dar conta da composicionalidade semântica destas construções?* Todas essas categorias serão retomadas, neste trabalho, no capítulo de análise dos dados, e servem como um guia para a tarefa de classificação dos compostos encontrados no *corpus* de estudo.

### 3 METODOLOGIA

O objetivo deste capítulo é apresentar os procedimentos metodológicos que foram seguidos para a realização da análise proposta no presente estudo. Conforme já exposto anteriormente, o propósito deste trabalho é estudar a semântica dos compostos nominais por meio de sua tradução do inglês para o português. Para tanto, foi necessário inicialmente construir um *corpus* paralelo próprio, para, em um segundo momento, extrair os compostos NN a serem analisados.

A utilização de recursos computacionais para o estudo da língua, como *corpora*, lista de palavras e concordâncias, faz parte de uma área chamada Linguística de *Corpus*. Na próxima seção, serão apresentadas as principais características dessa área, bem como os conceitos pertinentes a este trabalho. Em seguida, os critérios que nortearam a escolha do *corpus* e as etapas do seu processamento serão especificados.

Como nem sempre é possível encontrar um *corpus* disponível e apropriado para os objetivos de pesquisa, optou-se pela compilação<sup>20</sup> de um *corpus*. Os principais passos para a compilação do *corpus* serão apresentados na seção 3.3. Para se obterem, no *corpus* compilado, as informações pertinentes ao estudo proposto, utilizaram-se ferramentas computacionais gratuitas disponíveis on-line, e outras precisaram ser criadas especialmente para este estudo. Entre as ferramentas utilizadas, estão o itemizador, o etiquetador morfológico, o extrator e o alinhador<sup>21</sup>, necessários para a organização de um *corpus* paralelo e a extração dos compostos. A utilização de cada um desses recursos será apresentada na seção 3.4. Este capítulo finalizará com a apresentação dos resultados obtidos a partir da aplicação dessas ferramentas, ou seja, a lista de compostos encontrados no *corpus*, cuja análise será apresentada no próximo capítulo.

#### 3.1 Linguística de *Corpus*

Uma das formas mais práticas e rápidas de se analisar a língua em uso é por meio dos recursos computacionais. Quando um professor, um aprendiz ou até mesmo um simples

---

<sup>20</sup> Compilar, para a área de Linguística de *Corpus*, refere-se à tarefa de reunir textos para a confecção de um *corpus*. Berber Sardinha (2002) refere-se à compilação como a criação de *corpus*.

<sup>21</sup> Para a confecção do itemizador e do extrator de sequências de expressões formadas por dois substantivos, foi fundamental a colaboração do aluno Lucas Lermen, bolsista de Apoio Técnico do projeto FrameCorp (coordenado por Rove Luiza de Oliveira Chishman) em 2008.



falante de uma língua qualquer tem dúvidas sobre o uso de alguma expressão ou palavra, basta consultar o computador mais próximo. A sua fonte de informações pode ser um dicionário digital, enciclopédias on-line, *websites* especializados ou até mesmo os resultados de um *site* de busca.

No meio acadêmico, um dos principais recursos utilizados pelo linguista para o estudo da língua tem sido o *corpus*, que pode ser definido como “um conjunto de textos reunidos de acordo com critérios explícitos de planejamento para uma proposta específica” (ATKINS e CLEAR *apud* GRANGER, 1998, p. 7). Apesar de os computadores pessoais terem se popularizado somente a partir da década de 1980, o estudo de *corpus* existe há muito mais tempo, porém era realizado manualmente.

Um dos primeiros estudos que tiveram como base os dados de um *corpus*, conforme informações de Berber Sardinha (2000), foi realizado em 1921 por Thorndike. Considerando-se os recursos tecnológicos de que dispomos hoje em dia, o objetivo do autor parece bastante simples. No seu trabalho, ele identificou as palavras mais frequentes da língua inglesa por meio do levantamento manual em um *corpus* de 4,5 milhões de palavras. Listas com palavras e a sua frequência deram origem a novos materiais didáticos de ensino de língua inglesa como língua materna e estrangeira.

Uma das dificuldades da Linguística de *Corpus* no seu início era a demora e a falta de confiabilidade no trabalho realizado por seres humanos. Por meio do computador, esse problema é facilmente resolvido. Inclusive, várias pesquisas surgiram em função dos avanços na área da informática. Há uma vasta demanda de estudos multidisciplinares, tanto da linguística contribuindo para a informática quanto dos recursos computacionais servindo de auxílio à análise da língua. Uma das principais colaborações da linguística para a informática diz respeito à área do PLN, em que se busca fornecer informações sobre a língua para a máquina, de forma que sistemas possam analisar, sumarizar ou produzir textos automaticamente, entre outras tarefas. Já o acesso a *corpora* eletrônicos é um exemplo de recurso computacional que facilita o trabalho do linguista.

Para Berber Sardinha (2000), um dos fatos mais importantes na história da Linguística de *Corpus* foi a publicação de *Syntactic Structures*, de Noam Chomsky, no final da década de 1950. Com a consagração do gerativismo, os estudos empiristas e de *corpus* caíram em total descrédito. A língua em uso não merecia mais atenção, já que o que passou a importar foi o estudo da competência linguística, dos universais.

Conforme relata Berber Sardinha (2000), o próprio Chomsky se manifestou quanto à Linguística de *Corpus* ridicularizando os estudos na área por meio da frase *I live in Dayton*,

*Ohio*. O que Chomsky argumentou foi que, por haver mais pessoas em Nova Iorque do que em Dayton, uma frase como a dele seria menos frequente do que a frase *I live in New York*. Na realidade, segundo Berber Sardinha (2000), o que Chomsky pareceu não entender é que questões como diferença demográfica não interessam para os estudos da linguagem.

De acordo com Berber Sardinha (2000), uma teoria mais compatível com os estudos de *corpus* é o funcionalismo de Halliday (1994), que, diferentemente do gerativismo chomskiano, considera a linguagem como probabilidade. As regularidades lexicais da língua podem ser identificadas por meio da interpretação dos dados presentes em combinações recorrentes de palavras (colocação) e em condições linguísticas que envolvem o uso de determinadas palavras (coligação). Berber Sardinha (2000, on-line) define colocação como “associação entre itens lexicais, ou entre o léxico e campos semânticos”. Ou seja, a partir da interpretação das colocações, podem ser identificados certos padrões de uso de uma determinada palavra. Beaugrande (2000) afirma que, mesmo tendo-se um *corpus* extenso, não há como todas as possíveis combinações estarem presentes, mas, pensando-se que uma parte representa o todo, os resultados apresentarão ampla cobertura. Como exemplos de colocações, podem ser citadas as palavras que geralmente são encontradas após o verbo “causar”, como “problemas, danos, morte, prejuízos, impacto, polêmica, mal, doenças”<sup>22</sup>. A partir dessas colocações encontradas em um *corpus*, pode-se chegar à conclusão de que os complementos do verbo *causar* são palavras com sentido negativo.

Por coligação, entende-se a relação entre os itens lexicais e os gramaticais. É nas coligações que são percebidos os usos típicos de determinados termos, ou seja, as suas preferências gramaticais. Um exemplo clássico de coligação, na língua inglesa, refere-se aos verbos que aceitam um complemento com *-ing*. Caso após o verbo *enjoy* ocorra outro verbo, geralmente este estará no seu gerúndio: *enjoy dancing, enjoy winning, enjoy writing*<sup>23</sup>. Assim, pode-se afirmar que uma das coligações possíveis do verbo *enjoy* é *enjoy + v-ing*.

É possível perceber, portanto, que na Linguística de *Corpus* o conceito de regras linguísticas é mais amplo, pois o que encontramos são padrões de uso no discurso real, padrões esses interpretados a partir das colocações e coligações.

Uma característica importante dos estudos na Linguística de *Corpus* é que não há separação entre a gramática e o léxico; eles são estudados juntos. Essa separação se justifica, em outras áreas da linguística, por vários motivos, principalmente econômicos, pois materiais

<sup>22</sup> Dados de Berber Sardinha, 2004, p. 239.

<sup>23</sup> Dados obtidos no *corpus Collins Cobuild*, disponível em: <<http://www.collins.co.uk/Corpus/CorpusSearch.aspx>>.

como gramáticas e dicionários são vendidos separadamente. Porém, o léxico muda rapidamente e apresenta ordens funcionais muito mais diversas das encontradas na gramática, por isso se justifica o estudo de língua que apresenta o léxico e a gramática juntos. Ao se ter acesso a essas informações, podem-se obter dados bem específicos quanto ao domínio discursivo em questão e às variedades linguísticas encontradas em um *corpus*.

Diferentemente de análises formalistas, que inicialmente deixaram de lado a semântica e a pragmática e posteriormente começaram a estudá-las a partir da sintaxe, na Linguística de *Corpus*, a semântica, a pragmática e a sintaxe podem ser facilmente relacionadas. Por meio das colocações, podem-se comparar as combinações com o objetivo de estudar questões relacionadas à semântica. A pragmática também é favorecida por esse tipo de análise, já que dados sobre os contextos e os participantes são recuperados com facilidade.

Por meio dos recursos tecnológicos da Linguística de *Corpus*, os estudos da linguagem tornam-se mais confiáveis, pois lidam com dados reais, e mais abrangentes, já que com o computador pode-se lidar com uma grande quantidade de dados. As pesquisas nessa área podem trazer conclusões totalmente inusitadas sobre a língua a partir da observação de colocações, chegando até mesmo à contestação de alguns pressupostos do gerativismo.

Desde o início dos estudos de *corpus*, discussões sobre o tamanho mais adequado de um *corpus* têm sido constantes e provavelmente ainda estão longe de serem resolvidas. Mesmo tendo-se uma lista de todas as palavras presentes no *corpus*, não podem ser obtidas todas as combinações possíveis entre essas palavras. No entanto, quanto maior for o *corpus*, maior será a quantidade e a qualidade dos seus dados.

Há vários tipos de *corpora*, conforme diversos critérios, como finalidade, conteúdo ou autoria. Para este trabalho, foi utilizado um *corpus* paralelo, com textos em inglês e português, a ser descrito na seção 3.3.

O *corpus* paralelo consiste basicamente em dois textos, um original e sua tradução, organizados de forma que cada linha do texto de uma língua esteja alinhada com o seu correspondente no texto da segunda língua. As principais aplicações desse tipo de recurso estão relacionadas aos estudos de tradução, quando é necessário comparar semelhanças e diferenças entre original e traduções ou até mesmo comparar as diferentes traduções de uma mesma obra. Esse tipo de *corpus* também é chamado *corpus de traduções* na tradição da linguística contrastiva.

Entre as aplicações de *corpora* paralelos, McEnery e Wilson (1993) citam a tradução automática e a criação de léxicos. A partir de um *corpus* paralelo como base de dados, servindo como *corpus* de treinamento, podem-se criar métodos probabilísticos que auxiliem a

tarefa de tradução automática. É possível extrair de um *corpus* paralelo as palavras correspondentes em mais de uma língua ou até mesmo expressões multivocabulares, que podem ser incluídas em um léxico ou em um dicionário multilíngue. Por meio de estudos de frequência, é possível construir uma base de dados terminológica de textos especializados.

Em língua portuguesa, ainda há poucos *corpora* desse tipo. Entre eles, há o COMPARA<sup>24</sup> (FRANKENBERG-GARCIA, SANTOS, 2002), que apresenta romances em português europeu e brasileiro e suas traduções para o inglês. Há também traduções no sentido inverso, do inglês para o português.

Como o COMPARA é um recurso disponível e com textos traduzidos e originais nas línguas inglesa e portuguesa, pensou-se inicialmente em utilizá-lo neste trabalho. No entanto, nesse *corpus* encontram-se traduções que soam estranhas para um falante de português brasileiro contemporâneo, como *rés-do-chão* (*ground floor*), expressão portuguesa para andar térreo, e *algibeira da rabona* (*pocket jacket*), que se refere ao bolso do casaco em um texto de 1887.

O que merece destaque no uso de um *corpus* paralelo como o COMPARA é o fato de que, a partir de dois textos, original e tradução, pode-se obter o alinhamento de cada frase de forma automática. Isso facilita o trabalho do linguista que analisa os correspondentes de tradução em um determinado texto. Para realizar a investigação proposta nesta dissertação, ou seja, o estudo da semântica dos compostos nominais por meio de sua tradução do inglês para o português, decidiu-se organizar um *corpus* paralelo próprio, formado por edições em inglês e em português da revista *National Geographic*. Dessa forma, o *corpus* organizado apresenta linguagem atual – publicação entre 2007 e 2008 – e a variante brasileira da língua portuguesa.

No entanto, para a seleção dos compostos nominais, partiu-se do *corpus* em inglês, já que a direção da tradução é do inglês para o português. Assim, neste capítulo, chama-se de *corpus* de estudo o *corpus* formado apenas por reportagens da versão inglesa da revista *National Geographic*, de onde os compostos formados por dois substantivos (compostos NN) foram extraídos.

Além da utilização de um *corpus* paralelo formado por edições da revista *National Geographic*, outros recursos na área de Linguística de *Corpus* mostraram-se importantes para este trabalho: a lista de palavras e as concordâncias. Por meio da lista de palavras, foi possível identificar os elementos mais frequentes presentes nos compostos NN do *corpus* de estudo (maiores descrições dessa etapa serão apresentadas na seção 3.4). Por meio das

---

<sup>24</sup> Disponível em: <<http://www.linguateca.pt/COMPARA/>>.

concordâncias, uma busca por todas as ocorrências de determinados compostos, foi possível obter o contexto de uso de cada composto, facilitando assim a sua interpretação.

O programa *WordSmith Tools*, versão 5.0, possibilita a realização da busca por uma palavra nas formas singular e plural. Para pesquisar as ocorrências no singular e no plural de uma palavra, pode-se digitar, por exemplo, *monkey/monkeys*. Além disso, o programa pode informar a origem de cada frase, ou seja, em qual edição da revista as frases são encontradas. Esses recursos são importantes para este trabalho; além de realizar buscas pelas palavras tanto no singular quanto no plural, é interessante saber se uma determinada expressão ocorre em mais de uma edição da revista, fato que pode caracterizá-la como uma expressão mais geral, pois é usada em diferentes contextos.

### 3.2 Escolha do *corpus*

Nesta dissertação, inicialmente, pensou-se em fazer um estudo da semântica dos compostos NN do inglês, já que esse fenômeno por si só já representa um desafio para trabalhos de diferentes áreas, tanto da linguística quanto da informática, conforme visto no capítulo 2 deste trabalho. Como o objetivo do trabalho era compreender as relações entre os elementos de compostos desse tipo, seria necessário trabalhar com um *corpus* em língua inglesa suficientemente grande e com variados gêneros textuais, para se abarcar a linguagem em geral, e não alguma linguagem especializada.

Porém, considerando-se que os compostos NN em língua inglesa geralmente não mantêm essa mesma estrutura ao serem traduzidos para o português, também procurou-se analisar quais são as escolhas de traduções para essas expressões, conforme os dados da edição brasileira da revista *National Geographic*. Para tanto, o trabalho de Johnston e Busa (1999) serviu de grande inspiração. Conforme apresentado no capítulo 2 desta dissertação, esses autores, por meio da representação da estrutura *qualia*, relacionaram o tipo de relação semântica entre os elementos dos compostos com as traduções mais frequentes para o italiano. No caso do presente estudo, as traduções dos compostos feitas para o português demonstram como os tradutores representaram em língua portuguesa as relações semânticas dos compostos NN originalmente em inglês.

Assim, como o objetivo deste trabalho foi ampliado, era necessário um *corpus* paralelo, com textos em inglês e em português. A escolha das edições da revista *National Geographic* se deu basicamente porque as mesmas matérias são publicadas em inglês e em português, sendo o original a edição em língua inglesa. Outro fator importante para a escolha

foi que as matérias são traduzidas na sua íntegra, fato que colabora para que os correspondentes dos compostos NN sejam encontrados em língua portuguesa. A disponibilização do material também é simples: basta acessar a página on-line da revista, pois já estão em formato eletrônico.

A opção por uma revista foi motivada pela afirmação de Biber (1993, p. 233) de que os artigos de revistas “incluem uma grande variedade de propósitos e mostram vastas diferenças linguísticas entre os textos do registro”. Já que o propósito deste trabalho não é se comprometer com nenhuma linguagem específica, os artigos de revista pareceram poder fornecer dados ricos para um estudo da semântica dos compostos. Embora se tivesse em mente algo próximo da língua geral, ao optar-se por uma revista como a *National Geographic*, foram encontrados compostos relacionados aos assuntos abordados pela revista.

Nas próximas seções, será descrita a forma como foram feitos a coleta e o alinhamento do *corpus*, assim como serão especificados os critérios utilizados para selecionar os compostos.

### 3.3 Coleta e alinhamento do *corpus*

Para o estudo dos compostos, foi compilado um *corpus* paralelo formado por 10 edições da revista *National Geographic*, publicadas entre 2007 e 2008. A tradução dos textos de cada edição é encontrada na mesma edição da revista, porém na sua versão brasileira. As duas versões das edições foram obtidas nos *sites* da revista<sup>25</sup>.

Durante a compilação, foram copiados e colados, no aplicativo Bloco de Notas, os arquivos correspondentes às edições entre agosto de 2007 e maio de 2008. Optou-se por arquivos em formato *txt*, pois esse formato é pré-requisito para o seu processamento em diversos programas utilizados para o estudo de *corpora*. Cada edição constituiu um arquivo diferente, sendo que os nomes dos arquivos das edições em inglês iniciaram com a letra “e” seguida pelo mês e pelo ano de publicação, e os nomes dos arquivos das edições em português iniciaram com a letra “p”, trazendo os mesmo dados. Na tabela 7, a seguir, encontram-se as principais informações dos arquivos:

---

<sup>25</sup> *National Geographic Magazine*, disponível em <<http://ngm.nationalgeographic.com/>>, e *National Geographic Brasil*, disponível em <<http://viajeaqui.abril.uol.com.br/ng/>>.

TABELA 7: Dados dos *corpora*

Arquivos	<i>Tokens</i>	<i>Types</i>	Frases	Arquivos	<i>Tokens</i>	<i>Types</i>	Frases
e_august07.txt	19.795	4.455	1.036	p_agosto07.txt	19.745	4.835	953
e_september07.txt	17.528	4.343	872	p_setembro07.txt	17.001	4.526	781
e_october07.txt	30.068	5.862	1.644	p_outubro07.txt	28.036	6.423	1.420
e_november07.txt	19.090	4.407	1.047	p_novembro07.txt	19.040	5.044	1.030
e_december07.txt	18.544	4.682	1.026	p_dezembro07.txt	17.617	4.847	979
e_january08.txt	20.350	4.632	1.231	p_janeiro08.txt	20.798	5.252	1.211
e_february08.txt	17.443	4.245	926	p_fevereiro08.txt	16.711	4.559	887
e_march08.txt	24.222	5.201	1.272	p_marco08.txt	24.574	5.943	1.272
e_april08.txt	26.126	5.984	1.529	p_abril08.txt	25.502	6.395	1.397
e_may08.txt	19.359	4.384	1.144	p_maio08.txt	19.174	4.934	1.122
<b>Total</b>	212.535	20.265	11.727	<b>Total</b>	208.201	24.327	11.052

Ao comparar a quantidade de *tokens* e *types* nos arquivos do mesmo mês, percebe-se que a diferença é pouco significativa, fato que leva a crer que os textos foram traduzidos na sua íntegra. Isso também aumenta as possibilidades de que se encontrem correspondentes em português para os compostos NN em inglês. No entanto, o número de *tokens* é menor em língua portuguesa na maioria dos arquivos, com exceção das edições de janeiro e março, em que há mais *tokens* na versão em português. Isso pode indicar que algumas frases do original tenham sido omitidas na tradução, hipótese que é reforçada ao comparar-se o número de frases em cada arquivo, em todos há menos frases na língua portuguesa. O fato de haver mais *types* na edição traduzida, pode ter a relação com a crença de que em português utiliza-se mais palavras para expressar uma idéia. No entanto, também pode ter uma relação com a questão cultural, já que alguns tópicos são conhecidos de leitores falantes nativos do inglês, enquanto que para os brasileiros é necessário fornecer mais informações sobre o assunto. De qualquer forma, o que interessa para este trabalho é a possibilidade de se encontrar os correspondentes de tradução nos textos traduzidos e, a partir destes dados, acredita-se que a chance de isto acontecer seja grande.

No *site* da revista não há informações sobre os tradutores dos artigos, mas, considerando-se que a edição brasileira é publicada pela editora Abril, que está há anos no mercado, é possível pressupor que os tradutores sejam profissionais qualificados. Procurou-se manter a formatação apresentada no *site*, ou seja, a separação dos parágrafos, mas desconsideraram-se as imagens, os *links* e quaisquer outros dados que não fossem texto.

Coletados os dois *corpora*, em inglês e em português, passou-se para a tarefa de alinhar cada frase do original com a sua tradução, por meio de uma ferramenta que faz isso automaticamente, o alinhador. Os *corpora* em inglês e em português precisaram ser itemizados, já que esse formato é pré-requisito para o alinhador *Vanilla Aligner*. O programa

que formata o texto em uma palavra por linha, chamado aqui de itemizador, utiliza a arquitetura Java J2SE. A principal vantagem dessa arquitetura é o fato de ser multiplataforma, ou seja, independente de sistema operacional, podendo funcionar em *linux*, *windows*, entre outros.

Para separar cada palavra, o programa identifica os espaços em branco e os substitui por um símbolo de nova linha. Dessa forma, os sinais de pontuação não são separados das palavras. Expressões compostas separadas por hífen e siglas também são mantidas na mesma linha, conforme mostra a figura 2. O itemizador salva o texto em itens no mesmo diretório em que o texto-fonte estiver localizado, apenas acrescentando “Itemizado” ao nome do novo arquivo.

O único pré-requisito para o processamento dos textos é que eles devem estar no formato ANSI da extensão *txt*. Além disso, é necessário possuir o *Java Runtime Environment (JRE)* instalado no computador. O arquivo de saída deve ser aberto por meio do *WordPad*. A figura 2, a seguir, mostra como fica o texto itemizado:

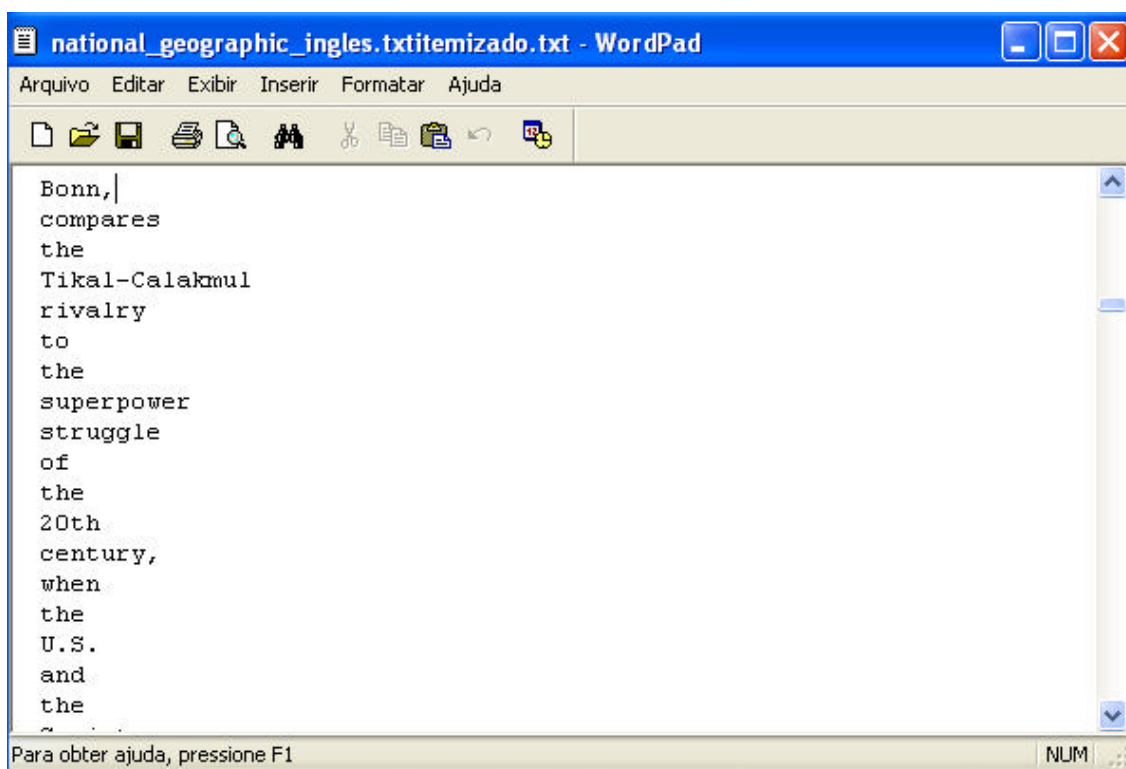


Figura 2: *Corpus* itemizado

O itemizador apresentou bons resultados, separando adequadamente cada palavra por linha. No entanto, quando havia um travessão entre duas palavras, o programa identificava



como um hífen, mantendo as duas palavras na mesma linha, tal como em: *shields—the, legacy—whether, cats—most*. Ao fazer isso, o itemizador causou alguns problemas para a fase de identificação dos compostos, pois duas palavras foram consideradas uma única e cada ocorrência desse tipo precisou ser conferida.

O alinhador *Vanilla Aligner* (DANIELSSON e RIDINGS, 1997) foi escolhido por ser gratuito, por estar disponível on-line e por sua precisão ser alta. A função do *Vanilla* é alinhar frases de *corpora* bilíngues, ou seja, identificar as frases correspondentes em uma língua e na outra em conjuntos de textos disponíveis em múltiplas línguas. Esse modelo probabilístico tem obtido bons resultados, atingindo, de acordo com os dados dos desenvolvedores do programa, Danielsson e Ridings (1997), uma média de 4% de erro em um *corpus* trilíngue formado por 15 artigos de economia do *Union Bank of Switzerland*, nos idiomas inglês, francês e alemão.

Conforme Danielsson e Ridings (1997), para se trabalhar com o texto no alinhador *Vanilla*, é necessário um pré-processamento, que consiste em, além de itemizar o texto, anotar as sentenças e os parágrafos. Entende-se por sentenças as unidades que um tradutor traduziria de uma só vez, incluindo, dessa forma, títulos ou frases dentro de um parágrafo. Para esse alinhador, cada unidade de uma língua, uma frase, por exemplo, corresponde a uma unidade de tamanho semelhante na outra língua. Entende-se por tamanho a quantidade de caracteres. Conforme sugerido no *site* do LAEL<sup>26</sup>, onde o programa é disponibilizado, optou-se por anotar manualmente o *corpus* utilizando as etiquetas .EOS para final de sentença e .EOP para final de parágrafo.

Quanto ao alinhamento, este é feito em duas etapas. Primeiramente os parágrafos são alinhados e, após, as frases de cada parágrafo são alinhadas. Para o programa funcionar, os textos devem apresentar o mesmo número de parágrafos. Isso dificultou um pouco a tarefa, já que há diferença de quantidade de parágrafos entre uma versão e outra das edições da *National Geographic*. Foi feita uma adaptação manual para que o alinhador funcionasse, tendo como referência a versão em língua inglesa.

Utilizando o alinhador *Vanilla* do *site* do LAEL, obtém-se como resultado uma lista com as relações entre as unidades (denominadas *links*), que podem ser do tipo 1-1, em que uma frase é alinhada com outra da tradução, 2-1, em que duas frases da primeira língua são alinhadas com apenas uma na segunda língua, e assim por diante. Entre as outras possibilidades, há: 0-1, 1-0, 1-2 e 2-2.

---

<sup>26</sup> Disponível em: <<http://www2.lael.pucsp.br/corpora/alinhador/index.html>>.

TABELA 8: Alinhamentos do *corpus* paralelo – total: 9.514

Alinhamento	Exemplo	Frequência
1-1	<i>She might have the best memory in the world.</i> / Ela talvez tenha a melhor memória do mundo.	2.745 – 28,85%
0-1	Além do apoio moral, Fogo Novo provavelmente contribuiu com sua pequena força expedicionária.	1.221 – 12,83%
1-0	<i>But one day, I was under a bridge, where a lot of addicts live and shoot up heroin, when suddenly I heard the loud wail of police sirens and saw two police cars coming at high speed.</i>	1.294 – 13,60%
1-2	<i>All her colleagues were enrolled in outside classes; even the vice-principal took a weekend class on educational technology.</i> / Todas as suas colegas estavam tendo aulas particulares. Até a vice-diretora tinha aulas de fim de semana em tecnologia educacional.	1.968 – 20,68%
2-1	<i>As epigraphers finally learned to read the Maya glyphs, a darker picture emerged, of warring dynasties, court rivalries, and palaces put to the torch. Maya history became a tapestry of precise dates and vividly named personages.</i> / Mas quando os especialistas em epigrafia decifraram os hieróglifos maias, descortinou-se um quadro de dinastias em guerra, rivalidades violentas e palácios incendiados.	1.978 – 20,79%
2-2	<i>I drove there with Larry Squire, a neuroscientist and memory researcher at the University of California, San Diego, and the San Diego VA Medical Center, and Jen Frascino, the research coordinator in Squire's lab who visits EP regularly to administer cognitive tests. Even though Frascino has been to EP's home some 200 times, he always greets her as a stranger.</i> / Fui até lá de carro com Larry Squire, neurocientista e pesquisador da memória da Universidade da Califórnia em San Diego e do Centro Médico dos Veteranos de San Diego, e com Jen Frascino, a coordenadora de pesquisa do laboratório de Squire. Ela visita EP regularmente para aplicar testes cognitivos, mas, embora já tenha ido à casa de EP umas 200 vezes, ele sempre a recebe como uma estranha, como se jamais a tivesse visto.	308 – 3,23%

Em relação aos erros cometidos pelo programa, Gale e Church (1993) afirmam que há uma maior precisão nos casos de resultados 1-1, em que uma frase da primeira língua corresponde a apenas uma na segunda. Os casos mais problemáticos são os de 1-0, em que uma frase em uma língua não possui correspondente na outra. Nos exemplos apresentados pelos autores, todas as relações classificadas como 1-0 estavam erradas de acordo com a

avaliação de seres humanos. Entre as conclusões dos autores, está o fato de que, quanto mais semelhantes forem as línguas, com maior frequência ocorrerão casos de 1-1, logo o programa será mais eficiente.

O *corpus* paralelo deste trabalho foi alinhado por meio do *Vanilla* e constatou-se que, de fato, o caso mais frequente foi 1-1, com case 29% dos alinhamentos totais, e os alinhamentos desse tipo foram feitos corretamente, conforme mostra a figura 3, a seguir:

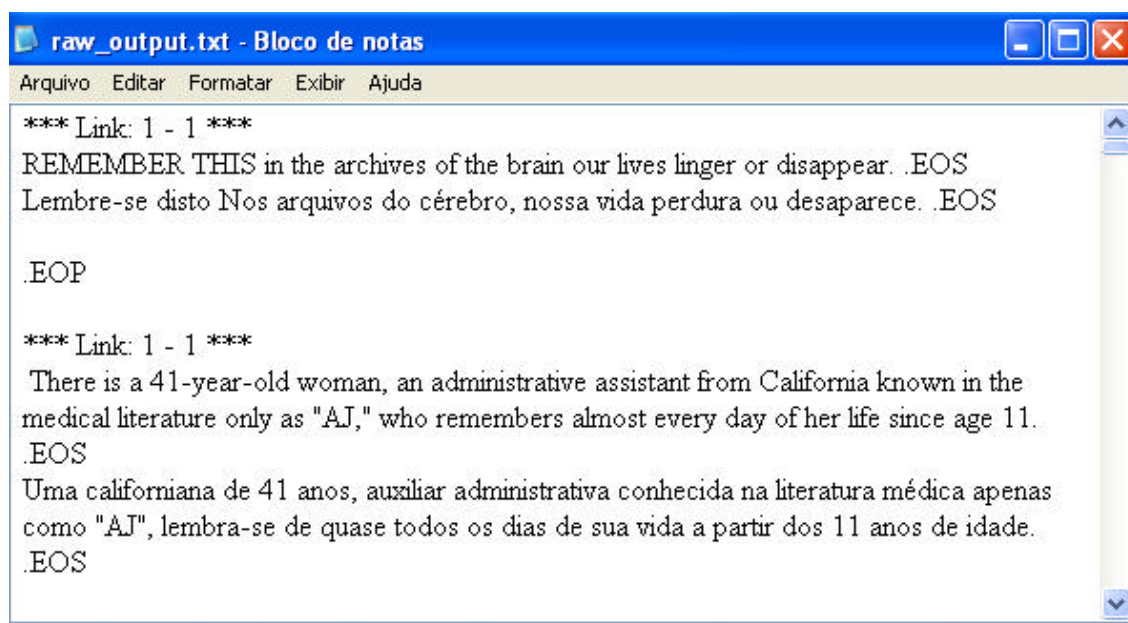


Figura 3: *Corpus* alinhado

O *Vanilla* apresenta alguns erros de alinhamento, e mesmo que nas relações 1-1 haja bons resultados, tornou-se necessária uma conferência manual dos resultados relacionados aos compostos. Provavelmente os erros no alinhamento ocorreram porque algumas informações do texto original, às vezes um parágrafo inteiro, não apareciam na versão traduzida. Quando essas omissões ocorrem, geralmente a parte do texto original traz alguma especificidade cultural que não faz sentido para os brasileiros. Em outras ocorrências, acredita-se que o tradutor tenha considerado a informação irrelevante e realizado a omissão pressupondo que isso não atrapalharia a compreensão do leitor.

Um exemplo de erro no alinhamento pode ser visto no seguinte trecho: “*Now warriors from the Petén, equipped with stone axes and short stabbing spears, swelled their ranks. As armor, many wore cotton vests stuffed with rock salt*”. Como resultado de alinhamento, obteve-se a seguinte tradução: “Quando se aproximaram da nascente, os soldados desembarcaram e marcharam seja pela margem, seja pela borda do cânion onde corria o rio”.

Nesse trecho, para identificar o correspondente de tradução adequado, foi necessário procurar pela frase no *corpus* em português, ou seja, a correção foi feita manualmente.

Mesmo que tenham sido identificados vários problemas ao longo do *corpus*, procurou-se corrigir apenas as ocorrências em que os compostos analisados ocorreram. Conferir manualmente todo o *corpus*, além de ser uma tarefa extremamente demorada, não se mostrou necessário, pois apenas uma parte do *corpus* foi aproveitada, as frases com os compostos. Caso esse *corpus* seja utilizado futuramente, uma revisão manual das ocorrências ainda não conferidas se fará necessária.

A seguir, será descrito o processo de como os compostos NN foram selecionados para a fase da análise semântica.

### 3.4 Extração e busca no *corpus* paralelo

Como este estudo trata dos compostos nominais formados por dois substantivos na língua inglesa e dos seus correspondentes de tradução para o português, era preciso extrair do *corpus* em inglês uma sequência de dois substantivos sem que houvesse outro substantivo antes ou depois. Também se tornou necessário obter uma lista de todas as expressões seguidas pela quantidade de vezes em que elas ocorrem no *corpus*. Para o levantamento desses dados, é necessário que o *corpus* esteja anotado morfológicamente, pois só assim é possível fazer uma busca por expressões formadas por substantivos. Como o *corpus* deste estudo não estava etiquetado, ele precisou passar por esse processamento, pois só assim outro programa poderia identificar a informação necessária.

Optou-se pelo etiquetador *TreeTagger* (SANTORINI, 1990) para a língua inglesa, por ser uma ferramenta gratuita e com bons resultados, com uma média de 96% de precisão. O etiquetador, assim como o alinhador, exige que o *corpus* esteja itemizado. Com as anotações morfológicas, o *corpus* estava pronto para a extração das sequências de dois substantivos. Como não foi encontrado nenhum extrator apropriado e que fosse gratuito, foi necessário criar essa ferramenta, que tem como base as etiquetas do *TreeTagger* e utiliza a mesma arquitetura do itemizador, Java J2SE. Essas três ferramentas, o itemizador, o etiquetador e o extrator, foram utilizadas para o processamento do *corpus* em língua inglesa.

O *TreeTagger*<sup>27</sup> é um etiquetador de *part-of-speech* (POS), ou seja, é um sistema que faz automaticamente o reconhecimento das categorias morfosintáticas. Ele foi desenvolvido

---

<sup>27</sup> Disponível em: <<http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>>.

na Universidade de Stuttgart, na Alemanha, e é utilizado em mais de 10 idiomas diferentes, dentre eles o inglês, o francês, o alemão e o italiano.

Como o objetivo era conseguir extrair do *corpus* sequências formadas por dois substantivos, consultou-se o manual de etiquetas do *TreeTagger* (SANTORINI, 1990). Entre as etiquetas pertinentes, foram encontradas as seguintes: NNS (substantivo comum, plural) e NN (substantivo comum, singular). Para diferenciar o singular do plural, o programa identifica o verbo que acompanha o substantivo. É o verbo que vai determinar se o substantivo está no plural ou no singular. Isso resolve o problema de substantivos que no singular terminam em “s”, como *linguistics*. Há etiquetas referentes aos nomes próprios, mas estas não foram consideradas para o presente trabalho, já que foram analisados apenas os compostos nominais formados por substantivos comuns.

Uma informação importante quanto às escolhas do etiquetador, encontrada no mesmo manual, refere-se ao fato de que um substantivo modificador será etiquetado como substantivo, e não como adjetivo. Essa escolha é importante, pois, caso o etiquetador anotasse um substantivo modificador como adjetivo, a presente pesquisa seria muito prejudicada, já que busca-se por sequências formadas por dois substantivos. Por outro lado, as cores também são etiquetadas como substantivos, embora nem sempre sejam consideradas substantivos em determinadas frases.

O *TreeTagger*, segundo dados de Schmid (1994), utiliza um modelo probabilístico baseado em árvores binárias de decisões; isto é, a partir de trigramas, sequências de três palavras encontradas em um *corpus*, determinam-se as relações morfosintáticas entre as palavras. Para chegar à conclusão se determinada palavra é um substantivo ou um adjetivo, é necessário responder afirmativa ou negativamente a perguntas quanto às palavras que aparecem ao seu redor. À medida que cada resposta afirmativa é dada, as informações na árvore são conectadas chegando-se a uma resposta, à “folha da árvore”, conforme exemplo apresentado na figura 4. O etiquetador também possui um léxico que foi criado a partir de uma parte do *corpus Penn Treebank*. Dois milhões de palavras desse *corpus* foram etiquetados e serviram de treinamento; ou seja, a partir dos dados obtidos nesse *corpus*, criaram-se regras probabilísticas que são utilizadas na tarefa de etiquetagem de quaisquer outros *corpora*.

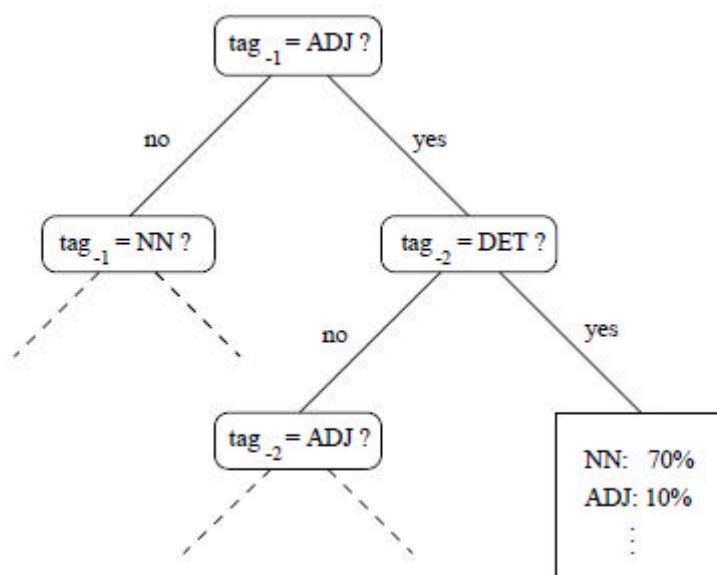


Figura 4: Exemplo de árvore de decisão (SCHMID, 1994, p. 46)

Entre os resultados relatados por Schmid (1994), o *TreeTagger* atinge em torno de 96% de precisão, mostrando-se um etiquetador bastante eficiente. Na figura 5, a seguir, é possível observar uma parte do *corpus* desta dissertação etiquetado, exibido em três colunas: a primeira com as palavras conforme são encontradas no texto, a segunda com as etiquetas morfológicas e a terceira com a forma canônica da palavra, como o infinitivo do verbo ou o substantivo no singular e sem marca de gênero:

Palavra	Etiqueta	Forma Canônica
doomed	JJ	doomed
splendor	NN	splendor
of	IN	of
the	DT	the
Maya	NP	Maya
unfolded	VVD	unfold
against	IN	against
the	DT	the
backdrop	NN	backdrop
of	IN	of
the	DT	the
rain	NN	rain
forests	NNS	forest

Figura 5: *Corpus* etiquetado

Foram obtidos bons resultados com o etiquetador, ainda que algumas palavras não tenham sido classificadas corretamente, como verbos e adjetivos que foram etiquetados como substantivos. Algumas palavras apareceram equivocadamente anotadas como substantivos, mas isso não dificultou a análise, pois quando examinou-se cada composto, o seu contexto de uso também foi considerado.

Conforme mencionado anteriormente, após o *corpus* estar etiquetado, era preciso extrair as sequências formadas por dois substantivos com o objetivo de se chegar aos compostos nominais. Para esse fim, foi desenvolvido um extrator com a mesma arquitetura utilizada pelo itemizador, Java J2SE.

Esse extrator busca pelas sequências de dois substantivos a partir das etiquetas do *TreeTagger*. Assim, ele busca por: NN NN, NN NNS, NNS NN e NNS NNS. Durante o desenvolvimento do extrator, houve também uma preocupação de que não ocorresse um substantivo antes ou depois dessa sequência, pois o foco deste trabalho são apenas os compostos formados por dois substantivos. Assim, quando três substantivos consecutivos ocorrem, o programa verifica isso e descarta.

Como saída, o programa oferece uma lista com possíveis compostos nominais e o seu número de ocorrências no *corpus*, ou seja, a sua frequência, conforme mostra a figura 6, a seguir:

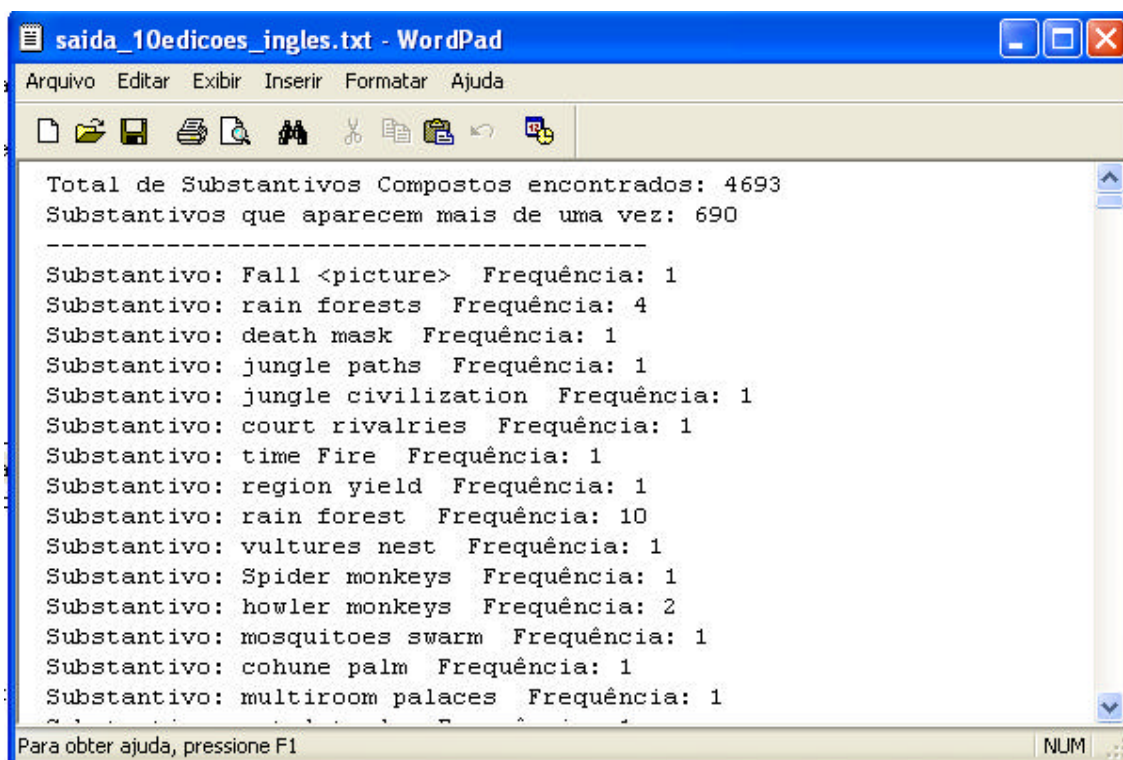


Figura 6: Resultados do extrator

Cada resultado do extrator precisou ser conferido, pois não havia garantias de que as palavras listadas pelo programa fossem realmente compostos nominais, já que erros de etiquetagem ou de formatação poderiam ter ocorrido. Entre os prováveis compostos, chamados de candidatos a compostos, é possível observar, na figura 6, expressões como *fall <picture>* e *time Fire*, que não são compostos de fato. No primeiro caso, a palavra *<picture>* aparece porque, no processo de compilação do *corpus*, ela foi incluída no lugar onde originalmente havia figuras no *site*. Optou-se por incluir essa palavra nos textos porque, mesmo que a análise proposta neste trabalho não incluía as imagens, pareceu interessante informar quando elas ocorrem nos textos, caso essa informação pudesse ser de alguma ajuda futuramente na interpretação dos compostos. No segundo caso, a palavra *Fire* é um nome próprio, referindo-se ao maia Fogo Novo, e não deveria ter sido extraída.

### 3.5 Resultados da extração

Nesta seção, apresenta-se a maneira como foi possível obter uma lista de candidatos a compostos, ou seja, de sequências de dois substantivos que poderiam constituir um composto. A partir dos resultados do extrator de sequências NN, chegou-se a 4.693 candidatos a



compostos, sendo que, desse total, 690 ocorreram mais de uma vez no *corpus* em língua inglesa. A grande maioria dos possíveis compostos ocorreu apenas uma vez no *corpus* todo, fenômeno que é chamado de *hapax*. Isso já era previsto, pois há pouca probabilidade de a mesma combinação de duas palavras específicas ocorrer repetidamente.

Tendo-se como resultado uma grande quantidade de dados, resolveu-se selecionar algumas sequências para este estudo. Como critério, foram adotadas as *core words*, ou seja, aquelas palavras que aparecem como elementos de mais de um composto. Um exemplo de *core word* no presente *corpus* é *day*, pois foram encontrados compostos em que essa palavra é utilizada tanto como núcleo quanto como modificador em uma construção composta, tais como *day care*, *day jobs*, *day laborer*, *day pack*, *election day*, *feast days*, *field day* e *harvest days*. Este é o mesmo critério utilizado por Ryder (1994) no seu estudo. Partindo-se de *core words*, é mais fácil chegar a possíveis *templates* que servem como um indicativo para as relações entre os elementos de um composto no presente *corpus*. Dessa forma, também os casos de *hapaxes* não representam um problema para a análise, já que um mesmo substantivo pode fazer parte de diversos compostos NN.

Foram selecionadas as *core words* que ocorrem 10 ou mais vezes, da lista de 4.693 candidatos a compostos, no *corpus* deste trabalho, com o objetivo de se poder caracterizar os *templates* mais frequentes no *corpus*. Não se levou em consideração, nesse primeiro momento, se a palavra ocorre no primeiro ou segundo substantivo. Utilizou-se a ferramenta *Wordlist* (lista de palavras) do *WordSmith Tools* para se chegar aos dados, apresentados na tabela 9, a seguir:

**TABELA 9:** Frequências de *core words* de compostos no *corpus National Geographic*

<i>Core Words</i>	Frequência	<i>Core Words</i>	Frequência
<i>water/waters</i>	41	<i>space</i>	40
<i>day/days</i>	37	<i>family/families</i>	34
<i>home/homes</i>	29	<i>tree/trees</i>	28
<i>plant/plants</i>	27	<i>field/fields</i>	25
<i>school/schools</i>	24	<i>ethanol</i>	24
<i>life/lives</i>	24	<i>oil/oils</i>	24
<i>group/groups</i>	23	<i>memory/memories</i>	23
<i>car/cars</i>	23	<i>government/governments</i>	23
<i>time/times</i>	22	<i>energy</i>	21
<i>forest/forests</i>	20	<i>fuel/fuels</i>	20
<i>workers/worker</i>	20	<i>system/systems</i>	20
<i>research</i>	20	<i>people</i>	19
<i>computer</i>	19	<i>sea/seas</i>	19
<i>state/states</i>	19	<i>year/years</i>	19

<i>dozen</i>	18	<i>building/buildings</i>	18
<i>gorilla/gorillas</i>	18	<i>chimp/chimps</i>	18
<i>buddhist</i>	17	<i>food</i>	17
<i>scientsts/scientist</i>	17	<i>today</i>	17
<i>stone/stones</i>	16	<i>construction</i>	16
<i>program/programs</i>	16	<i>century/centuries</i>	16
<i>cattle</i>	15	<i>business</i>	15
<i>winter</i>	15	<i>security</i>	15
<i>particle/particles</i>	15	<i>city/cities</i>	15
<i>animal/animals</i>	15	<i>team/teams</i>	15
<i>company/companies</i>	15	<i>farm/farms</i>	15
<i>village/villages</i>	15	<i>line/lines</i>	15
<i>thing/things</i>	15	<i>force/forces</i>	14
<i>science</i>	14	<i>corn</i>	14
<i>floor/floors</i>	14	<i>half</i>	14
<i>phone/phones</i>	14	<i>brain/brains</i>	14
<i>rights/right</i>	14	<i>class/classes</i>	14
<i>body/bodies</i>	14	<i>island/islands</i>	14
<i>way/ways</i>	13	<i>mountain</i>	13
<i>health</i>	13	<i>facility/facilities</i>	13
<i>metal/metals</i>	13	<i>gas/gases</i>	12
<i>power/powers</i>	12	<i>carbon</i>	12
<i>use/uses</i>	12	<i>police</i>	12
<i>head</i>	12	<i>fire/fires</i>	12
<i>species</i>	12	<i>summer/summers</i>	12
<i>high/highs</i>	12	<i>work/works</i>	12
<i>word/words</i>	12	<i>center/centers</i>	12
<i>border</i>	11	<i>electronics</i>	11
<i>country</i>	11	<i>rice</i>	11
<i>moon</i>	11	<i>side</i>	11
<i>rock</i>	11	<i>drug/drugs</i>	11
<i>trade</i>	11	<i>fruit/fruits</i>	11
<i>cane</i>	10	<i>conservation</i>	10
<i>culture</i>	10	<i>desert</i>	10
<i>ice</i>	10	<i>paper</i>	10
<i>skin</i>	10	<b>Total</b>	<b>1.641</b>

A partir dos substantivos, ou palavras que receberam a etiqueta de substantivo pelo etiquetador, foi realizada a conferência manual de cada candidato a composto. Para saber se cada sequência de dois substantivos constitui de fato um composto NN, foram analisadas as 1.641 ocorrências por meio do concordanceador do *WordSmith Tools*. Em caso de dúvida, conferiu-se a sua tradução no *corpus* paralelo.

Por exemplo, a palavra *people* apareceu nos dados de extração como um elemento de 19 candidatos a compostos, sendo estes os seguintes:

TABELA 10: Candidatos a compostos

<u>PEOPLE + N</u>	<u>N + PEOPLE</u>
<i>people archaeologists</i>	<i>bog people</i>
<i>people cash</i>	<i>days people</i>
<i>people distrust</i>	<i>dozen people</i>
<i>people today</i>	<i>lab people</i>
<i>people's eyes</i>	<i>round people</i>
<i>people's kindnesses</i>	<i>something people</i>
<i>people's place</i>	<i>stages people</i>
<i>people's sensibilities</i>	<i>thing people</i>

Na tabela 10, há apenas 16 candidatos a compostos formados com a *core word people*. Três foram desconsiderados por apresentarem erros visíveis de extração, sendo estes: *viruses-with people*, *people can't* e *people-entire families*. No primeiro e no terceiro casos, houve um problema de formatação, pois, no lugar do hífen, deveria aparecer um travessão; ou seja, as duas primeiras palavras são separadas e não constituem candidatos a compostos. Já em *people can't*, deve ter ocorrido um erro de etiquetagem, pois *can't* é um verbo.

Quase todas as sequências com *people* ocorrem uma única vez no *corpus*. Dessa forma, para se ter certeza de que se trata de um composto, conferiu-se a frase em que cada uma ocorre e analisou-se também a sua tradução. O primeiro candidato a composto é *people archaeologists*. Ao se buscar pela sua frase, obteve-se o seguinte:

*It is the oldest cemetery ever found in the Pacific islands, and it harbors the bones of an ancient people archaeologists call the Lapita, a label that derives from a beach in New Caledonia where a landmark cache of their pottery was found in the 1950s.*

Na leitura dessa frase, a primeira impressão foi de que *people* se refere a *ancient*, e não a *archaeologists*. Essa suspeita foi confirmada quando se buscou pela sua tradução:

*Trata-se do mais antigo cemitério jamais encontrado nas ilhas do Pacífico, e abriga os ossos de um povo antigo que os arqueólogos chamam de lapita, nome tirado de uma praia da Nova Caledônia onde um depósito da cerâmica desse povo, de importância fundamental, foi descoberto nos anos 1950.*

O primeiro critério semântico para confirmar se dois substantivos constituem um composto é a unidade do referente. Para que *people archaeologists* fosse um composto, os dois substantivos deveriam apresentar o mesmo referente. Como isso não acontece nesse caso,

*people archaeologists* foi descartado do grupo de compostos formados com a *core word people*.

Já o candidato *bog people* ocorre em: *Bodies still lying undiscovered in the bogs of northern Europe will yield more clues about how and why the bog people met their ends. Bog people*, que é traduzido como “seres do pântano”, é um composto, pois o referente é um só e também porque não é possível incluir um modificador entre os dois elementos, como *\*bog young people*.

Por meio dessa conferência manual de cada composto, o número de compostos diminuiu bastante, chegando a 842 compostos NN . No caso dos compostos com *people*, o número de compostos diminuiu quando se percebeu que, entre os candidatos formados por um substantivo antecedido por *people*, há apenas ocorrências de casos no genitivo. Como neste trabalho a proposta é analisar somente os compostos formados por dois substantivos, não foram consideradas as expressões no genitivo, pois há um elemento a mais entre as palavras, o apóstrofo.

Assim, do total de 19 candidatos a compostos com a *core word people*, apenas dois realmente são compostos formados por dois substantivos: *bog people* e *lab people*. Para a realização dessa tarefa de conferência, as concordâncias fornecidas pelo *WordSmith Tools* se mostraram extremamente importantes. Ao solicitar-se a concordância pela palavra *people*, é possível encontrar todas as frases em que a palavra ocorre, conforme mostra a figura 7, a seguir:

N Concordance

283 a bad harvest, the researchers **say**. **People** were hungry, reduced to eating  
 284 of the larger world," Thinley **says**. "**People** are searching for a new definition  
 285 has lost all credibility," he **says**. "**People** look at Musharraf and they see  
 286 It's not nice to needlessly **scare** **people**, but it's much less nice to  
 287 what happens when lots of **scared** **people** try to do the same thing in a big  
 288 humans in nine countries and **scaring** **people** all over the world. A quick bit of  
 289 All you can do then is sell it to **scrap** **people**, he says. "What they do with it  
 290 has revealed an ancient **seafaring** **people**, the distant ancestors of today's  
 291 Continuing through the hills, I **see** **people** tearing apart car batteries,  
 292 bringing people to these **seminaries**. **People** are seeking refuge and  
 293 in use throughout Pakistan **serve** **people** in need. Transportation and  
 294 sparked the usual commotion; **several** **people** began throwing rocks at them,  
 295 in her eyes was cataracts. **Several** **people** had already told me she was the  
 296 torso into an old rice bag. **Several** **people** saw him carrying the sack into  
 297 don't—at least not right away, **since** **people** who hijack ships (or at least  
 298 north, there aren't a lot of people, **so** **people**—like Islamic groups—come [to  
 299 to change my bad nail-biting habits, **so** **people** don't call me the Nail- Biting  
 300 here." What happens is that **some** **people** cash in on their property and  
 301 patchwork recovery, with **some** **people** raising their homes to protect

Figura 7: Parte da concordância de *people*

O recurso que torna mais fácil a localização do candidato a composto é a organização alfabética das palavras à esquerda (em vermelho na figura) ou à direita do núcleo. Na maioria dos casos, por meio da simples leitura das ocorrências nas concordâncias é possível identificar se o candidato é um composto ou não. Em caso de dúvida, consulta-se a sua versão para o português no *corpus* paralelo.

Após a conferência manual, do total de 1.641 candidatos a compostos chegou-se a uma lista com 842 compostos. A relação completa dos compostos encontra-se no anexo B deste trabalho e está organizada de acordo com as *core words*. No quadro a seguir, tem-se os filtros aplicados para atingir o número de compostos do *corpus*.

**TABELA 11:** Filtros para a seleção dos compostos

<b>Filtro</b>	<b>Resultado</b>
Utilização do extrator de sequências NN	4.693 candidatos a compostos
Seleção das <i>core words</i> que ocorrem 10 ou mais vezes no <i>corpus</i>	1.641 candidatos a compostos
Conferência manual através de análise de concordâncias	842 compostos NN

Retomando-se a terceira questão de pesquisa apresentada no início deste trabalho, *de que forma as ferramentas relacionadas ao estudo de corpus podem ser utilizadas para uma análise semântica dos compostos NN*, pode-se afirmar que as ferramentas da área de Linguística de *Corpus* podem facilitar de fato o estudo da semântica dos compostos, por meio: de uma lista de candidatos a compostos de forma automática; de uma lista com a frequência de *core words* no *corpus*; do acesso a concordâncias para auxiliar na conferência e na análise dos compostos; e do alinhamento de *corpus* paralelo, que facilita a identificação do correspondente de tradução.

Durante a aplicação das etapas da metodologia, percebeu-se que as ferramentas apresentaram alguns problemas, como erro na etiquetagem morfológica do *corpus*, erro de configuração e falhas no alinhamento automático. Mesmo com parte da análise tendo de ser feita manualmente, as ferramentas facilitaram o trabalho. Sem o auxílio dessas ferramentas não haveria como se realizar a busca por *core words* e se obter a sua frequência de forma precisa e rápida. O concordanceador do *WordSmith* também foi extremamente útil para a conferência dos candidatos a compostos.

Considerando os resultados de extração, conseguiu-se chegar a uma lista com mais de 800 compostos. Isso possibilitou e agilizou a escolha pelos compostos a serem analisados. A principal contribuição do *corpus* paralelo para o estudo de tradução é a possibilidade de identificar o equivalente de tradução a partir das frases alinhadas. Mesmo com os problemas de alinhamento, após a correção manual, é possível organizar todas as opções de tradução de um mesmo composto ou compostos com relações semânticas semelhantes por meio de uma busca nos dados do *corpus* paralelo.

Enfim, todo esse aparato computacional serviu para se organizarem os dados de estudo e se chegar a uma informação mais específica, os compostos NN, além de tornar possível a relação da expressão em inglês com o seu equivalente de tradução no *corpus* da revista *National Geographic*. A partir desses dados, foi possível iniciar a análise da semântica dos compostos, etapa em que também foram utilizados alguns desses recursos computacionais, conforme será descrito no próximo capítulo.

## 4 ANÁLISE E DISCUSSÃO DOS DADOS

A análise da semântica dos compostos NN encontrados no *corpus National Geographic* é apresentada neste capítulo. Partindo-se dos resultados de extração apresentados no capítulo anterior, foram selecionados 200 compostos aleatoriamente.

Os dados são apresentados de acordo com as suas etapas de análise, a saber: (i) uma análise inicial mais rasa em que os compostos são classificados de acordo com características como nuclearidade, produtividade, lexicalização, nominalização e endocentrismo; (ii) interpretação da relação semântica entre os substantivos por meio de uma abordagem co-composicional; (iii) identificação de padrões ou *frames* recorrentes nos compostos do *corpus*, relacionados às informações obtidas nas etapas anteriores; (iv) estudo dos equivalentes de tradução no *corpus* paralelo, em busca de algum padrão de tradução.

Esta análise tem como objetivo principal compreender as relações semânticas entre os elementos de compostos NN em língua inglesa. Um dos propósitos é reconhecer quais são as propriedades semânticas presentes nos compostos do *corpus* de estudo. O estudo dos equivalentes de tradução tem como objetivo apresentar as escolhas dos tradutores associando-as à relação entre os elementos do composto na língua inglesa. A tradução diz respeito a um contexto bem específico, a revista *National Geographic*, e servirá para ilustrar como as ideias presentes nos compostos são expressas em língua portuguesa, sem o compromisso com a descrição da língua geral ou de uma língua especializada.

### 4.1 Os compostos nominais em língua inglesa

Esta primeira análise dos compostos NN busca confrontar as principais características apresentadas na fundamentação teórica com os dados presentes no *corpus*. O propósito aqui é verificar como os substantivos que compõem essas construções do tipo NN se relacionam semanticamente. Para tanto, serão considerados os seguintes conceitos: produtividade, transparência semântica, nuclearidade, composição exocêntrica e endocêntrica, lexicalização e nominalização.

Considerando o estudo teórico apresentado no capítulo 2, este capítulo procura responder à última questão de pesquisa, já apresentada na Introdução do trabalho, ou seja, *quais as relações semânticas mais frequentes encontradas nos compostos nominais do corpus paralelo analisado?* No entanto, esta questão é desmembrada em várias outras questões menos centrais a serem apresentadas ao longo desta análise. As primeiras são as seguintes:

- É possível identificar claramente o núcleo semântico de cada composto NN? Em quais casos o núcleo do composto é o N1? Quais as características dos compostos em que não é possível identificar um núcleo?
- Qual a frequência de compostos endocêntricos no *corpus*, em que o composto é um tipo do seu núcleo? Quais os compostos cuja interpretação semântica por meio da relação de hiponímia é satisfatória? Quais as características dos compostos exocêntricos?
- Qual a relação entre produtividade e transparência semântica? De que forma isso é expresso nos compostos NN no *corpus* de estudo?
- Qual a frequência de compostos NN formados por meio de nominalizações? Como o verbo de origem pode contribuir para a sua interpretação semântica?

Parte dessas questões será respondida nesta seção de forma superficial a partir das conclusões obtidas por meio de uma análise mais rasa. Com um estudo mais aprofundado a ser apresentado na próxima seção, em que as relações semânticas e os elementos formadores de compostos serão identificados em padrões ou *frames*, as mesmas questões serão revistas e as conclusões serão reformuladas.

Esta análise inicia com uma lista de 200 compostos escolhidos de forma aleatória, que servirão como dados de estudo para esta pesquisa. Esses compostos são apresentados na tabela 12, a seguir:

**TABELA 12:** Compostos do *corpus National Geographic* selecionados para a análise

Compostos NN			
<i>algae plant</i>	<i>country boy</i>	<i>island home</i>	<i>salt water</i>
<i>animal bonés</i>	<i>country school</i>	<i>jet fuel</i>	<i>school play</i>
<i>animal track</i>	<i>day pack</i>	<i>judgment day</i>	<i>school uniform</i>
<i>banana trees</i>	<i>desert beetle</i>	<i>killing field</i>	<i>sea creatures</i>
<i>baseball field</i>	<i>drainage system</i>	<i>lab people</i>	<i>sea raiders</i>
<i>biomimetics research</i>	<i>drug addiction</i>	<i>law school</i>	<i>sheep brain</i>
<i>body cloak</i>	<i>drug users</i>	<i>leopard skin</i>	<i>siesta time</i>
<i>bog people</i>	<i>education center</i>	<i>life force</i>	<i>skin cancer</i>
<i>bog waters</i>	<i>election day</i>	<i>life histories</i>	<i>snack food</i>
<i>border guards</i>	<i>electronics shops</i>	<i>living thing</i>	<i>sound system</i>
<i>brain cells</i>	<i>energy farm</i>	<i>manufacturing center</i>	<i>space boots</i>
<i>brain infection</i>	<i>ethanol company</i>	<i>marble floors</i>	<i>space expert</i>



<i>brain science</i>	<i>ethanol plants</i>	<i>math class</i>	<i>space exploration</i>
<i>building blocks</i>	<i>ethanol production</i>	<i>memory championship</i>	<i>space station</i>
<i>business licenses</i>	<i>ethanol tanks</i>	<i>memory drugs</i>	<i>space suit</i>
<i>cane cutter</i>	<i>eucalyptus forests</i>	<i>memory loss</i>	<i>space tourist</i>
<i>cane ethanol</i>	<i>family friend</i>	<i>memory researcher</i>	<i>spring day</i>
<i>cane juice</i>	<i>family group</i>	<i>metal armor</i>	<i>state banks</i>
<i>canola oil</i>	<i>family home</i>	<i>metals buyers</i>	<i>stone axes</i>
<i>car accidents</i>	<i>family member</i>	<i>monitoring groups</i>	<i>stone houses</i>
<i>car batteries</i>	<i>family mosque</i>	<i>moon expedition</i>	<i>storage facility</i>
<i>car bombs</i>	<i>family planning</i>	<i>mountain air</i>	<i>sugar cane</i>
<i>car keys</i>	<i>farm kid</i>	<i>mountain bicycle</i>	<i>sugarcane workers</i>
<i>car park</i>	<i>farm tools</i>	<i>mountain lion</i>	<i>summer afternoon</i>
<i>carbon emissions</i>	<i>feast day</i>	<i>nest construction</i>	<i>summer climbing</i>
<i>carbon fiber</i>	<i>feed corn</i>	<i>night school</i>	<i>summer day</i>
<i>cattle guards</i>	<i>field chores</i>	<i>nursing home</i>	<i>summer grass</i>
<i>cattle owners</i>	<i>fire brigade</i>	<i>oil company</i>	<i>tool use</i>
<i>cattle trailers</i>	<i>food crops</i>	<i>oil consumption</i>	<i>trade routes</i>
<i>cedar tree</i>	<i>food sources</i>	<i>opium trade</i>	<i>trade union</i>
<i>cell phone</i>	<i>forest fire</i>	<i>orchid farm</i>	<i>train cars</i>
<i>childhood memories</i>	<i>fruit bat</i>	<i>oyster business</i>	<i>tree branches</i>
<i>chimp feces</i>	<i>fruit pulp</i>	<i>palm trees</i>	<i>tree holes</i>
<i>chimpanzee group</i>	<i>fruit tree</i>	<i>paper bag</i>	<i>village gossip</i>
<i>church floor</i>	<i>fuel alcohol</i>	<i>particle accelerators</i>	<i>village streets</i>
<i>city engineer</i>	<i>gas station</i>	<i>phone number</i>	<i>warming seas</i>
<i>city government</i>	<i>gas tanker</i>	<i>plant director</i>	<i>water bottles</i>
<i>city parks</i>	<i>gorilla families</i>	<i>plant fuel</i>	<i>water bowl</i>
<i>coffin tree</i>	<i>gorilla health</i>	<i>plant oils</i>	<i>water conservation</i>
<i>community schools</i>	<i>government buildings</i>	<i>police cars</i>	<i>water hole</i>
<i>commuter car</i>	<i>government efficiency</i>	<i>police sirens</i>	<i>water ice</i>
<i>computer classes</i>	<i>group culture</i>	<i>police station</i>	<i>water plants</i>
<i>computer keyboard</i>	<i>group house</i>	<i>recycling company</i>	<i>water pollution</i>
<i>computer monitors</i>	<i>group members</i>	<i>research coordinator</i>	<i>water tanks</i>
<i>computer scientist</i>	<i>guerrilla groups</i>	<i>research sites</i>	<i>weekend class</i>
<i>conservation efforts</i>	<i>head scarves</i>	<i>research vessel</i>	<i>winter day</i>
<i>construction jobs</i>	<i>health care</i>	<i>rice bag</i>	<i>winter mountaineering</i>
<i>construction workers</i>	<i>home prices</i>	<i>rice fields</i>	<i>wood construction</i>
<i>cooking oil</i>	<i>home village</i>	<i>rock salt</i>	<i>workers rights</i>
<i>corn ethanol</i>	<i>island culture</i>	<i>rocket scientist</i>	<i>youth culture</i>

Após a seleção aleatória dos 200 compostos apresentados na tabela 12, foi feita uma análise de cada composto. Essa análise, cujos dados aparecem de forma completa nos anexos

C<sup>28</sup> e D deste trabalho, será explicada ao longo deste capítulo, em que serão abordadas as relações semânticas entre os elementos dos compostos. Nesta seção, serão considerados apenas os resultados gerais dessa análise.

Os dados numéricos desses resultados gerais são apresentados na tabela 13, a seguir:

**TABELA 13:** Resultados gerais da análise inicial dos 200 compostos selecionados

<b>200 compostos</b>			
<i>Nuclearidade</i>	<i>Composição exocêntrica e endocêntrica</i>	<i>Transparência semântica e lexicalização</i>	<i>Nominalização</i>
195 compostos permitem a identificação do núcleo semântico	199 compostos endocêntricos	198 compostos com transparência total	44 compostos constituídos por algum tipo de nominalização
5 compostos apresentam dificuldades quanto à identificação do núcleo semântico	1 composto exocêntrico	1 composto com transparência parcial	156 compostos totalmente nominais
		1 composto lexicalizado	
Total: 200	Total: 200	Total: 200	Total: 200

A tabela 13 indica a ocorrência de quatro características dos compostos analisados, mostrando que, dos 200 compostos: (i) 195 permitem a identificação do núcleo semântico e 4 apresentam dificuldades quanto à identificação do núcleo, a serem apresentadas na próxima subseção; (ii) 199 são endocêntricos e 1 é exocêntrico; (iii) 198 apresentam transparência total, 1 apresenta transparência parcial e 1 é lexicalizado; e (iv) 44 são constituídos por algum tipo de nominalização e 156 são totalmente nominais. Os dados indicados na tabela 13 serão explicitados nas subseções a seguir.

#### 4.1.1 Nuclearidade

Conforme já mencionado, entre os 200 compostos analisados, em 195 foi possível identificar o núcleo semântico do composto, sendo que ele é representado no N2. Em *banana trees*, por exemplo, o seu núcleo é *trees*, pois se refere a uma árvore; o núcleo de *car*

<sup>28</sup> No anexo C deste trabalho, são apresentadas as informações concernentes à análise rasa dos 200 compostos apresentados na tabela 12. É apresentada uma tabela com as seguintes informações sobre cada um dos 200 compostos: (i) núcleo; (ii) se é endocêntrico ou exocêntrico; (iii) o seu grau de transparência semântica, podendo ser total, parcial ou lexicalizado; (iv) se um dos seus elementos é uma nominalização e qual é o elemento ou os elementos nominalizados.

*accidents* é também o seu N2, já que o seu referente é *accidents*. Entre os 200 compostos analisados, apenas cinco apresentaram dificuldades quanto à identificação do seu núcleo. Cada caso será comentado a seguir.

O primeiro composto que causa dúvida é *car bombs*, pois, considerando-se que o seu núcleo é o N2, diz-se que um carro-bomba é um tipo de bomba. No entanto, essa interpretação parece depender do ponto de vista. Pode-se dizer que um carro-bomba é uma bomba em forma de carro ou que é um carro que é utilizado como bomba. Como há outros compostos formados pela palavra bomba que surgem a todo momento na mídia, tais como homem-bomba e bicicleta-bomba, isso leva a crer que o núcleo mais adequado a essa construção é bomba. De qualquer forma, estamos diante de um composto aditivo em que o referente é as duas coisas ao mesmo tempo, trata-se de um carro e de uma bomba.

Foram identificados também três compostos em que não se pode afirmar que o núcleo é o N2, *water ice*, *island home* e *home village*. Parece que a interpretação desses compostos está condicionada ao ponto de vista do leitor. *Water ice* é um gelo constituído por água ou água em forma de gelo? *Island home* refere-se a uma ilha onde fica o lar de alguém ou a casa de alguém que fica em uma ilha? Partindo-se do princípio de que o gelo é geralmente feito da água no seu estado sólido, informar que o gelo contém água parece ser irrelevante, ferindo até mesmo uma das máximas de Grice (1971), a da relevância. Considerando-se como núcleo o N1, tem-se como referente a *água* e o modificador (*ice*) informa o seu estado, conforme a ocorrência no *corpus*: *A mission to look for water ice on Jupiter's moons and two missions to detect Earthlike planets around other stars have been canceled or put on indefinite hold, while budgets for in-house research and analysis, individual research projects, and future low-budget space missions have all been cut.* Ao procurar-se pela solução adotada pelo tradutor da revista, encontra-se como correspondente *água congelada*. Aceitando essa conclusão, pode-se afirmar que *water ice* é um composto cujo núcleo é o N1.

Para compreender o significado de *island home*, pode-se imaginar um conjunto, em que um lugar como uma ilha pode conter diversas coisas, entre elas o lar ou a casa de alguém. Não se pode dizer que a ilha toda é o lar de alguém, mas que apenas um pedaço da ilha é o seu lar, onde fica a sua casa. Pensando dessa forma, o seu núcleo seria o N2, um lar que fica em uma ilha. Ao comparar essa conclusão com a tradução da versão em português do artigo, identifica-se a expressão *ilha natal*. Se *home* é considerado um modificador, então o núcleo é ilha; assim, pode-se criar a paráfrase que explica essa relação: “uma ilha onde alguém nasceu ou foi criado”. O contexto pode ajudar a se optar por um dos núcleos, sendo que a frase em que esse composto ocorre é a seguinte: *Can the endangered Philippine eagle survive in the*

*shrinking forests of its island home?* Por se tratar de uma águia, o seu lar não se resume a uma casa, mas ela vive apenas na parte da ilha onde tem floresta. Mesmo assim, parece que, de forma geral, as duas interpretações são possíveis. Se se entende que o referente é indicado pelas duas coisas ao mesmo tempo, ou seja, que o lar é uma ilha e que a ilha é um lar, não há como optar por um núcleo. Para confirmar essa hipótese, um falante nativo da língua inglesa foi consultado e ele concordou que se trata de um composto aditivo, ou seja, não possui núcleo. Essa interpretação pode causar estranhamento para um falante nativo do português, pois, na tradução, *ilha natal*, temos como núcleo claramente a palavra *ilha*. O composto *home village* também é um aditivo e a sua interpretação é semelhante a de *island home*.

Em um composto NN, não foi possível identificar um núcleo, principalmente porque não se trata de um composto composicional. Ao se tentar compreender o que é um *water hole*, não se pode simplesmente traduzir literalmente, pois obter-se-ia *buraco de água*. Para encontrar-se um correspondente adequado na língua portuguesa, foi necessário consultar um dicionário, ou, nesse caso, as traduções do *corpus*, obtendo-se *nascente* para *water hole*. Assim, soa estranho dizer que uma nascente é um buraco. Trata-se de uma expressão lexicalizada, aspecto que será analisado em uma das subseções a seguir.

#### 4.1.2 Compostos endocêntricos

Retomando-se o que já foi exposto sobre os compostos endocêntricos, eles são os que constituem um hipônimo do seu núcleo. Como foi possível identificar um núcleo na maioria dos 200 compostos analisados, também é possível afirmar que eles sejam endocêntricos. Um *fruit bat* é um tipo de morcego, assim como uma *rain forest* é um tipo de floresta. A relação do composto com o seu núcleo, sem dúvida, indica um tipo de informação semântica, ainda que não diretamente relacionada à natureza composicional da expressão. Parece relevante informar que o composto é endocêntrico quando este é um exemplar de uma categoria maior como em *fruit bat* e *rain forest*. No entanto, também se pode dizer que uma *gorilla family* é um tipo de família e que uma *business licence* é um tipo de licença. Mas é possível trazer mais detalhes sobre a relação entre *gorilla* e *family* que a relação de hiponímia não abrange. O gorila é um membro da família, faz parte deste grupo, enquanto que uma licença para abrir negócios tem um propósito específico, uma função, serve para abrir negócios. Essa questão será retomada na próxima seção.

Apenas um composto não foi considerado endocêntrico. O composto *water hole* é denominado exocêntrico, pois é lexicalizado. A expressão *car park* causou dúvidas quanto à

sua classificação como endocêntrica. Traduzindo-se de forma geral, *car park* é um estacionamento de carros. No entanto, na nossa cultura raramente se utiliza a expressão estacionamento de carros, até porque o estacionamento geralmente não serve só para carros; tanto motos quanto bicicletas ocupam o mesmo lugar dos carros.

O principal motivo que leva à dúvida de que um *car park* é um tipo de estacionamento está relacionado a um aspecto cultural. Em alguns países como China e França, as bicicletas são muito utilizadas, e por isso há um estacionamento destinado somente a elas. No Brasil, embora exista estacionamento de bicicletas, muitas vezes carros e bicicletas são estacionados no mesmo lugar. Realizando-se uma busca na Internet por estacionamentos, encontrou-se um *site* em que um usuário de bicicletas reclama justamente de ter de usar um estacionamento de carros para deixar a sua bicicleta. Transcreve-se a seguir a manifestação dessa pessoa:

Eu não acho “adequado” ser obrigado a descer num estacionamento barulhento e fedorento a gases de carros com minha *bike* para depois ter que andar um monte ou pegar elevador pra finalmente chegar na loja. (...) Eu acho [que] ser obrigado a compartilhar um estacionamento com motos ou carros é contra o espírito de liberdade que eu busco andando de bicicleta pela cidade (KISSER, 2005<sup>29</sup>).

Esse depoimento reforça a ideia de que um estacionamento de carros não é um lugar destinado somente para carros na maioria das vezes; contudo, pelo seu significado a partir dos dois substantivos, ele não deixa de ser um composto endocêntrico.

#### 4.1.3 Transparência semântica e produtividade

Dos 200 compostos analisados, 199 são transparentes. Eles também são produtivos, pois se trata de expressões livres e há outras construções formadas a partir do núcleo. Como exemplo, consideremos o composto *baseball field*: além de *baseball field*, é possível encontrar na mídia *hockey field*, *soccer field*, entre outros. Vê-se que há uma relação entre a transparência semântica e a produtividade. Os compostos lexicalizados, como *water hole*, apresentam um grau menor de produtividade, já que não há outros compostos do mesmo tipo.

Um composto presente no *corpus* que causa dúvidas se é um composto composicional ou não é *cell phone*. Essa palavra é usada com bastante frequência, e os dois substantivos estão tão relacionados que muitas vezes não nos referimos a esse aparelho utilizando o seu nome completo, *telefone celular*, é mais comum falarmos simplesmente *celular*. Não há como

---

<sup>29</sup> Disponível em: <<http://apocalipsemotorizado.net/2005/07/14/lugar-para-bicicletas-nao-e-estacionamento-de-carros/>>.

negar que se trata de um tipo de telefone, um telefone móvel. Porém, são poucas as pessoas que sabem o porquê de esse aparelho possuir esse nome. No entanto, o conhecimento comum das pessoas sobre o composto não é considerado quando a questão é a sua transparência. Afinal, qual a relação entre telefone e célula? Para se chegar a essa resposta, consultou-se um *site* especializado em explicar como determinados aparelhos funcionam, o *HowStuffWorks*<sup>30</sup>. Conforme as informações do *site*, o aparelho é chamado de celular porque a transmissão é feita através de um mapeamento de um lugar, uma cidade, por células. Cada célula ou região possui a sua própria antena, o que faz com que o aparelho funcione em lugares diferentes. Se há uma relação entre os dois substantivos, então, *cell phone* é composicional.

O composto *rock salt* é parcialmente composicional, pois se trata de um sal, mas não há pedras no sal. O modificador *rock* foi utilizado para informar que o sal encontra-se em forma de pedra. Em língua portuguesa não se utiliza esse mesmo modificador, pois, ao nos referirmos a esse tipo de sal, utilizamos a expressão *sal grosso*.

Os compostos parcialmente transparentes podem ser interpretados de alguma forma, a partir do ponto de vista escolhido pelo analisador, pois há alguma relação entre os dois substantivos. Já nos compostos lexicalizados, como *water hole* não há uma transparência evidente.

#### 4.1.4 Nominalização

Compreende-se por compostos nominalizados aqueles em que um dos substantivos é formado por um verbo e um sufixo nominal. Entre os 200 compostos escolhidos para esta análise, 44 apresentam um elemento nominalizado. Na tabela 14, a seguir, são apresentados os sufixos deverbais identificados no *corpus National Geographic*:

**TABELA 14:** Sufixos deverbais

Sufixo	Nominalização	Composto
-ER	<i>cut – cutter</i>	<i>cane cutter</i>
-ING	<i>kill - killing</i>	<i>killing field</i>
-OR	<i>coordinate - coordinator</i>	<i>research coordinator</i>
-AGE	<i>store - storage</i>	<i>storage facility</i>
-ATION	<i>explore – exploration</i>	<i>space exploration</i>
-MENT	<i>govern – government</i>	<i>city government</i>
-ION	<i>construct - construction</i>	<i>construction efforts</i>

<sup>30</sup> Disponível em: <<http://www.howstuffworks.com/>>.

A partir desta primeira análise, podem-se apenas apontar os casos de nominalização, mas não há como afirmar de que maneira o verbo de origem do nome contribui para a interpretação do composto. Nesta primeira etapa da análise, o propósito é simplesmente identificar os compostos deverbais; uma análise mais profunda quanto o papel do verbo de origem para o significado dos compostos será feita na próxima seção.

Considerando-se as questões de pesquisa propostas no início desta seção e repetidas a seguir, chega-se às seguintes conclusões quanto ao núcleo e à relação entre lexicalização e produtividade dos compostos analisados:

- *É possível identificar claramente o núcleo semântico de cada composto NN?* Na maioria das vezes sim, pois basta identificar qual substantivo está sendo caracterizado e qual exerce o papel de modificador. Causam dúvidas os casos em que o núcleo não é o N2, e os compostos aditivos não apresentam um núcleo.
- *O núcleo em língua inglesa sempre será o N2?* Geralmente, o núcleo é o N2, mas em um dos compostos analisados o núcleo é o N1. Em *water ice*, o referente é a água e não o gelo.
- *Quais as características dos compostos em que não é possível identificar um núcleo?* De acordo com os dados do *corpus* analisado, isso ocorre com compostos aditivos, como em *island home*, e com os compostos lexicalizados.
- *Qual a relação entre produtividade e transparência semântica? De que forma isso é expresso por meio do uso de expressões compostas no corpus de estudo?* Uma das formas de identificar se um composto é transparente é procurando-se por outras construções semelhantes, com o mesmo núcleo. No *corpus* analisado, há várias expressões em que o modificador de *tree* indica um produto que é produzido pela árvore, confirmando o *template* sugerido por Ryder (1994), como em *banana trees*, *fig trees*, *fruit tree* e *hazelnut tree*. Se o composto é produtivo, há grande possibilidade de ele ser transparente. Dessa forma, há uma forte relação entre a produtividade e a transparência semântica, conforme afirmado por Smarsaro (2004).

Ficam em aberto, para serem retomadas na próxima seção, as questões concernentes às nominalizações e aos compostos endocêntricos.

## 4.2 Padrões semânticos

O principal objetivo desta seção é apresentar a análise da semântica dos 200 compostos nominais do *corpus National Geographic* de forma que se possa chegar a um grupo de padrões semânticos, ou seja, relações semânticas recorrentes entre os elementos dos compostos NN em língua inglesa. Para chegar a este fim, partiu-se de algumas relações consideradas em trabalhos de outros autores (conforme capítulo 2). A análise apresentada parte do princípio de que a semântica dos dois substantivos contribui para a semântica do composto como um todo. Para auxiliar esta análise, também contou-se com os dados presentes na base lexical *FrameNet*. Por meio dos resultados da análise, é possível expor quais as relações semânticas presentes nos compostos do *corpus* de estudo, oferecendo-se uma lista de padrões recorrentes.

Assim como na seção anterior, apresentam-se aqui algumas questões de pesquisa, retomando-se alguns tópicos abordados em 4.1 que ainda merecem maior discussão e sugerindo-se outros questionamentos:

- Quais as relações semânticas encontradas nos compostos do *corpus National Geographic*? Algumas relações são mais frequentes do que outras? Como estabelecer limites entre as categorias?
- Quais os compostos cuja interpretação semântica por meio da relação de hiponímia é satisfatória?
- Como o verbo de origem pode contribuir para a interpretação semântica nos compostos formados a partir de nominalizações?
- De que forma a base de dados do projeto *FrameNet* pode contribuir para uma análise co-composicional da semântica dos compostos?

Partindo das conclusões referentes ao estudo teórico e dos *templates* sugeridos por Ryder (1994), chegou-se a um conjunto de 26 relações, como as de função, instrumento, material, local, tempo, animal, humano, substância, artefato, produto, planta, parte do corpo, arma etc. (todas as relações estão esquematizadas na tabela 15). Iniciou-se com uma classificação para cada substantivo a partir dos dados encontrados nos *templates* de Ryder (1994) e em outros estudos, como Pustejovsky (1995), Warren (1978 *apud* RYDER, 1994), Marchand (1969 *apud* RYDER, 1994) e Levi (1978 *apud* DOWNING, 1977).



Além dessas relações, considerou-se importante incluir informações mais específicas, como doença, vírus, fonte de energia, possuidor, possuído e profissão. A partir do predicado CAUSE de Levi (1978 *apud* DOWNING, 1977), percebeu-se que os exemplos sempre traziam algo de negativo, como uma doença. Durante a análise, foi possível encontrar compostos em que a relação semântica é de causa, como em *skin cancer*. Ao realizar-se uma busca no *FrameNet* pelo item lexical *cancer*, chegou-se ao *frame Medical\_conditions* (condições médicas), que possui os seguintes elementos *frame*: doença, paciente, parte do corpo, causa e grau. Os dados do *FrameNet* contribuíram para uma maior compreensão sobre a relação semântica dos compostos. Assim, utilizando-se a etiqueta CAUSE, pode-se parafrasear o composto *skin cancer* como *um câncer causado na pele*.

Um tipo de composto apresentado por Ryder (1994) chamou a atenção, sendo exemplificado por *steamboat*. Nesse caso, o N1 indica a fonte de energia do barco. Ryder (1994) não apresenta essa relação entre os seus *templates*, mas nesse composto há uma relação semântica diferente das outras, pois não parece possível encaixá-lo em nenhum outro padrão. Durante a análise, houve apenas um caso desse tipo, *life force*, ou seja, a força vem da vida, do corpo de alguém.

A relação entre posse e possuído, mesmo podendo ser interpretada por meio do papel constitutivo, pareceu ser mais bem descrita separadamente. Um *family car* é um carro possuído pela família, mas soa estranho dizer que o carro faz parte da família. Entre os exemplos encontrados em Warren (1978 *apud* RYDER, 1994) sobre os compostos aditivos, há apenas casos em que o referente possui duas profissões. Procura-se, neste trabalho, aumentar o número de características de substantivos que podem fazer parte de um composto aditivo.

Os dados sobre cada substantivo encontrado no *FrameNet* também trazem informações desse tipo. Além disso, cada palavra, ou elemento lexical, faz parte de um *frame* que está relacionado com alguns verbos. É o verbo que vai instanciar o *frame*, relacionando os substantivos. Ryder (1994) também utiliza verbos para expressar a relação entre os substantivos de um composto. Ao optar por um verbo, chegou-se a uma única palavra que expressa a semântica do composto e que pode servir como uma etiqueta semântica a ser utilizada em tarefas de processamento da língua.

Na primeira coluna da tabela 15, a seguir, vê-se cada verbo que caracteriza a relação semântica do composto e, na coluna da direita, têm-se as características dos substantivos que podem exercer essas relações. Como não há uma direção fixa para cada relação, já que a localização de algo pode ser expressa tanto no N1 quanto no N2, não foi especificada a

direção nesta “chave” que serve de guia para a análise. Após a análise dos dados, será retomada a questão da direção (caracterizando tanto o N1 como o N2 de cada relação), expondo-se como os dados do *corpus* apresentam esse aspecto.

**TABELA 15:** Chave de análise dos compostos

<b>Relação / Verbo</b>	<b>Baseada no(s) estudo(s)</b>	<b>Características dos substantivos</b>	<b>Exemplo</b>
serve para	Jespersen (1909), Marchand (1969), Adams (1973), Warren (1978), FOR de Levi (1978), papel télico de Pustejovsky (1995)	artefato – função	<i>wineglass</i>
dá origem a	Jespersen (1909), FROM de Levi (1978), papel agentivo de Pustejovsky (1995)	instrumento – x	<i>gunshot</i>
causa	CAUSE de Levi (1978), papel agentivo de Pustejovsky (1995)	substância, vírus – doença	<i>drug death</i>
vem de, é feito a partir de	FROM de Levi (1978), papel agentivo de Pustejovsky (1995)	substântica, artefato – origem	<i>olive oil</i>
funciona a partir de	USE de Levi (1978), papel agentivo de Pustejovsky (1995)	x – fonte de energia	<i>voice vote</i>
é feito de	Jespersen (1909), Marchand (1969), Adams (1973), papel constitutivo de Pustejovsky (1995), MADOF de Ryder (1994)	artefato – material	<i>gold ring</i>
possui	Jespersen (1909), Marchand (1969), Adams (1973), Warren (1978), HAVE de Levi (1978), papel constitutivo de Pustejovsky (1995), HAS de Ryder (1994)	todo – parte integrante	<i>stone fruit</i>
possui	Warren (1978), papel constitutivo de Pustejovsky (1995)	todo – característica abstrata	<i>car speed</i>
contém	CON de Ryder (1994), papel constitutivo de Pustejovsky (1995)	recipiente – conteúdo	<i>matchbox</i>
é localizado em	Jespersen (1909), Marchand (1969), Adams (1973), Warren (1978), IN de Levi (1978), LOC de Ryder (1994)	pessoa, artefato, animal, planta – local	<i>field mouse</i>
ocorre em	Marchand (1969), Warren (1978), IN de Levi (1978)	pessoa, artefato, animal, planta – tempo	<i>summer people</i>
tem	Warren (1978), Ó Séaghdha (2007)	possuidor – possuído	<i>family car</i>
é parecido com	Marchand (1969), Adams (1973), BE de Levi (1978), LIK de Ryder (1994)	animal, flor – animal/flor	<i>bell-flower</i>
é um	Warren (1978), BE de Levi (1978)	hipônimo – hiperônimo	<i>man-servant</i>
o referente é os dois ao mesmo tempo	Warren (1978), BE de Levi (1978), EQUIB de Ryder (1994)	profissão – profissão	<i>poet-painter</i>
o composto é um tipo de núcleo	papel formal de Pustejovsky (1995), compostos endocêntricos	modificador – núcleo	<i>tuna fish</i>
cria, cuida, treina	TEND de Ryder (1994)	humano – animal doméstico	<i>horse man</i>
dirige, opera	RIDE/OPER de Ryder (1994)	humano – veículo, máquina	<i>boatman</i>

usa para trabalhar	WORKW de Ryder (1994)	humano – instrumento, ferramenta	<i>cameraman</i>
usa	USE/USE BY de Ryder (1994)	humano – arma	<i>axman</i>
faz, vende	MAK/SELL de Ryder (1994)	humano – produto	<i>flower girl</i>
é usado em	WEAR de Ryder (1994)	produto – parte do corpo	<i>face cream</i>
come, caça	EAT/HUNT de Ryder (1994)	animal carnívoro, animal	<i>bird dog</i>
é usado no animal	USE ON de Ryder (1994)	produto – animal doméstico	<i>dog collar</i>
é usado em	USE ON de Ryder (1994)	produto – x	<i>baby powder</i>
produz	MAK de Ryder (1994)	planta – produto	<i>apple tree</i>

Algumas das relações apresentadas na tabela 15 não trazem muitas características sobre o substantivo. Quando não é possível trazer características específicas sobre o substantivo que pode ocupar o lugar de um dos elementos, utiliza-se a letra X. A intenção é que após a análise possam-se descrever os elementos que ocupam a posição X no *corpus*. Há também duas relações reservadas para os casos de hiponímia: uma em que a relação entre os substantivos é bidirecional e a outra em que o composto como um todo é um hipônimo do seu núcleo. É importante destacar a diferença entre as duas relações. Dentro da etiqueta É UM, encontraram-se os casos em que se pode afirmar que o N1 é o N2 ou que o N2 é o N1. Encontrou-se apenas um composto desse tipo no *corpus*, *family group*, pois podemos dizer que família é um tipo de grupo, ou seja, o hipônimo de grupo (N1 é N2). Já na relação em que o composto é um tipo de núcleo, temos compostos já comentados anteriormente, como *fruit bat*, que é um tipo de morcego, e *rain forest*, um tipo de floresta (NN é um tipo de N2).

Para a análise, foram utilizadas duas expressões semelhantes, artefato e produto, que apresentam algumas diferenças e merecem um esclarecimento quanto ao uso a ser dado. Por produto, entende-se que seja “qualquer coisa cuja produção ou desenvolvimento é influenciado por pessoas” (RYDER, 1994, p. 101). Dessa forma, Ryder (1994) inclui no grupo dos produtos palavras como *banana*, *cloak*, *cane*, *scarves*, ou seja, tanto itens naturais quanto sintéticos que são produzidos por pessoas. Já quando nos referimos a artefatos, o que queremos ressaltar é a característica de que um objeto foi criado para desempenhar uma função específica. De qualquer forma, o artefato é um tipo de produto, mas reserva-se a palavra produto para um item natural, enquanto o artefato está diretamente relacionado com um item sintético que possui uma função.

A seguir, serão apresentados alguns padrões semânticos recorrentes nos compostos do *corpus* de estudo. Na tabela 16, é apresentada a frequência de cada relação encontrada no

*corpus*.<sup>31</sup> O número de ocorrências refere-se aos compostos diferentes, sem contar os casos de repetições. Por exemplo, mesmo que *memory drugs* tenha ocorrido três vezes no *corpus*, para esta tabela, ele foi contabilizado uma única vez.

**TABELA 16:** Relações encontradas no *corpus*

<b>Relação / Verbo</b>	<b>Types</b>	<b>Exemplo</b>
serve para	25	<i>memory drugs</i>
dá origem a	2	<i>car accidents</i>
é causada em	2	<i>brain infection</i>
vem de, é feito a partir de	9	<i>cane juice</i>
funciona a partir de	1	<i>life force</i>
é feito de	7	<i>metal armor</i>
possui – parte integrante	21	<i>church floor</i>
possui – característica abstrata	5	<i>island culture</i>
contém	7	<i>rice bag</i>
é localizado em	40	<i>school play</i>
ocorre em	9	<i>night school</i>
tem	11	<i>family mosque</i>
é parecido com	1	<i>rock salt</i>
é um	1	<i>family group</i>
o referente é os dois ao mesmo tempo	3	<i>car bombs</i>
o composto é um tipo de núcleo	49	<i>palm trees</i>
cria, cuida, treina	1	<i>cattle guards</i>
dirige, opera	0	X
usa para trabalhar	0	X
usa	0	X
faz, vende, corta, compra	2	<i>cane cutter</i>
é usado em	2	<i>head scarves</i>
come, caça	0	X
é usado no animal	0	X
é usado em	0	X
produz	2	<i>banana trees</i>

Pode-se dividir cada padrão em grupos de acordo com a natureza dos elementos: pessoas, animais, plantas, artefatos, substâncias e eventos. Cada um destes padrões é apresentado nas subseções a seguir.

<sup>31</sup> No anexo D deste trabalho, os compostos analisados são agrupados de acordo com a sua relação semântica. São apresentadas também algumas características específicas de cada relação.

#### 4.2.1 Padrões de pessoas

O substantivo que acompanha *pessoas* em um composto pode se referir à função dessa pessoa, ao local de trabalho ou de origem, ao grupo do qual alguém faz parte, a algo que é possuído ou a um tipo de pessoa. A seguir, podem ser observados exemplos de cada um dos casos:

- função + pessoas: *monitoring groups* = um grupo de pessoas cuja função é monitorar algo;
- local + pessoa: *country boy* = menino cuja origem é o campo; *lab people* = pessoas que trabalham em um laboratório;
- grupo + pessoa: *family member* = pessoa que faz parte deste grupo;
- pessoa + posse: *family mosque* = mesquita que pertence à família;

O “tipo de pessoa” pode indicar a profissão da pessoa (*construction workers*), a sua área de atuação (*rocket scientist*) ou algum vício que possui (*drug users*). É importante lembrar que as conclusões apresentadas nesta seção dizem respeito somente ao *corpus* analisado (*National Geographic*) e que a proposta aqui não é estabelecer generalizações sobre a língua. Os *templates* descritos aqui são os que foram identificados no *corpus*, e acredita-se que uma análise mais ampla possa chegar a outros padrões não descritos aqui. Destaca-se que há um outro caso específico em que uma pessoa é vista como parte de um grupo ou pertencente a um grupo que desempenha uma função específica, como *monitoring groups*, ou seja, o N1 expressa o papel télico do composto.

#### 4.2.2 Padrões de animais

As construções com animais trazem informações sobre o local onde o animal vive, partes do seu corpo, o animal como membro de um grupo, possuidor de algo e alguns tipos de animais. Exemplifica-se cada relação a seguir:

- local + animal: *desert beetle* = besouro que vive no deserto;
- animal + parte do corpo: *leopard skin* = pele de leopardo;
- animal + grupo: *gorilla families* = família da qual o gorila faz parte;

- animal + posse: *gorilla health* = saúde do animal.

Em relação a tipos de animais, o modificador traz alguma característica específica do bicho, como *fruit bat*, um morcego que se alimenta de frutas. Uma ocorrência no *corpus* chamou a atenção por não estar dentro de nenhum padrão recorrente: *chimp feces*. Não se pode dizer que o chimpanzé é o dono ou possuidor das fezes que produziu; entende-se que o animal é a fonte, a origem das fezes. Só se encontrou um caso desse tipo.

#### 4.2.3 Padrões de plantas

Nos compostos formados por plantas também foi possível identificar alguns padrões recorrentes, sendo estes: função, parte de um grupo, possuidor de partes, o seu local de origem ou cultivo e tipo. Esses padrões foram esquematizados da seguinte forma:

- função + planta: *feed corn* = milho destinado à alimentação de animais;
- planta + grupo de plantas: *eucalyptus forest* = floresta constituída por eucaliptos;
- planta + parte: *tree branches* = galhos que fazem parte da árvore;
- local + planta: *water plants* = plantas que nascem na água;
- planta + local de cultivo: *orchid farm* = fazenda onde orquídeas são cultivadas.

Quanto ao último esquema apresentado, que se refere a plantas que são cultivadas em um determinado local, buscou-se essa informação no *FrameNet*. Ao realizar-se a consulta pelo item lexical *farm*, obteve-se a informação de que ele faz parte do *frame Locale\_by\_use* (localização pelo uso). Esse *frame* possui como elementos principais um local e o seu uso, ou seja, o local é descrito a partir da forma como ele é usado. Voltando ao exemplo *orchid farm*, não se pode dizer que a fazenda é a localização natural das orquídeas, mas que a fazenda é um local utilizado para o cultivo, a plantação de orquídeas. Com o auxílio do *FrameNet*, foi possível identificar tipos diferentes de local, o local destinado a um uso específico e o local como localização, que informa onde algo se encontra.

Por tipos de plantas, deparou-se basicamente com tipos de árvores no *corpus*, sendo estas *cedar tree* e *palm trees*. Pode-se afirmar que cedro e palmeiras são tipos de árvores, mas não há como identificar alguma relação específica entre *cedar*, *palm* e *trees*, já que o N1, *cedar* e *palm*, refere-se à árvore por si mesma.

#### 4.2.4 Padrões de artefatos

O grupo dos artefatos é o que mais tem recebido a atenção dos estudos sobre a semântica dos compostos (conforme capítulo 2). Acredita-se que a grande quantidade de pesquisas quanto à natureza dos artefatos deve-se ao fato de que são inúmeras as relações que eles podem estabelecer com outros substantivos. Nos compostos no *corpus* paralelo, foram encontradas as seguintes relações acerca dos artefatos: função, material, parte, localização quanto ao uso e tipo. A relação mais frequente é a de função, ocorrendo na seguinte ordem:

- função + artefato: *car keys* = chaves usadas para abrir o carro.

Um aspecto interessante quanto à função dos artefatos e que vai ao encontro das questões propostas aqui refere-se ao fato de que, no caso das nominalizações, é o verbo de origem da palavra que expressa, na maioria das vezes, a função do artefato. A seguir, alguns exemplos do *corpus*:

- *Building blocks* são peças utilizadas para construir algo (*to build* – construir);
- *Commuter car* é um carro que serve para o transporte diário (*to commute* – ir ao trabalho utilizando um transporte);
- *Education center* é um local, um centro destinado a atividades educacionais (*to educate* – educar).

A partir desses exemplos, acredita-se que, em grande parte dos casos de nominalização, o verbo de origem do substantivo carrega informações relevantes para a interpretação semântica do composto como um todo. Além dessa relação, os compostos formados por artefatos também podem trazer as seguintes informações:

- material + artefato: *stone axes* = machados feitos de pedra;
- artefato + parte: *computer keyboard* = o teclado faz parte do computador;
- artefato + localização: *electronics shops* = lojas onde aparelhos eletrônicos são vendidos.

Como exemplo de tipo de artefato, temos o *cell phone*, pois um telefone celular se diferencia dos outros por apresentar a característica de ser móvel.

#### 4.2.5 Padrões de substâncias

O quinto tipo de padrão identificado no *corpus* diz respeito a substâncias, produto natural ou líquido em que outro elemento do composto pode expressar a sua origem, a sua função, o seu tipo ou se ele é o conteúdo de algum recipiente. Por origem, compreende-se de onde a substância vem ou onde ela foi produzida, sendo que, na maioria das vezes, essa origem refere-se a um tipo de planta. Seguiu-se para os esquemas:

- origem + substância: *corn ethanol* = etanol ou álcool feito a partir do milho;
- função + substância: *cooking oil* = óleo destinado o uso culinário;
- produto + recipiente: *rice bag* = saco que contém arroz.

Entre os tipos de substâncias, líquidos ou produtos, os compostos em que a relação entre os substantivos é mais geral, encontraram-se no *corpus* os seguintes: *water ice* e *warming seas*. Esses dois compostos são considerados hipônimos do seu núcleo, pois a água congelada é um tipo de água, a água na sua forma sólida, e mares aquecidos também trazem uma característica específica, a sua temperatura.

#### 4.2.6 Padrões de eventos

Por fim, puderam ser identificados padrões semânticos relacionados a eventos. Informações como local e tempo em que um evento ocorre aparecem nos compostos NN. Além disso, há também tipos de eventos. Estes são alguns exemplos de compostos eventivos encontrados no *corpus*:

- evento + local: *baseball field* = um jogo que ocorre em um lugar específico, o campo;
- tempo + evento: *weekend class* = aula que ocorre no final de semana.



*Opion trade, ethanol production e computer classes* são tipos de eventos, pois os substantivos *trade, production e classes* denotam uma série de ações específicas em contextos variados, como comércio, indústria e aula.

Feita a análise dos compostos, podem ser retomadas as questões apresentadas no início desta seção. Os comentários acerca de cada questão são apresentados em tópicos:

a) **Compostos endocêntricos – características e frequência.** Considerando que buscou-se realizar nesta seção uma análise mais profunda da semântica dos compostos, informar que um composto é endocêntrico significa que ele estabelece uma relação mais geral com outro item lexical. Um *fruit bat* é um composto endocêntrico, pois refere-se a um tipo específico de morcego, que é diferenciado dos outros por meio de uma característica particular, ele come frutas. No entanto, se fossem especificadas todas as características que são informadas pelos modificadores dos compostos endocêntricos do *corpus*, haveria uma quantidade enorme de relações semânticas no interior do compostos e correr-se-ia o risco de confrontação com diversas interpretações possíveis para um mesmo composto. Assim, partiu-se das relações semânticas apontadas como mais recorrentes em outros estudos, e os casos em que não é possível estabelecer uma relação específica entre os elementos foram considerados endocêntricos. Os compostos que constituem uma instância do seu núcleo trazem no modificador características bem específicas, tais como:

- o assunto, sobre o que se trata determinado evento: *biomimetics research, carbon emissions, conservation efforts, water pollution, math class*;
- uma subclasse ou exemplar que faz parte da classe: *snack food, palm trees, law school, fire brigade*;
- órgão ou grupo social responsável por uma determinada instituição: *community schools, state banks, city government, city parks*.

Entre os 200 compostos, 46 foram classificados apenas como endocêntricos, fato que indica que a maioria dos compostos analisados, mais de 75%, apresenta alguma relação semântica mais específica entre os seus elementos, tais como função, localização e parte/todo. Refletindo-se sobre a afirmação de Pustejovsky *et al.* (1993) de que os compostos endocêntricos são mais frequentes em textos especializados, de áreas específicas do conhecimento, chegou-se à conclusão de que o *corpus* da revista *National Geographic* é bastante variado. Encontraram-se alguns compostos de áreas especializadas como da biologia,

em relação a árvores, animais, plantas e pesquisas. Porém, há casos de palavras de uso geral, como telefone celular, lanche e número de telefone. De qualquer forma, a maior quantidade de compostos endocêntricos pode ser associada a alguma área do conhecimento, conclusão que confirma a hipótese de Pustejovsky *et al.* (1993) sobre a sua ocorrência.

b) **Nominalizações – contribuição do verbo de origem para a interpretação.** O verbo de origem do elemento nominalizado do composto geralmente traz alguma informação referente ao significado do composto. Neste estudo, uma nominalização como *cooking oil* é interpretada pela relação *serve para*, em que o N1 é a função e o N2 é o produto. Se a interpretação fosse mais específica, sugerindo uma paráfrase diferente para cada composto, teríamos algo como “óleo que serve para cozinhar”. O verbo *cook* (cozinhar) é a função do núcleo do composto. Como o objetivo deste trabalho não é chegar a esse tipo de detalhamento, listam-se apenas algumas das relações entre as nominalizações e o significado do composto:

- função: *building blocks, commuter car, cooking oil, drainage system, education center, nursing home, particle accelerator*;
- pessoa que possui algo: *workers rights, cattle owners*;
- o cargo de alguém em um local de trabalho: *plant director, construction workers, cane cutter*;
- um evento que é caracterizado de alguma forma pelo outro elemento: *wood construction, winter mountaineering, summer climbing, space exploration, judgment day*;
- o uso que é dado a determinado local: *storage facility, recycling company, manufacturing center*.

Se fosse aumentada a quantidade de dados analisados ou se analisasse outro *corpus*, provavelmente as interpretações possíveis para cada caso de nominalização seriam diferentes. A grande questão é que, por se tratar de verbos, somente uma base lexical construída a partir dos dados de um *corpus* extremamente amplo poderia prever as diversas relações possíveis entre as nominalizações e os substantivos em casos de composição.

c) **Relações semânticas – frequência e limites das categorias.** Nos 200 compostos analisados da revista *National Geographic*, algumas relações semânticas apareceram com maior frequência, conforme dados da tabela 16, sendo estas: o composto é um tipo de núcleo (49), é localizado em (40), serve para (25), possui (26), tem (11), vem de/é feito a partir de (9), ocorre em (9), é feito de (7) e contém (7). Essas relações talvez tenham sido mais frequentes porque, nas edições da revista, era necessário diferenciar os membros de determinados grupos através de alguma característica importante. Isto não impede que estas mesmas relações sejam encontradas em outros *corpora* mais gerais.

Como boa parte dos compostos traz alguma informação sobre eventos, parece lógico que algumas informações sejam mais frequentes. Quando pensamos em um evento, há alguns elementos que são geralmente associados a eles, como tempo (ocorre em) e local (é localizado em). Viu-se anteriormente que a função (serve para) também é descrita muitas vezes por meio de um evento.

Há vários artefatos entre os compostos analisados, e o que diferencia um do outro, além da sua função, pode ser o material do qual é feito e as suas partes constituintes. Como os artigos da revista *National Geographic* constituem um gênero que privilegia assuntos relacionados a ciência ou biologia, há uma grande quantidade de compostos endocêntricos e de informações sobre certas substâncias, como a sua origem ou qual o recipiente em que elas são tipicamente armazenadas. Outro uso frequente encontrado no *corpus* é também encontrado na língua geral, em que se unem dois substantivos em um composto na língua inglesa para expressar a relação de posse.

As seguintes relações não ocorreram no *corpus*: dirige/opera, usa para trabalhar, usa (arma), come/caça, é usado no animal e é usado em. Todas elas são inspiradas no trabalho de Ryder (1994) e estão relacionadas com padrões específicos sobre pessoas, animais e produtos. Não foi possível identificá-las entre os 200 compostos analisados, mas não se descarta que elas possam ocorrer em um estudo mais amplo.

A relação possui foi bastante frequente, ocorrendo 26 vezes nos dados (21 vezes com todo e parte integrante e cinco vezes com uma característica abstrata), mas conseguiram-se identificar apenas quatro casos de posse de algo abstrato, como *island culture* e *mountain air*. Por meio de uma busca rápida no conjunto total de compostos extraídos do *corpus*, é possível encontrar diversos casos de posse abstrata. Vemos essa relação em *animal cognition*, *chimp behaviors*, *construction stages*, *desert temperatures*, *family tradition*. Assim, mesmo havendo dois casos entre os 200 analisados, esse tipo de relação é bastante frequente no *corpus* todo.

d) **O uso do *FrameNet* para uma análise co-composicional.** Mesmo que a base de dados do *FrameNet* não tenha por objetivo estabelecer relações entre os elementos de uma expressão multivocabular, a consulta aos *frames* de que certos elementos lexicais fazem parte contribuiu para a presente análise.

Uma das contribuições diz respeito aos tipos de relações de local. Conforme já visto, o *frame* relacionado a *farm* é o *Locale\_by\_use* (localização pelo uso), e aplicável para uma *orchid farm*, que é o local destinado à cultura de orquídeas. De forma semelhante, o elemento lexical *field* também pode fazer parte do *frame Locale\_by\_use* se nos referirmos a um campo que serve para o cultivo de alguma espécie, como em *rice fields*. No entanto, *field* também evoca outro *frame*, *Locale\_by\_event* (localização por evento), em que a função desse local é de servir de “palco” para um determinado evento, como em *baseball field*.

Outro *frame* considerado para esta análise foi *Medical\_conditions* (condições médicas). Partiu-se da ideia de que a relação *causar* relaciona uma doença com o seu causador, podendo ser uma substância ou um vírus. Porém, esse *frame* pode possuir outros elementos, conforme os dados do *FrameNet*, sendo estes: doença, paciente, parte do corpo, causa, grau, sintoma e nome próprio da doença. Nos dados analisados, a relação identificada foi entre parte do corpo e doença. *Brain infection* é uma doença que afeta uma parte do corpo, o cérebro, e em *skin cancer*, a relação também é a mesma, a doença afeta a pele.

O uso do *FrameNet* contribuiu para que as relações semânticas fossem analisadas mais detalhadamente. Por meio das ricas descrições dos *frames*, puderam-se relacionar mais facilmente os elementos dos compostos.

#### 4.3 Tradução dos compostos para a língua portuguesa

Nesta seção, são analisados os dados do *corpus* paralelo, a fim de se identificarem os correspondentes de tradução dos compostos em língua inglesa para o português. O objetivo dessa comparação é identificar de que forma a semântica dos compostos em inglês sugere certa previsibilidade quanto à estrutura em português dessas expressões. Não se buscou refletir sobre questões de tradução. Esta análise foi inspirada no trabalho de Busa e Johnston (1999), que comparam os compostos do inglês e do italiano a partir das relações semânticas.

Ao considerarem-se as possíveis traduções dos compostos NN da língua inglesa para o português, nem sempre se encontra um equivalente de tradução formado por dois substantivos na língua portuguesa. De uma forma geral, os compostos em português são formados por um núcleo sucedido por um modificador. As classes desses elementos são variadas, ocorrendo

inclusive casos de NN, como em exemplos de Sandmann (1997): *fortaleza-símbolo* e *data-limite*. O autor também traz outras combinações possíveis: V+N (*pega-rapaz*), N+A ou A+N (*bola-murcha*, *pouca-telha*), N+NUM ou NUM+N (*camisa-dez*, *sete-barbas*), N+de+N (*pé-de-meia*, *pé-de-galinha*), entre outros. De acordo com os resultados de um estudo preliminar (TEIXEIRA; CHISHMAN, 2008) realizado a partir dos dados de um *corpus* paralelo, o COMPARA (FRANKENBERG-GARCIA; SANTOS, 2002), constatou-se que os principais equivalentes de tradução dos compostos NN do inglês para o português são: N+de+N, N+A ou um único substantivo. Com base nesse primeiro experimento, procurou-se verificar aqui se os equivalentes são os mesmos do estudo anterior, se há alguma combinação nova e até que ponto pode-se afirmar que as traduções obedecem a algum padrão.

A ordem de dois substantivos em um composto em português é semelhante à sequência de formada por um substantivo seguido por um adjetivo. O adjetivo é usado após o substantivo, como em *mulher bonita*, *casa cara* e *homem pobre*. No composto o núcleo é o primeiro elemento, que é seguido pelo seu modificador – *seguro-desemprego*. No entanto, há casos em que o modificador ocorre antes do núcleo, como em *egitomania*. Para Sandmann (1997), os compostos que apresentam a estrutura de modificador e núcleo são hoje tão produtivos quanto os outros, formados por núcleo e modificador.

A tradução dos compostos do *corpus National Geographic* para a língua portuguesa apresenta uma estrutura diferente do inglês, ou seja, se o composto é formado por dois substantivos em inglês, raramente possui um correspondente de tradução também formado por dois substantivos. Nesta seção, os equivalentes de tradução para o português são analisados quanto à sua estrutura. Foram selecionadas as traduções dos compostos tomando-se como base as oito relações semânticas mais frequentes no *corpus*: *serve para*, *vem de/é feito a partir de*, *é feito de*, *possui (parte/todo)*, *contém*, *é localizado em*, *ocorre em* e *tem (posse)*. O objetivo é identificar se há alguma regularidade quanto à tradução dos compostos para a língua portuguesa de acordo com a relação semântica expressa entre os substantivos. Para se realizar a análise, consultaram-se as concordâncias de cada composto no *corpus* paralelo. Cada ocorrência, com a frase de origem do composto em inglês e a sua tradução, é apresentada no anexo E deste trabalho. Na próxima tabela, são expostos os dados gerais com as frequências de cada padrão de tradução subdividido pelas relações semânticas. Assim como no estudo preliminar de Teixeira e Chishman (2008), feito a partir de dados de outro *corpus*, os equivalentes de tradução mais frequentes em português para os compostos NN do inglês são N de N (com artigo ou não), totalizando 91 das 165 ocorrências analisadas; Nadj com 28 das ocorrências; e também há casos em que o composto em português é traduzido com um

único substantivo, 16 casos. Outras possibilidades menos frequentes de tradução envolvem as estruturas N para N, N em N, N d' N, N V N, N de V e N para V. Cada caso é relacionado com cada relação semântica entre os elementos do compostos e a análise é apresentada a seguir.

A análise feita é dividida de acordo com cada relação semântica selecionada para este estudo e a quantidade de ocorrências tem como base os *tokens*, ou seja, considera-se também as repetições dos compostos, já que, para cada ocorrência, uma tradução diferente pode ter sido feita. Desta forma os números de frequência informados nesta seção são diferentes dos números da tabela 16 (seção 4.2).

**TABELA 17:** Dados gerais das frequências de padrões de tradução

<b>Padrão</b>	<b>Relações semânticas</b>	<b>Total de ocorrências</b>
N de N	SERVE PARA – 14 VEM DE, É FEITO A PARTIR DE – 12 FEITO DE – 6 POSSUI – 17 CONTÉM – 1 É LOCALIZADO EM – 13 OCORRE EM – 6 TEM – 1	70
N adjetivo	SERVE PARA – 13 VEM DE, É FEITO A PARTIR DE – 6 POSSUI – 3 É LOCALIZADO EM – 4 OCORRE EM – 1 TEM – 1	28
N de+artigo N	SERVE PARA – 4 VEM DE, É FEITO A PARTIR DE – 1 POSSUI – 5 É LOCALIZADO EM – 2 OCORRE EM – 2 TEM – 7	21
N	SERVE PARA – 4 VEM DE, É FEITO A PARTIR DE – 2 POSSUI – 5 CONTÉM – 1 É LOCALIZADO EM – 3 TEM – 1	16
SEM TRADUÇÃO	SERVE PARA – 4 POSSUI – 5 É LOCALIZADO EM – 5 TEM – 2	16
N para N	SERVE PARA – 4 VEM DE, É FEITO A PARTIR DE – 1	5

N em N	FEITO DE – 1 POSSUI – 1 OCORRE EM – 1	3
N d' N	CONTÉM – 2	2
N de V	SERVE PARA – 2	2
N V N	É LOCALIZADO EM – 1	1
N para V	SERVE PARA – 1	1

a) **Papel télico – Relação SERVE PARA.** Entre os 25 tipos de compostos analisados, totalizando 46 ocorrências, a tradução<sup>32</sup> da maioria foi feita relacionando-se os dois substantivos por meio da preposição *de* (*peças de montar, grupos de monitoramento*) ou traduzindo-se o modificador como um adjetivo (*ferramentas agrícolas, árvore funerária*). Em alguns casos a preposição é acompanhada de um artigo (*lar dos idosos, chave do carro*). Refletindo-se sobre essas opções, podem ser feitas algumas considerações:

- I. das 18 construções N de N (com ou sem artigo), quatorze são formadas por um substantivo eventivo deverbal (*nursing home, drainage system*). Parece que há uma tendência em utilizar a preposição “de” nesses casos;
- II. entre as traduções não deverbais para N de N, as quatro restantes, um dos casos chama a atenção: *food crop*. *Crop* é também considerado um eventivo em sua tradução, *produção de alimentos*;
- III. nos casos em que preposição é utilizada com um artigo definido (do, dos), as quatro ocorrências, as mesmas expressões também poderiam ocorrer sem o artigo, como de fato ocorre na tradução *lar de idosos*, sem grande alteração no seu sentido geral;
- IV. as doze ocorrências em que os substantivos modificadores do inglês são traduzidos por um adjetivo em português (*árvore funerária, uniforme escolar*) também aceitariam ser unidas pela preposição *de*, como em *árvore de funeral* ou *uniforme de escola*;
- V. outra preposição também identificada nas traduções é *para*, provavelmente por estar diretamente relacionada com a função télica do modificador; uma *business license* é uma licença para abrir negócios, este é o seu propósito;
- VI. outras opções que estão relacionadas com a telicidade do modificador são aquelas em que um verbo é utilizado para expressar a relação entre os dois substantivos em português (*alimento usado na alimentação de animais*).

<sup>32</sup> Todas as traduções mencionadas nesta seção correspondem às traduções encontradas na versão em português do *corpus*.

Em alguns casos, o tradutor optou por traduzir apenas o núcleo, deixando o modificador de fora. Isso provavelmente se deu por não ser comum expressar uma relação desse tipo em português. *Mountain bicycle* é traduzido apenas como bicicleta, provavelmente porque não nos é costumeiro aqui no Brasil nos referirmos a um tipo de bicicleta utilizada para subir montanhas. Da mesma forma o correspondente para *memory drugs* é *drogas*. Não temos uma expressão comum e frequentemente utilizada para nos referirmos a um tipo de medicamento que auxilie pessoas com problema de memória. Mesmo que pudesse ter sido utilizada uma paráfrase explicando essa relação, a partir do contexto, pois se trata de uma reportagem sobre a memória, o leitor pode inferir que se trata de um medicamento para a memória.

A seguir, é apresentada uma tabela com a frequência em que cada uma das opções de tradução ocorre no *corpus*:

**TABELA 18:** Ocorrências da relação SERVE PARA<sup>33</sup>

<b>Exemplo</b>	<b>Estrutura</b>	<b>Frequência</b>	<b>Natureza dos Elementos</b>
acelerador de partícula	N de N	14	Função de pessoas, função de plantas, função de artefatos, função de substâncias.
centro educacional	N adjetivo	13	Função de plantas, função de artefatos, função de locais, função de substâncias.
botas dos astronautas	N de+artigo N	4	Função de artefatos, função de locais.
lar para idosos	N para N	4	Função de locais, função de substâncias.
mochila	N	4	Função de artefatos, função de substâncias.
peças de montar	N de V	2	Função de artefatos.
licença para abrir negócios	N para V	1	Função de artefatos.

Considerando-se a natureza dos elementos dos compostos para cada estrutura, pode-se identificar que a estrutura N de N ocorre com todos os cinco tipos de funções, de pessoas, plantas, artefatos, substâncias e locais. Nas três ocorrências em que um verbo é usado após uma preposição em português, tratam-se de funções de artefatos, ou seja, indicam a atividade na qual o objeto é utilizado.

<sup>33</sup> Quatro compostos não foram traduzidos na versão em português da revista.



É por meio das escolhas feitas pelo tradutor que se percebe se a telicidade expressa pelo modificador é relevante para o texto em português. Em boa parte dos equivalentes de tradução, a telicidade é o que motiva a criação das expressões; quando nos referimos às botas dos astronautas não estamos nos referindo a qualquer tipo de botas, mas somente àquelas que são utilizadas por astronautas. Se o tradutor se refere a uma *day pack* apenas como mochila, parece que informar que há um tipo de mochila que é usada diariamente não é relevante para um leitor brasileiro, por exemplo. As questões culturais influenciam muito nas escolhas do tradutor. Nós, brasileiros, não conseguiríamos classificar uma árvore como uma *árvore-caixão*. No entanto, para um grupo de pessoas, na sua cultura a árvore que fornecerá a madeira para a construção do seu caixão é escolhida logo que a criança nasce.

b) **Relação VEM DE, É FEITO A PARTIR DE.** Assim como no caso dos compostos téllicos, a maioria das traduções de compostos que expressam a relação de origem, de um total de 22 ocorrências (9 compostos diferentes), é formada com a preposição *de* entre os dois substantivos (*etanol de milho*) ou por um substantivo seguido de um adjetivo (*óleo vegetal*). Ao se analisarem as construções utilizadas para expressar as relações entre os componentes dos compostos, concluiu-se o seguinte:

- I. a construção mais frequente é N de N, totalizando 12 ocorrências, e em nenhum dos casos ocorre uma nominalização;
- II. em todos os casos em que *plant* foi usado como modificador, a sua tradução foi um adjetivo, *vegetal* (*óleo vegetal, equivalentes vegetais*); no entanto, usar uma preposição seguida de substantivo também seria possível, como *óleo de plantas*;
- III. o único composto em que a preposição *de* é usada com um artigo, *caldo da cana*, também poderia ser usado sem o artigo;
- IV. em *corn ethanol*, o milho é a origem, de onde o etanol é feito, e é um caso semelhante a *álcool de cana*, tanto que, em duas traduções, esse composto aparece como *etanol de milho*. No entanto, em dois dos equivalentes de tradução, a opção do tradutor foi expressar a relação de telicidade entre o modificador e o núcleo utilizando a preposição *para* ou uma oração indicando a função do milho: *milho para a produção de etanol* e *milho destinado à produção de etanol*.

Essas informações podem ser vistas na tabela 19, a seguir:

**TABELA 19:** Ocorrências da relação VEM DE, É FEITO A PARTIR DE

<b>Exemplo</b>	<b>Estrutura</b>	<b>Frequência</b>	<b>Natureza dos Elementos</b>
óleo de canola	N de N	12	Origem (planta) de substância, origem (animal) de substância, origem de substância.
combustíveis vegetais	N adjetivo	6	Origem (planta) de substância.
etanol	N	2	Origem (planta) de substância, origem de substância.
caldo da cana	N de+artigo N	1	Origem (planta) de substância.
milho para a produção de etanol	N para N	1	Origem (planta) de substância.

Ao observarmos a tabela 19, vemos que a outra opção de tradução possível é a escolha por um único substantivo para traduzir o composto, como *cana* e *etanol*. No entanto, a partir do contexto, o leitor pode saber que o etanol é de milho e a cana a que se refere o texto é de açúcar.

c) **Material – Relação É FEITO DE.** Quanto às construções em que o material do qual algo é feito, quase todas as traduções apresentam a estrutura N de N, 6 de 7. Na única tradução diferente, N em N, o núcleo é um eventivo. Na tabela 20, a seguir, os dois casos são exemplificados:

**TABELA 20:** Ocorrências da relação É FEITO DE

<b>Exemplo</b>	<b>Estrutura</b>	<b>Frequência</b>	<b>Natureza dos Elementos</b>
casas de pedra	N de N	6	Material do artefato.
construção em madeira	N em N	1	Material do artefato.

d) **Meronímia – Relação POSSUI.** Foram analisadas 36 ocorrências (26 compostos diferentes) em que ocorre a relação de parte e todo. Cinco desses casos não foram traduzidos na versão em português da revista. Na maioria dos compostos traduzidos, o equivalente de tradução apresenta a construção N de N, como em *pele de leopardo* e *tela de computador*. Quanto às outras ocorrências, pode-se afirmar o seguinte:

- I. tanto *teclado do computador* como *solo da igreja* foram traduzidos com o artigo definido porque o referente é também definido, pois não se trata de qualquer computador ou de qualquer igreja. Isso é confirmado por meio das seguintes ocorrências:

A missa da véspera de Natal em Belém, transmitida mundialmente em 24 de dezembro, na verdade é celebrada em uma igreja bem mais nova ao lado da Natividade: a igreja de Santa Catarina, que é dos católicos romanos. (...) Não importa a versão do cristianismo que se siga - ou mesmo que não se siga religião nenhuma -, parece haver algo significativo na gruta no subsolo da igreja, com seu cheiro de incenso e vela, iluminada por uma fiavel de lâmpadas nuas.

Bella ensina aos pais as gírias mais recentes e os apresenta a sites bacanas na internet. (...). Quando a garota era menor, eles passaram a desconectar o teclado e o mouse do computador para que ela não ficasse online sozinha em casa.

Nesses casos, a igreja é a de Santa Catarina e o computador é o de Bella, ou de sua família;

- II. apenas dois compostos foram traduzidos utilizando-se um adjetivo, *células cerebrais* e *grupos guerrilheiros*, que também poderiam ter sido traduzidos como N de N;
- III. os equivalentes de tradução formados por um único substantivo (*monitores*, *parentes*, *gorilas*, *primavera*, *vagões*) podem ser facilmente compreendidos pelo leitor, pois é possível estabelecer a relação com o outro substantivo por meio do contexto. Como a reportagem da revista trata de questões sobre os resíduos tecnológicos, ao se referir a monitor, nesse caso, só pode ser o monitor do computador.

**TABELA 21:** Ocorrências da relação POSSUI

<b>Exemplo</b>	<b>Estrutura</b>	<b>Frequência</b>	<b>Natureza dos Elementos</b>
ossos de animais	N de N	17	Parte do corpo de animal, parte de planta, parte de artefato, animal como membro de um grupo, planta como membro de um grupo, parte de tempo, parte de local, característica abstrata de pessoa.
teclado do computador	N de+artigo N	5	Parte de artefato, parte de local, característica abstrata de local.
células cerebrais	N adjetivo	3	Pessoa como membro de um grupo, parte de órgão, característica abstrata de pessoa.
monitores	N	5	Parte de artefato, animal como membro de um grupo, pessoa como membro de um grupo, parte de tempo.
alguém na sua família	N em N	1	Pessoa como membro de um grupo.

Observando-se a tabela 21, chama a atenção a estrutura N em N, pois esta ocorre uma única vez. Conferindo-se a ocorrência em inglês, imagina-se que esta foi uma opção do tradutor, pois poderia ter sido traduzida simplesmente como membro da família: *Yes, I think if*

*you have a family member with dementia you really understand how memory loss manifests in a person.*

e) **Conteúdo e recipiente – Relação CONTÉM.** Foram analisados sete casos em que essa relação pode ser identificada e duas delas não foram traduzidas. Essa expressão que ficou de fora da versão em português poderia ter sido facilmente traduzida, pois *ethanol tanks* são tanques de etanol. Consideremos a tabela 22, a seguir:

**TABELA 22:** Ocorrências da relação CONTÉM

<b>Exemplo</b>	<b>Estrutura</b>	<b>Frequência</b>	<b>Natureza dos Elementos</b>
saco de arroz	N de N	1	Recipiente e conteúdo.
garrafas d'água	N d' N	2	Recipiente e conteúdo.
carretas transportando gado	N V N	1	Recipiente e conteúdo.
cocho	N	1	Recipiente e conteúdo.

A construção que ocorreu duas vezes (N d' N) se deu apenas com o modificador *água* (*garrafas d'água* e *caixas-d'água*), provavelmente porque o substantivo *água* permite essa construção. Já a palavra *cocho* na sua versão em inglês refere-se a *tigela de água para animais*, conforme a seguinte ocorrência: *He filled a water bowl for the animals and scattered some food.* Não destoando dos outros tipos de compostos analisados até então, uma das ocorrências apresenta a estrutura N de N, *saco de arroz*.

f) **Relação É LOCALIZADO EM.** A relação de local é a mais numerosa no *corpus*, com 40 tipos diferentes de compostos. Dessa forma, foram selecionados 28 ocorrências para este estudo, dos quais cinco não foram traduzidos. A estrutura mais frequente nos compostos que indicam uma localização é N de N (*fazenda de algas*, *usinas de etanol*, *posto de gasolina*). Um composto como *posto de gasolina* é interpretado como um local (o posto) pelo seu uso (onde a gasolina é vendida). Considerando os dados da tabela 23 e as ocorrências, pode-se afirmar o seguinte:

- I. entre os 13 compostos traduzidos com a estrutura N de N, três apresentam algum elemento eventivo na sua construção em português, sendo eles: *empresa de produção de etanol*, *unidade de produção de etanol* e *campo de batalha*;
- II. além de *rapaz do interior*, o outro composto em que a preposição *de* é usada com um artigo é *águas do pântano*;

- III. a segunda construção mais frequente entre os compostos que indicam localização é com um substantivo seguido de adjetivo, sendo os seguintes: *incêndio florestal*, *expedição lunar*, *puma americano* e *centros manufactureiros*;
- IV. assim como nos outros casos analisados anteriormente, quando o composto é traduzido com apenas um substantivo, pode-se interpretá-lo a partir do contexto. *Posto* refere-se ao *posto de gasolina*, assim como os *técnicos* são os *técnicos de laboratório*.

**TABELA 23:** Ocorrências da relação É LOCALIZADO EM

Exemplo	Estrutura	Frequência	Natureza dos Elementos
quadra de basquete	N de N	13	Localização pelo uso, localização por evento, origem de pessoas.
rapaz do interior	N de+artigo N	2	Origem de pessoas, local de substância.
lojas que vendem equipamentos eletrônicos	N V N	1	Localização pelo uso.
posto	N	3	Localização pelo uso, origem de pessoas.
incêndio florestal	N adjetivo	4	Localização por evento, localização de algo, onde animal vive.

Um dos casos que chamam a atenção é a tradução de *electronics shops* com uma construção contendo um verbo, *vender*. Mesmo que se tenha utilizado um verbo para descrever o local, não há como negar que se trata de uma localização caracterizada pelo seu uso, uma loja onde eletrônicos são vendidos.

g) **Relação OCORRE EM.** Entre as 10 ocorrências em que a relação é de tempo (do total de 9 tipos), a maioria é traduzida com a estrutura N de N (*montanhismo de inverno*, *dias de festejos*, *escaladas de verão*). O que mais chama a atenção nesses compostos, inclusive nos em que *de* é seguido por um artigo, é que todos se referem a eventos. O tempo informa quando esses eventos ocorrem. Dois desses compostos são deverbais em inglês, *montanhismo de inverno*, do verbo *climb*, e *dia de julgamento*, do verbo *judge*. O único composto com N de N que não traz um elemento eventivo é *grama de verão*, no entanto podemos interpretar que o composto se refere à grama que cresce no verão, ou seja, o evento é relacionado de alguma forma.

**TABELA 24:** Ocorrências da relação OCORRE EM

<b>Exemplo</b>	<b>Estrutura</b>	<b>Frequência</b>	<b>Natureza dos Elementos</b>
dia de julgamento	N de N	6	Quando um evento ocorre, evento que ocorre em período determinado, planta que cresce em um período específico.
hora da sesta	N de+artigo N	2	Quando um evento ocorre.
aulas nos fins de semana	N em N	1	Evento que ocorre em período determinado.
escola noturna	N adjetivo	1	Local que funciona em horário determinado.

Os dois últimos casos apresentados na tabela 24 ocorrem apenas uma vez. Assim como as construções com a preposição *de* indicam quando um evento ocorre, a preposição *em* também é utilizada com esse fim. Já em *escola noturna*, o adjetivo também informa quando essa escola funciona, porém o período em que as aulas ocorrem pode estar relacionado com a modalidade de ensino oferecida e outras construções são possíveis. Nos dados do *corpus COMPARA*, uma das traduções para *night school* é *supletivo*; ou seja, uma escola que tem aulas à noite pode trazer outras características, além do tempo. Um supletivo é uma modalidade de ensino em que geralmente alunos com mais idade completam o estudo em um tempo reduzido.

h) **Relação TEM.** O que mais chama a atenção nos equivalentes de tradução dos compostos que indicam posse é que o uso da preposição com o artigo é o caso mais frequente (*amigo da família, mesquita da família, prédios do governo, direitos dos trabalhadores*). Provavelmente o uso do artigo se deu porque, quando se refere a família, é uma família específica, assim como o governo. No entanto, ao se referir aos direitos dos trabalhadores, a ideia é que o grupo de todos os trabalhadores possui direitos.

**TABELA 25:** Ocorrências da relação TEM

<b>Exemplo</b>	<b>Estrutura</b>	<b>Frequência</b>	<b>Natureza dos Elementos</b>
família de amigos	N de N	1	Pessoa que possui alguma relação com outra pessoa ou pessoas.
casa da família	N de+artigo N	7	Possuidor e possuído, pessoa que possui alguma relação com outra pessoa ou pessoas.
donos	N	1	Possuidor e possuído.
residência grupal	N adjetivo	1	Possuidor e possuído.

Foram analisados 12 ocorrências de compostos que indicam algum tipo de posse (de um total de 11 tipos), sendo que dois deles não foram traduzidos, *police sirens* e *police cars*, que podem ser entendidos como sirenes da polícia e carros da polícia ou viaturas.

#### 4.3.1 Sumário da análise do *corpus* paralelo

Retomando-se os resultados da análise dos equivalentes de tradução feita nesta seção, pode-se concluir que a construção N de N é de fato a mais utilizada por tradutores humanos para expressar a relação entre os elementos de um composto em língua portuguesa, totalizando 91 (incluindo as expressões com artigos) das 165 ocorrências analisadas, conforme dados da tabela 17. Chama a atenção o fato de que os eventivos, tanto télicos quanto temporais, são relacionados por meio dessa preposição. Na realidade, a preposição *de*, devido ao seu aspecto polissêmico, pode expressar quaisquer uma das relações analisadas nesta seção: função, origem, material, parte, conteúdo, local, tempo e posse. Uma das questões que ainda merece atenção é conseguir identificar em quais casos esta preposição é acompanhada por um artigo definido. Conclui-se que, em alguns casos, isso acontece porque o referente é velho, já é conhecido pelo leitor. No entanto, nem sempre isso ocorre.

Outra opção de tradução bastante frequente é a utilização de um adjetivo no lugar do modificador. Analisando todos os compostos traduzidos, percebe-se uma preferência em utilizar um adjetivo quando há um correspondente em português. São poucos os casos em que o adjetivo poderia ter sido usado, mas optou-se por um substantivo precedido pela preposição, como *rapaz do interior*, em vez de *rapaz interiorano*, *dia de primavera*, em vez de *dia primaveril*, e *caroços de frutas*, em vez de *caroços frutíferos*. A escolha por uma construção N de N pode estar relacionada com o objetivo do tradutor em utilizar expressões de forma natural para um leitor brasileiro. *Rapaz interiorano*, embora exista em nosso léxico, não é uma expressão muito recorrente.

Em quase todas as relações entre os elementos dos compostos, houve traduções em que o composto foi traduzido como uma palavra simples. No entanto, esta escolha não afeta a compreensão do leitor. Além de ser possível inferir o significado da palavra a partir do contexto, ao utilizar apenas um substantivo, há menos repetições vocabulares no texto, aspecto que é considerado importante nos textos em português, mas que em inglês não constitui um problema.

Esta tarefa serviu como um exercício de análise de como a semântica dos compostos nominais em língua inglesa pode ser expressa em português. Lembre-se que os dados aqui

analisados referem-se somente a um *corpus* específico, constituído por 10 edições da Revista *National Geographic*. Ainda há muito trabalho a ser feito nessa área, como analisar quais as estruturas utilizadas em textos de domínios específicos ou explorar uma forma de representar a polissemia das preposições, em especial da preposição *de*.

Concluiu-se que de fato a preposição *de* em português pode representar diversas relações, mas outros recursos também são frequentemente utilizados, como o uso de adjetivos ou de substantivos simples. Outro aspecto que parece ser determinante nas escolhas do tradutor é a influência cultural. Muitas vezes é este o fator que determina como os dois substantivos do composto em inglês serão traduzidos para o português.



## 5 CONSIDERAÇÕES FINAIS

O estudo da semântica dos compostos nominais é tema interdisciplinar, pois é recorrente em trabalhos da área da linguística e da computação. A grande diferença está na forma como cada área aborda este fenômeno linguístico. Enquanto os estudos semânticos sugerem que os tipos de relação entre os elementos de um composto NN são infinitos e não buscam esgotar as possibilidades de interpretação, os trabalhos na área de PLN procuram identificar um grupo limitado de relações semânticas procurando dar conta senão de todos, da maioria dos compostos NN. Neste trabalho, a partir do aprofundamento de diferentes perspectivas teóricas, chegou-se a uma proposta de análise dos compostos nominais. Esta proposta partiu das diferentes relações semânticas responsáveis pela combinação dos elementos, mas sem ser estanque, pois parte do princípio de que os dois substantivos exercem um papel específico a partir do seu uso.

Para expor as conclusões gerais deste estudo, as questões de pesquisa apresentadas na introdução são retomadas e comentadas a seguir.

*Quais as relações semânticas mais recorrentes entre os elementos de compostos NN, conforme estudos já realizados?* A partir da análise dos estudos clássicos sobre a composição em língua inglesa, constatou-se que algumas relações são comuns. Seleccionamos as relações mais recorrentes e as que possuem alguma relação com as teorias adotadas para este estudo: telicidade, agentividade, meronímia, localização, posse, comparação, cópula, adição e hiponímia.

*Considerando-se a diversidade de perspectivas teóricas para tratar da semântica dos compostos NN, que tipologia é descritiva e explanatoriamente adequada para dar conta da composicionalidade semântica dessas construções?* Além de identificar a relação entre os dois substantivos que fazem parte do composto, mostrou-se necessário, para uma análise composicional, explicitar as características destes elementos. Tendo em conta que tanto o N1 quanto o N2 fazem parte de um esquema ou *frame*, nada mais lógico do que expressar esta relação através de verbos. A proposta de análise deste trabalho é constituída por vinte e seis relações expressas através dos seguintes verbos: *serve para, dá origem a, causa/é causada, vem de/é feito a partir de, funciona a partir de, é feito de, possui, contém, é localizado em, ocorre em, tem, é parecido com, é um, o referente é o N1 e o N2, o composto é um tipo de núcleo, cria/cuida/treina, dirige/opera, usa para trabalhar, usa, faz/vende/compra, é usado em, come/caça, é usado em, é usado no animal e produz.*

*De que forma as ferramentas relacionadas ao estudo de corpus podem ser utilizadas para uma análise semântica dos compostos nominais?* Inicialmente, ao trabalhar com o *corpus* em inglês, o etiquetador, o itemizador e o extrator de seqüências formadas por dois substantivos serviram para uma extração automática dos candidatos a compostos. Em um segundo momento, o itemizador e o alinhador serviram para a construção de um *corpus* paralelo, instrumento que facilitou a identificação dos equivalentes de tradução dos compostos. Estas duas etapas serviram para que fosse possível se chegar aos dados a serem analisados. Com a utilização do concordanceador do *WordSmith Tools* e do *corpus* paralelo foi possível analisar a semântica dos compostos através de diversos recursos: frequência de ocorrência dos compostos e de *core words* no *corpus*; o contexto de uso de cada expressão tanto em inglês como em português; busca pelas ocorrências dos compostos. De fato, a utilização de todos estes recursos, em especial a identificação automática dos candidatos a compostos, contribuiu para que se chegasse aos dados de análise de forma eficiente. Outro aspecto proporcionado pelo uso destas ferramentas diz respeito à forma facilitada de identificação dos compostos e suas traduções no *corpus*. A possibilidade de se ter acesso ao contexto maior de cada ocorrência, ou seja, a parte do texto onde o composto é usado, também facilitou muito a tarefa de análise semântica.

*Quais as relações semânticas mais frequentes encontradas nos compostos nominais do corpus paralelo analisado?* As relações mais frequentes identificadas são nove: *é um tipo de* (49<sup>34</sup>), *é localizado em* (40), *possui* (26), *serve para* (25), *tem* (11), *vem de/é feito a partir de* (9), *ocorre em* (9), *é feito de* (7) e *contém* (7). Os compostos cujo núcleo refere-se a um humano não apareceram no *corpus*; os compostos que trazem alguma especificidade sobre plantas e animais, por sua vez, foram frequentes. Ao longo das dez edições da revista analisada, os autores das reportagens precisaram caracterizar ou nomear diferentes tipos de animais ou plantas, mas o mesmo não ocorreu quanto aos seres humanos.

A partir desta reflexão proporcionada pelas questões de pesquisa, outras considerações sobre o trabalho realizado merecem ser feitas. A escolha por uma revista para constituir o *corpus* paralelo se deu por dois motivos, a mesma edição da *National Geographic* em inglês é traduzida em sua íntegra para o português, e em cada exemplar há mais de um gênero textual, como entrevistas, reportagens, depoimentos, etc. Esta opção pareceu bastante acertada, pois foi possível extrair inúmeros compostos que são de uso cotidiano de falantes das duas línguas.

---

<sup>34</sup> O número que aparece após cada relação refere-se a quantidade de tipos encontrados no *corpus*.

Mesmo que a variedade de tipos de compostos estivesse condicionada às temáticas abordadas pela revista, foi possível testar a tipologia proposta a partir destas ocorrências.

Uma das grandes inovações deste trabalho foi relacionar uma teoria baseada em *templates*, como a de Ryder (1994), que propõe uma interpretação semântica para os compostos nominais, com a teoria de *frames* semânticos de Fillmore (2006). Essa associação foi feita durante a análise semântica dos compostos e o uso da base de dados *FrameNet* foi apresentado na seção 4.2. A consulta à base de dados do *FrameNet* trouxe informações baseadas no uso para confirmar a tipologia proposta neste trabalho. Mesmo que os *frames* não considerem as relações entre os elementos de expressões compostas, foi possível utilizar os conceitos da base de dados do *FrameNet* para confirmar as relações semânticas dos compostos deste estudo.

Quanto à análise das estruturas utilizadas para expressar as relações semânticas dos compostos em língua portuguesa, os resultados foram bastante diferentes dos de Busa e Johnston (1999), cujo trabalho, em que os compostos em inglês foram comparados com a sua tradução em italiano, serviu de inspiração para esta parte da pesquisa. Não foi possível identificar uma única estrutura em língua portuguesa que expressasse uma relação semântica específica. As estruturas em português mais utilizadas são N de N, N adjetivo ou apenas N. Estes resultados se justificam pelos seguintes motivos: a preposição *de* é extremamente polissêmica e pode expressar diversas relações entre os elementos que ela relaciona; como a construção NN em português não é tão produtiva como em inglês, muito frequentemente um adjetivo é utilizado como modificador; por questões de variedade lexical, algo que é expresso em inglês por duas palavras pode possuir um correspondente em português formado por apenas uma. As diferenças culturais expressas através da língua justificam muitas das opções dos tradutores.

Este trabalho trouxe sugestões de análise semântica para os compostos nominais em inglês, além de fazer uma reflexão sobre os correspondentes de tradução utilizados em um contexto específico, uma revista. No entanto, este estudo ainda pode ser ampliado. Um dos temas que merece aprofundamento é a polissemia das preposições. Outro diz respeito ao potencial dos *frames* para expressar as relações semânticas dos compostos ou até mesmo como fatores culturais interferem nas opções de tradução. Além de sugerir uma análise semântica bilíngue, espera-se que esta dissertação sirva de inspiração para outros estudos que contribuam para a solução de problemas computacionais relativos às expressões multivocabulares e a produção de materiais multilíngues.

## REFERÊNCIAS BIBLIOGRÁFICAS

AZEREDO, Suzana de. **Expressões Anunciadoras de Paráfrase em Manuais Acadêmicos de Química: Um Estudo Baseado em Corpus**. Porto Alegre: UFRGS, 2007. Dissertação de Mestrado, Instituto de Letras, Universidade Federal do Rio Grande do Sul, 2007.

BAKER, Collin F.; FILLMORE, Charles J.; LOWE, John B. The Berkeley FrameNet project. In: **Proceedings of the COLING-ACL**. Montreal, Canada, 1998.

BALDWIN, Timothy, BANNARD, Colin, TANAKA, Takaaki, WIDDOWS, Dominic. An Empirical Model of Multiword Expression Decomposability. In: **Proceedings of the ACL 2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment**, 2003.

BASILIO, Margarida. **Teoria Lexical**. 7. ed. São Paulo: Ática, 2004.

BEAUGRANDE, Robert de. **Functionalism and corpus linguistics in the 'Next Generation'**. 2000. Disponível em: <<http://www.beaugrande.com/Functionalism%20and%20Corpus%20Linguistics.htm>>. Acesso em: 20 jun. 2007.

BERBER SARDINHA, Antonio Paulo. Linguística de *corpus*: Histórico e problemática. **DELTA**, São Paulo, v. 16, n. 2, p. 323-367, 2000.

BERBER SARDINHA, Tony. Tamanho de Corpus. **The ESpecialist**, São Paulo, . 23, n. 2, p. 103-122, jul./dez. 2002.

BIBER, Douglas. Using Register-Diversified Corpora for General Language Studies. **Computational Linguistics**, Cambridge, v. 19, n. 2, p. 219-241, jun. 1993.

BOLINGER, Dwight. **Aspects of Language**. New York: Harcourt Brace Janovich, 1975.

BORBA, Francisco S. **Uma gramática de valências para o português**. São Paulo: Ática, 1996.

BUSA, Federica; JOHNSTON, Michael. Qualia Structure and the Compositional Interpretation of Compounds. In: VIEGAS, Evelyne (org.). **Breath and Depth of Semantic Lexicons**. London: Kluwer, 1999. p. 167-187.

CHISHMAN, Rove Luiza de Oliveira. A teoria do léxico gerativo: uma abordagem crítica. In: IBAÑOS, Ana Maria T.; SILVEIRA, Jane Rita Caetano (org.). **Na Interface Semântica/Pragmática**. Porto Alegre: EDIPUCRS, 2002. p. 51-82.

CHURCH, Kenneth Ward, HANKS, Patrick. Word Association Norms, Mutual Information, and Lexicography. **Computational Linguistics** Volume 16, Number 1, March 1990.

COMPARA. Disponível em: <<http://www.linguateca.pt/>> Acesso em: 22 dez. 2007.

COPESTAKE, Ann. Compounds revisited. In: **2<sup>nd</sup> International Workshop on Generative Approaches to the Lexicon, GL'2003**. Geneva, maio, 2003. CD-ROM.

COPESTAKE, Ann, LASCARIDES, Alex. Integrating Symbolic and Statistical Representations: The Lexicon Pragmatics Interface. In **Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics**, 1997.

CRUSE, D. A. **Lexical Semantics**. Cambridge: Cambridge, 1986.

DANIELSSON, Pernilla, RIDINGS, Daniel. **Practical Presentation of a 'Vanilla' Aligner**. Presentation held at the TELRI Workshop in alignment and exploitation of texts in Ljubljana, 1997.

DOWNING, Pamela. On the creation and use of English compound nouns. **Language, Journal of the Linguistic Society of America**, Baltimore, v. 53, n. 4, p. 810-842, dez. 1977.

FILLMORE, Charles J. Frame Semantics. In: GEERAERTS, Dirk. **Cognitive Linguistics: Basic Readings**. Berlin, Nova Iorque: Mouton de Gruyter, 2006. p. 373-400.

FRANKENBERG-GARCIA, Ana, SANTOS, Diana. COMPARA, um corpus paralelo de português e inglês na Web. In: TAGNIN, Stella E. O. (Org.). **Cadernos de Tradução: Corpora e Tradução**. Florianópolis: NUT, 2002, v. 1, n. 9, p. 61-79. Disponível em: <<http://www.cadernos.ufsc.br/online/9/ana.htm>>

GALE, William A., CHURCH, Kenneth W. A program for aligning sentences in bilingual corpora. In.: **Computational Linguistics**, n. 19, v. 1, p. 75-102, 1993.

GIRJU, Roxana, NAKOV, Preslav, NASTASE, Vivi, SZPAKOWICZ, Stan, TURNEY, Peter, YURET, Deniz. SemEval-2007 Task 04: Classification of Semantic Relations between Nominals, 2007. In: **Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007)**. Prague, p. 13-18, 2007.

GRANGER, Sylviane (org.). **Learner English on computer**. New York: Longman, 1998. 228p.

GRICE, Paul. Logic and Conversation. In **Philosophical Logic**, Oxford, Oxford University press, 1971, p. 39-48.

HALLIDAY, M. A. K. **An introduction to functional grammar**. 2. ed. London: Arnold, 1994.

JACKENDOFF, Ray. **The Architecture of the Language Faculty**. London: MIT, 1997.

KAY, Martin. Multilinguality. In: COLE, Ronald (ed.). Nova Iorque: Cambridge University Press. **Survey of the state of the art in human language technology**, 1996.

LANGACKER, Ronald W. **Foundations of cognitive grammar**. Volume I: Theoretical prerequisites. Standford: Standford University, 1987. 540p.

LAPATA, Mirella, LASCARIDES, Alex. Detecting novel compounds: The role of distributional evidence. In: **Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics**, p. 235-242, 2003.

LIEBER, Rochelle. **Morphology and Lexical Semantics**. Cambridge: Cambridge University, 2004. 196p.

MARTINS, Ronaldo Teixeira, PELLIZONI, Jorge Marques, HASEGAWA, Ricardo, VOLPE NUNES, Maria das Graças. Da tradução automática para a língua portuguesa: apontamentos de três experiências baseadas em interlíngua. In: **Palavra**, Rio de Janeiro, n. 12, p. 164-171, 2004.

MCDONALD, Scott. **Learning Compound Order: Towards a Functional Explanation**. Edinburgh, Scotland: 1995. Dissertação de Mestrado, Centre for Cognitive Science, University of Edinburgh, 1995. 48p.

MCENERY, Tony, WILSON, Andrew. **Corpora and Translation: Uses and Future Prospects**. Technical Report from the Unit for Computer Research on the English Language (UCREL). Lancaster, University of Lancaster, 1993. Disponível em: <<http://ucrel.lancs.ac.uk/papers/techpaper/vol2.pdf>>

**National Geographic Magazine**, novembro de 2007, National Geographic Society, disponível em: <<http://ngm.nationalgeographic.com/ngm/2007-11/tableofcontents.html>> Acesso em: 24 dez. 2007.

Ó SÉAGHDHA, Diarmuid. Annotating and Learning Compound Noun Semantics. In: **Proceedings of the ACL 2007 Student Research Workshop**. Prague, p. 73-78, 2007.

Ó SÉAGHDHA, Diarmuid. **Annotation Guidelines for Compound Noun Semantics**. Material não publicado. Disponível em: <<http://www.cl.cam.ac.uk/~do242/guidelines.pdf>> Acesso em: 02 jan. 2008.

PUSTEJOVSKY, James. The Generative Lexicon. **Computational Linguistics**, v. 17, n. 4, p. 409-440, dez. 1991.

\_\_\_\_\_. **The Generative Lexicon**. Cambridge: MIT, 1995. 298p.

PUSTEJOVSKY, James, BERGLER, Sabine, ANICK, Peter. Lexical Semantic Techniques for Corpus Analysis. In: **Computational Linguistics**, Association for Computational Linguistics, v. 19, n. 2, p. 331-358, 1993.

**Revista National Geographic Brasil**, novembro de 2007, editora abril, disponível em: <[http://nationalgeographic.abril.uol.com.br/home/index\\_0711.shtml](http://nationalgeographic.abril.uol.com.br/home/index_0711.shtml)> Acesso em: 24 dez. 2007.

RYDER, Mary Ellen. **Ordered Chaos: The Interpretation of English Noun-Noun Compounds**. Berkeley: University of California, 1994. 449p.

SAG, Ivan, A. *et al.* Multiword Expressions: A Pain in the Neck for NLP. In: **Proceedings of the Third International Conference on Computational Linguistics and Intelligent Text Processing**, p.1-15, fev. 2002.

SANDMANN, Antônio José. **Morfologia Geral**. 3. ed. São Paulo: Contexto, 1997. 79p.

SANTORINI, Beatrice. **Part-of-Speech Tagging Guidelines for the Penn Treebank Project**. Pennsylvania: Department of Computer & Information Science, 1990. Relatório Técnico.

SCHMID, Helmut. Probabilistic Part-of-Speech Tagging Using Decision Trees. In: International Conference on New Methods in Language Processing (NeMLaP-1), 1994, London. **Proceedings**. London: USL Press, 1994, p. 44-49.

SCOTT, Mike. **WordSmith Tools 5.0**. Disponível em: <<http://www.lexically.net/downloads/version5/HTML/index.html>> Acesso em: 22 dez. 2007.

SMARSARO, Aucione. Um estudo de palavras compostas com estrutura N de N para processamento automático. **Revista Palavra**, Rio de Janeiro, n. 12, p.164-171, 2004.

TEIXEIRA, Lílian Figueiró; CHISHMAN, Rove Luiza de Oliveira. Um estudo do cópua COMPARA: a semântica dos compostos nominais. In: COSTA, Luis; SANTOS, Diana; CARDOSO, Nuno. (Ed.). **Perspectivas sobre a Linguatca**: Actas do encontro Linguatca: 10 anos. Porto: Linguatca, 2008, p. 35-41.

VALE, Oto Araújo. **Expressões cristalizadas do português do Brasil**: uma proposta de tipologia. Araraquara: UNESP, 2001. Tese (Doutorado em Linguística e Língua Portuguesa), Faculdade de Ciências e Letras, Universidade Estadual Paulista Julio Mesquita Filho, 2001. 213p

**VANILLA ALIGNER**. Alinhador online de textos paralelos. CEPRIL, LAEL, PUC/SP. Disponível em: <<http://www2.lael.pucsp.br/corpora/alinhador/index.html>> Acesso em: 20 dez. 2007.

## ANEXO A – Traduções livres feitas pela autora deste trabalho

[Real tennis] club: [tênis autêntico] clube  
1948 season: temporada de 1948  
Accident at (an) airshow: acidente em um show aéreo  
Afternoon clerk: secretário da tarde  
Afternoon sun: sol da tarde  
Airshow accident: acidente no show aéreo  
Altarpiece: enfeite de altar  
Angler fish: peixe sapo  
Angora cat: gato angorá  
Angora rabbit: coelho angorá  
Ankle socks: meia soquete  
Ankle supporter: suporte do tornozelo  
Apple cake: bolo de maçã  
Apple orchard: pomar  
Apple pie: torta de maçã  
Apple tree: macieira  
Apple-juice seat: assento do suco de maçã  
Arms budget: orçamento para armas  
Attorney general: procurador geral  
Axman: machadeiro  
Baby powder: talco para bebê  
Baby-spoon: colher do bebê  
Bag lunch: almoço de pacote  
Banana fork: garfo da banana  
Barbershop: barbearia  
Bargeman: barqueiro  
Baseball field: campo de baseball  
Baseball personality: personalidade do baseball  
Basket woman: mulher da cesta  
Battle fatigue: cansaço da batalha  
Beach plum: ameixa da praia  
Beautiful: bonita  
Bedroom: quarto de dormir  
Bee fly: mosca abelha  
Bee louse: piolho de abelha  
Bee moth: mariposa abelha  
Beehive: colméia  
Beer drinker: bebedor de cerveja  
Bell-flower: flor em forma de sino  
Billboard: quadro de avisos  
Bird brain: cérebro de pássaro  
Bird dog: cão de pássaro  
Birdcage: gaiola  
Birdhouse: casa de pássaros  
Birth pains: dores do parto  
Birth spacing: intervalos de parto  
Biscuit factory: fábrica de biscoito  
Blackmail: chantagem



Blacksmith's hammer: martelo do ferreiro  
Blazing fire: fogo ardente  
Blockhead: cabeça-dura  
Boarhound: tipo de cachorro usado na caça de javalis  
Boathouse: casa barco  
Boatman: barqueiro  
Body stocking: meia utilizada no corpo todo  
Bonehead: cabeça de osso  
Bookcase: estante de livros  
Bowman: arqueiro  
Box bed: cama caixa  
Box containg polysterene: caixa que contém poliestireno  
Box office: bilheteria  
Box stall: banco caixa  
Box wine: vinho de caixa  
Boxcar: vagão  
Brakeman: responsável pelos freios de um trem  
Bread knife: faca de pão  
Breadboard: tábua de cortar pão  
Breastplate: peito de armas  
Broomstick: cabo de vassoura  
Buckhound: cão de caça  
Buffalo fly: mosca de búfalo  
Buffalo wolf: lobo de búfalo  
Bullet hole: buraco de bala  
Bullet wound: ferimento da bala  
Bus money: dinheiro do ônibus  
Bus stop: parada de ônibus  
Bushman: lenhador  
Busman: motorista de ônibus  
Butcher knife: faca do açougueiro  
Butterman: homem da manteiga  
By and large: em geral, geralmente  
C language: linguagem C  
Cabman: motorista de táxi  
Camera gear: mecanismo da câmera  
Cameraman: homem câmera  
Campstool: banco de acampamento  
Candlelight: luz da vela  
Car engine: motor do carro  
Car park: estacionamento  
Carboard box: caixa de papelão  
Cat collar: coleira para gato  
Cat instinct: instinto felino  
Catbird: pássaro gato  
Cattleman: homem que lida com gado  
Chalkboard: quadro negro  
Checkerboard: tábua de dama (jogo)  
Chin strap: faixa presa a um chapéu que passa por debaixo do queixo  
Cigar-case: cigarreira

Cinderman: homem da cinza  
City employee: empregado municipal  
City street: rua da cidade  
Class that teaches relaxation: aula que ensina relaxamento  
Clay pigeon: alvo flutuante para tiros  
Coachman: cocheiro  
Coal scuttle: recipiente para carvão  
Corkboard: aglomerado de cortiça  
Cornfield: milharal  
Countrywoman: mulher do campo  
Cow hair: pelo de vaca  
Cow pasture: pasto das vacas  
Cow shed: galpão das vacas  
Cowman: vaqueiro  
Cranberry bog: pântano de amoras  
Cranberry bush: arbusto de amoras  
Crime investigation: investigação de crime  
Crossbowman: arqueiro  
Cupboard: armário  
Customer account: conta de cliente  
Dairymaid: trabalhadora de leiteria  
Dairyman: trabalhador de leiteria  
Daisy chains: corrente de margaridas  
Day school: escola primária que funciona durante o dia  
Desktop computer: computador desktop  
Disease germ: germes de doença  
Dishpan: bacia para lavar louça  
Dog bed: cama do cachorro  
Dog biscuit: biscoito para cachorro  
Dog brush: escova de cachorro  
Dog chow: comida de cachorro  
Dog collar: coleira  
Dog dish: pote do cachorro  
Dogfish: esqualo, pleurotremado (tipo de tubarão)  
Doghouse: casa de cachorro  
Doughface: pessoa indolente  
Drainage improvement: melhora na drenagem  
Drillman: homem furradeira  
Driver's cab: táxi do motorista  
Drug death: remédio fatal  
Dry goods man: verdureiro  
Dustbin: lixeira  
Dustman: gari  
Earmuffs: protetor de ouvidos  
Earring: brinco  
Earthworm: minhoca  
Egg woman: mulher ovo  
Egghead: intelectual (cabeça de ovo)  
Elbow brace: braçadeira  
Elbow cuffs: proteção de cotovelo

Elephant seal: foca elefante  
Elkhound: tipo de cão  
Evening hours: horas da noite  
Evening song: canção da noite  
Eyeglasses: óculos  
Face cream: creme facial  
Face towel: toalha de rosto  
Facecloth: toalhinha de rosto  
Fairy tale: conto de fadas  
Fairyland: mundo encantado  
Fall colors: cores do outono  
Family car: carro da família  
Fan of jazz: fã de jazz  
Feather-bed: cama de penas  
Ferryman: barqueiro  
Fiberboard: fibra plástica em chapas  
Field mouse: rato do campo  
Finish the cigarette: terminar o cigarro  
Firehouse: posto de corpo de bombeiros  
Fishwife: vendedora de peixes  
Flagstaff: mastro de bandeira  
Flood losses: perdas com a enchente  
Floor vacuum: aspirador de pó  
Flower garden: jardim de flores  
Flower girl: vendedora de flores  
Foot brake: pedal do freio  
Foot pedal: pedal  
Football game: jogo de futebol  
Foothills: montanha ao pé de outras mais altas  
Footstool: escabelo  
Forest hut: cabana da floresta  
Foundation investigation: investigação da fundação  
Fox terrier: fox terrier (tipo de cachorro)  
Foxhound: cão de caça (para a caça de raposas)  
Fruit woman: mulher das frutas  
Fuel oil: querosene  
Garbage man: lixeiro  
Garden-party: festa no jardim  
Garage man: homem da garagem  
Gas station: posto de gasolina  
Gear wheel: roda traseira  
Giraffe bird: pássaro girafa  
Giraffe-cage: jaula da girafa  
Glass door: porta de vidro  
Goat milk: leite de cabra  
Goatboy: menino das cabras  
Gold ring: anel de ouro  
Goldfish: peixe dourado  
Golf-season: temporada de golfe  
Goosegirl: garota dos gansos

Granite outcroppings: afloramento de granito  
Grassland: pasto; savana  
Grocery store: quitanda  
Groceryman: verdureiro  
Gunman: pistoleiro  
Gunshot: tiro  
Hairbrush: pente  
Hairpiece: trança de cabelos postiços (usada em alguns tipos de penteado)  
Hamster wheel: roda do hamster  
Hand brake: freio de mão  
Hand drill: furadeira manual  
Hand soap: sabão de mão  
Handbag: carteira, bolsa de mão  
Handcart: carrinho de mão  
Handkerchief: lenço  
Handsled: trenó de mão  
Hard metal floor: chão de metal duro  
Hardwareman: vendedor de hardware  
Hatchet man: homem da machadinha  
Hayloft: palheiro  
Headband: faixa de cabeça  
Headboard: cabeceira da cama  
Hen house: galinheiro  
Hip boots: botas longas que vão até o quadril  
Home secretary: Ministro do Interior  
Honey bee: abelha de mel  
Honey moon: lua de mel  
Honeybee: abelha  
Horse doctor: veterinário de cavalos  
Horse trough: cocho do cavalo  
Horseback gathering: encontro do lombo do cavalo  
Horseman: cavaleiro  
Horsewoman: amazona  
Hotel desk: recepção do hotel  
House paint: tinta de casa  
Houseboat: casa flutuante  
Houseboy: servente, criado  
Housefly: mosca de casa  
Housemaid: servente, ajudante, empregada  
Humpback: corcunda (costas de corcova)  
Hunter-gatherer: caçador colhedor  
I can't find my bus money: não consigo encontrar o meu dinheiro do ônibus  
Ice bag: bolsa de gelo  
Ice legs: pessoa que se adapta ao gelo  
Ice-cream container: recipiente do sorvete  
Icehouse: fábrica de gelo  
Iceman: vendedor de sorvete; vendedor de gelo  
In short: em resumo  
Jail birds: pássaro de cativoiro  
Jazz fan: fã de jazz

John is scrubbing the fireplace: John está lavando a lareira  
John painted the door: John pintou a porta  
John walked through the door: John caminhou através da porta  
Kangaroo rat: rato canguru  
Keyhole: buraco da fechadura  
Knee boot: botas de cano alto  
Knee pants: corsário  
Knee plate: joelheira  
Knee socks: meias que chegam até a altura do joelho  
Knife on the fridge: faca na geladeira  
Laundry detergent: sabão de roupa  
Law court: tribunal de justiça  
Lemon juice: suco de limão  
Lemon peel: casca de limão  
Linen chest: peito de linho  
Lionheart: coração de leão  
Load tolerance: tolerância de carga  
Mailbox: caixa de correspondencia  
Mailman: carteiro  
Man servant: criado  
Marital sex: sexo conjugal  
Matchbook: caixa de fósforos  
Matchbox: caixa de fosforo  
Meadowlark: tipo de pássaro  
Meatman: açougueiro  
Milkmaid: leiteira, mulher que trabalha na retirada do leite das vacas  
Milkman: leiteiro  
Mole rat: rato toupeira  
Monkey business: negócio duvidoso  
Morning prayer: prece matinal  
Moth hole: buraco da mariposa  
Mother-in-law: sogra  
Mountain laurel: louro da montanha  
Mountain man: homem da monatanha  
Mouse hawk: gavião rato  
Mouse trap: ratoeira  
Mule deer: veado de orelhas longas  
Neck brace: colar cervical  
Neckband: gargantilha  
Neckcloth: cachecol  
Neckerchief: cachecol  
Necklace: colar  
Neckpiece: acessório para o pescoço  
Necktie: gravata  
Newspaperboy: jornaleiro  
Newspaperwoman: jornaleira  
Night school: escola noturna  
Night sky: céu noturno  
Nightclub: clube noturno  
Nightdress: vestido de festa

Nose drops: gotas para o nariz  
Oak tree: carvalho  
Oarsman: remador  
Olive oil: azeite de oliva  
Owl-house: casa da coruja  
Oysterman: homem das ostras  
Paper knife: faca de abrir correspondências  
Pastry man: pasteleiro  
Penman: escritor  
Pet shampoo: xampu de animais  
Picture book: álbum de fotos  
Pigpen: chiqueiro  
Pikeman: lanceiro  
Plastic box: caixa de plástico  
Plastic gun: arma de plástico  
Ploughboy: lavrador  
Plowman: lavrador  
Pocket handkerchief: lenço de bolso  
Pocket knife: canivete  
Poet-painter: poeta-pintor  
Polar bear: urso polar  
Polystyrene box: caixa de poliestireno  
Pond lily: lírio do lago  
Potbelly: pançudo (barriga de pote)  
Poultryman: palestrante  
Prairie dog: cão de pradaria  
Price war: preço da guerra  
Radioman: radialista  
Rat poison: veneno de rato  
Relaxation class: aula de relaxamento  
Reptilian scales: escamas de réptil  
Rice cooker: panela de arroz  
Rifleman: atirador  
Riverboat: barco  
Rock badger: etiqueta da pedra  
Room temperature: temperatura da sala  
Rose fertiliser: fertilizante para rosas  
Rose garden: jardim de rosas  
Rubber duck: pato de borracha  
Rubberneck: turista (pescoço de borracha)  
Sabre-cut: corte de sabre  
Saucepan: broto de soja  
Scarecrow: espantalho  
School dance: escola de dança  
School friends: amigos da escola  
Schoolroom: sala de aula  
Science lesson: aula de ciências  
Scrapman: espantalho  
Sea cow: vaca marinha  
Sea gull: albatroz

Sea horse: cavalo marinho  
Sea legs: pessoa que se adapta ao mar  
Sea lion: leão do mar  
Seahorse: cavalo marinho  
Seaman: homem do mar  
Seaweed: alga marinha  
Secretary-treasurer: secretário-tesoureiro  
Sharp bread knife: faca de pão afiada  
Sharp knife on the fridge: faca afiada na geladeira  
Sheepman: pastor de ovelas  
Shipyards: estaleiro  
Silicon breast: seio de silicone  
Similarity crystal: cristal de similaridade  
Snowman: boneco de neve  
Soccer game: jogo de futebol  
Songbird: pássaro cantor  
Spaceman: astronauta  
Spearman: lanceiro  
Spider monkey: macaco aranha  
Spider wasp: vespa aranha  
Spiderwoman: mulher aranha  
Spoon people: colheres em forma de pessoa  
Sportsman: desportista  
Spring squad: squad de primavera  
Stableboy: menino do estaleiro  
Stamp collection: coleção de selos  
Steam iron: ferro a vapor  
Steel knife: faca de aço  
Steel sword: espada de aço  
Stone fruit: fruta com caroço  
Stone wall: parede de pedra  
Street fights: luta de rua  
Street name: nome da rua  
Student athlete: estudante atleta  
Suitcase: valise  
Summer home: casa de veraneio  
Summer job: emprego de verão  
Summer people: veranistas  
Summertime: verão  
Sunday paper: jornal de domingo  
Surfboard: prancha de surfe  
Swordsman: espadachim  
Tail pipe: tubo de exaustão  
Tailgate: guarda traseira (de caminhão)  
Tax law: imposto  
Taximan: taxista  
Tea pot: pote de chá  
Teaman: homem do chá  
Teapot: bule  
Tear gas: gás de lágrimas

The smoke filled the fireplace: A fumaça encheu a lareira  
Tiger cowrie: tigre cauri  
Tiger shark: tubarão tigre  
Tillerman: lavrador  
Tomato plant: tomateiro  
Toothbrush: escova de dentes  
Townfolk: pessoas da cidade  
Toy box: caixa de brinquedo  
Traffic light: semáforo  
Trainman: homem do trem  
Tree ant: formiga da árvore  
Tree bear: urso da árvore  
Tree house: casa da árvore  
Trolleyman: operador do bonde elétrico  
Truck driver: motorista de caminhão  
Trumpet plant: planta da trombeta  
Tugboat man: rebocador de navios  
Tuna fish: peixe tuna  
Unix operating system: sistema de operação Unix  
Voice vote: voto oral  
Waist-belt: cinto  
Waistcoat: colete; paletó  
Warehouse: depósito  
Wastebasket: lixeira  
Water bug: percevejo aquático  
Water cress: agrião  
Water volume: volume de água  
Weekend guests: convidados do final de semana  
Whitemail: incentivo financeiro para que alguém faça o seu trabalho mais rapidamente e da melhor forma possível.  
Wine bottle: garrafa de vinho  
Wine glass: copo de vinho  
Wineglass: garrafa de vinho  
Winter air: ar do inverno  
Wolfhound: raça de cães  
Woman driver: motorista feminina  
Women officers: oficiais femininas  
Woodshed: cabana de madeira  
Wristband: punho de camisa  
Wristwatch: relógio de pulso  
Yachtsman: dono de iate  
Yard boy: menino da fazenda



ANEXO B – Grupos de compostos do *corpus National Geographic***WATER + N**

water bodies – corpos d'água  
 water bottles – garrafas d'água  
 water bowl – cocho de água  
 water buffalo – búfalo de rio (manso)  
 water column – coluna d'água  
 water conservation – conservação da água  
 water consumption – consumo de água  
 water crisis – crise hídrica  
 water flow – fluxo de água  
 water gorilla – gorila de água  
 water hole – nascente d'água  
 water ice – água congelada  
 water lines – linhas de água  
 water pipe – cachimbo de água  
 water plants – plantas aquáticas  
 water pollution – poluição da água  
 water pressure – pressão da água  
 water pump – bomba-d'água  
 water resources – recursos hídricos  
 water scarcity – escassez de água  
 water shortage – escassez de água  
 water sources – fontes de água  
 water supply – suprimento de água  
 water tables – lençol freático  
 water tanks – tanques de água  
 water tower – caixa-d'água

**N + WATER**

bog waters – águas do pântano  
 reservoir water – água de reservatório  
 salt water – água salgada  
 surface waters – superfície do mar

**SPACE + N**

space age – era espacial  
 space agency – agência espacial  
 space boots – botas dos astronautas  
 space capsule – cápsula espacial  
 space debris – escombros espaciais  
 space effort – corrida espacial  
 space expert – especialista em atividades espaciais  
 space exploration – exploração do espaço  
 space historian – historiador especializado em temas espaciais  
 space junk – lixo espacial

space mission – missão espacial  
 space observatories – observatórios espaciais  
 space probes – sonda espacial  
 space program – programa espacial  
 space race – corrida espacial  
 space science – ciência espacial  
 space shuttle – ônibus espacial  
 space station – estação espacial  
 space suit – traje espacial  
 space telescope – telescópio espacial  
 space tourist – turista espacial  
 space travel – viagem espacial  
 space voyages – viagens espaciais

### **N + SPACE**

work space – local de trabalho

### **DAY + N**

day care – creche  
 day jobs – trabalhos diários  
 day laborer – diarista  
 day pack – mochila diária  
 day traffic – tráfego diário

### **N + DAY**

election day – dia da eleição  
 feast days – dias de festejos  
 field day – dia de campo  
 harvest days – dias de colheita  
 judgment day – dia de julgamento  
 midsummer day – dia de verão  
 spring day – dia de primavera  
 summer day – dia de verão  
 tourist day – dia turístico  
 winter day – dia de inverno

### **FAMILY + N**

family affair – assunto de família  
 family dynamics – dinâmica familiar  
 family friend – amigo da família  
 family group – grupo familiar  
 family home – casa da família  
 family ironworks – metalúrgica da família  
 family life – vida em família  
 family member – membro da família  
 family mosque – mesquita da família

family name – nome da família  
 family place – berço da família  
 family planning – planejamento familiar  
 family portrait – retrato familiar  
 family ranch – rancho da família  
 family retainer – agregado da família  
 family sizes – tamanho das famílias  
 family spread – terras da família  
 family tombs – túmulos da família  
 family tradition – tradição da família  
 family wealth – riqueza da família

### **N + FAMILY**

gorilla families – famílias de gorila  
 pachycephalosaur family - família dos paquicefalossauros

### **HOME + N**

home base – base (QG)  
 home island – ilha natal  
 home ports – portos domésticos  
 home prices – preço das casas  
 home range – território  
 home state – estado natal  
 home village – vilarejo natal

### **N + HOME**

family home – casa da família  
 forest home – hábitat na floresta  
 island home – ilha natal  
 nursing home – lar para idosos

### **TREE + N**

tree branches – galhos de árvores  
 tree canopy – copa das árvores  
 tree ferns – fetos arbóreos  
 tree holes – buracos em árvores  
 tree limb – tronco de árvore  
 tree stump – tronco de árvore  
 tree trunk – tronco de árvore

### **N + TREE**

banana trees - bananeira  
 cedar tree – cedro  
 coffin tree – árvore fúnebre/ árvore funerária  
 cypress trees – ciprestes

elm tree – olmo/ ulmeiro  
 fig trees – figueiras  
 fir trees – abetos  
 fruit tree – árvore frutífera  
 hazelnut tree – aveleira  
 kino tree – árvore de kino  
 mulga trees – acácias  
 oak trees – carvalhos  
 palm trees – palmeiras  
 poplar trees – choupos  
 rubber trees – seringueiras  
 sapodilla trees – sapotizeiros  
 thorn tree – espinheiro

### **PLANT + N**

plant director – diretor da usina  
 plant fuel – combustível vegetal  
 plant material – matéria vegetal  
 plant matter – tecido vegetal  
 plant oils – óleos vegetais

### **N + PLANT**

alfalfa plant – alfafa  
 algae plant – fazendas de algas  
 corn plants – milho  
 ethanol plant – usina de etanol  
 marsh plants – manguezais  
 pilot plants – usinas experimentais  
 power plants – usinas elétricas  
 spider plant – clorófito  
 water plants – plantas aquáticas

### **FIELD + N**

field chores – tarefas do campo  
 field day – dia de campo  
 field director – diretor de campo  
 field notes – anotações de campo  
 field observations – observações de campo  
 field operations – operações de campo  
 field trials – testes em campo

### **N + FIELD**

baseball field – campo de beisebol  
 cane fields – canaviais  
 cotton fields – campos de algodão  
 football field – campo de futebol

harvesting fields – plantação  
 killing field – campo de batalha  
 oil fields – campos petrolíferos  
 potato fields – plantações de batata  
 quantum fields – campos quânticos  
 rice fields – plantações de arroz  
 wheat fields – plantações de trigo

### **SCHOOL + N**

school play – peça escolar  
 school uniform – uniforme escolar  
 school year – ano letivo

### **N + SCHOOL**

community schools – escolas comunitárias  
 country school – escola rural  
 ghost schools – escolas-fantasma  
 grade school – ensino fundamental  
 law school – escola de Direito  
 night school – escola noturna  
 parent schools – escolas para pais

### **ETHANOL + N**

ethanol backers – apoiadores do etanol  
 ethanol boom – prosperidade do etanol  
 ethanol company – empresa de produção de etanol  
 ethanol demand – demanda de etanol  
 ethanol distilleries – destilarias de etanol  
 ethanol enthusiasts – entusiastas do etanol  
 ethanol industry – indústria do etanol  
 ethanol plants – usinas de etanol  
 ethanol producers – produtores de etanol  
 ethanol production – produção de etanol  
 ethanol project – projeto de etanol  
 ethanol refineries – refinaria de etanol  
 ethanol subsidies – subsídios ao etanol  
 ethanol tanks – tanques de etanol  
 ethanol use – uso do etanol

### **N + ETHANOL**

cane ethanol – álcool de cana  
 corn ethanol – etanol de milho

### **LIFE + N**

life cycle – ciclo de vida

life expectancy – expectativa de vida  
 life force – força vital  
 life histories – histórias de vida  
 life situations – situações da vida  
 life spans – período de vida

### **N + LIFE**

cowboy life – vida de coubói  
 family life – vida em família  
 material lives – condições materiais de vida

### **OIL + N**

oil age – era do petróleo  
 oil company – empresa petrolífera  
 oil consumption – consumo de petróleo  
 oil embargo – embargo do petróleo  
 oil fields – campos petrolíferos  
 oil globules – glóbulos de óleo  
 oil imports – importações de petróleo  
 oil industry – refinaria  
 oil prices – preços do petróleo  
 oil production – produção de petróleo  
 oil refinery – refinaria de petróleo  
 oil security – segurança do petróleo  
 oil shipments – carregamentos de petróleo  
 oil subsidies – subsídios ao petróleo

### **N + OIL**

canola oil – óleo de canola  
 cooking oil – óleo de cozinha  
 mineral oil – óleo mineral  
 palm oil – azeite-de-dendê/ óleo de palmeira  
 peanut oil – óleo de amendoim  
 plant oils – óleos vegetais  
 rock oil – óleo de pedra

### **GROUP + N**

group culture – cultura de grupo  
 group house – residência grupal  
 group members – membros do grupo

### **N + GROUP**

activist groups – grupos ativistas  
 advocacy group – grupo de defesa  
 chimp group – grupo de chimpanzés

chimpanzee group – grupo de chimpanzés  
 community groups – grupos comunitários  
 conservation groups – grupos conservacionistas  
 family group – grupo familiar  
 guerrilla groups – grupos guerrilheiros  
 hunting groups – grupos de caçadores  
 island groups – grupos de ilhas  
 monitoring groups – grupos de monitoramento  
 sauropod group – grupo dos saurópodes

### **MEMORY + N**

memory aid – auxiliar de memória  
 memory championship – campeonato de memória  
 memory chips – placas de memória  
 memory contest – concurso de memória  
 memory disorder – distúrbio de memória  
 memory drugs – remédios para a memória  
 memory load – carga de memória  
 memory loss – perda de memória  
 memory medicines – remédios para a memória  
 memory palaces – palácios da memória  
 memory problems – problemas de memória  
 memory researcher – pesquisador da memória  
 memory systems – sistemas de memória  
 memory techniques – técnicas de memorização  
 memory tests – testes de memória  
 memory textbook – manual sobre a memória  
 memory training – treinamento da mente  
 memory treatises – tratados sobre a memória

### **N + MEMORY**

childhood memories – lembranças de infância

### **CAR + N**

car accidents – acidentes de carro  
 car batteries – baterias de carro  
 car bombs – carros-bomba  
 car company – empresa automobilística  
 car keys – chave do carro  
 car mechanic – mecânico de carros  
 car model – modelo de carro  
 car park – estacionamento  
 car windshields – pára-brisas de carros

### **N + CAR**

commuter car – carro de transporte

concept car – carro-conceito  
 flex cars – veículos flex  
 freight cars – vagões  
 police cars – carros de polícia  
 railroad cars – vagões  
 tanker cars – vagões-tanque  
 train cars – vagões de trem

### **GOVERNMENT + N**

government archaeologists – arqueólogos da agência governamental  
 government buildings – prédios do governo  
 government commission – comissão governamental  
 government efficiency – eficiência do governo  
 government forces – forças do governo  
 government jobs – emprego público  
 government ministers – ministros do governo  
 government money – recursos do governo  
 government newsletter – boletim governamental  
 government offer – oferta do governo  
 government offices – repartições públicas  
 government officials – funcionários do governo  
 government paperwork – papelada governamental  
 government prison – presídio do governo  
 government side – lado do governo  
 government subsidies – subsídios oficiais

### **N + GOVERNMENT**

city government – autoridades municipais  
 county government – governo do condado

### **TIME + N**

time delay – atraso de tempo  
 time machine – máquina do tempo  
 time stamp – carimbo do tempo

### **N + TIME**

siesta time – hora da sesta

### **ENERGY + N**

energy balance – equilíbrio energético  
 energy companies – empresas produtoras de energia  
 energy content – teor energético  
 energy crops – plantações de combustível  
 energy farm – fazenda de energia  
 energy frontier – fronteira energética



energy initiative – programa de pesquisa de energia  
 energy loss – perda de energia  
 energy potential – potencial energético  
 energy ratio – vantagem energética  
 energy security – segurança energética  
 energy shortages – apagões  
 energy transition – mudança da matriz energética  
 energy woes – problemas energéticos

### **N + ENERGY**

vacuum energy – energia do vácuo

### **FOREST + N**

forest canopy – cobertura da floresta  
 forest cover – cobertura florestal  
 forest crew – grupo florestal  
 forest deforestation – desmatamento de florestas  
 forest fire – incêndio florestal  
 forest floor – chão da floresta  
 forest habitat – hábitat na floresta  
 forest home – hábitat na floresta  
 forest raptor – ave de rapina da floresta  
 forest reserves – reservas florestais

### **N + FOREST**

eucalyptus forests – florestas de eucalipto  
 gallery forest – mata de galeria  
 hillside forest – floresta na encosta de uma montanha  
 rain forest – floresta tropical

### **FUEL + N**

fuel alcohol – álcool combustível  
 fuel initiative – programa de combustíveis  
 fuel source – fonte de combustível  
 fuel systems – sistemas de ignição  
 fuels mandate – programas compulsórios de pesquisa de combustíveis

### **N + FUEL**

algae fuel – alga-combustível  
 jet fuel – combustível para aviões a jato  
 motor fuel – combustível para automóveis  
 plant fuel – combustível vegetal  
 rocket fuel – combustível de foguete  
 transportation fuel – combustíveis

**WORKER + N**

workers rights – direitos dos trabalhadores

**N + WORKER**

aid workers – equipe médica

cane workers – cortadores de cana

construction workers – peões de obra

electronics workers – pessoas que se dedicam ao processamento da sucata eletrônica

factory worker – operário

farm workers – trabalhadores rurais

harbor workers – estivadores

migrant workers – trabalhadores migrantes

rig workers – petroleiros

security worker – segurança

sugarcane workers – pessoas que trabalham com cana-de-açúcar

textile worker – operário têxtil

**N + SYSTEM**

address system – sistema de comunicação

apartheid system – sistema de apartheid

calendar system – sistema de calendário

canal system – sistema de canais

communications systems – sistemas de comunicação

drainage system – sistema de drenagem

examination system – sistema examinatório

fuel systems – sistemas de ignição

levee system – sistema de barragens

memory systems – sistemas de memória

protection system – sistema de proteção

sound system – aparelho de som

**RESEARCH + N**

research center – centro de pesquisa

research coordinator – coordenador de pesquisa

research facilities – instalações de pesquisa

research hospitals – hospitais de pesquisa

research institute – instituto de pesquisa

research library – biblioteca de pesquisas

research projects – projetos de pesquisa

research report – relatório de pesquisa

research scientist – pesquisador

research sites – locais de pesquisa

research teams – equipes de cientistas

research vessel – embarcação de pesquisa

research veterinarian – veterinário

**N + RESEARCH**

biomimetics research – pesquisa biomimética

**N + PEOPLE**

bog people – seres do pântano

lab people – técnicos de laboratório

**COMPUTER + N**

computer cases – gabinetes de computador

computer classes – aulas de computação

computer donors – doadores de computador

computer games – jogos de computador

computer keyboard – teclado do computador

computer mice – ratos de computador (plural de *mouse*)

computer modeling – programas digitais de modelagem

computer models – simulações digitais/ modelos de computador

computer monitors – monitores de computador

computer motherboards – placas de computador

computer rendering – imagem computadorizada

computer scientist – cientista de computação

computer scrap – sucata eletrônica

computer screen – tela de computador

computer wire – fiação de computador

**SEA + N**

sea creatures – criaturas marinhas

sea fans – leques-do-mar

sea ice – gelo marítimo

sea level – nível do mar

sea mammals – mamíferos marinhos

sea otters – lontra-marinha

sea patrols – patrulhas marítimas

sea raiders – ladrões do mar

sea slug – lesma-do-mar

sea surface – superfície do mar

sea turtle – tartaruga marinha

sea voyages – viagens marítimas

**N + SEA**

warming seas – mares aquecidos

**STATE + N**

state banks – bancos estatais

state curriculum – currículo estadual

state enterprises – empresas estatais  
 state entities – departamentos estaduais  
 state governor – governador de estado  
 state historian – historiador local  
 state industry – indústria estatal  
 state laws – leis estaduais  
 state official – funcionário do estado

### **N + STATE**

garrison state – cidade-Estado  
 home state – estado natal  
 predator state – Estado predador

### **N + YEAR**

drought years – anos de seca  
 reform years – anos de reforma  
 school year – ano letivo

### **BUILDING + N**

building blocks – peças de montar/ peças de construir  
 building campaign – programa de obras civis  
 building code – código de construções  
 building materials – materiais de construção  
 building supplies – materiais de construção

### **N + BUILDING**

assembly building – centro espacial  
 capacity building – construção de instalações  
 government buildings – prédios do governo  
 pyramid building – construção de pirâmides  
 settlement buildings – casas das povoações

### **GORILLA + N**

gorilla carcasses – carcaças de gorilas  
 gorilla families – famílias de gorila  
 gorilla health – saúde dos gorilas  
 gorilla patriarch – patriarca gorila  
 gorilla sanctuary – refúgio de gorilas  
 gorilla subpopulation – subpopulação de gorilas  
 gorilla traits – traços dos gorilas

### **N + GORILLA**

lowland gorillas – gorilas das planícies  
 mountain gorillas – gorilas-das-montanhas

water gorilla – gorila de água

### **CHIMP + N**

chimp behaviors – comportamentos dos chimpanzés

chimp feces – excrementos de chimpanzé

chimp genome – genoma do chimpanzé

chimp group – grupo de chimpanzés

chimp nest – ninho de chimpanzés

chimp politics – política dos chimpanzés

chimp researchers – pesquisadores de chimpanzés

chimp vision – visão de chimpanzé

### **FOOD + N**

food cache – esconderijos de alimentos

food crops – cereais alimentícios

food prices – preço da comida

food security – oferta de alimentos

food shortages – escassez de víveres

food sources – fontes de alimento

food stand – barraca de comida

food stores – estoque de alimento

food supply – suprimento de alimentos

### **N + FOOD**

snack food – petisco

staple food – alimento básico

### **N + SCIENTIST**

climate scientists – estudiosos em climatologia

computer scientist – cientista de computação

materials scientist – cientista de materiais

research scientist – pesquisador

rocket scientist – cientista de foguetes

### **STONE + N**

stone axes – machados de pedra

stone edifice – edifício petrificado

stone fortress – fortaleza de pedra

stone fragments – fragmentos de pedra

stone houses – casas de pedra

stone monuments – monumentos de pedra

stone mosaic – mosaico de pedras

stone necklace – colar de pedras

stone statues – estátuas de pedra

stone temple – templo de pedra

stone tower – torre de pedra

### **CONSTRUCTION + N**

construction company – empresa de construção  
 construction cranes – guindastes de construção  
 construction fill – estruturas/ entulho  
 construction jobs – trabalhos na construção civil  
 construction projects – obras civis  
 construction sites – canteiros de obra  
 construction stages – etapas da construção  
 construction supplies – material de construção  
 construction worker – operário da construção civil

### **N + CONSTRUCTION**

nest construction – construção do ninho  
 wood construction – construção em madeira

### **N + PROGRAM**

alcohol program – programa do álcool  
 algae programs – programas de algas  
 breeding program – programa de reprodução  
 conservation programs – programas de conservação  
 crash program – programa de emergência  
 development program – programa de desenvolvimento  
 Islamization program – programa de islamização  
 pilot program – programa-piloto  
 release program – programa de soltura na natureza  
 space program – programa espacial  
 spaceflight program – programa espacial  
 tourism program – programa para turistas

### **CATTLE + N**

cattle business – criação de gado  
 cattle chute – calha de gado  
 cattle empire – império do gado  
 cattle guards – vigias de gado  
 cattle industry – indústria do gado  
 cattle kingdom – reino pecuário  
 cattle operations – operações pecuárias  
 cattle owners – donos de gado  
 cattle ranches – fazendas de gado  
 cattle trailers – carretas que transportam gado

### **BUSINESS + N**

business arrangement – arranjo comercial

business card – cartão de negócios  
 business end – extremidade funcional  
 business licenses – licença para abrir negócios  
 business model – modelo de negócio  
 business strategy – estratégia de negócios

### **N + BUSINESS**

cattle business – criação de gado  
 household business – negócio doméstico  
 import business – negócio de importação  
 oyster business – empresa de ostras

### **WINTER + N**

winter ascent – escalada de inverno  
 winter day – dia de inverno  
 winter enchainment – circuito de inverno  
 winter expedition – expedição de inverno  
 winter mountaineering – montanhismo de inverno  
 winter mountaineers – montanhistas de inverno  
 winter sunlight – luz do Sol de inverno  
 winter use – uso no inverno  
 winter winds – ventos invernais

### **SECURITY + N**

security barrier – barreira de segurança  
 security experts – especialistas em segurança  
 security fence – cerca de segurança  
 security guard – segurança  
 security picture – quadro de segurança (situação)  
 security procedures – procedimentos de segurança  
 security wall – muro de segurança  
 security worker – segurança

### **N + SECURITY**

energy security – segurança energética  
 food security – oferta de alimentos  
 oil security – segurança do petróleo

### **PARTICLE + N**

particle accelerators – acelerador de partículas  
 particle beams – feixes de partículas  
 particle collider – colisor de partículas  
 particle collisions – colisões de partículas  
 particle detectors – detectores de partículas  
 particle guns – armas de partículas

particle physicists – físicos que lidam com partículas  
 particle physics – física das partículas  
 particle zoo – zoológico das partículas

### **N + PARTICLE**

dust particles – partículas de poeira

### **CITY + N**

city engineer – engenheiro da cidade  
 city government – autoridades municipais  
 city official – autoridade da cidade  
 city ordinance – lei municipal  
 city parks – parques urbanos  
 city person – pessoa da cidade  
 city resident – cidadão  
 city sprawl – mancha urbana

### **N + CITY**

capital city – capital

### **ANIMAL + N**

animal bones – ossos de animais  
 animal cognition – cognição animal  
 animal husbandry – criação de animais  
 animal locomotion – locomoção animal  
 animal researchers – pesquisadores de animais  
 animal rights – direitos dos animais  
 animal shelf – prateleira de bichos  
 animal studies – estudos sobre animais  
 animal track – trilhas de animais

### **N + ANIMAL**

land animal – animal terrestre  
 pack animals – animais de carga

### **TEAM + N**

team doctor – médico da equipe  
 team leader – líder de equipe  
 team members – membros da equipe

### **N + TEAM**

football team – time de futebol  
 hockey team – time de hockey



lab team – grupo do laboratório  
 magazine team – equipe da revista  
 mountaineering team – time de montanhistas  
 research teams – equipes de cientistas  
 sports teams – esportes

### **COMPANY + N**

company president – presidente da empresa

### **N + COMPANY**

biotech companies – empresas de biotecnologia  
 car company – empresa automobilística  
 construction company – empresa de construção  
 drug companies – indústria farmacêutica  
 energy companies – empresas produtoras de energia  
 ethanol company – empresa de produção de etanol  
 oil company – empresa petrolífera  
 opera companies – companhias de ópera  
 power companies – empresas de eletricidade  
 recycling company – empresa de reciclagem  
 telecommunications company – empresa de telecomunicações

### **FARM + N**

farm kid – criança do campo  
 farm productivity – produtividade da fazenda  
 farm size – área das fazendas  
 farm tools – ferramentas agrícolas  
 farm workers – trabalhadores rurais

### **N + FARM**

algae farms – fazendas de algas  
 energy farm – fazenda de energia  
 orchid farm – sítio de orquídeas  
 pig farms – fazendas de criação de porcos  
 seed farm – campo de produção de sementes  
 sugarcane farm – plantação de cana  
 village farms – vilarejos rurais

### **VILLAGE + N**

village children – crianças do vilarejo  
 village farms – vilarejos rurais  
 village gossip – fofocas do vilarejo  
 village headman – chefe da aldeia  
 village hunters – caçadores do vilarejo  
 village kids – garotos do vilarejo

village mosque – mesquita do vilarejo  
village streets – ruas de vilarejos

### **N + VILLAGE**

cancer villages – povoados cancerosos  
clan villages – povoados do clã  
fishing villages – vilarejo de pescadores  
hillside village – vilarejo na encosta de uma montanha  
home village – vilarejo natal

### **N + LINE**

armistice line – linha de armistício  
assembly line – linha de montagem  
debris lines – linhas de dejetos  
disassembly line – linha de desmontagem  
freight line – linha de trem de carga  
gender lines – diferenças sexuais  
primate line – linhagem dos primatas  
railroad line – linha de trens  
sight line – linha de visão

### **NOUN + THING**

living thing – coisa viva

### **N + FORCE**

afternoon force – força vespertina  
coalition forces – forças de coalizão  
driving force – força motriz  
electroweak force – força eletrofraca  
government forces – forças do governo  
life force – força vital  
market forces – mercado  
police force – força policial

### **SCIENCE + N**

science budget – orçamento científico  
science editor – editor de ciência  
science geeks – nerds  
science writer – escritor de ciência

### **N + SCIENCE**

brain science – ciência do cérebro  
disease science – zoonoses  
Earth science – ciência da Terra

materials science – ciência dos materiais  
space science – ciência espacial

### **CORN + N**

corn belt – região produtora de milho  
corn ethanol – etanol de milho  
corn farmer – pessoas que plantam milho  
corn harvesters – ceifeiros de milho  
corn hooch – álcool de milho  
corn kernels – grãos de milho  
corn plants – milho  
corn prices – preços do milho

### **N + CORN**

feed corn – milho usado na alimentação de animais

### **N + FLOOR**

church floor – solo da igreja  
dirt floor – chão de terra  
factory floor – chão de fábrica  
forest floor – chão da floresta  
ground floor – andar térreo  
marble floors – piso de mármore  
nylon floor – chão de nylon  
temple floor – chão do templo  
tile floor – piso de lajotas

### **PHONE + N**

phone book – agenda telefônica  
phone call – ligação telefônica  
phone cards – cartões para ligações  
phone number – número de telefone

### **N + PHONE**

cell phone – celular  
hand phone – celular  
hotel phone – telefone do hotel  
pay phone – telefone público  
satellite phone – telefone via satélite

### **BRAIN + N**

brain cells – células cerebrais  
brain infection – infecção no cérebro  
brain inflammation – inflamação cerebral

brain matter – matéria cerebral  
 brain science – ciência do cérebro  
 brain tissue – tecido do cérebro

### **N + BRAIN**

adult brain – cérebro adulto  
 sheep brain – cérebro do carneiro

### **N + RIGHT**

animal rights – direitos dos animais  
 workers rights – direitos os trabalhadores

### **CLASS + N**

class cadres – representantes de classe  
 class elections – eleições de classe  
 class rank – nível dentro da classe  
 class reunions – reuniões de classe

### **N + CLASS**

computer classes – aulas de computação  
 essay class – aula de dissertação  
 gym class – aula de educação física  
 math class – aula de matemática  
 prep class – aula preparatória  
 ruling class – grupo dominante  
 weekend class – aulas nos fins de semana

### **BODY + N**

body cavities – orifícios dos corpos  
 body cloak – capa que cobre o corpo  
 body features – características físicas  
 body lengths – comprimento do corpo  
 body movements – movimentos do corpo  
 body parts – partes do corpo  
 body weight – peso do corpo

### **N + BODY**

bog body – corpo do pântano  
 camera bodies – corpos de câmera  
 rocket bodies – corpos de foguete  
 water bodies – corpos d'água

**ISLAND + N**

island culture – cultura da ilha  
 island groups – grupos de ilhas  
 island home – ilha natal  
 island nation – nação da ilha  
 island neighbors – vizinhos de ilhas  
 island peaks – picos insulares  
 island specks – ilhotas

**N + ISLAND**

barrier islands – ilhas de barreira  
 home island – ilha natal  
 mangrove islands – ilhas de mangue

**N + WAY**

buckaroo ways – jeito buckaroo de ser

**MOUNTAIN + N**

mountain air – ar da montanha  
 mountain bicycle – bicicleta de montanha  
 mountain debris – entulho montanhoso  
 mountain expedition – expedição na montanha  
 mountain gorillas – gorilas das montanhas  
 mountain lion – leão-da-montanha  
 mountain nation – nação montanhosa  
 mountain passes – gargantas montanhosas  
 mountain peaks – topo das montanhas  
 mountain slopes – encosta da montanha

**HEALTH + N**

health care – cuidados de saúde  
 health clinics – clínicas de saúde  
 health issues – questões relativas à saúde  
 health risks – riscos à saúde  
 health threats – ameaças sanitárias

**N + HEALTH**

gorilla health – saúde dos gorilas  
 livestock health – saúde dos rebanhos  
 wildlife health – saúde da fauna selvagem

**N + FACILITY**

detention facilities – campos de detenção

recuperation facility – centro de recuperação  
 recycling facility – instalações de reciclagem  
 research facilities – instalações de pesquisa  
 shipment facility – depósito  
 storage facility – unidade de armazenamento  
 support facility – plataforma de suporte  
 wedding facility – centro de casamento

### **METAL + N**

metal armor – armaduras de metal  
 metal crib – berço de metal  
 metal part – componente metálico  
 metal pipes – canos de metal  
 metal saws – serras metálicas  
 metal scrap – sucata metálica  
 metal spatula – espátula de metal  
 metal straw – cânula de metal  
 metal tools – ferramentas de metal  
 metal wheels – rodas de metal  
 metals buyers – compradores de metal

### **GAS + N**

gas mileage – milhagem da gasolina  
 gas pipeline – gasoduto  
 gas royalties – rendimentos do gás  
 gas station – posto de gasolina  
 gas stove – fogão à gas  
 gas tanker – caminhão tanque

### **N + GAS**

greenhouse gases – gases de efeito estufa

### **POWER + N**

power backups – backups de força  
 power brokers – proprietários de terras  
 power companies – empresas de eletricidade  
 power outages – apagões  
 power plants – usinas elétricas  
 power plugs – tomadas

### **N + POWER**

flight powers – capacidade de voar  
 man power – força humana  
 trading power – potência mercantil

**CARBON + N**

carbon cycle – ciclo de carbono  
 carbon dioxide – dióxido de carbono  
 carbon emissions – emissões de carbono  
 carbon fiber – fibra de carbono  
 carbon sink – sequestrador de carbono

**N + CARBON**

car carbon – conta de carbono de um carro  
 soil carbon – carbono no solo

**N + USE**

ethanol use – uso do etanol  
 gasoline use – consumo de gasolina  
 tool use – uso de instrumentos  
 winter use – uso no inverno

**POLICE + N**

police cars – carros de polícia  
 police chief – chefe de polícia  
 police constable – policial  
 police force – força policial  
 police inspector – delegado responsável  
 police officer – policial  
 police presence – presença da polícia  
 police sirens – sirenes da polícia  
 police station – delegacia

**N + POLICE**

migration police – polícia de imigração  
 riot police – tropas

**HEAD + N**

head scarves – lenços usados na cabeça  
 head shields – painéis/escudos

**N + HEAD**

railroad head – base ferroviária  
 tape head – cabeça de fita

**FIRE + N**

fire brigade – brigada de incêndio

fire retardants – retardantes de fogo  
 fire shrine – altar do fogo

### **N + FIRE**

evening fire – fogo da tarde  
 forest fire – incêndio florestal

### **SUMMER + N**

summer afternoon – tarde de verão  
 summer climbing – escalada de verão  
 summer day – dia de verão  
 summer grass – grama de verão  
 summer meltwater – água de degelo  
 summer monsoon – monções de verão  
 summer months – meses de verão  
 summer sun – sol do verão  
 summer tires – pneus de verão

### **N + SUMMER**

childhood summers – verões da infância

### **WORK + N**

work site – obra  
 work space – local de trabalho

### **N + WORK**

lab work – exames laboratoriais

### **WORD + N**

word choice – escolha de termo  
 word learning – aprendizado de palavras  
 word order – ordem dos termos

### **CENTER + N**

center beam – viga principal  
 center point – ponto central

### **N + CENTER**

command center – centro de comando  
 district center – centro distrital  
 education center – centro educacional  
 manufacturing center – centros manufatureiros



nerve center – centro nevrálgico  
 recruiting centers – centros de recrutamento  
 research center – centro de pesquisa  
 trading center – centro comercial

### **BORDER + N**

border guards – guardas de fronteira  
 border town – cidade fronteira

### **ELECTRONICS + N**

electronics graveyard – cemitério de eletrônicos  
 electronics manufacturers – fabricantes de eletrônicos  
 electronics shops – lojas de eletrônicos  
 electronics stores – lojas de eletrônicos  
 electronics workers – pessoas que se dedicam ao processamento da sucata eletrônica

### **N + ELECTRONICS**

junk electronics – detritos eletrônicos  
 scrap electronics – ferro-velho eletrônico

### **COUNTRY + N**

country boy – rapaz do interior  
 country school – escola rural

### **N + COUNTRY**

hill country – região montanhosa

### **RICE + N**

rice bag – saco de arroz  
 rice farmers – rizicultores  
 rice fields – plantações de arroz  
 rice fortune – adivinho que diz o futuro através do arroz  
 rice grains – grãos de arroz  
 rice shoots – brotos de arroz  
 rice wine – vinho de arroz

### **MOON + N**

moon base – base na lua  
 moon expedition – expedição lunar  
 moon landing – pouso na Lua  
 moon missions – missões a Lua  
 moon rockets – foguetes lunares

**N + SIDE**

government side – lado do governo

**ROCK + N**

rock circles – círculos rochosos  
 rock glacier – geleira rochosa  
 rock jocks – alpinistas  
 rock oil – óleo de pedra  
 rock overhang – saliência rochosa  
 rock salt – sal grosso  
 rock stars – celebridades da área  
 rock walls – muros de pedra

**DRUG + N**

drug addiction – dependência de drogas  
 drug addicts – dependentes químicos  
 drug cartel – cartel de drogas  
 drug companies – indústria farmacêutica  
 drug routes – rotas da droga  
 drug runners – "mulas" de drogas  
 drug scene – cenário das drogas  
 drug smugglers – traficantes de drogas  
 drug users – consumidores de drogas

**N + DRUG**

memory drugs – remédios para a memória

**TRADE + N**

trade artery – artéria comercial  
 trade routes – rotas mercantis  
 trade union – sindicato  
 trade winds – ventos alíseos

**N + TRADE**

opium trade – comércio de ópio  
 rodeo trade – mundo do rodeio  
 spice trade – comércio de especiarias  
 tourist trade – turismo  
 world trade – comércio mundial

**FRUIT + N**

fruit bat – morcego de fruta  
 fruit pulp – polpa de fruta

fruit seeds – sementes de frutos  
 fruit smoothies – vitaminas de frutas  
 fruit tree – árvore frutífera

### **N + FRUIT**

bambu fruit – frutas de bambu  
 mobei fruit – fruta de mobei  
 saba fruits – frutos de saba

### **CANE + N**

cane acreage – área de cultivo de cana  
 cane cutter – cortador de cana  
 cane ethanol – álcool de cana  
 cane fields – canaviais  
 cane juice – caldo da cana  
 cane trucks – caminhões que transportam cana  
 cane waste – o que resta da cana / bagaço  
 cane workers – cortadores de cana

### **N + CANE**

sugar cane – cana-de-açúcar  
 waste cane – bagaço

### **CONSERVATION + N**

conservation activities – atividades de conservação  
 conservation efforts – esforços de conservação  
 conservation groups – grupos conservacionistas  
 conservation issues – problemas ecológicos  
 conservation needs – necessidades de preservação  
 conservation policy – política de preservação  
 conservation programs – programas de conservação

### **N + CONSERVATION**

water conservation – conservação da água  
 wildlife conservation – conservação da fauna selvagem

### **N + CULTURE**

cowboy culture – cultura caubói  
 group culture – cultura de grupo  
 hunting culture – cultura de caça  
 island culture – cultura da ilha  
 youth culture – cultura jovem

**DESERT + N**

desert beetle – besouro do deserto  
 desert flora – flora do deserto  
 desert guide – guia do deserto  
 desert region – região desértica  
 desert road – estrada deserta  
 desert sands – areias do deserto  
 desert temperatures – temperaturas no deserto

**N + DESERT**

scrub desert – deserto semi-árido

**ICE + N**

ice age – era glacial  
 ice burner – queimador de gelo  
 ice conditions – condições do gelo  
 ice edge – borda do gelo  
 ice sheets – calotas de gelo

**N + ICE**

sea ice – gelo marítimo  
 water ice – água congelada

**PAPER + N**

paper bag – sacola de papel  
 paper clip – clipe de papel  
 paper industry – fabricantes de papel  
 paper mill – fábrica de papel  
 paper millionaires – milionários franciscanos  
 paper plate – folha de papel

**SKIN + N**

skin cancer – câncer de pele  
 skin tone – tons de pele

**N + SKIN**

bronze skin – pele bronzeada  
 devil skin – pele de diabo  
 leopard skin – pele de leopardo  
 lizard skin – pele de lagarto

## ANEXO C – Análise rasa dos 200 compostos

N°	Composto	Núcleo	E N	E X	C T	C P	L E	N o	ELDE
1	algae plant	plant	X		X				
2	animal bones	bones	X		X				
3	animal track	track	X		X				
4	banana trees	trees	X		X				
5	baseball field	field	X		X				
6	biomimetics research	research	X		X				
7	body cloak	cloak	X		X				
8	bog people	people	X		X				
9	bog waters	waters	X		X				
10	border guards	guards	X		X				
11	brain cells	cells	X		X				
12	brain infection	infection	X		X			X	infection
13	brain science	science	X		X				
14	building blocks	blocks	X		X			X	building
15	business licenses	licenses	X		X				
16	cane cutter	cutter	X		X			X	cutter
17	cane ethanol	ethanol	X		X				
18	cane juice	juice	X		X				
19	canola oil	oil	X		X				
20	car accidents	accidents	X		X				
21	car batteries	batteries	X		X				
22	car bombs	bombs	X		X				
23	car keys	keys	X		X				
24	car park	park	X		X				
25	carbon emissions	emissions	X		X				
26	carbon fiber	fiber	X		X				
27	cattle guards	guards	X		X				
28	cattle owners	owners	X		X			X	owners
29	cattle trailers	trailers	X		X				
30	cedar tree	tree	X		X				
31	cell phone	phone	X		X				
32	childhood memories	memories	X		X				
33	chimp feces	feces	X		X				
34	chimpanzee group	group	X		X				
35	church floor	floor	X		X				
36	city engineer	engineer	X		X				
37	city government	government	X		X			X	government
38	city parks	parks	X		X				
39	coffin tree	tree	X		X				
40	community schools	schools	X		X				
41	commuter car	car	X		X			X	commuter
42	computer classes	classes	X		X				
43	computer keyboard	keyboard	X		X				
44	computer monitors	monitors	X		X				

N °	Composto	Núcleo	E N	E X	C T	C P	L E	N o	ELDE
45	computer scientist	scientist	X		X				
46	conservation efforts	effort	X		X			X	conservation
47	construction jobs	jobs	X		X			X	construction
48	construction workers	workers	X		X			X	construction workers
49	cooking oil	oil	X		X			X	cooking
50	corn ethanol	ethanol	X		X				
51	country boy	boy	X		X				
52	country school	school	X		X				
53	day pack	pack	X		X				
54	desert beetle	beetle	X		X				
55	drainage system	system	X		X			X	drainage
56	drug addiction	addiction	X		X			X	addiction
57	drug users	users	X		X			X	users
58	education center	center	X		X			X	education
59	election day	day	X		X			X	election
60	electronics shops	shops	X		X				
61	energy farm	farm	X		X				
62	ethanol company	company	X		X				
63	ethanol plants	plants	X		X				
64	ethanol production	production	X		X			X	production
65	ethanol tanks	tanks	X		X				
66	eucalyptus forests	forests	X		X				
67	family friend	friend	X		X				
68	family group	group	X		X				
69	family home	home	X		X				
70	family member	member	X		X				
71	family mosque	mosque	X		X				
72	family planning	planning	X		X			X	planning
73	farm kid	kid	X		X				
74	farm tools	tools	X		X				
75	feast day	day	X		X				
76	feed corn	corn	X		X				
77	field chores	chores	X		X				
78	fire brigade	brigade	X		X				
79	food crops	crops	X		X				
80	food sources	sources	X		X				
81	forest fire	fire	X		X				
82	fruit bat	bat	X		X				
83	fruit pulp	pulp	X		X				
84	fruit tree	tree	X		X				
85	fuel alcohol	alcohol	X		X				
86	gas station	station	X		X				
87	gas tanker	tanker	X		X				
88	gorilla families	families	X		X				
89	gorilla health	health	X		X				
90	government buildings	buildings	X		X			X	government buildings

N°	Composto	Núcleo	E N	E X	C T	C P	L E	N o	ELDE
91	government efficiency	efficiency	X		X			X	government
92	group culture	culture	X		X				
93	group house	house	X		X				
94	group members	members	X		X				
95	guerrilla groups	groups	X		X				
96	head scarves	scarves	X		X				
97	health care	care	X		X				
98	home prices	prices	X		X				
99	home village	-----	X		X				
100	island culture	culture	X		X				
101	island home	-----	X		X				
102	jet fuel	fuel	X		X				
103	judgment day	day	X		X			X	judgment
104	killing field	field	X		X			X	killing
105	lab people	people	X		X				
106	law school	school	X		X				
107	leopard skin	skin	X		X				
108	life force	force	X		X				
109	life histories	histories	X		X				
110	living thing	thing	X		X			X	living
111	manufacturing center	center	X		X			X	manufacturing
112	marble floors	floors	X		X				
113	math class	class	X		X				
114	memory championship	championship	X		X				
115	memory drugs	drugs	X		X				
116	memory loss	loss	X		X				
117	memory researcher	researcher	X		X			X	researcher
118	metal armor	armor	X		X				
119	metals buyers	buyers	X		X			X	buyers
120	monitoring groups	groups	X		X			X	monitoring
121	moon expedition	expedition	X		X				
122	mountain air	air	X		X				
123	mountain bicycle	bicycle	X		X				
124	mountain lion	lion	X		X				
125	nest construction	construction	X		X			X	construction
126	night school	school	X		X				
127	nursing home	home	X		X			X	nursing
128	oil company	company	X		X				
129	oil consumption	consumption	X		X			X	consumption
130	opium trade	trade	X		X				
131	orchid farm	farm	X		X				
132	oyster business	business	X		X				
133	palm trees	trees	X		X				
134	paper bag	bag	X		X				
135	particle accelerators	accelerators	X		X			X	accelerators
136	phone number	number	X		X				

N°	Composto	Núcleo	E N	E X	C T	C P	L E	N o	ELDE
137	plant director	director	X		X			X	director
138	plant fuel	fuel	X		X				
139	plant oils	oils	X		X				
140	police cars	cars	X		X				
141	police sirens	sirens	X		X				
142	police station	station	X		X				
143	recycling company	company	X		X			X	recycling
144	research coordinator	coordinator	X		X			X	coordinator
145	research sites	sites	X		X				
146	research vessel	vessel	X		X				
147	rice bag	bag	X		X				
148	rice fields	fields	X		X				
149	rock salt	salt	X			X			
150	rocket scientist	scientist	X		X				
151	salt water	water	X		X				
152	school play	play	X		X				
153	school uniform	uniform	X		X				
154	sea creatures	creatures	X		X				
155	sea raiders	raiders	X		X			X	raiders
156	sheep brain	brain	X		X				
157	siesta time	time	X		X				
158	skin cancer	cancer	X		X				
159	snack food	food	X		X				
160	sound system	system	X		X				
161	space boots	boots	X		X				
162	space expert	expect	X		X				
163	space exploration	exploration	X		X			X	exploration
164	space station	station	X		X				
165	space suit	suit	X		X				
166	space tourist	tourist	X		X				
167	spring day	day	X		X				
168	state banks	banks	X		X				
169	stone axes	axes	X		X				
170	stone houses	house	X		X				
171	storage facility	facility	X		X			X	storage
172	sugar cane	cane	X		X				
173	sugarcane workers	workers	X		X			X	workers
174	summer afternoon	afternoon	X		X				
175	summer climbing	climbing	X		X			X	climbing
176	summer day	day	X		X				
177	summer grass	grass	X		X				
178	tool use	use	X		X				
179	trade routes	routes	X		X				
180	trade union	union	X		X				
181	train cars	cars	X		X				
182	tree branches	branches	X		X				



N°	Composto	Núcleo	E N	E X	C T	C P	L E	N o	ELDE
183	tree holes	holes	X		X				
184	village gossip	gossip	X		X				
185	village streets	street	X		X				
186	warming seas	seas	X		X			X	warming
187	water bottles	bottles	X		X				
188	water bowl	bowl	X		X				
189	water conservation	conservation	X		X			X	conservation
190	water hole	-----		X			X		
191	water ice	water	X		X				
192	water plants	plants	X		X				
193	water pollution	pollution	X		X			X	pollution
194	water tanks	tanks	X		X				
195	weekend class	class	X		X				
196	winter day	day	X		X				
197	winter mountaineering	mountaineering	X		X			X	mountaineering
198	wood construction	construction	X		X			X	construction
199	workers rights	rights	X		X			X	workers
200	youth culture	culture	X		X				

Legenda:

EN – endocêntrico

EX – exocêntrico

CT – composição total

CP – composição parcial

LE – lexicalizado

No – nominalização

ELDE – elemento deverbal

## ANEXO D – Relações semânticas dos 200 compostos analisados

## SERVE PARA

**Função + pessoas**

Monitoring groups

**Função + plantas**

Coffin tree

Feed corn

Food crops

**Função + artefatos**

Building blocks

Business licenses

Car batteries

Car keys

Commuter car

Day pack

Drainage system

Farm tools

Mountain bicycle

Particle accelerator

School uniform

Sound system

Space boots

Space suit

**Função + substâncias**

Cooking oil

Fuel alcohol

Jet fuel

Memory drugs

**Função + locais**

Car park

Education center

Nursing home

## DÁ ORIGEM A

**Animal dá origem ao N2 (efetuado)**

Animal track

**Instrumento dá origem ao N2 (evento)**

Car accidents

É CAUSADA EM

**Parte do corpo + doença**

Brain infection

Skin cancer

VEM DE, É FEITO A PARTIR DE

**Origem (planta) + substância**

Cane ethanol

Cane juice

Canola oil

Corn ethanol

Plant fuel

Plant oils

**Origem (animal) + substância**

Chimp feces

**Substância + origem**

Food sources

Sugar cane

FUNCIONA A PARTIR DE

**N1 é a fonte de energia de N2**

Life force

É FEITO DE

**Material + artefato**

Carbon fiber

Marble floors

Metal armor

Paper bag

Stone axes

Stone houses

Wood construction

POSSUI

Parte integrante

**Animal + parte do corpo**

Animal bones

Leopard skin

Sheep brain

**Planta + parte**

Fruit pulp

Tree branches

**Artefato + parte**

Computer keyboard

Computer monitors

Train cars

**Animal como membro de um grupo**

Chimpanzee group

Gorilla families

**Planta como membro de um grupo**

Eucalyptus forests

**Pessoa como membro de um grupo**

Family member

Group members

Guerrilla groups

**Tempo + parte**

Spring day

Summer afternoon

Summer day

Winter day

**Local + parte**

Church floor

Village streets

**Órgão + partes**

Brain cells

Característica abstrata

**Local + característica abstrata**

Island culture

Mountain air

Village gossip

**Pessoas + característica abstrata**

Group culture

Youth culture

**CONTÉM****Conteúdo + recipiente**

Cattle trailers  
 Ethanol tanks  
 Gas tanker  
 Rice bag  
 Water bottles  
 Water bowl  
 Water tanks

**É LOCALIZADO EM****Localização pelo uso**

Algae plant  
 Electronics shop  
 Energy farm  
 Ethanol company  
 Ethanol plants  
 Gas station  
 Oil company  
 Orchic farm  
 Oyster business  
 Rice fields  
 Storage facility

**Localização por evento**

Baseball field  
 Killing field  
 Manufacturing center  
 Moon expedition  
 Recycling company  
 Research sites  
 Research vessel  
 School play  
 Space exploration  
 Trade routes

**Origem de pessoas, onde vivem ou trabalham**

Bog people  
 Border guards  
 City engineer  
 Contry boy  
 Farm kid  
 Lab people  
 Plant director  
 Police station  
 Sea raiders

**Local + substância**

Bog waters

**Onde o animal vive**

Desert beetle

Mountain lion

Sea creatures

**Origem de planta**

Water plants

**Localização do N2**

Country school

Forest fire

Space station

Tree holes

Space tourist

**OCORRE EM****Evento + tempo**

Election day

Feast day

Judgment day

Siesta time

**Tempo + evento**

Summer climbing

Weekend class

Winter mountaineering

**Tempo + local**

Night school

**Tempo + planta**

Summer grass

**TEM****Possuído + possuidor**

Cattle owners

**Possuidor + possuído**

Family home

Family mosque

Gorilla health

Government buildings

Government efficiency

Group house

Police cars  
 Police sirens  
 Workers rights

**Pessoa + pessoa**  
 Family friend

É PARECIDO COM

**Mineral + produto**  
 Rock salt

É UM

**Hipônimo + hiperônimo**  
 Family group

O REFERENTE É OS DOIS AO MESMO TEMPO

**Artefato + artefato**  
 Car bombs

**Local + local**  
 Island home  
 Home village

O COMPOSTO É UM TIPO DE NÚCLEO

**Tipo de pessoa (profissão, área de atuação, vício)**  
 Computer scientist  
 Construction workers  
 Drug users  
 Memory researcher  
 Research coordinator  
 Rocket scientist  
 Space expert  
 Sugarcane workers

**Tipo de animal**  
 Fruit bat

**Tipo de planta**  
 Cedar tree  
 Palm trees

**Tipo de artefato**  
 Cell phone

**Tipo de substância**

Salt water  
Warming seas  
Water ice

**Tipo de evento**

Biomimetics research  
Carbon emissions  
Computer classes  
Conservation efforts  
Ethanol production  
Family planning  
Field chores  
Math class  
Nest construction  
Oil consumption  
Opium trade  
Tool use  
Water conservation  
Water pollution

**Exemplar de um grupo**

Brain science  
Childhood memories  
City government  
Construction jobs  
Drug addiction  
Fire brigade  
Health care  
Home prices  
Law school  
Live histories  
Memory championship  
Memory loss  
Phone number  
Snack food  
Trade union

**Instituição cujo responsável é um órgão ou grupo social**

City parks  
Community schools  
State banks

**Tipo não específico**

Living thing



CUIDA, CRIA, TREINA

**Animal + humano**

Cattle guards

FAZ, VENDE, CORTA, COMPRA

**Produto + humano**

Cane cutter

Metals buyers

É USADO EM

**Parte do corpo + produto**

Body cloak

Head scarves

PRODUZ

**Produto + planta**

Banana trees

Fruit tree

LEXICALIZADO

Water hole

ANEXO E – Equivalentes de tradução do *corpus National Geographic*

## 1 – Papel télico – Relação SERVE PARA

Physicists use particle accelerators not only to find the smallest <b>building blocks</b> of the universe but also to shed light on the biggest questions: What is the universe composed of? What laws govern it?	Os físicos usam aceleradores de partículas para encontrar as menores <b>peças de montar</b> do universo, mas também para lançar luz sobre questões maiores: Do que o universo é composto? Que leis o governam?
Physicists use accelerators to smash subatomic particles together to find the smallest <b>building blocks</b> of the universe.	Os físicos usam aceleradores para fazer com que partículas subatômicas entrem em colisão, para encontrar as menores <b>peças de construir</b> do universo.
Many, however, still live on the fringes of society, relegated to manual labor and barred from obtaining <b>business licenses</b> , government jobs, or access to higher education.	Mas muitos ainda vivem à margem da sociedade, relegados ao trabalho braçal e impedidos de obter <b>licença para abrir negócios</b> , de ter emprego público ou acesso à educação superior.
Continuing through the hills, I see people tearing apart <b>car batteries</b> , alternators, and high-voltage cable for recycling, and others hauling aluminum scrap to an aging smelter.	SEM TRADUÇÃO.
Tragedies and humiliations seem to be etched most sharply, often with the most unbearable exactitude, while those memories we think we really need—the name of the acquaintance, the time of the appointment, the location of the <b>car keys</b> —have a habit of evaporating.	Tragédias e humilhações parecem ficar gravadas mais nitidamente, enquanto aquelas informações que julgamos realmente precisar - o nome de um conhecido, a hora de um compromisso, o lugar onde deixamos a <b>chave do carro</b> - têm a mania de evaporar.
According to a decade's worth of "forgetting diaries" kept by his undergraduate students (the amount of time it takes to find the <b>car keys</b> , for example), Anderson calculates that people squander more than a month of every year just compensating for things they've forgotten.	Por dez anos, seus alunos da faculdade tiveram um "diário de esquecimentos" (onde anotavam, por exemplo, o tempo que perdiam procurando a <b>chave do carro</b> ). Com base nesses diários, Anderson calculou que as pessoas desperdiçam mais de um mês por ano tendo trabalho por causa de coisas que esqueceram.
Paradise nowadays is finding a free spot in the crowded <b>car park</b> .	SEM TRADUÇÃO.
The son never cut down his father's <b>coffin tree</b> to have it made into a coffin.	O filho não abatera a <b>árvore funerária</b> do pai para fazer um caixão.
Wu Lianlian shares a laugh with her mother beside her <b>coffin tree</b> , chosen for her at birth.	Wu Lianlian ri com a mãe ao lado de sua <b>árvore fúnebre</b> , escolhida quando ela nasceu.
Volkswagen introduced Brazil's first TotalFlex vehicle in 2003, modifying a small soccer ball of a <b>commuter car</b> called the Gol, which means—you guessed it—"goal!" It was an instant hit, and soon every other carmaker in Brazil followed suit.	Em 2003, a Volks lançou o primeiro carro TotalFlex no mercado nacional, introduzindo o sistema em um modelo básico, o Gol.

Sunup, sundown, winter, spring, summer, fall, they wait, hoping someone will hire them to use their carts to transport lumber, building materials, bags of wheat, cans of <b>cooking oil</b> , panes of glass, window frames, dishes for wedding receptions—something, anything—from one place to another.	Eles ficam ali à espera de que alguém precise de carroto para madeira, material de construção, sacos de trigo, latas de <b>óleo de cozinha</b> , molduras de janela, pratos para festa de casamento - alguma coisa, qualquer coisa.
Were the Buddhas still standing last May, they would have gazed down on a young man walking Bamian's main street, a bumpy unpaved tract with shops on both sides selling <b>cooking oil</b> , medicines, and building materials.	Se os Budas ainda existissem em maio passado, teriam fitado lá embaixo um moço andando pela rua principal de Bamian, um caminho sem calçamento ladeado por lojas que vendem <b>óleo de cozinha</b> , remédios e materiais de construção.
Here's what I discovered: First, empty the <b>day pack</b> of everything, except for the sandwich, trail mix, and water.	Eis o que eu descobri: primeiro, tire tudo da <b>mochila</b> , menos o seu sanduíche, seu lanche e água.
From June to September Kolkata can get torrential rains, and its <b>drainage system</b> doesn't need torrential rain to begin backing up.	De junho a setembro, chuvas torrenciais costumam se abater sobre Kolkata, sendo que o <b>sistema de drenagem</b> da cidade não necessita de nenhuma tempestade para regurgitar água.
Meanwhile, visitors to the foundation's <b>education center</b> can see more than a dozen eagles, some of which were rescued after they were trapped or shot.	Enquanto isso, os visitantes no <b>centro educacional</b> da fundação podem ver mais de uma dúzia dessas aves, das quais algumas foram salvas depois de baleadas ou capturadas em armadilhas.
The farmers tallied their losses: homes, pigs, <b>farm tools</b> , grain sheds, and the woven clothes and silver heirlooms of grandmothers and mothers.	Os agricultores avaliam seus prejuízos: casas, porcos, <b>ferramentas agrícolas</b> , depósitos de grãos, as roupas tecidas e as heranças de prata de suas avós e mães.
Nearly all the ethanol in the U.S. is brewed from yellow <b>feed corn</b> .	Quase todo o etanol dos Estados Unidos é destilado de <b>milho amarelo usado na alimentação de animais</b> .
Monoculture farming—growing one staple <b>food crop</b> that could be accumulated and stored for hard times or for trade—could not be sustained in the rain forest.	A monocultura - o cultivo intensivo de um <b>tipo básico de alimento</b> , que pudesse ser acumulado e guardado para as épocas de escassez ou para trocas - não era viável na floresta tropical.
Breaking up those chains and fermenting the sugars could yield a cornucopia of biofuels, without competing with <b>food crops</b> .	O rompimento dessas cadeias e a fermentação dos açúcares poderiam gerar uma variedade de biocombustíveis, sem com isso prejudicar a <b>produção de alimentos</b> .
Ethanol and biodiesel are now made from <b>food crops</b> like corn and soybeans, but in principle any plant material will do.	O etanol e o biodiesel, produzidos a partir de <b>cereais alimentícios</b> como milho e soja, em princípio poderiam ser feitos com qualquer vegetal.
In a living sea of green, a laborer takes a break from weeding on an organic sugarcane farm in central Brazil that	Em Sertãozinho, perto de Ribeirão Preto, cortador descansa em plantação de cana numa fazenda que produz 150 milhões de

produces 150 million liters of <b>fuel alcohol</b> each year.	litros de <b>álcool combustível</b> por ano.
It wasn't until 2000 that <b>fuel alcohol</b> staged a major comeback, largely as an additive in less polluting gasoline blends.	Só depois de 2000 o <b>álcool</b> voltou com toda a força como <b>combustível</b> , sobretudo sob a forma de aditivo em fórmulas menos poluentes de gasolina.
A cane cutter is part of a harvest army in Brazil, which refines half its sugar crop into <b>fuel alcohol</b> for cars.	Um cortador de cana faz parte do exército de colheita no Brasil, que refina metade da cana-de-açúcar produzida para transformar em <b>álcool</b> para carros.
Some algae make starch, which can be processed into ethanol; others produce tiny droplets of oil that can be brewed into biodiesel or even <b>jet fuel</b> .	Algumas espécies produzem amido, que por sua vez pode ser transformado em etanol; outras geram minúsculas gotas de um óleo que, refinado, se torna biodiesel ou mesmo <b>combustível para aviões a jato</b> .
In recent years, at least three companies have been formed with the express purpose of developing <b>memory drugs</b> .	Recentemente foram fundadas pelo menos três companhias com o objetivo expresso de criar <b>drogas</b> .
Despite these advances, the government banned some TV channels for broadcasting violent wrestling shows and other content regarded as being too sexually provocative, and the country has received low marks for press freedom from some international <b>monitoring groups</b> , including Reporters Without Borders and Freedom House.	Apesar desses avanços, o governo proibiu alguns canais de TV por transmitir programas de luta violentos e outros conteúdos considerados provocadores do ponto de vista sexual, e o país recebeu nota baixa por liberdade de imprensa em alguns <b>grupos</b> internacionais <b>de monitoramento</b> , incluindo Reporters Without Borders e Freedom House.
When they reached the Mexican side of the river, Armando unloaded the used <b>mountain bicycle</b> they had bought in Guatemala, and they waited while a uniformed soldier on the riverbank rifled indifferently through Jessenia's bag, explaining that he was looking for weapons or drugs.	Quando chegaram ao lado mexicano do rio, Armando descarregou a <b>bicicleta</b> , comprada na Guatemala.
Recently, Bella's father and his sister and cousins put their grandfather in a <b>nursing home</b> .	Recentemente, o pai de Bella e a irmã dele, junto com uns primos, puseram o avô comum em uma <b>casa para idosos</b> .
But Bella announced that she would one day put her parents in the best <b>nursing home</b> .	Bella, no entanto, já anunciou que irá pôr seus pais na melhor <b>casa para idosos</b> que houver.
"When we are old, we'll sell the house, take a trip and see the world, and enter the <b>nursing home</b> and live a quiet life there.	"Quando estivermos velhos, venderemos a casa, faremos uma viagem para ver o mundo e nos mudaremos para o <b>lar dos idosos</b> . E lá viveremos uma vida tranqüila.
It belongs to an 81-year-old man known as "HM," an amnesiac who lives in a <b>nursing home</b> in Connecticut.	Ele pertence a um homem de 81 anos conhecido como "HM", que tem amnésia e vive em um <b>lar para idosos</b> em Connecticut.
Surveying the whole panoply of physics,	Quando contemplamos a panóplia toda da

from quasars imaged by Hubble near the edge of the observable universe to the subatomic realms probed by <b>particle accelerators</b> , one increasingly gets the sense that science has as yet detected only the tip of an iceberg.	física, desde os quasares visualizados pelo Hubble junto à borda do universo observável até os domínios subatômicos vasculhados pelos <b>aceleradores de partículas</b> , somos tomados pela sensação cada vez mais forte de que até agora os cientistas detectaram apenas a ponta de um iceberg.
This technological netherworld is one very big scientific instrument, specifically, a <b>particle accelerator</b> -an atomic peashooter more powerful than any ever built.	Toda essa parafernália tecnológica faz parte de um único e imenso instrumento científico, um <b>acelerador de partículas</b> - o mais poderoso canhão de átomos já construído.
By the early 1930s Ernest Lawrence had invented the first circular <b>particle accelerator</b> , or "cyclotron."	Então, no início dos anos 1930, Ernest Lawrence construiu o primeiro <b>acelerador de partículas</b> em forma de círculo, o "ciclotron".
Below is a list of laboratories around the world conducting high-energy physics experiments with <b>particle accelerators</b> .	Abaixo, uma lista de laboratórios espalhados pelo mundo que conduzem experimentos de física de alta energia com <b>aceleradores de partículas</b> .
Physicists use <b>particle accelerators</b> not only to find the smallest building blocks of the universe but also to shed light on the biggest questions: What is the universe composed of? What laws govern it?	Os físicos usam <b>aceleradores de partículas</b> para encontrar as menores peças de montar do universo, mas também para lançar luz sobre questões maiores: Do que o universo é composto? Que leis o governam?
Strings are too small for current <b>particle accelerators</b> to detect, but physicists hope to find indirect evidence of their existence, such as superparticles, particles disappearing into other dimensions, or disturbances in the behavior of ordinary particles.	As cordas são pequenas demais para que os atuais <b>aceleradores de partículas</b> as detectem, mas os físicos esperam encontrar evidências indiretas a respeito de sua existência, tais como as superpartículas, partículas que desaparecem em outras dimensões, ou perturbações no comportamento de partículas comuns.
Physicists use <b>particle accelerators</b> to create quark-gluon plasma to investigate its properties and learn more about the early universe and the cores of neutron stars.	Os físicos usam <b>aceleradores de partículas</b> para criar plasma de quark-glúon para examinar suas propriedades e saber mais a respeito dos primórdios do universo e sobre o núcleo das estrelas de nêutrons.
This idea is known as the Grand Unification Theory, or GUT, and physicists hope <b>particle accelerators</b> will help them find evidence of unification.	Esta idéia é conhecida como a Grand Unification Theory (GUT - teoria da grande unificação), e os físicos esperam que os <b>aceleradores de partículas</b> os ajudem a encontrar evidências da unificação.
"He's wearing his <b>school uniform</b> ," says Shawrieh.	"Ele está de <b>uniforme escolar</b> ", diz Shawrieh.
The physicists won a Nobel Prize in 2007, and you got a nifty <b>sound system</b> that's smaller than a Hershey bar.	Por essa descoberta, os físicos receberam o Prêmio Nobel de 2007.
NASA was an early user, sending Velcro to	A NASA foi um dos primeiros usuários,

the moon on <b>space boots</b> and suits—including Apollo astronaut John Young's glove in 1972 (above)—on tabs to latch down loose items in zero gravity, and inside helmets as nose scratchers.	enviando à Lua fechados com velcro nas <b>botas</b> e nos trajes <b>dos astronautas</b> – incluindo a luva de John Young, membro de uma das missões Apollo em 1972 –, em placas que prendiam objetos flutuantes na gravidade zero e até na parte interna dos capacetes para se coçar o nariz.
In its early days the grabby fabric, inspired by prickly cockleburs the inventor peeled from his clothes after a hike, held together parts of Commander Alan Shepard's <b>space suit</b> (above, from the 1971 Apollo 14 moon mission).	Em seus primórdios, o tecido pegajoso, inspirado por carrapichos que o inventor tirou da roupa depois de uma caminhada, prendeu partes da <b>roupa espacial</b> do comandante Alan Shepard (acima, da missão à lua Apollo 14, de 1971).
They must have <b>space suits</b> that can cope with the abrasive effects of moon dust—the microscopic potpourri of jagged glass and rock that nearly froze the joints of Apollo <b>space suits</b> after only three days of moonwalks.	Eles necessitarão de <b>trajes espaciais</b> que suportem os efeitos abrasivos da poeira lunar - o pot-pourri de fragmentos afiados de gelo e de rocha que quase congelou as dobras das <b>roupas espaciais</b> do pessoal da Apollo após três dias de passeios pela superfície da Lua.
And robots don't need <b>space suits</b> , radiation shields, toilets, exercise bikes, a bail-out system during launch, or any consumables to speak of except energy.	Além disso, robôs não necessitam de <b>trajes espaciais</b> , escudos anti-radiação, banheiros, sistemas de escape em caso de acidente nem de comida, exceto energia.
Designed to withstand temperatures ranging from 240°F above zero (115°C) to 240°F below (-151°C), the Mark III <b>space suit</b> undergoes testing in the Johnson Space Center's lunar yard.	Designado para agüentar temperaturas de -151°C a 115°C, a <b>roupa espacial</b> Mark III passa por testes na quadra lunar do centro espacial Johnson.

## 2 – Relação VEM DE, É FEITO A PARTIR DE

Experts estimate that producing and burning <b>cane ethanol</b> generates anywhere from 55 to 90 percent less carbon dioxide than gasoline.	Segundo estimativas dos especialistas, a produção e a queima do <b>álcool de cana</b> geram cerca de 55% a 90% menos dióxido de carbono do que no caso da gasolina.
Though corn kernels and <b>cane juice</b> are the traditional sources of ethanol, you can also make it from stalks, leaves, and even sawdust—plant by-products that are normally dumped, burned, or plowed back under.	Embora os grãos de milho e o <b>caldo da cana</b> sejam fontes tradicionais de etanol, pode-se obter álcool a partir de galhos, de folhas e até de serragem.
Germany is the world's leading producer, relying on <b>canola oil</b> ;	A Alemanha é o maior produtor, concentrando-se no <b>óleo de canola</b> ;
We know this because most evenings at six o'clock research assistant Sally Macdonald sits down with a set of sieves and buckets, and one or two ziplock bags of the <b>chimp feces</b> that the researchers bring back most days.	Sabemos disso porque, quase sempre às 6 da tarde, a assistente de pesquisa Sally Macdonald senta-se com um conjunto de peneiras e baldes, e um ou dois sacos plásticos repletos de <b>excrementos de chimpanzé</b> , recolhidos pelos pesquisadores

	quase todos os dias.
And producing <b>corn ethanol</b> consumes just about as much fossil fuel as the ethanol itself replaces.	A própria produção de <b>etanol de milho</b> consome uma quantidade considerável de combustível fóssil - justamente o que ele vem substituir.
Some studies of the energy balance of <b>corn ethanol</b> —the amount of fossil energy needed to make ethanol versus the energy it produces—suggest that ethanol is a loser's game, requiring more carbon-emitting fossil fuel than it displaces.	Alguns estudos sobre o equilíbrio energético do <b>etanol de milho</b> - os quais estimam a quantidade de energia fóssil necessária para produzi-lo - apontam a irracionalidade do processo, que requer mais combustível fóssil emissor de carbono do que aquele que viria a substituir.
But however the accounting is done, <b>corn ethanol</b> is no greenhouse panacea.	Entretanto, seja qual for o cálculo, o <b>etanol de milho</b> não representa nenhuma panacéia contra o efeito estufa.
It's easy to lose faith in biofuels if <b>corn ethanol</b> is all you know.	É fácil perder a fé nos biocombustíveis quando se conhece apenas o <b>etanol de milho</b> .
But many experts believe the touted benefits of <b>corn ethanol</b> —from cutting carbon dioxide emissions to improving energy security—are overblown.	Mas muitos especialistas acreditam que os benefícios alegados do <b>etanol de milho</b> – desde a diminuição das emissões de carbono ao aumento da segurança energética – são exagerados.
Rising demand for <b>corn ethanol</b> and speculation by large producers drove up prices for both yellow and white corn last year, causing tortilla prices to more than double in some parts of Mexico and setting off angry protests.	O aumento da demanda de <b>milho para a produção de etanol</b> e a especulação por parte dos grandes produtores fez os preços do milho branco e do amarelo subirem no ano passado, fazendo com que o preço da tortilha mais que dobrasse em algumas partes do México, o que causou protestos ferrenhos.
A beaker of <b>corn ethanol</b> , aka grain alcohol, burns brightly, revealing its energy potential.	O <b>etanol</b> , ou álcool de cereais, queima reluzente e exhibe seu potencial energético.
The start-up ethanol company hopes to make a greener <b>corn ethanol</b> by creating a closed-loop system in which methane from the cattle manure fuels the ethanol plant and spent grain from the distillery is fed back to the cattle.	A nova empresa de produção de etanol espera produzir <b>etanol de milho</b> mais ecológico por meio de um sistema mais fechado, em que o metano do esterco do gado alimenta a unidade de produção de etanol e os restos não usados para a fabricação de combustível servem de alimento para o gado.
Nebraska corn farmer Dwayne Johnson is one of many midwestern growers benefiting from the highest corn prices in years, thanks to increasing demand for <b>corn ethanol</b> .	Dwayne Johnson, que planta milho no Nebraska, é um dos muitos cultivadores do meio-oeste norte-americano que estão se beneficiando dos altos preços do milho em anos, graças ao aumento da demanda para <b>milho destinado à produção de etanol</b> .
Honeybees have long amazed scientists with their social behaviors (dancing out	As abelhas de mel há muito surpreendem os cientistas com seus comportamentos sociais

directions to a <b>food source</b> , working in tandem with thousands of hive mates, taking on specialized jobs within and outside the hive).	(fazer uma dança para dar indicações de como chegar a uma <b>fonte de alimento</b> , trabalhar em sintonia com milhares de companheiras de colméia, assumir tarefas especializadas tanto dentro quanto fora da colméia).
Replacement peaks during peak memory load—when birds learn new songs, find new <b>food sources</b> , or meet new social partners.	A troca é máxima durante os picos de exigência de memória, quando as aves aprendem novos cantos, procuram novas <b>fontes de alimento</b> ou encontram novos parceiros sociais.
But both inventors soon discovered that "rock oil," when slightly refined, held far more bang per gallon than <b>plant fuel</b> , and was cheap to boot.	No entanto, ambos os inventores logo descobriram que o "óleo de pedra", depois de refinado, liberava muito mais potência por litro do que seus <b>equivalentes vegetais</b> , além de ser barato.
Oil soon left <b>plant fuels</b> in the dust.	Por isso, logo o petróleo fez com que os <b>combustíveis vegetais</b> fossem descartados.
The lure of <b>plant fuels</b> even seems to have reached the petroleum-rich sands of the Middle East, where the United Arab Emirates has launched a 250-million-dollar renewable energy initiative that includes biofuels—perhaps a sign that even the sheikhs now realize that the oil age won't last forever.	O fascínio pelos <b>combustíveis vegetais</b> parece ter chegado aos desertos ricos em petróleo do Oriente Médio. Os Emirados Árabes Unidos destinaram 250 milhões de dólares a um programa de pesquisa de energia renovável que inclui os biocombustíveis - talvez um sinal de que até os xeques começam a se dar conta de que a era do petróleo não vai durar para sempre.
The superslick <b>plant oil</b> has also quieted a little of that annoying diesel rattle.	O superviscoso <b>óleo vegetal</b> contribui, ainda, para reduzir aquela incômoda trepidação dos motores a diesel.
Chemically altering <b>plant oils</b> to make biodiesel takes less energy than distilling corn into ethanol; the fuel's main drawbacks are low yield and high cost.	A alteração química de <b>óleos vegetais</b> para obter biodiesel consome menos energia do que a produção de etanol a partir de milho.
I wanted to shoot the soybeans and <b>sugar cane</b> as specimens, but also in their natural environment.	Eu queria fotografar a soja e a <b>cana</b> como se fossem espécimes, mas também em seu ambiente natural.
I had no idea <b>sugar cane</b> could grow 14, 16 feet tall.	Eu não fazia idéia de que a <b>cana-de-açúcar</b> chegava a quase cinco metros de altura.

## 3 – Material – Relação É FEITO DE

" <b>Carbon fiber</b> outperforms fly chitin," he said, with a trace of self-satisfaction.	"A <b>fibra de carbono</b> tem desempenho melhor que a quitina das moscas", comenta ele, com evidente satisfação.
The day was warm, but the villa sat atop a hill, and a cool breeze swept in. The sound of the secretary's shoes echoed off <b>marble floors</b> and columns.	Era um dia quente, mas, como a vila estava no alto do morro, uma brisa fria entrou de roldão. O som dos sapatos do secretário ecoou no <b>piso</b> e nas colunas <b>de mármore</b> .
Eleven hundred years later, the Spanish	Onze séculos depois, quando se viram na



conquistadores shed their own <b>metal armor</b> in the sweltering rain forest in favor of these Maya "flak jackets."	abafada e úmida floresta tropical, os conquistadores espanhóis logo substituíram suas <b>armaduras de metal</b> por esses coletes protetores maias.
He holds a <b>paper bag</b> containing hummus and bread.	Ele tem na mão uma <b>sacola de papel</b> , contendo húmus e pão, e fuma um cigarro L&M.
Now warriors from the Petén, equipped with <b>stone axes</b> and short stabbing spears, swelled their ranks.	Os guerreiros de Petén, munidos de <b>machados de pedra</b> e lanças, engrossaram as fileiras da tropa de assalto.
Rabbi Menachem Froman lives in the Jewish settlement of Tekoa, perched on a mesa, a clean collection of bleached <b>stone houses</b> capped with red-tiled roofs, double strollers parked on several porches.	Menachem Froman, o rabino, mora no assentamento judaico de Tekoa, uma coleção de <b>casas de pedra</b> esbranquiçada, encimada por telhado vermelho, muitas delas com carrinhos duplos de bebê na varanda.
The county government subsidized traditional <b>wood construction</b> , though most villagers would rather have used brick.	O governo do condado subsidia a <b>construção</b> tradicional <b>em madeira</b> , embora a maioria dos moradores preferiria usar tijolos.

## 4 – Meronímia – Relação POSSUI

What little is known or surmised about them has been pieced together from fragments of pottery, <b>animal bones</b> , obsidian flakes, and such oblique sources as comparative linguistics and geochemistry.	O pouco que se sabe, ou se supõe, sobre eles vem-se articulando, peça a peça, com base em fragmentos de cerâmica, <b>ossos de animais</b> , lascas de obsidiana (um tipo de vidro vulcânico) e fontes indiretas, tais como as fornecidas pela lingüística comparada e a geoquímica.
They generate new neurons to replace lost <b>brain cells</b> .	gerar neurônios para substituir <b>células cerebrais</b> perdidas.
At 24 square miles, Fongoli is the largest home range of any habituated <b>chimpanzee group</b> ever studied.	Com 63 quilômetros quadrados, a região de Fongoli é o maior território já estudado de um <b>grupo de chimpanzés</b> habituado aos seres humanos.
But no matter your version of Christianity—or even if you're not religious at all—there seems to be something significant to the cave beneath the <b>church floor</b> , with its odor of incense and candle wax, lit by a string of bare bulbs.	Não importa a versão do cristianismo que se siga - ou mesmo que não se siga religião nenhuma -, parece haver algo significativo na gruta no <b>subsolo da igreja</b> , com seu cheiro de incenso e vela, iluminada por uma fileira de lâmpadas nuas.
When Bella was younger, her parents began unplugging the <b>computer keyboard</b> and mouse so she wouldn't go online when she was home alone, but they knew this wouldn't last.	Quando a garota era menor, eles passaram a desconectar o <b>teclado</b> e o mouse <b>do computador</b> para que ela não ficasse online sozinha em casa. Sabiam, porém, que isso não ia durar.
Some 16 million "dots" of texture per square millimeter all but eliminate the glare on the right half of a <b>computer monitor</b> .	Cerca de 16 milhões de “pontos” de textura por milímetro quadrado eliminam quase todo o reflexo na metade direita de uma <b>tela</b>

	<b>de computador.</b>
Workers strip picture tubes from thousands of used <b>computer monitors</b> piled at Monitex, a Grand Prairie, Texas, recycler.	Trabalhadores retiram o tubo de milhares de <b>monitores</b> usados na Monitex, indústria recicladora do Texas.
How to Help Americans toss out about two million tons of unwanted electronics annually, fouling landfills with toxic lead from old <b>computer monitors</b> , cadmium from leaking batteries, and more.	Os norte-americanos jogam fora, a cada ano, cerca de dois milhões de toneladas de aparelhos eletrônicos que não querem mais, enchendo aterros de chumbo tóxico de antigos <b>monitores de computador</b> , cádmio de baterias que vazam e outros compostos tóxicos.
Now such bug-inspired technology has been adapted to dampen reflection on <b>computer monitors</b> and solar cells.	Hoje, esse tipo de tecnologia inspirada em insetos foi adaptado para diminuir o reflexo em <b>monitores de computador</b> e em células solares.
Part of the answer is that human destruction of <b>eucalyptus forests</b> has disrupted the customary feeding and roosting habits of some flying foxes, forcing them toward shady suburbs, orchards, botanical gardens, city parks, and closer proximity to people.	Parte da resposta é que a destruição das <b>florestas de eucalipto</b> desorganizara os hábitos de alimentação e de descanso de espécies de raposas-voadoras, forçando-as a frequentar subúrbios arborizados, jardins botânicos e parques urbanos - ou seja, obrigando-as a viver mais perto das pessoas.
Yes, I think if you have a <b>family member</b> with dementia you really understand how memory loss manifests in a person.	Si, acho que quando há <b>alguém na sua família</b> que perdeu a memória, você aprende como isso se manifesta nas pessoas.
The janjaweed may toss your kids into vats of boiling water as they had done to children in another village, and the Sudanese Air Force may bomb your wretched fields as they had before, killing five of your <b>family members</b> .	Os assassinos janjaweeds podem jogar seus filhos em tinas de água fervente, como fizeram com crianças de outro povoado, e a Força Aérea sudanesa pode bombardear quando quiser suas roças miseráveis, como naquela vez em que matou cinco <b>parentes</b> seus.
<b>Family members</b> only a decade younger than Bella's parents inhabited another world.	SEM TRADUÇÃO.
Bats came to eat the figs in that solitary tree, and the poor mare, seeking shade, grazing too carelessly, evidently swallowed not just grass but also something of what they dropped, such as <b>fruit pulp</b> , feces, urine, afterbirth, and virus	Os morcegos vinham comer os figos na árvore isolada, e o pobre animal, em busca de sombra e pasto, deve ter engolido não apenas grama mas também algo que eles haviam deixado cair, como <b>caroços de frutas</b> , fezes, urina, placenta, e vírus.
He'd lost his <b>gorilla family</b> , and also members of his own family. It was very hard, Balo said.	Ele perdera aqueles <b>gorilas</b> e também gente de sua família. "Foi muito duro."
Lowland <b>Gorilla Family</b>	SEM TRADUÇÃO.
<b>Gorilla Family: How to Help</b>	Como ajudar
The total range of the Kingo family is about six square miles (15 square kilometers), sections of which overlap with the ranges of other <b>gorilla families</b> .	O território total da família de Kingo tem cerca de 15 quilômetros quadrados, com partes sobrepostas a de outras <b>famílias de gorila</b> .

Doran-Sheehy has shown that dominant males in western <b>gorilla families</b> may be related (such as brothers, half brothers, or fathers and sons), which may help explain their remarkable tolerance of one another.	Diane Doran-Sheehy mostrou que os machos dominantes das <b>famílias de gorila</b> do oeste podem ser aparentados (irmãos, meio-irmãos, pais ou filhos), o que talvez explique em parte sua notável tolerância mútua.
"It's now built into our <b>group culture</b> ."	"Agora ela faz parte de nossa <b>cultura de grupo</b> ."
Being social creatures, they could easily pass the infection among <b>group members</b> by mutual grooming, infant care, or trying to rouse their sick or their dead.	SEM TRADUÇÃO.
The 21st-century inheritors of their tradition continue to hunt these waters, mainly in three incarnations: gangs that board vessels to rob the crews; multinational syndicates that steal entire ships; and <b>guerrilla groups</b> that kidnap seamen for ransom.	Os herdeiros de tal tradição no século 21 continuam a caçar nessas águas - em pequenas gangues que abordam as embarcações e assaltam a tripulação, sindicatos do crime multinacionais que roubam navios inteiros e <b>grupos guerrilheiros</b> que seqüestram navegantes para pedir resgate.
Europeans introduced new diseases and made slaves of many islanders, striking terrible blows to <b>island culture</b> .	Os europeus introduziram novas doenças e transformaram muitos ilhéus em escravos, o que foi um golpe terrível à <b>cultura da ilha</b> .
Accompanied by a sacred ship bearing an image of the god Amun, Taharqa is robed in a priestly <b>leopard skin</b> and crowned with the double uraeus that declares him Lord of the Two Lands—ruler of both Nubia and Egypt.	Acompanhado por uma nau sagrada carregando a imagem do deus Amon, Taharqa veste uma <b>pele de leopardo</b> sacerdotal e está coroado com o uraeus duplo, que o torna Senhor das Duas Terras – soberano da Núbia e do Egito.
She loves the views, the <b>mountain air</b> , the settlers' tight sense of community.	Ela adora a vista, o <b>ar da montanha</b> , o forte espírito comunitário dos colonos.
How the <b>sheep brain</b> encodes facial identity and emotional cues may be relevant in some human disorders, he says.	SEM TRADUÇÃO.
I met EP at his home, a bright bungalow in suburban San Diego, on a warm <b>spring day</b> .	Conheci EP em sua residência, uma luminosa casa térrea nos subúrbios de San Diego, em um cálido <b>dia de primavera</b> .
On a mild, late <b>spring day</b> , Akbar irrigated a small plot of wheat just outside the village.	Em uma tarde amena no fim da <b>primavera</b> , Akbar irrigava um pequeno trigal na orla do povoado.
On a steamy <b>summer afternoon</b> with squalls in the offing, coastal scientists Paul Kemp of Louisiana State University and John Lopez of the Lake Pontchartrain Basin Foundation set out by boat to inspect the "funnel," formed east of town by the levees lining the MRGO and another channel that converges with it, the Gulf Intracoastal Waterway.	Em uma abafada e úmida <b>tarde de verão</b> que prenunciava tempestades, dois cientistas especializados em áreas costeiras - Paul Kemp e John Lopez - saíram de barco para inspecionar o "funil" formado a leste da cidade pelos diques que ladeiam o MRGO e outro canal convergente, o Gulf Intracoastal Waterway.
After a <b>summer day's</b> swim, a boy returns	Depois de nadar num <b>dia de verão</b> , um

to his new bike.	menino volta para sua bicicleta nova. Divertimentos modernos ganham atenção crescente no vilarejo.
No schedule is posted for the freights; locomotives arrive every few days, gather <b>train cars</b> , and roll out without advance notice, sometimes in the middle of the night.	Não há tabela de horários para as locomotivas de carga. Os trens aparecem a cada tantos dias, arrebanham <b>vagões</b> e partem sem aviso prévio, às vezes no meio da noite.
These chimps fashion weapons from <b>tree branches</b> and apply them as primitive spears or skewers to stab or impale bush babies [tiny nocturnal primates], which hide in tree holes.	Esses chimpanzés fazem armas com <b>galhos de árvores</b> e usam como lanças u espetos primitivos para empalar gálagos [pequenos primatas noturnos] que se escondem em buracos de árvores.
Sex jokes, <b>village gossip</b> , little wisps of song, rebukes to children—all of it lay scribbled in the eccentric lines of their crops.	Piadas maliciosas, <b>fofocas do vilarejo</b> , fiapos de canções, pito nos filhos está tudo ali, escrito nas excêntricas linhas de sua roça.
Some of the riders were absolutely reckless, so there was a sense of surrealism about how and where we rode: off steep eroded paths and on slick cobbles, down narrow <b>village streets</b> (vehicles not bothering to make way or slow down, often).	Alguns dos cavaleiros eram completamente imprudentes, então havia uma noção surreal a respeito de como e onde cavalgávamos: em trilhas íngremes erodidas e sobre pedregulhos escorregadios, por <b>ruazinhas estreitas de vilarejos</b> (os carros com frequência não se davam ao trabalho de dar passagem nem de diminuir a velocidade).
The man—or what was left of him—emerged from the Irish sod one <b>winter day</b> in 2003, his hair still styled the way he wore it during his last moments alive.	O homem - ou o que restou dele - emergiu do relvado irlandês num <b>dia de inverno</b> de 2003 com o cabelo ainda penteado como estivera nos últimos momentos de vida: cortado rente na nuca e com um topete de 20 centímetros arrebitado para trás e fixado por resina de pinho.
And when Norbu twirls to a stop in a no-hands headstand, his shirt rides up to reveal his homage to global <b>youth culture</b> : red Nike high-top sneakers, baggy Adidas sweatpants, and a temporary tattoo that spells out, in jagged English letters, the name he and his homeys have adopted—“B-Boyz.”	E quando Norbu faz uma parada de cabeça, sem as mãos, a camisa cai e revela sua homenagem à <b>cultura jovem</b> global: tênis Nike vermelho de cano longo, calça larga de agasalho Adidas e uma tatuagem temporária que anuncia, em letras inglesas denteadas, o nome que ele e sua turma adotaram, B-Boyz.

## 5 – Conteúdo e recipiente – Relação CONTÉM

The cowboys hold their passion in reserve—waiting till the snows melt and the <b>cattle trailers</b> pull up to the gates, followed by the year's first swell of hoofbeats.	Os caubóis preservam sua paixão na intimidade, à espera de que neve derreta nas pradarias e <b>carretas transportando gado</b> apareçam no portão da fazenda, tudo ao som do tropel de gado, o primeiro do ano.
The general heavily subsidized and financed new ethanol plants, directed the state-owned	Geisel concedeu financiamentos para a construção de usinas de álcool, fez com que

oil company, Petrobras, to install <b>ethanol tanks</b> and pumps around the country, and offered tax incentives to Brazilian carmakers to crank out cars designed to burn straight ethanol.	a estatal Petrobras instalasse bombas de álcool em seus postos por todo o país e ofereceu incentivos fiscais à indústria automobilística para que fabricasse veículos com motor adaptado ao uso do novo combustível.
Nine vessels chugged through the strait, including a liquefied natural <b>gas tanker</b> that towered over the others like a skyscraper.	SEM TRADUÇÃO.
The ghost of the old man, however, was not happy, according to the neighbors. After the fire, the Eldest Son stuffed his father's torso into an old <b>rice bag</b> .	Mas o fantasma do velho não está satisfeito, segundo os vizinhos. Depois do incêndio, o primogênito pôs o tronco do pai num <b>saco de arroz</b> surrado.
Everything is frozen solid—boots, socks, sunscreen, <b>water bottles</b> —as if left over from some ghastrly ice age.	Tudo está duro de gelo - botas, meias, <b>garrafas d'água</b> -, feito sobras de alguma tenebrosa idade glacial.
He filled a <b>water bowl</b> for the animals and scattered some food.	Pôs água no <b>cocho</b> e espalhou comida para os animais. .EOS
But of course there are cowgirls, too, like Jodi Miner, a clear-eyed woman with a formidable handshake who grew up on a ranch near Dell, Montana, doctoring calves, repairing <b>water tanks</b> , and mending fences.	Mas há também as cowgirls, é claro, como Jodi Miner, uma mulher de olhos claros com um possante aperto de mão, criada numa fazenda perto de Dell, Montana, onde cresceu cuidando de bezerros e novilhas, consertando <b>caixas-d'água</b> e cercas. Nos intervalos da faculdade, em Bozeman, Jodi ocupava-se de bezerros, período em que dormia nos barracões, cercada de homens roncando.

## 6 – Relação É LOCALIZADO EM

Hobbs says he has fielded dozens of calls from power companies interested in building an <b>algae plant</b> of their own to scrub emissions and help meet their renewable fuels mandate.	Hobbs afirma que foi procurado dezenas de vezes por empresas de eletricidade interessadas em instalar <b>fazendas de algas</b> para limpar emissões e contribuir para seus programas compulsórios de pesquisa de novos combustíveis.
The neighboring <b>baseball field</b> is named Field of Dreams.	A <b>quadra de basquete</b> ao lado se chama "A Quadra dos Sonhos".
Bodies still lying undiscovered in the bogs of northern Europe will yield more clues about how and why the <b>bog people</b> met their ends.	SEM TRADUÇÃO.
Previous x-rays of his body were hard to read—the bones, demineralized by acidic <b>bog waters</b> , looked like glass.	As radiografias anteriores eram difíceis de interpretar, pois os ossos, desmineralizados pelas <b>águas ácidas do pântano</b> , pareciam de vidro.
Instead, they call themselves <b>border guards</b> , reconnaissance brigades, even mujahideen--holy warriors.	Em vez disso, eles se autodenominam <b>guardas de fronteira</b> , brigadas de reconhecimento e até mesmo mujahideen —

	guerreiros sagrados.
But in the early 1900s a brilliant <b>city engineer</b> , A.	SEM TRADUÇÃO.
This son of Hazarajat is the proverbial <b>country boy</b> who came to the big city and made good.	Esse filho de Hazarajat é o proverbial <b>rapaz do interior</b> que foi para a cidade grande e se deu bem.
All around the city, the sidewalks are choked with used <b>electronics shops</b> .	Por toda a cidade, as ruas estão repletas de <b>lojas que vendem equipamentos eletrônicos</b> de segunda mão.
The <b>energy farm</b> , as GreenFuel calls it, isn't much to look at, just a cluster of shipping containers and office trailers next to a plastic greenhouse structure longer than a football field and perhaps 50 feet (15 meters) wide.	A <b>fazenda de energia</b> , para usar o termo adotado pela GreenFuel, não passa de um agrupamento de contêineres e trailers com escritórios ao lado de uma grande estufa coberta de plástico.
Christine Wietzki, a former <b>farm kid</b> from western Nebraska, is technical manager for one of the newest and most advanced ethanol plants in the country, the E3 BioFuels plant in tiny Mead, Nebraska, population 564.	Christine Wietzki responde pela gerência técnica de uma das mais avançadas usinas de etanol dos Estados Unidos, a E3 BioFuels, localizada na minúscula Mead, em Nebraska, com população que se resume a 564 pessoas.
The start-up <b>ethanol company</b> hopes to make a greener corn ethanol by creating a closed-loop system in which methane from the cattle manure fuels the ethanol plant and spent grain from the distillery is fed back to the cattle.	A nova <b>empresa de produção de etanol</b> espera produzir etanol de milho mais ecológico por meio de um sistema mais fechado, em que o metano do esterco do gado alimenta a unidade de produção de etanol e os restos não usados para a fabricação de combustível servem de alimento para o gado.
This year it's heading for a nearby <b>ethanol plant</b> .	Neste ano, seu destino é virar <b>álcool</b> .
The start-up ethanol company hopes to make a greener corn ethanol by creating a closed-loop system in which methane from the cattle manure fuels the <b>ethanol plant</b> and spent grain from the distillery is fed back to the cattle.	A nova empresa de produção de etanol espera produzir etanol de milho mais ecológico por meio de um sistema mais fechado, em que o metano do esterco do gado alimenta a <b>unidade de produção de etanol</b> e os restos não usados para a fabricação de combustível servem de alimento para o gado.
By this summer, with Nebraska's 16 <b>ethanol plants</b> gearing up to consume a third of the state's crop, corn prices had doubled, briefly topping four dollars a bushel, and growers were looking forward to the best profits in memory.	Em meados deste ano, com as 16 <b>usinas de etanol</b> de Nebraska se preparando para processar um terço da safra do estado, os preços do milho haviam dobrado e os agricultores estavam ansiosos para embolsar os lucros mais polpudos que tinham visto em toda a vida.
But more <b>ethanol plants</b> are popping up all the time.	SEM TRADUÇÃO.
Christine Wietzki, a former farm kid from western Nebraska, is technical manager for	Christine Wietzki responde pela gerência técnica de uma das mais avançadas <b>usinas</b>

one of the newest and most advanced <b>ethanol plants</b> in the country, the E3 BioFuels plant in tiny Mead, Nebraska, population 564.	<b>de etanol</b> dos Estados Unidos, a E3 BioFuels, localizada na minúscula Mead, em Nebraska, com população que se resume a 564 pessoas.
Most <b>ethanol plants</b> burn natural gas or, increasingly, coal to create the steam that drives the distillation, adding fossil- fuel emissions to the carbon dioxide emitted by the yeast.	A maioria das <b>usinas de etanol</b> depende da queima de gás natural ou, cada vez mais, de carvão para gerar o vapor utilizado na destilação, adicionando emissões de combustíveis fósseis ao dióxido de carbono produzido pela levedura.
The general heavily subsidized and financed new <b>ethanol plants</b> , directed the state-owned oil company, Petrobras, to install ethanol tanks and pumps around the country, and offered tax incentives to Brazilian carmakers to crank out cars designed to burn straight ethanol.	Geisel concedeu financiamentos para a construção de <b>usinas de álcool</b> , fez com que a estatal Petrobras instalasse bombas de álcool em seus postos por todo o país e ofereceu incentivos fiscais à indústria automobilística para que fabricasse veículos com motor adaptado ao uso do novo combustível.
The Melanophila beetle, which lays its eggs in freshly burned wood, has evolved a structure that can detect the precise infrared radiation produced by a <b>forest fire</b> , allowing it to sense a blaze a hundred kilometers away.	O besouro Melanophila, que costuma pôr seus ovos em madeira recém-queimada, é capaz de identificar a exata radiação infravermelha emitida por um <b>incêndio florestal</b> , permitindo-lhe localizar queimadas distantes a até uma centena de quilômetros.
Whenever they rented out a vehicle, they made sure it had just enough fuel to make it to a <b>gas station</b> .	Essa era a estratégia de negócios da Prosperous Automobile: sempre entregavam o veículo com gasolina suficiente apenas para se chegar ao <b>posto</b> mais próximo.
But some truck drivers at a nearby <b>gas station</b> point us toward a shipment facility just up the street, where they say computers are often unloaded.	Os guardas nos impedem de entrar, mas motoristas de caminhão, em um <b>posto de gasolina</b> ali perto, apontam um depósito na mesma rua que, segundo eles, costuma receber carregamentos de computadores obsoletos.
Viewed this way, the Sahel represents the oldest <b>killing field</b> in human history.	Visto desse modo, o Sahel representa o mais antigo <b>campo de batalha</b> da história humana.
At first the <b>lab people</b> called it equine morbillivirus, meaning a horse virus closely related to measles.	No início, os <b>técnicos</b> o batizaram de "morbillivírus equino", imaginando tratar-se de um vírus bastante próximo do patógeno do sarampo.
"The U.S. right now is shipping large quantities of leaded materials to China, and China is the world's major <b>manufacturing center</b> ," Weidenhamer says.	"Neste exato momento, os Estados Unidos estão enviando enormes quantidades de materiais com chumbo à China, um dos principais <b>centros manufatureiros</b> do mundo", diz Weidenhamer.
NASA is also planning a much larger cargo rocket called Ares V, which will be able to lift as much as 150 tons (136 metric tons) to	A Nasa também planeja um foguete de carga muito maior, o Ares V, capaz de colocar 150 toneladas em órbita, incluindo o

orbit, including the booster rocket, lander, and other hardware needed for a <b>moon expedition</b> .	propulsor, o módulo lunar e demais equipamentos necessários a uma <b>expedição lunar</b> .
Moose have also returned, along with the <b>mountain lion</b> .	Os alces também voltaram à paisagem, assim como o <b>puma americano</b> .
The general heavily subsidized and financed new ethanol plants, directed the state-owned <b>oil company</b> , Petrobras, to install ethanol tanks and pumps around the country, and offered tax incentives to Brazilian carmakers to crank out cars designed to burn straight ethanol.	Geisel concedeu financiamentos para a construção de usinas de álcool, fez com que a estatal Petrobras instalasse bombas de álcool em seus postos por todo o país e ofereceu incentivos fiscais à indústria automobilística para que fabricasse veículos com motor adaptado ao uso do novo combustível.
Amid dense palm and rubber plantations that stretched for miles, we suddenly passed a small <b>orchid farm</b> .	Entre palmeiras e plantações de seringueiras densas que se estendem por quilômetros, de repente passamos por um pequeno <b>sítio de orquídeas</b> .

## 7 – Relação OCORRE EM

<b>Election day</b> , like everything in school, ended with a moral.	O <b>dia da eleição</b> , como tudo mais na escola, terminou com uma lição de moral: "Não fique mal se você perdeu desta vez", resume a professora.
Percussionists in a Brihadishvara mural play the same instruments that are heard today during <b>feast days</b> in the temple's vast courtyard.	Percussionistas de um mural de Brihadishvara tocam os mesmos instrumentos que hoje se escuta em <b>dias de festejos</b> no amplo pátio do templo.
Thursdays were <b>judgment day</b> in El Fasher.	Quinta-feira é <b>dia de julgamento</b> em El Fasher.
Lu Yan had just gotten her four-year college degree at <b>night school</b> and planned to study English next.	Lu Yan tinha acabado de terminar seu quarto ano do colegial na <b>escola noturna</b> e planejava estudar inglês a seguir.
After lunch, it's <b>siesta time</b> .	Depois do almoço, é <b>hora da sesta</b> .
But <b>winter mountaineering</b> , involving exponentially more pain and suffering than <b>summer climbing</b> —frostbite, hypothermia, avalanches—became an obsession of the Poles.	Assim, o <b>montanhismo de inverno</b> , que envolve, de forma exponencial, mais dores e sofrimentos que a <b>escalada de verão</b> , virou obsessão para os poloneses.
It is a scene of stark beauty: rolling hills blanketed by sprouts of <b>summer grass</b> ; herds of yaks and sheep grazing on distant slopes; and in the foreground a clear, shallow stream that is the beginning of the Yellow River.	A cena é deslumbrante: elevações ondulantes forradas de brotos de <b>grama de verão</b> , rebanhos de iaques e ovelhas pastando nas encostas distantes e, em primeiro plano, um ribeirão cristalino e raso, que é o começo do rio Amarelo. "
All her colleagues were enrolled in outside classes; even the vice-principal took a <b>weekend class</b> on educational technology.	Todas as suas colegas estavam tendo aulas particulares. Até a vice-diretora tinha <b>aulas de fim de semana</b> em tecnologia educacional.
Bella's mother took on more administrative	A mãe de Bella assumiu mais



responsibilities at work and enrolled in a <b>weekend class</b> to qualify to study for a master's degree.	responsabilidades administrativas no trabalho e passou a ter <b>aulas nos fins de semana</b> visando uma pós-graduação.
--	---

## 8 – Relação TEM

But there is a bottom line, and it comes at the end of October, when the 4,100 head he is paid to tend are herded into corrals, and the <b>cattle owners</b> roll up in their dusty pickups to count and inspect their property.	Para Miner, todos os anos isso tem prazo para acabar, o fim de outubro, quando as 4,1 mil cabeças que lhe pagam para cuidar são recolhidas, à espera dos <b>donos</b> .
A company run by a <b>family friend</b> went bankrupt, and his daughter, who was Bella's age, started buying clothes at discount stalls.	Uma firma tocada por um <b>amigo da família</b> foi à falência, e sua filha, da idade de Bella, passou a comprar roupas baratas em bancas.
She is the daughter of <b>family friends</b> , a Sayed Shiite who traces her lineage to the prophet Muhammad. .EOS	Ela vem de uma <b>família de amigos</b> xiitas da seita Sayed que descendem do profeta Muhammad.
He was in a hut next to the <b>family home</b> because people understand how contagious it is.	Ele estava em uma cabana ao lado da <b>casa da família</b> porque as pessoas compreendem como a doença é contagiosa.
Drifting to sleep on a prayer rug outside Ademu's <b>family mosque</b> , it was possible to imagine that there was nothing in the world that could not be reclaimed.	Adormecendo num tapete para oração às portas da <b>mesquita da família</b> de Ademu, é possível imaginar que não havia nada no mundo a ser corrigido.
Billy Karesh himself had visited the same area in 2000, before Ebola struck, to gather baseline data on <b>gorilla health</b> .	Karesh já visitara essa área em 2000, antes do ataque do ebola, para colher dados sobre a <b>saúde dos gorilas</b> .
Thick black smoke would hang over the city, as crowds flipped cars, set fire to offices, and threw stones at <b>government buildings</b> , demanding more democratic representation.	Com a cidade coberta por uma fumaceira negra, a multidão tombaria carros, atearia fogo em escritórios e apedrejaria <b>prédios do governo</b> , exigindo mais representação democrática.
New technology may improve <b>government efficiency</b> , but the changes democracy activists are seeking go deeper.	A tecnologia pode aumentar a <b>eficiência do governo</b> ; contudo, mudanças democráticas são mais profundas.
Like any <b>group house</b> , honeybee hives can get too crowded.	Assim como qualquer <b>residência grupal</b> , colméias de abelhas podem ficar lotadas demais.
But one day, I was under a bridge, where a lot of addicts live and shoot up heroin, when suddenly I heard the loud wail of <b>police sirens</b> and saw two <b>police cars</b> coming at high speed.	SEM TRADUÇÃO.
They are singing about <b>workers rights</b> , the man told me in broken English.	Cantavam sobre <b>direitos dos trabalhadores</b> , o homem me disse em inglês capenga.