

REPRESENTAÇÃO DA SEMÂNTICA DOS COMPOSTOS NOMINAIS EM LÍNGUA INGLESA ATRAVÉS DE *TEMPLATES*: UM ESTUDO DE *CORPUS* PARALELO

Lílian Figueiró TEIXEIRA*
Rove Luiza de Oliveira CHISHMAN**

ABSTRACT: *This paper presents an investigation about the semantics of the elements of the noun compounds in English, which are formed by two nouns (NN compounds) and are found in ten editions of the National Geographic Magazine. Through the methodological resources from Corpus Linguistics – such as compilation and alignment of a parallel corpus – it was possible to identify translation correspondents in Portuguese, which gave hints about the semantics of each construction. The identification of the semantic patterns was based on the cognitive study of Ryder (1994), who proposes a classification according to linguistic templates, which are schemas that include semantic characteristics of the components and of the compound structure. This theory reminds the frames by Fillmore (1982), which are a description of a specific context, made through the identification of actors and lexical items related.*

KEYWORDS: *noun compounds; corpus linguistics; cognitive linguistics; frame semantics; lexical semantics.*

1. Introdução

O estudo apresentado neste artigo é parte de uma dissertação de Mestrado (TEIXEIRA, 2009) que tratou da semântica das construções compostas formadas por dois substantivos, doravante compostos NN (*noun-noun compounds*). Partiu-se de compostos encontrados em língua inglesa na revista *National Geographic*, com o objetivo de identificar como os elementos constituintes de cada composto se relacionam semanticamente. Como a revista selecionada também apresenta uma edição em português, foi possível construir um *corpus* paralelo, em que os textos em inglês são alinhados com seus correspondentes em português. A identificação dos correspondentes de tradução dos compostos contribuiu para a análise da semântica de cada construção.

As teorias que serviram de base para este estudo são apresentadas na próxima seção, sendo elas os *templates* de Ryder (1994) e os *frames* de Fillmore (1982). Para expressar os resultados da pesquisa, apresentamos alguns *templates* identificados no *corpus* na seção 3. Como uma tentativa de incluir expressões compostas nos *frames*, tendo como base os dados projeto *FrameNet*, sugerimos a inclusão de dados encontrados no *corpus* em alguns *frames*, proposta inédita para este trabalho.

2. *Templates* linguísticos e *frames* semânticos

Nesta seção, 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*¹ (BAKER *et al*, 1998) também serão apresentados.

* Mestre em Linguística Aplicada; Universidade do Vale do Rio dos Sinos.

** Doutora em Linguística e Letras; Universidade do Vale do Rio dos Sinos.

¹ Disponível em: <<http://framenet.icsi.berkeley.edu/>>

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. 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.

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.

É possível organizar os compostos NN em esquemas através de duas formas: (i) a combinação de palavras evoca um único esquema; (ii) 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, 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”.

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. 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.

Os *templates* de Ryder (1994) também podem ser chamados de *frames*, por isso foi feita a relação com o *FrameNet* neste trabalho. Fillmore (1982) 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 é feita a partir da descrição do papel semântico e dos seus argumentos.

Segundo Fillmore (1982), 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 (1982), 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 (1982). As principais características desse projeto são o uso de evidência empírica, de *corpus*, para a elaboração de generalizações sintáticas e semânticas; e 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): (i) o léxico, com uma lista de itens que são relacionados aos elementos *frames* e a frases de exemplo extraídas de um *corpus*; (ii) 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; (iii) 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.

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 [rel] 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

Fig. 1 - *Frame* Localização_por_evento da base de dados do *FrameNet*. Disponível em: http://framenet.icsi.berkeley.edu/index.php?option=com_wrapper&Itemid=118&frame=Locale_by_event&.

3. Análise do *corpus*

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”². 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.

Para a seleção dos compostos a serem analisados, foi necessário um pré-processamento do *corpus*, que pode ser dividido em dois momentos: coleta e alinhamento do *corpus*; e extração e busca no *corpus* paralelo.

Os dados foram coletados nos *sites* da revista e são constituídos por dez edições publicadas entre agosto de 2007 e maio de 2008. Os textos em inglês e em português foram copiados e colados em um arquivo *.txt*, pois esse formato é pré-requisito para o processamento de textos em diversos programas. Coletado o *corpus*, 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. O *corpus* precisou ser itemizado, já que esse formato é pré-requisito para o alinhador *Vanilla Aligner*³ (DANIELSSON; RIDINGS, 1997). O itemizador, que utiliza a arquitetura Java J2SE, formata o texto em uma palavra por linha. 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.

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.

Optou-se pelo etiquetador *TreeTagger* (SANTORINI, 1990) para a língua inglesa, por ser uma ferramenta gratuita e com bons resultados. 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 J2SE⁴. Essas três ferramentas, o itemizador, o etiquetador e o extrator, foram utilizadas para o processamento do *corpus* em língua inglesa.

Como saída, o programa oferece uma lista com possíveis compostos NN e o seu número de ocorrências no *corpus*, ou seja, a sua frequência. A partir dos resultados do

² Tradução livre por Teixeira, L. F.

³ O programa é disponibilizado gratuitamente no *site* do LAEL da PUCSP: <<http://www2.lael.pucsp.br/corpora/alinhador/index.html>>.

⁴ 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.

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. 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 (RYDER, 1994). Partindo-se de *core words*, é mais fácil chegar a possíveis *templates*. 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 uma ferramenta chamada *Wordlist* (lista de palavras) do programa *WordSmith Tools* (SCOTT, 2008) para se chegar aos dados. Esse recurso traz uma lista com todas as palavras que ocorrem no *corpus*, seguidas pelo número de vezes que elas são utilizadas. Algumas das *core words* mais frequentes são as seguintes, seguidas pelo número de ocorrências: *water/waters* (41), *space* (40), *day/days* (37), *family/families* (34), *home/homes* (29), *tree/trees* (28), *plant/plants* (27), *field/fields* (25).

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 (número obtido pelo filtro das *core words* mais frequentes) por meio do concordanceador do *WordSmith Tools*. O concordanceador permite a visualização de todas as frases, do *corpus*, em que uma palavra de busca ocorre. Em caso de dúvida, conferiu-se a sua tradução no *corpus* paralelo.

Por meio dessa conferência, o número de compostos diminuiu bastante, chegando a 842 compostos NN. Partindo-se dos resultados de extração, para a fase de análise, foram selecionados 200 compostos aleatoriamente. No quadro a seguir, são apresentadas as relações mais frequentes identificadas no *corpus*, com exemplos e a tradução identificada no *corpus* paralelo. Alguns compostos não apresentam um correspondente de tradução no texto em português porque algumas frases ou partes de frases no original foram suprimidas na tradução.

Relação	Types	Exemplo	Tradução
serve para	25	<i>memory drugs</i>	<i>drogas</i>
vem de, é feito a partir de	9	<i>cane juice</i>	<i>Sem tradução</i>
é feito de	7	<i>metal armor</i>	<i>armaduras de metal</i>
possui (parte integrante)	21	<i>church floor</i>	<i>solo da igreja</i>
possui (característica abstrata)	5	<i>island culture</i>	<i>cultura da ilha</i>
contém (recipiente)	7	<i>rice bag</i>	<i>saco de arroz</i>
é localizado em	40	<i>school play</i>	<i>Sem tradução</i>
ocorre em (tempo)	9	<i>night school</i>	<i>escola noturna</i>
tem (posse)	11	<i>family mosque</i>	<i>mesquita da família</i>

Fig. 2 - Relações mais frequentes encontradas no *corpus*

As relações identificadas na figura 2 foram adaptadas a partir dos esquemas de Ryder (1994) e de outros estudos apresentados pela mesma autora. Optou-se por expressar a relação através de um verbo para que assim fosse possível estabelecer uma integração com a base de

dados do *FrameNet*, já que o elemento principal de cada *frame* é um verbo, que evoca toda a situação. Essa tentativa de integração dos dados é apresentada na próxima seção.

Até mesmo Ryder (1994) utilizou verbos para expressar seus esquemas. A relação “serve para” ou *purpose* não faz parte dos esquemas de Ryder (1994), mas é sugerida por diversos estudos apresentados pela autora, tais como Jespersen (1909), Marchand (1969), Warren (1978) e Adams (1973) (*apud* RYDER, 1994). Optou-se por considerar essa relação, pois foram encontrados exemplos deste tipo no *corpus*, como *food crops*, *farm tools*, *fuel alcohol* e *education center*. Nesse tipo de relação, pode-se dizer que o N1 indica para que serve o N2. A mesma relação é identificada em um estudo bilíngue (inglês e italiano) sobre a semântica dos compostos (BUSA; JOHNSTON, 1999), que representam as relações através da estrutura *qualia* de Pustejovsky (1995). A relação “serve para” corresponde ao papel télico da estrutura *qualia*, em é possível identificar a função ou o propósito de algo.

A relação “vem de/é feito a partir de” também não é apresentada por Ryder (1994), mas traz uma característica interessante que pôde ser identificada em 9 compostos do *corpus*, tais como: *cane ethanol*, *canola oil*, *plant fuel*, e *sugar cane*. Essa relação assemelha-se ao predicado FROM de Levi (1978). Levi (1978 *apud* RYDER, 1994) propõe que por trás de cada composto há uma estrutura subjacente formada por predicados apagados. Esses predicados apagados podem fazer parte da forma subjacente de um composto NN, sendo uma forma de expressar a relação entre os dois substantivos. O predicado FROM indica uma relação de direção, de onde algo veio.

Do esquema “*X is made of Y*” de Ryder (1994), em que o Y é o primeiro elemento e X é o segundo, foi feita a tradução da relação “é feito de”. Esse esquema se mostrou relevante já que foram encontrados 7 compostos desse tipo no *corpus*, todos indicando o material do qual um determinado item é feito: *carbon fiber*, *marble floors*, *metal armor*, *paper bag*, *stone axes*, *stone houses* e *wood construction*. O mesmo acontece com a relação “possui (parte integrante)” que é uma tradução dos esquemas “*X has Y as a part/component*” e “*X is a part of Y*”. Neste estudo não houve a preocupação em estabelecer relações diferentes, neste caso, para diferenciar se a parte é o primeiro ou o segundo elemento. Entende-se que o verbo “possuir” pode representar a mesma relação tanto se a parte for o primeiro ou o segundo substantivo. Em *church floor*, a parte é o segundo, enquanto que em *gorilla families* a parte é o primeiro elemento. Pode-se afirmar que a igreja “possui” um piso e que a família “possui” gorilas, no sentido de que os gorilas são membros de um grupo. Há, segundo Warren (1978 *apud* RYDER, 1994), diversas relações de parte e todo, sendo que puderam ser identificadas neste estudo as seguintes: parte integrante, membro de grupo, parte integrante de local ou tempo, uma característica abstrata e posse. Como entendemos que a característica abstrata como em *room temperature* e a posse, como em *family car* (exemplos de Warren) são relações que diferem dos outros casos, em que é evidente a relação entre um todo e a sua parte, decidimos estabelecer duas categorias diferentes para esses casos, “possui (característica abstrata)” e “tem (posse)”. A relação de posse pode ser identificada nos esquemas “*X is owned by Y*” e “*X owns Y*” de Ryder (1994), que também sentiu necessidade em evidenciar a relação de posse. As três relações, “possui (parte integrante)”, “possui (característica abstrata)” e “tem (posse)” podem ser identificadas com o papel constitutivo de Pustejovsky (1995), que estabelece a relação entre um objeto e suas partes.

Ryder (1994) também estabelece a relação entre conteúdo e recipiente através do verbo contém, expresso pelos esquemas “*X contains Y*” e “*X is contained in Y*”. Já a relação de localização no espaço é expressa pelos esquemas “*X is located within Y*” e “*X is the location of Y*”. Ryder (1994) utiliza, nos esquemas de localização espacial, preposições que indicam se o objeto está dentro, sobre ou abaixo do referente. Como para este trabalho, não

pareceu relevante expressar a localização exata, já que buscávamos uma relação mais geral, sugerimos apenas uma relação.

Mesmo não tendo em Ryder (1994) um esquema específico para a localização no tempo, achamos importante expressar essa relação de alguma forma, já que foram encontrados 9 casos no *corpus*, tais como *judgment day*, *weekend class*, *night school* e *summer grass*.

Além da identificação das relações semânticas mencionadas, é possível dividir os compostos em grupos de acordo com a natureza dos elementos: pessoas, animais, plantas, artefatos, substâncias e eventos. Cada um desses grupos apresenta padrões semânticos que são explicitados a seguir.

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 ou a algo que é possuído. Exemplos: *country boy* = menino cuja origem é o campo; *family member* = pessoa que faz parte deste grupo; *family mosque* = mesquita que pertence à família.

As construções com **animais** trazem informações sobre o local onde o animal vive, partes do seu corpo, o animal como membro de um grupo ou possuidor de algo. Exemplos: *desert beetle* = besouro que vive no deserto; *leopard skin* = pele de leopardo; *gorilla families* = família da qual o gorila faz parte.

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. Exemplos: *feed corn* = milho destinado à alimentação de animais; *eucalyptus forest* = floresta constituída por eucaliptos; *tree branches* = galhos que fazem parte da árvore; *orchid farm* = fazenda onde orquídeas são cultivadas.

O grupo dos **artefatos** é o que mais tem recebido a atenção dos estudos sobre a semântica dos compostos. 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 e localização quanto ao uso. A relação mais frequente é a de função, ocorrendo na seguinte ordem: função + artefato: *car keys* = chaves usadas para abrir o carro. Outros exemplos: *stone axes* = machados feitos de pedra; *computer keyboard* = o teclado faz parte do computador;

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 ou se ele é o conteúdo de algum recipiente. Por origem, compreende-se de onde a substância vem ou onde ela foi produzida. Exemplos: *corn ethanol* = etanol ou álcool feito a partir do milho; *rice bag* = saco que contém arroz.

Por fim, puderam ser identificados padrões semânticos relacionados a **eventos**. Informações como local e tempo em que um evento ocorre aparecem nos compostos NN. Estes são alguns exemplos de compostos eventivos encontrados no *corpus*: *baseball field* = um jogo que ocorre em um lugar específico, o campo; *weekend class* = aula que ocorre no final de semana.

4. Proposta de *frame* com compostos nominais

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 a figura 1, apresentada anteriormente, pode-se ver que um dos elementos lexicais que evoca o *frame*

Locale_by_event (localização por evento) é a palavra *field*. Nesse caso, a função desse local é de servir de “palco” para um determinado evento, como em *baseball field*. No entanto, o mesmo item lexical pode evocar outro frame, o *Locale_by_use* (localização pelo uso) se nos referirmos a um campo que serve para o cultivo de alguma espécie, como em *rice fields*. De forma semelhante, o elemento lexical *farm* também pode fazer parte do frame *Locale_by_use*, sendo aplicável para uma *orchid farm*, que é o local destinado à cultura de orquídeas.

Com os dados identificados no *corpus* de estudo, poderíamos incluir compostos NN na lista de itens lexicais nos dois *frames* citados. Entre os exemplos de localização por evento, temos: *baseball field*, *moon expedition*, *research sites*, *research vessel*, *school play*, *space exploration* e *trade routes*. Já os itens que denotam uma localização pelo uso são os seguintes: *algae plant*, *electronics shop*, *energy farm*, *ethanol company*, *ethanol plants*, *gas station*, *oil company*, *orchid farm*, *rice fields* e *storage facility*.

Comparando os dois *frames*, percebe-se que a localização por evento sempre traz uma palavra que remete diretamente a um evento, como jogo, peça de teatro ou pesquisa. Já, na localização pelo uso, a palavra relacionada ao local denota algo que é produzido ou, de certa forma, depositado no local.

Mesmo que através do referente do composto, como *field* e *farm*, possamos chegar ao *frame*, o que contribui para a compreensão da relação semântica entre os elementos de um composto, certamente, uma base de dados lexical que pudesse incluir expressões compostas seria muito mais rica e útil, tanto para uma análise semântica feita por humanos quanto para uma possível automatização deste processo.

Referências

- BAKER, Collin F.; FILLMORE, Charles J.; LOWE, John B. The Berkeley FrameNet project. In: *Proceedings of the COLING-ACL*. Montreal, Canada: ACL, 1998.
- BIBER, Douglas. Using Register-Diversified Corpora for General Language Studies. *Computational Linguistics*, Cambridge, v. 19, n. 2, p. 219-241, jun. 1993.
- 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.
- 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.
- FILLMORE, Charles J. Frame Semantics. In: LINGUISTICS Society of Korea (ed.). *Linguistics in the Morning Calm*. Seoul: Hanshin, 1982. p. 111-137.
- FRAMENET. Disponível em: <<http://framenet.icsi.berkeley.edu/>>. Acesso em: 23 jan. 2010.
- LANGACKER, Ronald W. *Foundations of cognitive grammar*. Volume I: Theoretical prerequisites. Stanford: Stanford University, 1987. 540p.
- NATIONAL GEOGRAPHIC MAGAZINE. Disponível em: <<http://ngm.nationalgeographic.com/ngm/2007-11/tableofcontents.html>>. Acesso em: 24 dez. 2007.
- PUSTEJOVSKY, James. *The Generative Lexicon*. Cambridge: MIT, 1995. 298p.
- REVISTA NATIONAL GEOGRAPHIC BRASIL. Disponível em: <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.



Anais do IX Encontro do CELSUL
Palhoça, SC, out. 2010
Universidade do Sul de Santa Catarina

TEIXEIRA, LÍlian Figueiró. *A semântica dos compostos nominais: um estudo de corpus paralelo inglês/português*. 2009. 209 f. Dissertação (Mestrado em Linguística Aplicada) - Programa de Pós-Graduação em Linguística Aplicada, Universidade do Vale do Rio dos Sinos, São Leopoldo, RS, 2009. Orientação de Rove Luiza de O. Chishman.

VANILLA ALIGNER. Alinhador online de textos paralelos. CEPRIIL, LAEL, PUC/SP.

Disponível em: <<http://www2.lael.pucsp.br/corpora/alinhador/index.html>> Acesso em: 20 dez. 2007.

SCOTT, M. *WordSmith Tools 5.0*. Disponível em:

<<http://www.lexically.net/downloads/version5/HTML/index.html>>. Acesso em: 22 dez. 2007.