

# ANÁLISE ONTOLOLÓGICA APLICADA AO DESENVOLVIMENTO DE *FRAMES*

Alexandra MOREIRA \*

Maria Margarida Martins SALOMÃO \*\*

- **RESUMO:** Semântica de Frames é uma teoria que busca estabelecer a semântica de um item lexical pela evocação da cena conceitual a ele associada. A FrameNet é uma base léxico-semântica fundamentada na Semântica de Frames, constituída por um conjunto de *frames* relacionados por ligações semânticas. A metodologia de descoberta e incorporação de *frames* à base obedece a critérios semiformais e incorporam alguns critérios ontológicos. Neste artigo, buscamos o aprofundamento dessa análise ontológica de forma a permitir maior formalização do processo. Para isso, analisamos sentenças do Corpus do Português contendo o lexema *viagem* e alinhamos os elementos de *frames* com os conceitos pertencentes a ontologias de topo. Em seguida, detectamos as sentenças que incorporavam o sentido prototípico apresentado na FrameNet. Além disso, identificamos o uso metafórico do lexema e seu uso com o sentido de transporte a partir da distinção do tipo ontológico do agente. Assim, a partir da análise ontológica foi possível encontrar um novo *frame* ainda não definido na base FrameNet. Essas evidências sugerem possíveis benefícios da metodologia proposta, podendo facilitar, futuramente, uma formalização na ligação de *frames* distintos motivados pela polissemia de um lexema. Acreditamos, dessa forma, favorecer a inferência textual automática e o uso da rede por pessoas.
- **PALAVRAS-CHAVE:** Semântica de Frames. Frames. Ontologias.

## Introdução

A Semântica de Frames foi proposta por Charles Fillmore (1977) com o objetivo de explicar as variações de sentidos e de valência dos itens lexicais. Nas palavras de Croft (2009, p.1), a Semântica de Frames é “uma teoria de como os conceitos são organizados e representados na mente.” Na Semântica de Frames, o conceito fundamental é o conceito de *frame*: estruturas conceituais estabelecidas na memória permanente, frutos de nossa interação com o mundo e da consolidação de nossa experiência diária. A grande contribuição

---

\* UFJF – Universidade Federal de Juiz de Fora. Pós-graduação em Linguística. Juiz de Fora – MG – Brasil. 36036-330 – xandramoreira@yahoo.com.br

\*\* UFJF - Universidade Federal de Juiz de Fora. Departamento de Linguística. Juiz de Fora – MG – Brasil. 36036-330 – mm.salomao@uol.com.br

da Semântica de Frames para a Linguística é propor que os significados das expressões linguísticas sejam interpretados de acordo com o *frame* de fundo que é evocado no momento do enunciado. Assim, segundo Ohara e Nikiforidou (2009), é rejeitada qualquer dicotomia entre semântica e pragmática enquanto desenvolvem-se os meios para uma análise do significado que abarca parâmetros pragmáticos, sociais e de discurso. De fato, o que é rejeitada é a distinção dicotômica entre conhecimento do mundo e semântica linguística. A distinção entre a semântica e a pragmática é na verdade a distinção entre conceptualização (memória permanente) e uso (processamento na mente). Os resultados das pesquisas relacionadas à Semântica de Frames transcendem as fronteiras da Linguística tendo impacto em outras áreas do conhecimento, tais como o Processamento da Linguagem Natural (PNL), a Psicologia e a Inteligência Artificial. Notadamente na área do PNL, destacam-se as pesquisas de busca de informação com uso de *corpus* para análise de texto para identificação de cenários associados, tradução e inferência textual.

Fillmore (1985) descreve o modelo de Semântica de Frames como um modelo semântico do entendimento em oposição à semântica formal, que é baseada na ocorrência de condições que tornam uma sentença verdadeira. A Semântica de Frames veio como resposta à incapacidade da semântica tradicional de explicar as interpretações de lexemas em diversas situações, tais como explicar porque não é adequado caracterizar o Papa como um homem solteiro (KATZ; FODOR, 1964). Este é o exemplo clássico do lexema do inglês *bachelor*, usado em várias argumentações (FILLMORE, 1977; LAKOFF, 1990; PETRUCK, 1996) sobre a insuficiência da abordagem semântico composicional, em que os conceitos são definidos por propriedades mínimas e necessárias. No entanto, para entender o conceito de “casado” é necessário entender uma série de estruturas conceituais inter-relacionadas, como a instituição do casamento no mundo ocidental, a noção das funções típicas de um homem casado e o que é estar apto a exercer essas funções. Só assim é possível aplicar adequadamente o termo “solteiro” a alguém. Existem vários termos como esse na língua natural, cujo significado só pode ser compreendido pela estrutura como um todo (*gestalt*) e não pela análise de suas partes.

A semântica baseada em *frames* tem se mostrado útil na interpretação de outros estudos do uso da linguagem tais como anáforas e metáforas. Como salientado por Croft e Cruse (2004), a abordagem cognitiva não só revela novos aspectos da linguagem, como também aborda preocupações tradicionais dos gramáticos e semanticistas de uma forma mais satisfatória.

Com base na Semântica de Frames, foi implementada uma base lexical, denominada FrameNet (RUPPENHOFER et al., 2010). A FrameNet (BAKER; FILLMORE; LOWE, 1998) é um recurso lexical on-line para a língua inglesa

implementado com base na Semântica de Frames e apoiado por um *corpus*. Segundo Salomão (2009, p.174), o “[...] projeto FrameNet surge do cruzamento da Semântica de Frames com a Lexicografia através da cooperação entre Fillmore e Atkins no princípio da década de noventa.” O projeto tem por objetivo registrar as possibilidades combinatórias semânticas e sintáticas (valências) de cada lexema (nomes, adjetivos e verbos) em cada um dos seus sentidos. O projeto utiliza sistemas computacionais para auxiliar na anotação das sentenças, para tabular os resultados e para exibir os resultados da anotação.

Segundo os critérios propostos por Brachman (1979), a representação de conceitos e suas relações estão relacionadas ao nível epistemológico. Nas palavras de Brachman, esse nível se refere à “estrutura formal de unidades conceituais e seus relacionamentos como unidades conceituais.” (BRACHMAN, 1979, p.30). Segundo Guarino (1995), nesse nível, assumem-se algumas escolhas segundo bases cognitivas e computacionais, que refletem um conjunto de comprometimentos ontológicos implícitos. É no nível ontológico que esses comprometimentos são explicitados por meio de restrições semânticas, que podem ser operacionalizadas pelo enquadramento em uma ontologia de nível topo (do inglês, *top-level ontology*) ou genéricas. Segundo o mesmo pesquisador, o uso combinado dos vários níveis de conhecimento gera benefícios no entendimento e na construção de sistemas de representação de conhecimento. Nesta pesquisa, é adotada a posição da Linguística Cognitiva que se fundamenta na hipótese da continuidade entre a linguagem e a cognição humana, segundo Salomão (2009, p.171), ou seja, a linguagem humana é indissociável da cognição. No entanto, a visão modular se presta a uma análise no estilo “dividir para conquistar”, de modo a compreender as diversas facetas do significado de um item lexical. Portanto, apenas por uma questão metodológica foi estabelecida essa divisão.

As ontologias desenvolvidas no âmbito da Ciência da Computação podem ser compreendidas como uma representação parcial das estruturas conceituais que descrevem os objetos de um domínio ou as tarefas executadas sobre os objetos dentro de um domínio. As ontologias também são classificadas de acordo com o grau de generalidade das classes que englobam os objetos do domínio. Assim, as classes que são independentes de domínio, denominadas de categorias topo, ocorrem nas chamadas ontologias fundamentais ou de nível topo. Já as classes específicas de determinados domínios ocorrem nas ontologias de domínio.

Trabalhos que unem esses sistemas de representação conceitual podem ser verificados tanto na Linguística como na Ciência da Computação (CHISHMAN et al., 2006; BÄRZDINS et al., 2008; SCHEFFCZYK et al., 2006; CHOW; WEBSTER, 2007; BURCHARDT; PENNACCHIOTTI, 2008; OVCHINNIKOVA et al., 2010). A

maior parte dessas pesquisas visa combinar a base de *frame* por meio de sua correspondente base lexical, a FrameNet (BAKER; FILLMORE; LOWE, 1998), com ontologias no intuito de ampliar a utilização da FrameNet no PLN. No entanto, não foi detectada na literatura a aplicação de ontologias para o aperfeiçoamento da metodologia de desenvolvimento de uma FrameNet. Este artigo apresenta uma alteração no processo de desenvolvimento de uma FrameNet, incluindo uma análise ontológica, de forma a permitir um melhor estabelecimento das relações entre *frames*, o que pode favorecer a inferência textual automática e o próprio uso da rede por pessoas. As evidências linguísticas apresentadas neste trabalho foram obtidas do Corpus do Português (DAVIES; FERREIRA, 2006) de livre acesso do portal da BRIGHAM YOUNG UNIVERSITY (DAVIES, 2008). Esse *corpus* possui em torno de 45 milhões de palavras obtidas de textos que vão desde o século XIV até o século XX. A escolha desse *corpus* deveu-se à grande quantidade de sentenças associadas ao domínio em estudo: o domínio de *viagem*. O *corpus* possui 2375 sentenças com o nominal *viagem* e 1587 sentenças com o verbo *viajar*.

Na próxima seção, comentamos sucintamente a Semântica de Frames. Em seguida, apresentamos uma proposta de alteração da metodologia de desenvolvimento de *frames*. Posteriormente, descrevemos os princípios da Teoria do Léxico Gerativo (PUSTEJOVSKY, 1991), empregados na nova metodologia. Apresentamos, em seguida, definição do lexema *viagem* na FrameNet. Na seção ulterior, analisamos a cena de VIAGEM sob a metodologia proposta. Finalmente, apresentamos as conclusões do trabalho.

## **Semântica de Frames**

O entendimento de um conceito não é obtido isoladamente e sim a partir de seu posicionamento no sistema de conceitos, como pode ser observado nas palavras de Fillmore (2006, p.373): “Com o termo *frame* o que eu tenho em mente é um sistema de conceitos relacionados de tal maneira que para entender qualquer um deles é preciso que se compreenda toda a estrutura na qual ele se encaixa.” É um conhecimento construído pela vivência e experimentação, formando um corpo de conhecimento estereotipado. A ideia de *frame* propõe um modelo alternativo de semântica para os elementos sintáticos em um enunciado: a Semântica de Frames. Na Semântica de Frames, o significado das expressões linguísticas é interpretado de acordo com o *frame* de fundo que representa a cena no momento do enunciado. Por exemplo, o lexema na língua inglesa *risk* (risco) pode significar “assumir um risco” ou “estar em risco”, dependendo do *frame* de fundo aplicado (FILLMORE, 2008):

“(a) *I’m going to risk a swim in the sea. (Daring)*”

“(b) *Buildings in California risk destruction by earthquakes. (Being\_at\_risk)*”

O que define o significado efetivo do lexema é o *frame* evocado pelo lexema e pelos outros elementos linguísticos que ocorrem simultaneamente no mesmo enunciado. Assim, é possível estabelecer que o sentido de *risk* em (a) difere do sentido em (b). Na cena do primeiro enunciado, temos um **protagonista** (*I*), uma **ação** (*swim in the sea*), um **dano** (de se afogar, omitido na sentença) e um **recurso** (a vida, omitido na sentença). Já a cena do segundo enunciado não prevê a ocorrência de uma **ação**.

A Semântica de Frames atribui funções temáticas mais específicas (microtemáticas) do que os papéis temáticos tradicionais (agente, paciente, tema, etc.). Assim, cada *frame* pode possuir um conjunto de papéis temáticos específicos para a situação descrita. Por exemplo, o *frame* COMMERCE\_BUY, possui como papéis temáticos centrais o *Comprador* (*Buyer*) e o *Bem* (*Goods*). Enquanto que, segundo o enfoque anterior, haveria os papéis de **Agente** e de **Tema**.

## A FrameNet

Como já foi mencionado, a Semântica de Frames deu origem a um projeto de desenvolvimento de uma base léxico-semântica, denominada FrameNet (BAKER; FILLMORE; LOWE, 1998). Os conceitos básicos que fundamentam o projeto FrameNet são os conceitos de *frames*, as relações entre *frames*, os elementos de *frames* (EF) e as unidades lexicais (UL).

Uma **unidade léxica** (UL) é o pareamento de uma palavra com um significado (RUPPENHOFER et al., 2010). Ainda segundo o mesmo autor, cada sentido de um lexema polissêmico pertence a um *frame* semântico diferente. Cada significado pertence a uma estrutura de informação, projetada de acordo com a visão de *frames*, que descreve uma determinada situação, um objeto ou um evento e os elementos participantes. Essa estrutura de informação é o *frame*. Atualmente, a FrameNet possui mais de 10.000 unidades léxicas (UL), sendo que 5249 foram finalizadas durante o projeto FrameNet II, 1276 criadas durante o projeto FrameNet I, 4273 foram criadas, mas ainda não foram associadas com evidência de *corpus*. As ULs evocam os *frames*, por exemplo, a ocorrência do lexema *buy* em uma sentença evoca o evento de uma compra comercial, capturado pelo *frame* COMMERCE\_BUY.

Os **Elementos de frame** (EF) correspondem aos papéis microtemáticos mencionados no início desta seção. Segundo Ruppenhofer et al. (2010), os EF são divididos em três grupos: centrais (core), periféricos e extratemáticos. Este

último grupo denota os EF que se acredita não pertencerem conceitualmente aos *frames* em que aparecem.

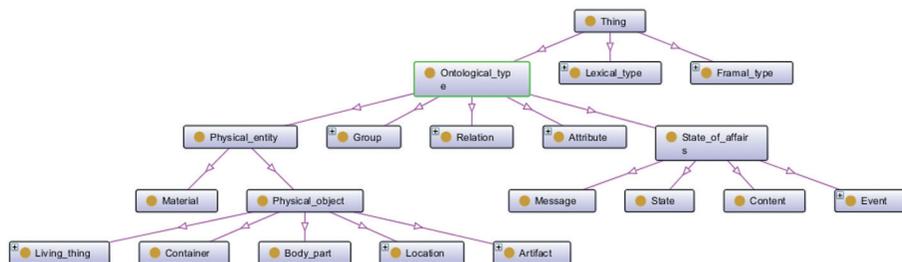
Os *frames* se relacionam por meio de relações semânticas tais como *herança*, *uso*, *subframe* e *perspectiva sobre*. Isso os diferencia de outras bases lexicais, como os tesouros. Os tesouros se baseiam em relações lexicais, tais como sinonímia, antonímia, hiponímia, e meronímia, relações essas que ocorrem apenas entre os lexemas. Já as relações da FrameNet relacionam *frames*, que são estruturas de conhecimento. Portanto, as relações introduzidas na FrameNet tem um status semântico.

Os *frames* criados podem ser posteriormente divididos com bases em alguns critérios, como descrito por Ruppenhofer et al. (2010), tais como diferença de perspectiva, variação na estrutura argumental, alternância causativo-incoativa e distinção ontológica dos EFs. Neste último caso são utilizados tipos semânticos que são descritos a seguir.

## Tipos semânticos

Alguns elementos de *frames* são associados a determinados tipos semânticos. Os tipos semânticos atribuem restrições às instâncias dos elementos de *frames* de modo que todas as instâncias devem ser consideradas instâncias dos tipos semânticos. O objetivo dos tipos semânticos é adicionar informação tipológica aos elementos de *frames*. Por exemplo, o tipo semântico *sentient* (ser sensível) é atribuído para o EF *Perpetrador* no frame de PIRATARIA. A Figura 1 apresenta o topo da hierarquia de tipos semânticos da FrameNet. A apresentação gráfica de todos os elementos tornaria a figura muito complexa.

**Figura 1 – Hierarquia topo de tipos semânticos da FrameNet.**



**Fonte:** Elaboração própria.

A FrameNet não oferece nenhuma formalização para os tipos semânticos e nem discrimina o conjunto completo de tipos que podem ser atribuídos a

todo elemento de *frame*. No entanto, o conjunto dos tipos semânticos pode ser considerado uma ontologia inicial de nível topo. A formalização é a expressão das definições dos elementos de um objeto de estudo em uma linguagem formal, cuja semântica é claramente definida e sem ambiguidade, como é o caso da lógica matemática. A formalização permite uma comunicação não ambígua e a criação de sistemas automáticos de manipulação do objeto de estudo.

## **Métáfora na FrameNet**

Segundo Ruppenhofer et al. (2010), a abordagem da FrameNet para as metáforas está fundamentada no fato bem conhecido de que as metáforas diferem em sua novidade/pré-empacotamento. Com relação à metáfora, a FrameNet faz apenas uma distinção binária entre “produtivos” e “lexicalizados”, apesar de existir uma gradação entre esses dois extremos. Essa distinção é implementada por meio da anotação do domínio fonte de uma metáfora (o *frame* literal) ou do domínio alvo (o *frame* que mais diretamente codifica o que o locutor estava tentando dizer). Ainda segundo Ruppenhofer, a metáfora produtiva é sinalizada pelo rótulo “*metaphor*” no nível de sentença e o domínio de origem é anotado somente se:

- Todos os sinônimos do *frame* corrente e termos relacionados possuem uma alternância correspondente entre os usos literal e metafórico;
- Todos os elementos de *frames* do domínio alvo podem ser mapeados para o domínio fonte; e
- A sentença só pode ser compreendida por evocar, de certa forma conscientemente, o domínio fonte (RUPPENHOFER et al., 2010, p.90).

Os autores admitem que este último critério é vago e, segundo eles, nunca foi usado na prática como um critério de decisão, uma vez que está fortemente correlacionado com os outros critérios.

Nos casos desses critérios terem sido satisfeitos, então uma UL seria anotada no domínio alvo da metáfora e deveria ser criada uma relação *Frame* a *Frame* metafórica entre os domínios, mas, segundo Ruppenhofer et al. (2010, p.91), esses links ainda não foram adicionados à FrameNet. Apenas foi criado, experimentalmente, um *frame* metafórico, denominado “EMOTION\_HEAT”, que representa o *blend* semântico da metáfora EMOÇÃO É CALOR. Portanto, conclui-se que existe uma proposta de tratamento da metáfora na FrameNet, porém essa proposta ainda não foi implementada. Apesar da existência dessa proposta, a mesma não considera a anotação simultânea dos domínios alvo e fonte, uma vez que, segundo Ruppenhofer et al. (2010), isso mereceria, por si só, todo um programa de pesquisa.

## Inferência na FrameNet

A FrameNet fornece uma quantidade significativa de informação útil para a realização de inferências em linguagem natural, tais como as relações entre *frames*, mapeamento de elementos de *frames* e ULs. No entanto, como é reconhecido por Ruppenhofer et al. (2010), a FrameNet não fornece toda a informação que um sistema automatizado de inferências pode requerer. Nesse caso, é necessário recorrer ao auxílio de outras bases de informação, como a WordNet (FELLBAUM, 1998).

## Limitações

A FrameNet possui o mérito de promover o entendimento de um lexema a partir das cenas em que é empregado. Essa base é um rico repositório de informações semânticas com um nível de detalhamento fino (nível fino de granularidade). Porém, a FrameNet possui algumas limitações, que dificultam seu uso em certas aplicações. Uma das limitações da FrameNet é a falta de formalização. A FrameNet é armazenada em uma base textual semiestruturada em HTML (*HyperText Markup Language*), o que dificulta seu uso por sistemas computacionais. Existe uma formalização em OWL DL<sup>1</sup> (SCHEFFCZYK, 2006), mas sem dispor de uma axiomatização, ou seja, foram listados em OWL-DL os tipos de elementos que ocorrem na FrameNet, mas não foram definidas, por meio de axiomas, as restrições e as propriedades associadas a esses tipos. Outra queixa comum a respeito da FrameNet é a sua baixa cobertura lexical, inferior à da *WordNet* (12000 lexemas contra 155000).

Ovchinnikova et al. (2010) também levantaram algumas limitações da FrameNet: baixa cobertura lexical, incompletude das relações, inconsistências na herança de propriedades e falta de axiomatização. Particularmente, com relação à incompletude das relações, os autores alegam que ligações que deveriam existir entre certos *frames* estão ausentes, por exemplo, os autores acreditam que os *frames* SURVIVING e RECOVERY deveriam estar conectados, uma vez que (t) implica em (h)<sup>2</sup>:

(t) *[people]SURVIVOR who [survive]SURVIVING [Sars]DANGEROUS SITUATION ...*

(h) *[Those]PATIENT who [recovered]RECOVERY [from Sars]AFFLICTION ...*

---

<sup>1</sup> Conferir W3C (2009).

<sup>2</sup> Os pares de sentenças t e h são utilizados em estudos de inferência textual, onde t (texto) implica em h (hipótese).

A falta de axiomatização também impede a realização de algumas inferências. Por exemplo, seja o par de sentenças *t* e *h* abaixo:

(*t*) [*Three leaders...*]*ACCUSED* were [*charged*]*NOTIFICATION OF CHARGES* [*with illegally diverting money to their organization*]*CHARGES* ...

(*h*) [*Three leaders ...*]*SUSPECT* are [*suspected*]*SUSPICION* [*of stealing money*]*INCIDENT* .

Nesse exemplo, os autores afirmam que os elementos do *frame* *SUSPICION* podem ser mapeados para os elementos do *frame* *NOTIFICATION OF CHARGES*, segundo a seguinte cadeia lógica: *NOTIFICATION OF CHARGES subframe* do *frame* *ARRAIGNMENT* que é *subframe* do *frame* *CRIMINAL PROCESS* que é precedido pelo *frame* *CRIMINAL INVESTIGATION* que é usado pelo *frame* *SUSPICION*. Contudo, ainda segundo os autores, esse caminho não garante a vinculação. Para que ocorra a inferência desejada, é preciso axiomatizar a relação de precedência e substituir a relação de uso por uma relação de dependência axiomatizada.

## Proposta de alteração

Na construção de uma FrameNet para o português brasileiro, existem diversas estratégias possíveis. Primeiramente, pode-se adotar a metodologia da FrameNet de Berkeley, sem aproveitar a base de *frames* desenvolvida na FrameNet de Berkeley. A segunda estratégia é aproveitar a metodologia e a base de *frames* já desenvolvida, realizando os ajustes necessários em função das evidências de *corpus* distintas. Esta é a estratégia adotada na construção da FrameNet para a língua sueca (BORIN et al., 2009) e também a adotada no projeto FrameNet Brasil (SALOMÃO, 2009).

Propomos aqui uma abordagem alternativa que leva em consideração as relações ontológicas entre os elementos que compõem uma cena. A análise ontológica permite uma melhor análise das relações entre *frames*, o que pode favorecer a inferência textual automática e o uso da rede por pessoas. Isso pode levar a uma rede de *frames* que contorna algumas das limitações apontadas por alguns pesquisadores (OVCHINNIKOVA et al., 2010).

Segundo Ruppenhofer et al. (2010), as etapas típicas do desenvolvimento de um *frame* são as seguintes: extrair de um *corpus* sentenças associadas a um conjunto de lexemas que se supõe que tenham alguma sobreposição semântica. Estas sentenças, que exercem o papel de atestações, são divididas em grupos segundo critérios semiformais, tais como estrutura argumental, tipos semânticos dos argumentos, sentenças causativas *versus* incoativas etc. Uma descrição mais detalhada está disponível em Ruppenhofer et al. (2010). Em seguida, os grupos

são reagrupados em grupos maiores, com tamanho suficiente para desenvolver os *frames*. Esta última fase não é apresentada em detalhes por Ruppenhofer et al. (2010).

As etapas utilizadas pelo grupo que desenvolve a FrameNet do espanhol<sup>3</sup> são mais claramente explicitadas:

1. **identificar os elementos pertencentes ao frame subjacente a cada lexema;**
2. **encontrar as sentenças que instanciam os EFs nas sentenças que devem ser anotadas;**
3. **anotar os EF com nomes apropriados; e**
4. **reunir as sentenças (descritas semântica e sintaticamente) e exibi-las com a descrição construcional do lexema.**

Propomos aqui acrescentar as alterações nas etapas do processo acima. As alterações estão destacadas em negrito.

1. Identificar os elementos pertencentes ao frame subjacente a cada lexema;
2. Encontrar as sentenças que instanciam os EFs nas sentenças que devem ser anotadas;
3. Anotar os EF com nomes apropriados;
4. **Anotar os elementos de *frames* e unidades lexicais, segundo a perspectiva de instrução de uma ontologia;**
5. **Agrupar as sentenças de acordo com a natureza de ontologia;**
6. **Montar a rede de sentidos dos *frames*;**
7. Reunir as sentenças **por frame** (descritas semântica e sintaticamente) e exibi-las com a descrição construcional do lexema.

Na metodologia predominante na elaboração de *frames*, parte-se de um sentido para a obtenção dos itens lexicais associados ao significado. Claro que, nesse processo, alguns sentidos relacionados são detectados e *frames* distintos, porém relacionados, são criados. As evidências da necessidade para o particionamento de um *frame* são mencionadas em Ruppenhofer et al. (2010) e incluem: (a) alteração no número de argumentos, (b) perfilamento diferente dos EFs, (c) relações entre os EFs distintas, e (d) alteração no tipo semântico dos EFs. Este último critério é o embrião de nossa proposta, que aprofunda a análise do tipo semântico para uma análise ontológica mais ampla, uma

---

<sup>3</sup> Confira SPANISH FRAMENET (2011).

vez que os tipos semânticos apresentados na FrameNet não são aplicados a todos os elementos. Isso ocorre porque sua função não é a realização de uma análise ontológica, mas sim, como ressalta Ruppenhofer et al. (2010), registrar as informações que não são representadas na hierarquia de *frames* e elementos de *frames*. Sendo assim, as várias naturezas de um item lexical correlacionadas semanticamente, obtidas por meio de evidência de *corpus*, auxiliam no estabelecimento de uma rede de *frames* associados de forma mais abrangente que a metodologia corrente.

Cabe aqui esclarecer que não estamos propondo a volta da visão clássica em que as categorias são percebidas como conjuntos com fronteiras bem delimitadas e cuja pertinência de um conceito é verificada através da obediência a condições mínimas e necessárias. Porém, em um sentido específico, os elementos de *frames* devem possuir certas propriedades ontológicas para atender às expectativas do sentido.

Para realizar a análise ontológica, é preciso selecionar uma ontologia de nível topo cujas categorias serão usadas para anotar os elementos das sentenças. No entanto, cada ontologia de nível topo expressa uma visão do que existe no mundo, e essa visão pode não contemplar todos os aspectos necessários para interpretação do mundo. De fato, uma ontologia que contenha todos os conceitos e todas as relações entre os conceitos de nível topo de modo a explicar os sentidos dos itens lexicais é impossível. Por exemplo, uma ontologia pode estabelecer uma ligação entre o tempo e o espaço (ambas são dimensões), unindo eventos e objetos físicos, como é o caso da ontologia proposta por Russell e Norvig (2010). Por outro lado, ontologias como a DOLCE (descrita mais adiante) podem colocar eventos e objetos físicos em hierarquias distintas. Desse modo, a ontologia de Russell e Norvig ajudaria a entender porque as pessoas utilizam verbos associados a deslocamentos no espaço para mencionar deslocamentos no tempo, enquanto que a ontologia DOLCE não. Já as associações estabelecidas na ontologia DOLCE podem ajudar a explicar outros fenômenos linguísticos. Por essa razão, foram utilizadas, neste trabalho, três ontologias para fornecer a base mais ampla para as análises ontológicas.

Dentre as ontologias utilizadas temos a *Generalized Upper Model* (GUM) (FARRAR; BATEMAN, 2005; BATEMAN, 2008; BATEMAN et al., 2010), que é uma ontologia baseada em conceitos linguísticos, independente de domínio e tarefa, e que foi desenvolvida com o objetivo de organizar a informação para a expressão em linguagem natural. Por conceitos linguísticos, queremos dizer de conceitos que foram observados em enunciados de linguagem natural e são designados lexicalmente. A escolha da ontologia deve-se justamente à sua aderência ao enfoque linguístico. A ontologia GUM está disponível na Web no formato OWL (W3C, 2009).

A segunda ontologia que será utilizada é a ontologia DOLCE (*Descriptive Ontology for Linguistic and Cognitive Engineering*) (GAIO et al., 2010). A ontologia DOLCE também possui inspiração linguística, mas opera em um nível de abstração maior do que a ontologia GUM. A ontologia DOLCE possui várias versões e a versão que utilizaremos é a DOLCE+DnS Ultralite (DUL), também chamada de DOLCE Spray (OLTRAMARI et al., 2010). Os termos utilizados na DUL são mais intuitivos, o que a torna mais fácil de ser aplicada.

A terceira ontologia é a ontologia SIMPLE-CLIPS<sup>4</sup> (*Semantic Information for Multifunctional Plurilingual Lexica-Corpora e Lessici dell'Italiano Parlato e Scritto*) (LENCI et al., 2000; RUIIMY; TORAL, 2008). A ontologia SIMPLE-CLIPS (doravante SIMPLE) consiste de 157 tipos semânticos organizados por meio de relações conceituais hierárquicas e não hierárquicas. A base teórica para o desenvolvimento da SIMPLE é a Teoria do Léxico Gerativo (PUSTEJOVSKY, 1995), que será abordada na próxima seção. No momento, é suficiente mencionar que sua adoção se deve ao fato de permitir adicionar dimensões de análise a um item lexical. Na ontologia SIMPLE, as unidades lexicais se referem a diferentes tipos de entidades (entidades de primeira ordem, entidades de segunda ordem e entidades de terceira ordem), de acordo com o preconizado por Lyons (1977). Segundo Paradis (2005), de um modo geral, as entidades de primeira ordem são objetos físicos tais como: animais, pessoas, plantas, veículos. Essas entidades são relativamente estáveis do ponto de vista perceptivo. Elas existem no espaço tridimensional, em qualquer ponto do tempo, e são publicamente observáveis. O status ontológico das entidades de segunda e de terceira ordem é vago uma vez que não está associado ao mesmo número de propriedades estáveis como as entidades de primeira ordem. As entidades de segunda e de terceira ordem são mais variáveis e, portanto, também mais difícil de definir e conseqüentemente, mais controversa. Entidades de segunda ordem são acontecimentos, processos e estados, tais como “vitória”, “discussão” e “felicidade”, respectivamente. Essas entidades estão localizadas no tempo e a tendência é dizer que elas ocorrem em oposição a dizer que elas existem. Finalmente, as entidades de terceira ordem são entidades abstratas que estão fora do espaço e do tempo. São entidades como “conceitos”, “ideias”, “possibilidades” e “proposições”, “dia” e “ano”, referidas por Schmid (2000) como SHELLS (concha). A denominação de SHELLS ocorre, segundo Paradis (2005), em função das entidades de terceira ordem não possuírem um conjunto estável de propriedades, sendo que as propriedades são incorporadas à medida que se enquadram em determinada noção abstrata ou pelas necessidades pragmáticas das pessoas. Um exemplo de SHELL apresentado por Schmid (2000, p.3) seria o lexema *problem* (problema) na sentença “*the (big) problem was that I had no money*”, que serve de SHELL para “*I had no money*”.

---

<sup>4</sup> Disponível em: <<http://www.ilc.cnr.it/clips/Ontology.htm>>. Acesso em: 10 mar. 2011.

## Teoria do léxico gerativo

A anotação de um item lexical sob o ponto de vista ontológico não é uma tarefa trivial. O que se busca é anotar algum aspecto de sua denotação, que pode variar a partir do contexto em que está inserido o item. Por exemplo, nas sentenças a seguir, o lexema **viagem** pode assumir diferentes naturezas.

- a) A **viagem** demorou cerca de duas horas e meia - e a comitiva foi recebida com uma chuva miudinha e oito graus de temperatura. (19N:Pt:Jornal). (**processo**)
- b) A viagem custou uns R\$ 15,00/R\$ 20,00.  
(site [www.mochileiros.com/rio-de-janeiro-em-um-fds-t43496.html](http://www.mochileiros.com/rio-de-janeiro-em-um-fds-t43496.html)). (**evento**)

Na sentença (a), **viagem** denota um processo com etapas que podem ser mencionadas e descritas. Já na sentença (b), **viagem** é vista como um evento atômico, sem etapas. Da mesma forma, um objeto, como um carro, pode ser visto sob diversas perspectivas: como um objeto concreto; como um artefato; um objeto estruturado; ou como um objeto definido por sua função (veículo de transporte). A relação de herança e outras relações ontológicas (parte-todo, por exemplo) podem oferecer uma saída para descrever os vários aspectos da denotação de um item lexical, mas o ideal seria utilizar uma estrutura que fundamentasse a ocorrência desses vários aspectos e sua seleção em um enunciado. Uma possibilidade é o uso da Teoria do Léxico Gerativo de Pustejovsky (1995), doravante, (TLG).

Na TLG, os sentidos de um item lexical não são enumerados e tratados separadamente. Os sentidos são tratados como aspectos de um mesmo todo estruturado, que são destacados a partir da interação com determinados aspectos dos elementos que fazem parte de sua estrutura argumental.

Observa-se uma distinção marcante entre as duas abordagens. Enquanto a Semântica de Frames trata cada tipo semântico como uma entrada separada para cada unidade lexical, a TLG propõe um único tipo complexo (*dotted type*) para o elemento cujos aspectos são selecionados por operações gerativas. Em outras palavras, a Semântica de Frames lida com unidades léxicas e TLG lida com lexemas.

Assim, segundo Chishman (2003), a TLG propõe uma forma distinta de composicionalidade, em que a semântica do predicado é definida em conjunto com seus argumentos. Um exemplo típico apresentado por Pustejovsky (1995, p.91) trata sobre os sentidos do lexema *livro*:

- a) Pensei: bom, vou correndo até em casa pegar um **livro** dele e trazer para ele assinar. (19Or:Br:Intrv:ISP) (objeto físico)

b) Também escrevi um **livro** sobre anjos. (19Or:Br:Intrv:ISP) (informação)

Segundo Pustejovsky, a habilidade de um item lexical agrupar diversos sentidos é o que ele e Anick (1988) chamam de **Paradigma Lexical Conceitual** (lcp), que preconiza que existe algo inerente à semântica de um nome que o capacita a projetar qualquer um dos sentidos designados pelo nome de acordo com o ambiente sintático-semântico onde ocorre.

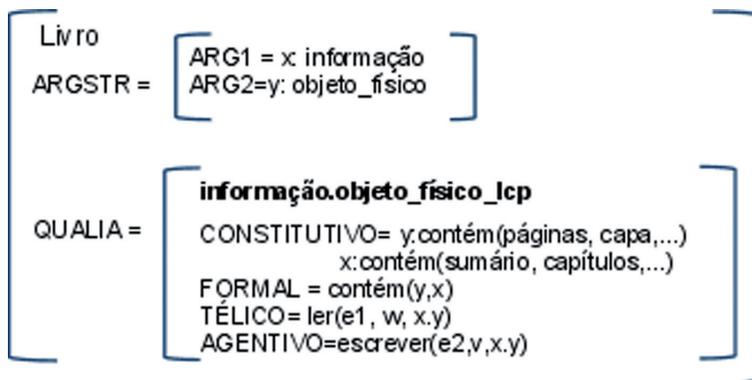
A TLG estrutura a informação de um item lexical em quatro níveis de representação: estrutura argumental, estrutura de eventos, estrutura *qualia* e estrutura de herança lexical. A estrutura argumental especifica o número, o tipo lógico dos argumentos e como eles ocorrem sintaticamente. A estrutura de eventos define o tipo de evento de um item lexical ou sentença. A estrutura *qualia* descreve a natureza da denotação através de seus atributos fundamentais organizados nas dimensões **formal, constitutivo, télico e agentivo**. Finalmente, a estrutura de herança lexical captura como a estrutura lexical está relacionada a outras estruturas em um reticulado de tipos.

A estrutura *qualia* é o nível mais proeminente da TLG e o mais utilizado no PLN (ZAVAGLIA; GREGHI, 2003; GONZALEZ; LIMA, 2004; PARADIS, 2005; CIMIANO; WENDEROTH, 2005; MENDES; CHAVES, 2001; AMARO et al., 2010). Para a proposição da estrutura *qualia*, Pustejovsky inspirou-se na interpretação de Moravcsik (1973 apud PUSTEJOVSKY, 1995, p.76) dos modos de explicação de Aristóteles. As definições dos componentes da estrutura *qualia* são as seguintes:

- **Formal** - é aquilo que o distingue em um domínio mais amplo. Trata-se do aspecto de uma entidade como um tipo.
- **Constitutivo** - trata da relação de uma entidade e suas partes constituintes.
- **Agentivo** - trata dos fatores envolvidos em sua origem.
- **Télico** - seu propósito e função.

Para ilustrar como é feita a descrição de um item lexical nesses níveis de representação, é mostrado a seguir (Figura 2) a descrição da palavra *livro*.

**Figura 2 – Representação do lexema *livro* na TLG.**

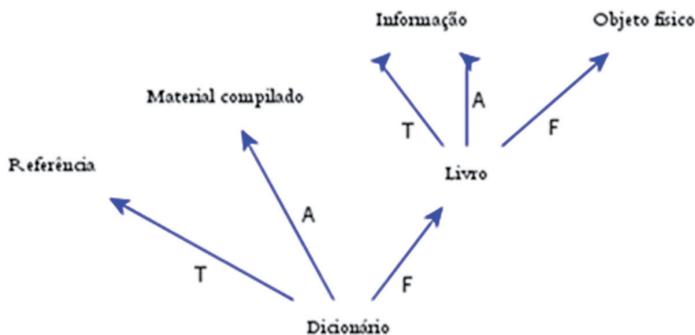


**Fonte:** Elaboração própria adaptado de Pustejovsky (1995, p.101).

A Figura 2 mostra as estruturas *qualia* e argumental do nome *Livro*. A estrutura argumental define quais os tipos que o objeto pode assumir: um objeto físico ou uma informação. Sendo considerado um tipo complexo, esses sentidos não são tratados separadamente e o *qualia* FORMAL define como um argumento se relaciona com o outro (o objeto físico contém informação). A etiqueta “**informação.objeto\_físico\_lcp**” indica tratar-se da representação de um tipo complexo (lcp), cuja referência da instância complexa é representada por “x.y”. O *qualia* CONSTITUTIVO descreve as partes que formam os sentidos do lexema. O *qualia* TÉLICO registra, na forma de um predicado de evento<sup>5</sup>, que a função do livro (ser lido em um evento e1). Finalmente, o *qualia* GENTIVO registra que sua criação envolveu um evento de escrever. A Figura 3 mostra o relacionamento desta estrutura lexical com outras estruturas. As etiquetas T, A e F junto aos relacionamentos indicam relacionamentos tético, agentivo e formal, respectivamente.

<sup>5</sup> Predicado de evento é usado no Cálculo de eventos, que é um formalismo para raciocinar sobre situações que mudam com o tempo. Neste tipo de predicado, o primeiro argumento denota o evento.

**Figura 3 – Rede de herança para “livro” na TLG.**



**Fonte:** Elaboração própria adaptado de Pustejovsky (1995, p.145).

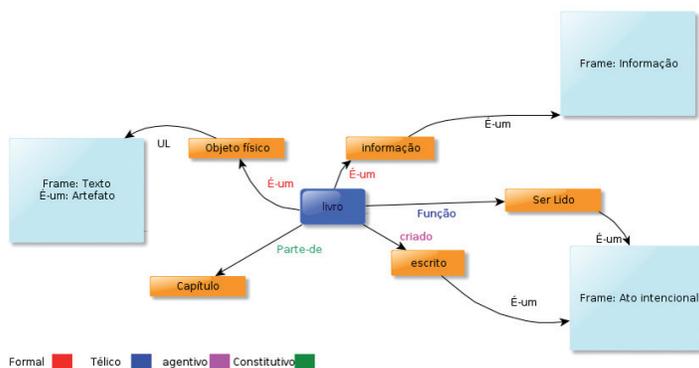
A possibilidade de agrupar os diversos sentidos de um lexema em uma estrutura formal que pode ser manipulada, posteriormente, de forma automática, é bastante atrativa e tem sido aplicada na construção de léxicos, de ontologias e na interpretação de textos. A estrutura *qualia* foi utilizada no desenvolvimento da EuroWordNet (VOSSSEN, 1998), da WordNet para o português europeu (MENDES; CHAVES, 2001), da WordNet para o Dinamarquês (PEDERSEN et al., 2006) e no projeto SIMPLE (LENCI et al., 2000). Paradis (2005) aplicou a estrutura *qualia* em um *framework* para interpretação da semântica de itens lexicais. O *framework* proposto por Paradis analisou o significado lexical em termos de ontologias e interpretações. Segundo ela, é por meio das operações de interpretações sobre o material ontológico que o significado das expressões lexicais surge. Kasama et al. (2010) utilizou a estrutura *qualia* na representação ontológica do domínio da nanociência e nanotecnologia. Chishman (2009) aplicou a estrutura *qualia* para integrar léxicos semânticos e ontologias.

Nos trabalhos citados, a estrutura *qualia* foi utilizada para organizar os conceitos e para estabelecer ligações semânticas. No trabalho de Kasama et al. (2010), a dimensão *formal* é definida pelo enquadramento do conceito em uma classe da ontologia de domínio. As dimensões *constitutivo* e *télico* são caracterizadas por meio de relações semânticas entre os conceitos (*compõe, é-composto-por, utiliza, utilizado-em*).

A abordagem da TLG pode ser unida à Semântica de Frames para obtermos uma forma mais vantajosa de analisar a semântica de itens lexicais. Segundo Ruppenhofer et al. (2010), em vários *frames* que possuem nominais que denotam artefatos, alguns elementos de *frames* foram definidos, refletindo um tipo de estrutura *qualia*, por exemplo, modificadores que denotam o material do artefato (*qualia* constitutivo). Ruppenhofer et al. (2010) também mencionam que a divisão

de *frames* em função de inconsistências no tipo semântico facilita a busca de dados para estudar as coerções Pustejovskiana. A ideia de que o significado de um item lexical é função do relacionamento dos demais itens lexicais da sentença com determinadas dimensões da estrutura *qualia* do item pode permitir o estabelecimento de novos relacionamentos na estrutura de *frames* (uma das deficiências apontadas na FrameNet). Assim, a TLG fornece uma base teórica para a extensão dos relacionamentos da FrameNet. Se esses relacionamentos são suficientes para fornecer uma explicação completa para a polissemia de um item lexical é algo que merece uma análise mais detalhada. A Figura 4 mostra o mapeamento de um item lexical, segundo a TLG, com os *frames* relacionados.

**Figura 4 – Mapeamento do lexema *livro*, segundo a TLG, com os *frames* relacionados.**



**Fonte:** Elaboração própria.

Na Figura 4, podemos observar que o lexema *livro* possui diversos aspectos que são explicitados pela estrutura *qualia*. Esses aspectos são perfilados em cenas distintas, descritas pelos frames. Dessa forma, podemos ter um entendimento claro de como a TLG pode se relacionar com os *frames*.

Neste trabalho, utilizaremos a estrutura *qualia* para atribuir as dimensões da ontologia a um conceito, de forma a melhor caracterizar um conceito e revelar sua natureza. A ontologia utilizada será a ontologia GUM, mencionada anteriormente, e complementada com as ontologias SIMPLE<sup>6</sup> e DOLCE. Essas ontologias foram selecionadas por sua inspiração linguística e por possuírem um nível de abstração mais próximo dos conceitos que serão analisados. Para exemplificar a proposta em questão, iremos apresentar um caso de estudo envolvendo a cena de *viagem*.

<sup>6</sup> Disponível em: <<http://www.ilc.cnr.it/clips/Ontology.htm>>. Acesso em: 10 mar. 2011.

## A cena VIAGEM na FrameNet de Berkeley

Na FrameNet de Berkeley, a cena de VIAGEM é representada pelo *frame* TRAVEL. A definição apresentada pela FrameNet para esse *frame* é a seguinte:

*In this frame a Traveler goes on a journey, an activity, generally planned in advance, in which the Traveler moves from a Source location to a Goal along a Path or within an Area. The journey can be accompanied by Co\_participants and Baggage. The Duration or Distance of the journey, both generally long, may also be described as may be the Mode\_of\_transportation. Words in this frame emphasize the whole process of getting from one place to another, rather than profiling merely the beginning or the end of the journey.<sup>7</sup>*

A partir da definição do *frame* TRAVEL, é possível notar que se trata de um deslocamento geralmente longo tanto em distância quanto em duração. Deslocamentos curtos não podem ser considerados como uma viagem. Essa é uma distinção ontológica fundamental do *frame* em relação ao *frame* de SELF-MOTION. Mas como alguém caracteriza um deslocamento como uma viagem? Parece que o deslocamento espacial é essencial. Apesar das Unidades Léxicas poderem ser usadas em deslocamentos puramente temporais, como viagem no tempo, este é um uso metafórico (TEMPO COMO ESPAÇO) (YU, 1998), como exemplificado em (1) e (2). Note que o lexema *time* é incluído no termo para enfatizar que o deslocamento é no tempo. As sentenças em inglês usadas nos exemplos a seguir foram extraídas do *The Corpus of Contemporary American English* (DAVIES, 2008). A identificação da fonte junto a cada sentença foi extraída do *corpus*.

(1) “*Many fine writers are masters of this kind of **time travel**. Slowing down time, recapturing the past, reframing the past [...]*” (WILLIS, M. S. Improve your fiction with film techniques. *Write*, v.123, n.4, p.26-28, Apr. 2010. 1 color photograph).

(2) “*Mendelsohn then leads Sigge through a series of corridors and stairways, a trek that appears to be a **journey in time** as well, since it eventually takes them to the palace of Saladin, the twelfth-century Sultan of Jerusalem*” (WRIGHT, R. Textual dialogue and the humanistic tradition. *Scandinavian Studies*, v.72, n.3, p.279, 2000).

Em outros casos, a viagem pelo tempo está expressa metaforicamente como uma viagem na dimensão temporal (3).

(3) “*A **Journey** through the Layers of Time.*” (MUSEUMS. Washington Post, p.WE 39, 2010).

---

<sup>7</sup> FrameNet. Disponível em: <<http://framenet.icsi.berkeley.edu>>. Acesso em: 10 mar. 2011.

Outra característica é que o deslocamento é geralmente longo, ou seja, deve ser maior que os deslocamentos diários de uma pessoa. A fronteira entre um deslocamento curto e um longo tende a ser difusa e relativizada ao contexto da cena. As sentenças (4) e (5) mostram enunciados típicos de cenas de viagens.

(4) “*The Saudi jihadists presumably choose to **travel** to Iraq through Syria because Assad tolerates what the Saudi leadership will not.*” (RUBIN, M. Syria’s Path to Islamist terror. *Middle East Quarterly*, v.17, n.1, p.27-37, 2010).

(5) “*He was then selected to **travel** to the USA to play at an Adidas camp.*” (GARCIA, M. Cal’s Zhang seeks personal growth; Berkeley opens up new world to 7-3 center from China. *Sports*, 2010, p.1C).

Isso reforça que é necessário o estudo ontológico, não só dos elementos de *frames* e unidades lexicais, como também do próprio *frame*, sendo que parte dessa análise ontológica é dada pelo posicionamento do *frame* na rede de *frames* (Figura 5). Na análise ontológica do *frame* TRAVEL, a qualidade *Distance*, associada ao *Path*, possui papel determinante na detecção do *frame*, apesar de não ser, na maioria das vezes, lexicalizado.

O lexema *travel* pode ocorrer em outros *frames*, mas com outro sentido, sendo, portanto, outra unidade lexical. Por exemplo, em (6) a unidade léxica *travel* é utilizada para evocar o sentido de um deslocamento simples, não evocando, desta forma, o sentido de viagem.

(6) “*Neurons in my retina convert these light waves into electrical impulses that **travel** farther back into my brain.*” (LONG, P. My brain on my mind: a thrumming, plastic mystery allows us to think, feel, and remember. *American Scholar*, v.79, n.1, p.20-37, 2010. 2 color photographs).

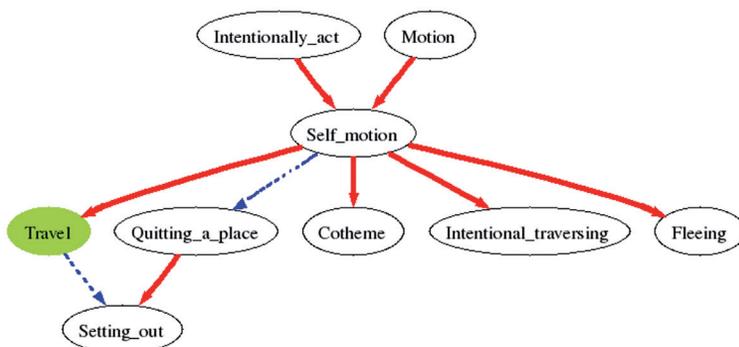
Mostraremos agora uma situação associada à de herança entre *frames*. Neste tipo de relação, o *frame* filho é uma especialização do *frame* pai e herda os elementos do *frame* pai, sendo que tudo que é estritamente verdadeiro no *frame* pai, também deve ser verdadeiro no *frame* filho (RUPPENHOFER et al., 2010), se comportando, desse modo, de forma monotônica em relação ao acréscimos de propriedades. Por exemplo, o *frame* TRAVEL herda as propriedades do *frame* SELF-MOTION (movimento próprio), que, por sua vez, herda as propriedades do *frame* MOTION (movimento) e do *frame* INTENTIONALLY\_ACT (ato intencional), sendo este último, um *frame* abstrato, ou seja, sem realização lexical. Essas relações podem ser observadas graficamente com o uso da ferramenta gráfica da FrameNet, o FrameGrapher<sup>8</sup> (Figura 5). A hierarquia de *frames* é muito

---

<sup>8</sup> Disponível em: <<https://framenet.icsi.berkeley.edu/fndrupal/FrameGrapher>>. Acesso em: 10 mar. 2011.

importante, pois revela muito da natureza do conceito de *viagem*, ou seja, é um movimento da própria entidade, feito de forma consciente. Aparentemente, existe uma inconsistência na herança entre o *frame* SELF-MOTION e o *frame* TRAVEL. O *frame* SELF-MOTION é definido como um movimento com meios próprios de locomoção, mas, em vários exemplos de viagem, o movimento é realizado por meio de veículos: *Barney used to travel by bus a lot*. Por isso, em função da semântica da relação de herança que estabelece que tudo que é verdade para o *frame* pai deve ser verdade para o *frame* filho, não se pode dizer que a *Viagem* é um movimento próprio.

**Figura 5 – Grafo com as principais relações do *frame* TRAVEL.**



**Fonte:** (RUPPENHOFER et al, 2010).

Uma *viagem*, pela definição da FrameNet, é uma atividade e isso é próximo ao enquadramento ontológico realizado na ontologia DOLCE. O enquadramento ontológico do conceito dentro da ontologia DOLCE revela que VIAGEM é um *accomplishment*<sup>9</sup> (**realização completa**), apesar de o termo denotar um *frame* distinto (porém, relacionado através do uso do *frame* INTENTIONALLY\_ACT) na FrameNet.

## A cena VIAGEM no português

Para desenvolver o *frame*, é preciso buscar atestação de *corpus* para os lexemas que se supõem que estão ligadas à cena. A cena típica que se deseja descrever é a viagem realizada por uma pessoa ou grupo de pessoas a determinada localidade, por meios próprios ou por um meio de transporte e com algum propósito. Lexemas candidatos a unidades léxicas nessa cena são: (a) nominais, por exemplo, *viagem, excursão, safari, peregrinação, expedição*

<sup>9</sup> Um *accomplishment* ocorre quando um agente consegue atingir seu objetivo.

e (b) verbais, por exemplo, *viajar, excursionar, peregrinar, visitar*. No entanto, em função das limitações de espaço, restringiremos nossa análise aos lexemas *viagem* e *viajar*. Analisaremos, então, os enunciados em que ocorrem esses elementos sob o ponto de vista ontológico. Seguem alguns exemplos de enunciados com esses lexemas. No *Corpus* do Português (DAVIES; FERREIRA, 2006), existem 3940 ocorrências do lexema *viagem*, sendo que, em alguns casos, ocorre em um sentido não correlacionado, como é o caso de ocorrências em nomes próprios, como em “Festa de Nossa Senhora de Boa Viagem”. Já o verbo *viajar* ocorre 1864 no mesmo *corpus*, sendo que muitas vezes de forma metafórica.

O *corpus* analisado é um *subcorpus* do *Corpus* do Português, contendo 2294 sentenças do século XX do português brasileiro e europeu. As sentenças selecionadas contêm o nome *viagem* e a seguintes flexões do verbo *viajar*: *viajar, viajou, viaja, viajado, viajava, viajam, viajavam, viajado, viajei, viajaram* e *viajo*. Como mencionado anteriormente, o objetivo é capturar os vários sentidos desses lexemas a partir de uma análise ontológica do contexto em que ocorrem. O nominal *viagem* e o verbo *viajar* ocorreram basicamente como *NonAffectingDirectedMotion* (movimento direcionado não afetante). A definição desse conceito, segundo a ontologia GUM, é *Movimento que inclui tanto uma fonte, o destino, ou indicação do caminho ou inclui uma direção*. Exemplo da ocorrência deste sentido pode ser visto a seguir (Figura 6).

**Figura 6 - Exemplo de anotação ontológica<sup>10</sup> de sentença.**

<i>Camada</i>	Na sua <b>viagem</b> para Maputo, o avião militar português transportou 27 militares moçambicanos. (19N:Pt:Público)								
<i>papel</i>	<table style="width: 100%; border: none;"> <tr> <td style="border: none; text-align: center;"> ----- </td> </tr> <tr> <td style="border: none; text-align: center;">UL</td> <td style="border: none; text-align: center;">destino</td> <td style="border: none; text-align: center;">Modo_de_transporte</td> <td style="border: none; text-align: center;">transportado</td> </tr> </table>	-----	-----	-----	-----	UL	destino	Modo_de_transporte	transportado
-----	-----	-----	-----						
UL	destino	Modo_de_transporte	transportado						
<i>ontológico</i>									
<i>DUL</i>	região_espacial								
<i>GUM</i>	mov_não_afetante								
<i>SIMPLE</i>	[transporte(telec)]      [grupo(constitutivo)]								

**Fonte:** Elaboração própria.

Note que, na sentença, o viajante não é um agente animado. Em função disso, deve-se questionar se essa sentença evoca o *frame* VIAGEM. Como explica Ruppenhofer et al. (2010, p.46), no caso do tipo semântico do EF ser distinto do sugerido no *frame*, deve-se verificar se o EF não ocorre metonimicamente, substituindo o EL que possui o tipo semântico adequado. Caso contrário, é

<sup>10</sup> Anotação ontológica é a tarefa de indicar qual o tipo ontológico que iremos atribuir a um determinado lexema.





**Figura 10 – Exemplo de anotação ontológica de sentença.**

<i>Camada</i>	As duas sondas percorrerão uma trajetória particularmente complicada em sua <b>viagem</b> de quase sete anos. (19N:Br:Recf)						
<i>papel</i>	<table style="width: 100%; border: none;"> <tr> <td style="border: none; width: 33%; text-align: center;"> _____  viajante</td> <td style="border: none; width: 33%; text-align: center;"> _____  caminho</td> <td style="border: none; width: 33%; text-align: center;"> _____  UL</td> </tr> <tr> <td style="border: none; text-align: center;"> _____  duração</td> <td style="border: none;"></td> <td style="border: none;"></td> </tr> </table>	_____  viajante	_____  caminho	_____  UL	_____  duração		
_____  viajante	_____  caminho	_____  UL					
_____  duração							
<i>ontológico</i>							
<i>DUL</i>	região_espacial						
<i>GUM</i>	mov_não afetante						
<i>SIMPLE</i>	artefato(agentivo) região_temp						

**Fonte:** Elaboração própria.

Isso pode ser explicado pelo fato de que qualquer entidade que apresente um movimento, próprio ou causado, pode ser encarada como um viajante, desde que o deslocamento seja relativamente longo. No entanto, isso não é um sentido prototípico do lexema. Ainda mais distante do uso prototípico é o emprego de itens lexicais associados à cena de VIAGEM para designar deslocamento de entidades abstratas, como exemplificado a seguir.

**Figura 11 – Exemplo de anotação ontológica de sentença.**

<i>Camada</i>	A imaginação de Verne <b>viajou</b> por todos os quadrantes do planeta e até para bem longe dele. (19N:Br:SP)			
<i>papel</i>	<table style="width: 100%; border: none;"> <tr> <td style="border: none; width: 33%; text-align: center;"> _____  viajante</td> <td style="border: none; width: 33%; text-align: center;"> _____  UL</td> <td style="border: none; width: 33%; text-align: center;"> _____  caminho</td> </tr> </table>	_____  viajante	_____  UL	_____  caminho
_____  viajante	_____  UL	_____  caminho		
<i>ontológico</i>				
<i>DUL</i>	região_espacial			
<i>GUM</i>	mov_não afetante			
<i>SIMPLE</i>	fato_cognitivo			

**Fonte:** Elaboração própria.

Não somente o viajante pode ser abstrato, como também o plano de deslocamento, como mostra o exemplo a seguir.

**Figura 12 – Exemplo de anotação ontológica de sentença.**

<i>Camada</i>	Agora, o velho negro está a <b>viajar</b> pelo seu passado, dentro da toada que vem de longe. (19:Fic:Pt:Sorromenho:Terra)
<i>papel</i>	<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;"> <span style="border: 1px solid black; padding: 2px 10px;"> </span>  <i>viajante</i> </div> <div style="text-align: center;"> <span style="border: 1px solid black; padding: 2px 10px;"> </span>  <i>UL</i> </div> <div style="text-align: center;"> <span style="border: 1px solid black; padding: 2px 10px;"> </span>  <i>caminho</i> </div> </div>
<i>Ontológico</i>	
<i>DUL</i>	região_espacial
<i>GUM</i>	mov_não afetante
<i>SIMPLE</i>	humano

**Fonte:** Elaboração própria.

Metáforas complexas, como as mostrada na Figura 12 são estruturadas a partir de metáforas básicas, arraigadas no pensamento e na linguagem, como TEMPO É ESPAÇO. Podem ser detectadas por meio de variação nos tipos ontológicos, como no caso do agente no sentido do frame VIAGEM (consciente) e no *frame* TRANSPORTE (sem frame correspondente na FrameNet). Representar essa ligação é fundamental para a compreensão dos processos cognitivos e para a aplicação mais precisa da FrameNet em sistemas de PLN, permitindo uma melhor compreensão de sentenças metafóricas e polissêmicas.

## Conclusões

A interpretação de um item lexical é uma tarefa complexa e que envolve vários aspectos. Paradis (2005) sintetiza o que preconiza a Linguística Cognitiva ao mencionar que o significado de um item lexical é delimitado pelo conhecimento enciclopédico, mapeamentos convencionalizados entre itens lexicais e conceitos, modos convencionais de pensamento em contextos distintos e enquadramentos situacionais. Portanto, o significado do item lexical não é inerente ao item e sim evocado por ele, como verificado pelos pesquisadores pioneiros da Linguística Cognitiva: Fillmore, Fauconnier, Turner e outros. Além disso, suas interpretações são múltiplas e dinâmicas. Sendo assim, não é possível construir um léxico semântico para uma língua sem adotar mais de um tipo de análise. A análise ontológica é um desses aspectos, sendo que uma das mais fortes evidências disso são as pesquisas realizadas por pesquisadores do Instituto que abriga a FrameNet, o *International Computer Science Institute*, tais como Scheffczyk e Lönneker-Rodman. A associação de tipos semânticos aos elementos da FrameNet é, de certa forma, uma análise ontológica, mas, como ressalta Scheffczyk et al. (2006), os tipos semânticos

são limitados, não formalizados e não foram atribuídos a todos os elementos de *frames*.

A teoria do Léxico Gerativo proposta por Pustejovsky (1995) caminha no sentido de permitir a construção de base lexicais com multicamadas e multifacetadas, como é o caso do projeto SIMPLE-CLIPS. Mas, os projetos que adotam essa abordagem deixam de lado um componente importante do significado lexical, que é a cena onde o item ocorre, o que é capturado pela Semântica de Frames de Fillmore. A junção de mais esse aspecto à semântica lexical permitirá a construção de léxicos mais completos.

A proposta deste trabalho foi mostrar que a análise ontológica pode auxiliar a verificar se o sentido de uma determinada sentença se enquadra em determinado *frame*. Além disso, ela pode ajudar a explicar a ligação entre diferentes *frames* com sentidos próximos, por meio da detecção de variação na propriedade ontológica dos elementos de *frame*. A proposta inclui uma alteração na metodologia de desenvolvimento de *frames*. Baseamo-nos, também, na ideia de que nenhuma ontologia é capaz de capturar todas as naturezas dos objetos expressos pela linguagem, por isso, foram usadas mais de uma ontologia. Para exemplificar a proposta, a metodologia foi aplicada a itens lexicais associadas ao *frame* de VIAGEM para o português. A análise mostrou que o lexema *viagem* é usado em português no sentido de uma cena de VIAGEM prototípica (deslocamento de pessoa), no sentido metafórico (deslocamento numa dimensão abstrata) e no sentido de transporte (sem *frame* correspondente na FrameNet) e, por isso, não evocaria a cena de VIAGEM. Esse sentido pode ser evidenciado pela distinção do tipo ontológico do agente que, nesse caso, exerce o papel de transportador. A ligação entre os dois *frames* (VIAGEM e TRANSPORTE) por meio de um link de distinção ontológica do agente (consciente versus artefato) pode ajudar na compreensão dos fenômenos relacionados à polissemia. A variação do tipo ontológico também sinaliza a possibilidade do lexema estar sendo utilizado em um sentido metafórico, evocando, dessa forma, uma cena distinta da evocada quando usado em seu sentido prototípico.

Estudos mais abrangentes devem ser realizados para comprovar os benefícios da abordagem metodológica.

## **Agradecimentos**

Agradecemos a CAPES e a FAPEMIG pelo financiamento desta pesquisa.

MOREIRA, A.; SALOMÃO, M. M. M. Ontological analysis applied to frame development. *Alfa*, São Paulo, v.56, n.2, p.491-521, 2012.

- **ABSTRACT:** *Frame semantics is a theory that aims to establish the semantics of a lexical item by the evocation of the associated conceptual scene related to it. FrameNet is a lexical-semantic database developed according to the frame semantics frames and is constituted of a set of frames related by semantics links. The methodology for the discovery and incorporation of frames is based on semi-formal criteria and incorporates some ontological criteria. This article seeks to deepen this ontological analysis in order allow greater formalization of the process. We analyzed sentences from the "Corpus do Português" corpus containing the lexeme "viagem" and aligned the frame elements types with top ontologies. We detected the prototypical sense presented in FrameNet. Moreover, we identified the metaphorical use of the lexeme and its use in the transportation sense based on the distinction of the ontological type of agent. Thus, from the ontological analysis was possible to find a new frame yet to be included in the FrameNet database. Such evidence suggests possible benefits of the proposed methodology and may facilitate, in the future, a formal link in distinct frames motivated by the lexeme polysemy. We believe that this may establish better relations among frames, which can facilitate the automatic textual inference and the actual network usage by people.*
- **KEYWORDS:** *Frame semantics. Frames. Ontology..*

## REFERÊNCIAS

AMARO, R. et al. Lexical-conceptual relations as qualia role encoders. In: INTERNATIONAL CONFERENCE ON TEXT, SPEECH AND DIALOGUE (TSD'10), 13, 2010, Berlin. *Proceedings...* Heidelberg: Springer-Verlag, 2010. p.29-36.

BAKER, C. F.; FILLMORE, C. J.; LOWE, J. B. The Berkeley FrameNet project. In: CONFERENCE COLING-ACL, 1998, Montreal. *Proceedings...* Montreal, [s.n.], 1998. p.86-90.

BARZDINS, G. et al. Multidimensional ontologies: integration of Frame semantics and ontological semantics. In: EURALEX INTERNACIONAL CONGRESS, 13, 2008, Barcelona. *Proceedings...* Barcelona: Universitat Pompeu Fabra, 2008. p.277-284. (Serie Activitats, v.20).

BATEMAN, J. *The generalized upper model*. 8 Sept. 2008. Disponível em: <<http://www.fb10.uni-bremen.de/anglistik/langpro/webSPACE/jb/gum/index.htm>>. Acesso em: 20 jan. 2011.

BATEMAN, J. A. et al. A linguistic ontology of space for natural language processing. *Artificial Intelligence*, [S.l.], v.174, n.14, p.1027-1071, Sep. 2010.

BORIN, L. et al. Thinking green: toward swedish FrameNet++. In: FRAMENET MASTERCLASS AND WORKSHOP, 2009, Milano. *Proceedings...* Milano: [s.n.], 2009. Disponível em: <[http://tlt8.unicatt.it/allegati/Session\\_I\\_3.pdf](http://tlt8.unicatt.it/allegati/Session_I_3.pdf)>. Acesso em: 20 jan. 2011.

BURCHARDT, A.; PENNACCHIOTTI, M. FATE: a FrameNet-Annotated corpus for textual entailment. In: LANGUAGE RESOURCES AND EVALUATION

CONFERENCE, 2008, Marrakech. *Proceedings...* Marrakech: [s.n.], 2008. Disponível em: <[http://www.marcopennacchiotti.com/pro/publications/LREC\\_2008\\_2.pdf](http://www.marcopennacchiotti.com/pro/publications/LREC_2008_2.pdf)>. Acesso em: 20 jan. 2011.

BRACHMAN, R. J. On the epistemological status of semantic networks. In: FINDLER, N.V. (Ed.) *Associative Networks: representation and use of knowledge by computers*. [S.l.]: Academic Press, 1979. p.3-50.

CHISHMAN, R. L. O. Integrando léxicos semânticos e ontologias. *Informação & Informação*, Londrina, v.14, p.103-124, 2009.

\_\_\_\_\_. O sentido polissêmico dos verbos eventivos segundo a teoria do léxico gerativo. *Fórum Linguístico*, Florianópolis, v. 3, n. 2, p.49-62, 2003.

CHISHMAN, R. L. et al. O conhecimento semântico representado em ontologias aplicadas à busca. In: SIMPÓSIO NACIONAL, 11; SIMPÓSIO INTERNACIONAL DE LETRAS E LINGÜÍSTICA, 1, 2006, Uberlândia. *Anais...* Uberlândia: [s.n.], 2006. p.334-345.

CHOW, I. C.; WEBSTER, J. J. Integration of linguistic resources for verb classification: FrameNet Frame, WordNet verb and SUMO. In: COMPUTATIONAL LINGUISTICS AND INTELLIGENT TEXT (CICLING'07), 2007, Mexico City. *Proceedings...* Mexico City: LNCS, 2007. p.1-11.

CIMIANO, P.; WENDEROTH, J. Automatically learning qualia structures from the web. In: ACL-SIGLEX WORKSHOP ON DEEP LEXICAL ACQUISITION, 2005, Ann Arbor. *Proceedings...* Ann Arbor: Association for Computational Linguistics, 2005. p.28-37.

CROFT, W. Connecting frames and constructions: a case study of 'eat' and 'feed'. *Constructions and Frames*, [S.l.], v.1, n.1, p.7-28, 2009.

CROFT, W.; CRUSE, D. A. *Cognitive linguistics*. Cambridge: Cambridge University Press, 2004.

DAVIES, M. *Corpus of contemporary american English: 450 million words, 1990-present*. Brigham Young University, 2008. Disponível em: <<http://corpus.byu.edu/coca/>>. Acesso em: 20 jan. 2011.

DAVIES, M.; FERREIRA, M. *Corpus do português: 45 million words, 1300s-1900s*. National Endowment for the humanities, 2006. Disponível em: <<http://www.corpusdoportugues.org>>. Acesso em: 20 jan. 2011.

FARRAR, S.; BATEMAN, J. A. *Linguistic ontology baseline: collaborative research center for spatial cognition*. Germany: University of Bremen, 2005.

- FAUCONNIER, G.; TURNER, M. Conceptual integration networks. *Cognitive Science*, [S.l.], v.22, n.2, p.133-187, 1998. Disponível em: <<http://ssrn.com/abstract=1292966>>. Acesso em: 20 jan. 2011.
- FELLBAUM, C. *WordNet*: an electronic lexical database. Cambridge: MIT, 1998.
- FILLMORE, C. J. *FrameNet*: the lexicon. ppt. 2008. Disponível em: <[http://www.hf.uib.no/forskingskole/FillmoreJune2\(a\).ppt](http://www.hf.uib.no/forskingskole/FillmoreJune2(a).ppt)>. Acesso em: 20 jan. 2011.
- \_\_\_\_\_. Frame semantics. In: GEERAERTS, D. (Ed.). *Cognitive linguistics: basic readings*. Berlin: Mouton de Gruyter, 2006. p.373-400.
- \_\_\_\_\_. Frames and the semantics of understanding. *Quaderni di Semantica*, [S.l.], v.5, n.2, p.222-54, Dec. 1985.
- \_\_\_\_\_. Scenes-and-frames semantics, linguistic structures processing. In: ZAMPOLLI, A. (Ed.). *Fundamental studies in computer science*. Amsterdam: North Holland Publishing, p.55-88, 1977.
- GAIO, S. et al. Un'introduzione all'ontologia DOLCE. *AIDAinformazioni*, [S.l.], Ano 28, n.1-2, p.107-125, 2010.
- GEERAERTS, D. *Cognitive linguistics: basic readings*. Berlin: Mouton de Gruyter, 2006.
- GONZALEZ, M.; LIMA, V. L. S. Redefining traditional lexical semantic relations with qualia information. *Revista Palavra*, [S.l.], n.12, p.25-36, 2004.
- GUARINO, N. Formal ontology, conceptual analysis and knowledge representation. *International Journal of Human-Computer Studies*, [S.l.], v.43, n.5/6, 1995.
- KASAMA, D. Y. et al. Do termo à estruturação semântica: representação ontológica do domínio da nanociência e nanotecnologia utilizando a estrutura Qualia. *Linguamática*, Braga, v.2, n.3, p.43-58, 2010.
- KATZ, J. J.; FODOR, J. A. The structure of a semantic theory. *Language*, [S.l.], v.39, n.2, p.170-210, 1964.
- LAKOFF, G. The invariance hypothesis: is abstract reason based on imageschemas? *Cognitive Linguistics*, Berlin, v.1, n.1, p.39-74, 1990.
- LENCI, A. et. al. *SIMPLE work package 2*: linguistic specifications, deliverable D2.1. Pisa: ILC-CNR, March 2000.
- LYONS, J. *Semantics*. Cambridge: Cambridge University Press, 1977.
- MENDES, S.; CHAVES, R. P. Enriching WordNet with Qualia information. In: WORKSHOP ON WORDNET AND OTHER LEXICAL RESOURCES: APPLICATIONS, EXTENSIONS, E CUSTOMIZATIONS AT THE II NAACL MEETING, 2001, Pittsburgh. *Proceedings...* Pittsburgh: Carnegie Mellon University, 2001. p.108-112.

- OHARA, K.; NIKIFORIDOU, K. Background: the constructional view of language. *Constructions and Frames*, [S.l.], v.1, n.1, p.7-28, 2009.
- OLTRAMARI, A. et al. Senso comune. In: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION, 7., 2010, Valletta. *Proceedings...* Valletta: LREC, 2010. p. 3873-3877.
- OVCHINNIKOVA, E. et al. Data-driven and ontological analysis of FrameNet for natural language processing. In: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION, 7., 2010, Valletta. *Proceeding...* Valletta: LREC, 2010. p.3157-3164.
- PARADIS, C. Ontologies and construals in lexical semantics. *Axiomathes*, [S.l.], v.15, p.541-573, 2005.
- PEDERSEN, B. S. et al. DanNet: a WordNet for danish. In: INTERNATIONAL CONFERENCE ON GLOBAL WORNETS, 3, 2006, Jeju. *Proceedings...* Jeju: [s.n.], 2006. p.329-330.
- PETRUCK, M. R. L. Frame semantics. In: VERSCHUEREN, J. et al. (Ed.). *Handbook of pragmatics*. Philadelphia: John Benjamins, 1996. p.01-13.
- PUSTEJOVSKY, J. *The generative lexicon*. Cambridge: MIT, 1995.
- \_\_\_\_\_. The generative lexicon. *Computing Linguist*, [S.l.], v.17, n.4, p.409-441, Dec. 1991.
- PUSTEJOVSKY, J.; ANICK, P. On the semantic interpretation of nominals. In: COLING, 1988, Budapeste. *Proceedings...* Budapeste: University of Busapest, 1988. v.2. p.518-523.
- RUIMY, N.; TORAL, A. More semantic links in the SIMPLE-CLIPS database. In: LANGUAGE RESOURCES AND EVALUATION CONFERENCE, 2008, Marrakech. *Proceedings...* Marrakech: LREC, 2008. 1 CD-ROM.
- RUPPENHOFER, J. et al. *FrameNet II: extended theory and practice*. 2010. Disponível em: <<http://framenet.icsi.berkeley.edu/>>. Acesso em: 12 jan. 2011.
- RUSSELL, S.; NORVIG, P. *Artificial intelligence: a modern approach*. 3.ed. Upper Saddle River: Prentice Hall, 2010.
- SALOMÃO, M. M. M. FrameNet Brasil: um trabalho em progresso. *Calidoscópico*, São Leopoldo, v.7, n.3, p.171-182, set/dez. 2009.
- SCHEFFCZYK, J. et al. Linking FrameNet to the suggested upper merged ontology. In: FORMAL ONTOLOGY IN INFORMATION SYSTEMS, 2006. *Proceedings...* IOS Press, 2006, p.289-300.

SCHMID, H. *English abstract nouns as conceptual shells: from corpus to cognition*. Berlin: Mouton de Gruyter, 2000.

SPANISH FRAMENET: an on-line lexical resource and its application to Spanish NLP. Disponível em: <<http://gemini.uab.es:9080/SFNsite/spanish-framenet-process>>. Acesso em: 12 jan. 2011.

W3C. *OWL 2 web ontology language: document overview*. 27 Oct. 2009. Disponível em: <<http://www.w3.org/2009/pdf/REC-owl2-overview-20091027.pdf>>. Acesso em: 17 jan. 2011.

VOSSSEN, P. *EuroWordNet: a multilingual database with lexical semantic networks*. New York: Springer, 1998.

YU, N. *The contemporary theory of metaphor: a perspective from Chinese*. Amsterdam: John Benjamins, 1998. (Human Cognitive Processing series, v.1).

ZAVAGLIA, C.; GREGHI, J. G. Homonymy in natural language processes: a representation using Pustejovsky's Qualia structure and ontological information. In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL PROCESSING OF THE PORTUGUESE LANGUAGE, 6., 2003, Berlin. *Proceedings...* Berlin: Springer-Verlag, 2003. p.86-93.

Recebido em 23 de setembro de 2011.

Aprovado em 20 de agosto de 2012.

