

## LUDI: UM *FRAMEWORK* PARA DESAMBIGUAÇÃO LEXICAL COM BASE NO ENRIQUECIMENTO DA SEMÂNTICA DE FRAMES

*Ely Edison da Silva Matos (UFJF)<sup>1</sup>, Maria Margarida Martins Salomão (UFJF)<sup>2</sup>*

### RESUMO

Enquanto no âmbito da Sintaxe as aplicações em Processamento da Língua Natural estão relativamente bem estabelecidas, no âmbito da Semântica não se observa ainda a mesma maturidade. Este trabalho é uma contribuição para os estudos em Semântica Computacional, com base na Linguística Cognitiva, reunindo ferramentas computacionais e recursos linguísticos. O objeto de estudo é a desambiguação de Unidades Lexicais. O resultado é uma aplicação computacional, chamada *Framework* LUDI (*Lexical Unit Discovery through Inference*). O *framework* pode ser integrado em ferramentas para recuperação de informação, sumarização, *parsers* semânticos e em processos de Etiquetagem de Papéis Semânticos.

**PALAVRAS-CHAVES:** Desambiguação do Sentido de Lexemas, Semântica Computacional, Semântica de Frames, Léxico Gerativo.

### ABSTRACT

While in the field of Syntax, applications in Natural Language Processing are relatively well established, the situation is different for the field of Semantics. This work is a contribution to the studies in Computational Semantics, using ideas from Cognitive Linguistics and bringing together computational tools and linguistics resources. The object of study is disambiguation of Lexical Units. The result is an application, called *Framework* LUDI (*Lexical Unit Discovery through Inference*). The framework can be integrated into information retrieval and summarization tools, semantic parsers and into Semantic Role Labeling process.

**KEYWORDS:** Lexeme Sense Disambiguation, Computational Semantics, Frame Semantics, Generative Lexicon.

1 Universidade Federal de Juiz de Fora

2 Universidade Federal de Juiz de Fora

## 1. INTRODUÇÃO

Os maiores sucessos do Processamento da Língua Natural (PLN) estão na área da Sintaxe, apoiados em ideias da Linguística Gerativa. O crescente uso de modelos estatísticos e probabilísticos em PLN expõe claramente, no entanto, que não podemos esperar que a linguagem esteja presa apenas a princípios e parâmetros. Por outro lado, na área da Semântica Computacional, cujo objetivo é a captura do significado de expressões em língua natural (BOS, 2011), o instrumento por excelência tem sido a lógica formal, em especial a Lógica de Primeira Ordem. Mas neste campo a dificuldade não é menor, pois fenômenos como a estrutura de eventos, a aspectualidade, os contextos modais, a quantificação e a resolução de anáforas resistem a uma formalização.

Este trabalho é fruto da tese de doutorado apresentada em Matos (2014), com o objetivo de implementar um *framework* para a desambiguação de itens lexicais. O trabalho pode ser enquadrado na área do PLN, por estar relacionado às técnicas de WSD (*Word Sense Desambiguation*). Porém, ele se afasta do *mainstream* em três aspectos principais: (i) adota-se a filosofia, os conceitos e a terminologia proposta pela Linguística Cognitiva, (ii) apoia-se na Semântica de Frames, uma abordagem que privilegia a compreensão da linguagem em lugar de sua formalização e (iii) é utilizado um Modelo Conexionista Estruturado, sem apelar (diretamente) para uma abordagem baseada em probabilidades.

O resultado do trabalho é corporificado na forma de uma aplicação computacional, chamada Framework LUDI (*Lexical Unit Discovery through Inference*), composta por algoritmos e estruturas de dados usados na desambiguação. O *framework* é uma aplicação de Compreensão da Língua Natural, que pode ser integrada em ferramentas para recuperação de informação, sumarização, *parsers* semânticos e em processos de Etiquetagem de Papéis Semânticos.

O texto foi dividido em cinco seções, além desta Introdução. A seção 2 aborda a questão do significado das expressões linguísticas, além de alguns pressupostos e hipóteses adotadas no trabalho. A seção 3 apresenta aspectos linguísticos e computacionais usados na concepção do *framework* e na sua implementação. A seção 4 descreve o *framework* LUDI, abordando o modelo conexionista e a construção da rede semântica. A seção 5 mostra a aplicação do *framework* no estudo da polissemia do verbo <abrir.v>. Finalmente, a última seção apresenta algumas conclusões e sugestões de trabalhos futuros.

## 2. A QUESTÃO DO SIGNIFICADO

As seguintes sentenças não devem apresentar dificuldades de compreensão para um leitor ou ouvinte humano:

- i. João abriu a janela.
- ii. João abriu sua boca.
- iii. João abriu o livro.

- iv. João abriu as cortinas.
- v. O cirurgião abriu o peito de João.

Nestes exemplos o significado de <abrir> é função do conhecimento que temos de como as entidades e objetos de diferentes tipos são “abertos” e do contexto da enunciação. A contribuição semântica de <abrir>, portanto, é bastante variável. Porém, se consideramos usos mais figurativos a dependência do contexto se torna ainda mais marcante:

- i. João abriu uma conta no banco.
- ii. João abriu a reunião.
- iii. João está aberto ao diálogo.
- iv. O policial abriu fogo.
- v. “Aquarela do Brasil” abre a coletânea.
- vi. Finalmente ele se abriu com ela.

O significado de <abrir> agora se relaciona com várias ações, eventos, situações e agentes. Apesar desta diversidade, que pode incluir usos idiomáticos, metonímicos e metafóricos, um leitor humano ainda teria pouca ou nenhuma dificuldade para compreensão das sentenças.

A questão muda se considerarmos a legibilidade por máquinas. Desprovidas de um “conhecimento do mundo” e de posse de um contexto limitado (quando este é fornecido), as aplicações computacionais têm muita dificuldade para compreender um texto, se comparadas com os seres humanos. Estas dificuldades podem estar associadas a dois fatos, que se destacam no PLN: o foco em questões relacionadas à Sintaxe (com abordagem gerativista) ou, quando a Semântica é considerada, o uso quase exclusivo de teorias associadas a Semântica Formal (baseada em Lógicas).

Existem alternativas. A abordagem cognitivista da linguagem propõe que o processo de significação, ao invés de ser um processo em que uma “realidade” é “representada”, é um processo em que são evocadas estruturas cognitivas complexas, que se articulam em redes e que categorizam ou analisam a experiência do mundo. Um exemplo desta abordagem é a Semântica da Compreensão (*Understanding Semantics*, ou *U-semantics*) (FILLMORE, 1985). A *U-semantics* destaca algumas limitações das chamadas “condições de verdade” e da categorização baseada em condições “necessárias e suficientes”, pontos chave das semânticas de orientação formalista. Ela estabelece que nenhum aspecto do significado, que seja relevante para a compreensão, deve ser excluído da análise apenas porque as premissas metodológicas da teoria escolhida não consideram o conhecimento de fundo (ZIEM, 2014).

A proposta em Fillmore (1985) é estruturar este conhecimento de fundo através de *frames*. Os *frames* são estruturas cognitivas, formando uma rede, usadas para organizar nossas experiências e, ao mesmo

tempo, funcionar como ferramentas para a descrição e explicação dos significados lexical e gramatical. Assim, a chamada Semântica de Frames (*Frame Semantics*) (FILLMORE, 1982; FILLMORE, 2006), propõe que um item lexical só pode ter seu significado esclarecido quando relativizado a um *background*, expresso pelo *frame*.

No nosso caso, o uso da Semântica de Frames em uma aplicação real (TORRENT et al., 2014), que implementa um dicionário multilingual baseado em *frames*, apresentou uma questão eminentemente prática: se um lexema polissêmico está associado a mais de um *frame*, qual *frame* deve ser exibido ao usuário quando ele consultar o dicionário informando um lexema polissêmico, como no caso de <abrir>?

A resposta a esta questão deveria ser óbvia: devem ser exibidos todos os *frames* associados com aquele lexema. Porém, neste caso, o dicionário baseado em *frames* não se distinguiria dos dicionários comuns, que exibem todos os sentidos de um lexema na mesma entrada ou criam entradas repetidas com o lexema para cada sentido. Caberia ao usuário o processo da desambiguação (o que não deixa de ser paradoxal, pois se ele já conhece o sentido do que procura, o dicionário pouco pode lhe ajudar).

Porém, a questão motivadora do trabalho é outra: se o usuário fornecer “pistas” a respeito do que procura, por exemplo informando mais uma ou duas palavras, não é possível fornecer a ele uma resposta mais precisa? Afinal, se cada palavra pode evocar um ou mais *frames* e os *frames* estão relacionados entre si em uma rede, estas relações poderiam ajudar a estabelecer o *background* proposto pela Semântica de Frames, contra o qual o sentido mais correto (ou mais próximo) da palavra poderia ser descoberto. O processamento no *framework* LUDI é justamente a travessia desta rede, com o propósito de descobrir qual sentido (qual *frame*) está associado ao item lexical e, portanto, promover a desambiguação.

Além dos pressupostos teóricos inerentes à Semântica de Frames, o trabalho adota alguns pressupostos mais específicos:

- a) Os lexemas só tem significação quando associados a um *frame*. Mais claramente, somente Unidades Lexicais (um lexema associado a um *frame*) possuem significação. Esta questão é importante, já que as teorias sobre léxico postulam, quase unanimemente, uma significação para os lexemas independentemente do uso (GEERAERTS, 2009);
- b) O léxico é visto como uma extensão dos modelos cognitivos (no caso, dos *frames*). As Unidades Lexicais são tratadas igualmente como dispositivos cognitivos com uma estrutura interna. Ainda que esta estrutura seja insuficiente para atribuir um significado por si só, é ela que permite a composição da Unidade Lexical em um cenário mais amplo;
- c) O significado de uma sentença é alcançado através de um processo dinâmico de composicionalidade em duas vias: não apenas o significado da sentença está relacionado com a composição dos significados das palavras que a formam, mas o próprio significado das palavras

dentro da sentença é resultado da composicionalidade. Ou seja, enquanto o significado da sentença é função dos *frames* evocados pelos lexemas presentes na sentença, o *frame* escolhido para cada lexema é aquele que melhor se encaixa na composição.

Estes pressupostos fundamentam as seguintes hipóteses básicas do trabalho:

- a) Se o significado é dependente do uso do item lexical, os testes devem ser feitos com dados obtidos a partir da língua em uso, ou seja, dados de corpora;
- b) Se o significado pode ser obtido como produto da composição dos *frames* e esta composição se dá usando as relações existentes entre os *frames*, aumentar o número destas relações pode fornecer uma compreensão melhor do significado;
- c) Se a variação da valência de um verbo é um critério para avaliar a variação do sentido de um lexema (com a definição de um novo *frame*), a interface sintaxe-semântica deve ser considerada no processo de desambiguação;
- d) Se o item lexical possui uma estrutura interna, é necessário o apoio de uma teoria que descreva essa estrutura.

### 3. ASPECTOS LINGÜÍSTICOS E COMPUTACIONAIS

Esta seção apresenta, de forma bastante resumida, aspectos linguísticos e computacionais que foram considerados na elaboração do trabalho. Estes aspectos ajudam a contextualizar o trabalho nos cenários da Linguística Cognitiva e da Linguística Computacional.

#### 3.1. CATEGORIZAÇÃO E POLISSEMIA

Os lexemas estão associados a conceitos. Se os conceitos não existem de forma isolada na mente, mas estão organizados de alguma forma, o léxico também deve possuir uma estrutura. O léxico não é apenas uma longa lista de lexemas isolados e desconectados, sendo possível reconhecer vários tipos de relações e conexões entre os lexemas que nos ajudam a estabelecer regularidades e reconhecer padrões. Fundamentalmente, podemos considerar que as relações lexicais representam relações entre conceitos e que, inversamente, a categorização de conceitos também deve se refletir na categorização dos lexemas que os evocam.

A Teoria Clássica de conceitos propõe que estes são categorizados com base em definições, que codificam as condições necessárias e suficientes para a sua aplicação. Um conceito pertencerá a uma categoria se apresentar o conjunto de características daquela categoria. Por exemplo, a categoria <pássaro> pode ser definida pelas características: um ANIMAL, que TEM BICO e VOA.

A abordagem cognitivista levanta vários questionamentos contra esta categorização formalista. Em primeiro lugar, os estudos feitos por Rosch e descritos em Lakoff (1987) mostram que as categorias são tanto graduadas em termos de tipicidade quanto estruturadas internamente. Em uma categoria, alguns membros são mais típicos que outros, funcionando como protótipos. Em segundo lugar, as categorias cognitivas possuem limites imprecisos, ou seja, há gradações no pertencimento de um elemento a uma dada categoria. Uma terceira questão diz respeito à completude das condições necessárias e suficientes. Um pássaro empalhado, por não voar, deixa de ser classificado como pássaro? É preciso definir uma nova categoria de pássaros ou ampliar/detalhar as condições de classificação como <pássaro>?

Estas questões se refletem na categorização dos lexemas: um mesmo lexema pode apontar para mais de um conceito. Neste caso, pode se tratar de homonímia ou de polissemia. Embora não exista um consenso na diferenciação dos dois casos (LYONS, 1979), pode-se considerar, simplificada, que na homonímia não há relação semântica entre os conceitos evocados pela palavra - é comum afirmar que na realidade são duas palavras diferentes que calharam de ter as mesmas propriedades formais (CRUSE, 2010). Na polissemia, os sentidos evocados pela palavra estão de alguma forma relacionados. Mas qual a estrutura destas relações?

Dentre as estruturas propostas, destacam-se as categorias radiais, conforme o modelo descrito em Brugman e Lakoff (1988) e Lakoff (1987), o Modelo de Redes Esquemáticas (LANGACKER, 1987) e as Redes Semânticas (SOWA, 2015). Independente da estrutura adotada, não devem ser ignorados os usos metonímicos e metafóricos dos lexemas que, de certa forma, estendem os seus sentidos centrais.

Outro caminho para se estudar a distinção entre sentidos é a análise de dados em corpora (GRIES, 2006). O uso de dados reais permite perceber que a contribuição semântica de uma dada palavra está associada à situação interpretativa no contexto em que ela é usada. Além disso é possível observar as variações de sentido associadas aos padrões sintáticos e a influência de outras palavras em situação de coocorrência. Este cenário apresenta uma grande semelhança com a abordagem proposta pela Semântica de Frames.

### 3.2. SEMÂNTICA DE FRAMES E FRAMENET

A Semântica de Frames é uma abordagem para estudo do significado lexical baseado nos trabalhos de Fillmore e seus colaboradores ao longo dos últimos 35 anos (FILLMORE, 1977), (FILLMORE, 1982), (FILLMORE, 1985), (FILLMORE, 2006), (PETRUCK, 1996), (FILLMORE; BAKER, 2010), (GAWRON, 2008). A ideia central da Semântica de Frames é que o significado de uma palavra deve ser descrito em relação a um *frame* semântico. Um *frame*, neste contexto, é uma representação esquemática de estruturas conceituais e padrões de crença, práticas, instituições, etc. que provê a base para uma interação significativa em uma comunidade de fala.

Um *frame* representa um sistema de conceitos relacionados entre si de tal forma que, para compreender um deles é necessário compreender toda a estrutura onde ele se situa. O significado de uma expressão

linguística é interpretado de acordo com o *frame* que representa a cena no momento da enunciação. Por exemplo, o lexema <quebrar> pode significar <partir, romper> (“O copo quebrou”) ou <falir> (“A empresa quebrou”), dependendo do *frame* “evocado” e dos elementos linguísticos que ocorrem simultaneamente no mesmo enunciado.

O tratamento da polissemia no contexto da Semântica de Frames envolve a escolha entre manter significados distintos agrupados em uma só acepção (*lumping*) ou separar os significados em diferentes acepções (*splitting*). Embora a Semântica de Frames não direcione esta escolha, a abordagem que considera cada Unidade Lexical (UL) como o pareamento de um lexema com um *frame* possibilita o processo de *splitting*.

Este é um ponto fundamental para o trabalho descrito aqui, uma vez que a melhoria da interpretação por máquinas pressupõe a possibilidade de desambiguação dos lexemas polissêmicos. O fato de cada UL possuir seu próprio sentido, ou seja, evocar um *frame* específico já configura uma restrição para a rede semântica.

Em termos linguísticos, os elementos de um *frame* emergem na valência lexical e construcional das expressões que evocam este *frame* e podem ser considerados, segundo Salomão (2009), funções microtemáticas, postuladas em relação ao *frame* caracterizados por elas. Os elementos do *frame* são importantes no processo de desambiguação de sentenças como em (1) (adaptadas de Salomão (2009)), onde <substituir> evoca diferentes *frames*.

- (1) (a) Dunga <sub>[técnico]</sub> substituiu o atacante Robinho jogador no último jogo.  
(b) Neymar <sub>[jogador]</sub> substituiu o atacante Robinho jogador no último jogo.

As ideias de Fillmore em relação a Semântica de Frames deram origem ao projeto FrameNet. O projeto FrameNet (FILLMORE; JOHNSON; PETRUCK, 2003), uma iniciativa do International Computer Science Institute (ICSI) de Berkeley, Califórnia (EUA), vem, desde 1997, implementando computacionalmente uma base de dados lexical fundada na Semântica de Frames. O objetivo da FrameNet é documentar as possibilidades combinatórias de cada palavra, em cada um dos seus sentidos, ou seja, sua valência sintática e semântica. Para isso, a FrameNet criou uma base de dados com centenas de *frames* (relacionados semanticamente entre si) e associou cada sentido de um lexema (uma Unidade Lexical) a um *frame*. Através da anotação de sentenças obtidas em corpora, são obtidas as valências de cada Unidade Lexical. Estas informações são apresentadas em vários relatórios web.

### 3.3. COMPOSICIONALIDADE

Um dos pressupostos apresentados para o trabalho é que a interpretação de expressões linguísticas é governada pelo Princípio da Composicionalidade (CRUSE, 2010). Este princípio estabelece que o significado de uma expressão semanticamente complexa é uma função composicional dos significados de seus constituintes semânticos.

Os defensores da composicionalidade direta argumentam que qualquer operação sintática tem uma operação semântica correlacionada e, conseqüentemente, qualquer expressão bem formada em termos sintáticos tem um significado predizível. Este é o chamado Princípio da Composicionalidade Forte.

No entanto, a Linguística Cognitiva mostra que a questão é muito mais complexa: uma língua pode apresentar diversos padrões de combinações de palavras, onde cada combinação (das mesmas palavras) pode apresentar um sentido diferente; o mesmo item lexical pode se referir a conceitos diferentes; a alternância dos verbos pode provocar uma mudança na leitura; existem significados que não são sintaticamente transparentes; existem expressões que são não-composicionais. Além disso, como dito antes, fatores contextuais, conceptuais, discursivos e pragmáticos, além de um “conhecimento do mundo” limitado, podem comprometer seriamente a compreensão do enunciado.

Não obstante estas dificuldades, adotamos como pressuposto que o processo de composicionalidade é fundamental na busca do significado da sentença. Para obter este significado, é essencial olhar além das palavras individualmente, reconhecendo que o significado de uma dada palavra tipicamente inclui conceitos que não são únicos para nenhum item linguístico (DODGE, 2010). Também é preciso reconhecer que o significado não está associado apenas com o lexema, mas também com os padrões construcionais onde estes lexemas ocorrem.

O uso da Semântica de Frames permite que nos afastemos da hipótese forte da composicionalidade, que não admite ambigüidade, sinonímia, homonímia e vagueza. O uso de Frames provê o acesso a um parcial conhecimento de mundo e o entendimento de que a soma do significado das partes pode não ser igual ao significado do todo. A Semântica de Frames possibilita inferências explicadas (ou motivadas) pelo *frame* evocado que, associadas à integração de conceitos através de processos cognitivos, usando as relações entre *frames*, permite uma abordagem mais abrangente para a construção do significado.

Como um recurso teórico complementar à Semântica de Frames, no que se refere especificamente a entidades, bem como para atender a pressuposto que os itens lexicais possuem uma estrutura interna que pode ajudar no processo da composicionalidade, estamos utilizando também a Teoria do Léxico Gerativo.

### **3.4. TEORIA DO LÉXICO GERATIVO E ONTOLOGIA SIMPLE**

A Teoria do Léxico Gerativo (TLG), apresentada em Pustejovsky (1991) e desenvolvida em Pustejovsky (1995), considera o léxico como um sistema complexo, que forma uma parte crucial das línguas naturais e fornece as informações necessárias para o tratamento das propriedades polimórficas dos itens lexicais e sua permeabilidade ao contexto. Assim, as Unidades Lexicais são representadas por estruturas de informação, seguindo um conjunto finito de mecanismos gerativos que possibilitam a descrição do significado no contexto e a descrição da interface sintaxe-semântica (AMARO, 2009).

Um léxico gerativo pode ser caracterizado como um sistema envolvendo pelo menos quatro estruturas:



estrutura argumental (A); estrutura de eventos (E); estrutura qualia (Q) e estrutura de herança lexical (I). Um item lexical (IL) é então representado pela informação nestes níveis distintos por  $IL = \langle A, E, Q, I \rangle$ . Os mecanismos gerativos conectam estes quatro níveis, provendo a interpretação composicional das palavras em contexto (PUSTEJOVSKY, 1995).

A TLG considera que o significado das palavras está associado à estrutura qualia, na base de quatro fatores gerativos, chamados “papéis qualia”. Cada papel qualia captura como os seres humanos entendem os objetos e as relações no mundo e provê uma explicação mínima para o comportamento linguístico dos itens lexicais. A determinação dos valores de cada papel qualia não é obrigatória; para um dado item lexical, somente os valores dos papéis pertinentes devem ser definidos. (PUSTEJOVSKY, 1995) define os seguintes papéis qualia:

- 1 Formal: valores que diferenciam um dado objeto dentro do seu domínio semântico; tipicamente é a descrição de sua categoria básica.
- 2 Constitutivo: valores que expressam a relação entre um dado objeto e seus constituintes ou suas partes, tais como o material, o peso ou partes características.
- 3 Télico: valores relativos a informação sobre a função ou o propósito do objeto, tais como a intenção de um agente executando uma dada ação ou a função intrínseca do objeto.
- 4 Agentivo: valores que determinam a origem do objeto, tais como seu criador, tipo de origem (natural ou artificial) ou sua causa inicial.

Dado o viés computacional da TLG, os valores dos papéis qualia são especificados através de expressões lógicas, com tipos bem definidos e estruturas relacionais, indicando a associação apropriada com o termo predicador. O sistema de tipos, usado para categorizar os itens lexicais, é assim uma questão importante na TLG. A categorização é feita via herança ortogonal. Neste modelo de herança, ao invés de uma única estrutura organizacional para os tipos, diferentes estruturas são usadas ortogonalmente, baseadas na estrutura qualia. Um item lexical herda informação de acordo com a estrutura qualia que ele carrega. Por exemplo, <livro> é-um <objeto físico> (qualia formal), mas <livro> é-uma <informação> (qualia télico). Os tipos para um dado item lexical podem então ser unificados. Esta estratégia permite a criação de conceitos cada vez mais específicos. Tipos unificados podem ser vistos como estruturados por dimensões ortogonais (ou perspectivas).

Uma implementação computacional que usa (e estende) os conceitos da TLG é a ontologia SIMPLE (LENCI et al., 2000). SIMPLE é um projeto de larga escala, patrocinado pela União Europeia, com o objetivo de desenvolver um léxico semântico computacional, de ampla cobertura e múltiplos propósitos, para doze línguas da Europa. A ontologia SIMPLE foi projetada neste contexto e desenvolvida para garantir a uniformidade e consistência entre os dicionários monolíngues, provendo um modelo comum harmonizado, que codifica tipos semânticos estruturados e um conjunto de templates (análogos aos *frames* semânticos).

O vocabulário básico da SIMPLE se apoia em uma extensão da estrutura qualia, para organizar os tipos semântico/conceptuais como um mecanismo representacional que expresse a multidimensionalidade do significado. Isto fornece ao modelo um alto grau de generalidade, uma vez que os mesmos mecanismos são usados para conceitos de diferentes áreas semânticas (entidades, eventos, nomes abstratos, etc).

Neste trabalho foi utilizada a implementação da ontologia SIMPLE em OWL, chamada SIMPLE-OWL (TORAL; MONACHINI, 2007). Na SIMPLE-OWL os tipos semânticos foram implementados como classes. As relações são modeladas como propriedades de objeto e os traços (*features*) como propriedades de dados. Uma característica importante desta ontologia é que ela consiste tanto de Tipos Simples, que identificam somente o aspecto monodimensional do significado expresso por relações hiperonímicas, quanto de Tipos Unificados, que expressam os aspectos multidimensionais, combinando as relações de subtipo e as dimensões semânticas ortogonais.

### **3.5. MODELO CONEXIONISTA ESTRUTURADO**

Os Modelos Conexionistas são caracterizados como modelos computacionais com habilidades de adaptar, aprender, generalizar, agrupar ou organizar dados, nos quais a operação é baseada em processamento paralelo. Neste modelo, uma rede é formada por um conjunto de unidades de processamento simples, chamadas de nós, células ou neurônios (donde a denominação de Redes Neurais Artificiais). Estas unidades comunicam entre si enviando sinais de uma para outra, através de conexões ponderadas, através da técnica Ativação Propagada (Spread Activation).

Porém, uma questão se levanta: como unidades tão simples podem representar algum conhecimento? A análise de Feldman e Ballard (1982) indica a possibilidade de construção de redes com um nível mais alto de abstração, em que as unidades podem possuir algum tipo de estrutura, necessária para a organização do conhecimento na rede. Este modelo é chamado de Modelo Conexionista Estruturado.

Shastri e Ajjanagadde (1989) apresentam a proposta de adotar o Modelo Conexionista Estruturado como uma abordagem para a implementação de redes semânticas que representem conhecimento e inferências. Este é o modelo usado no trabalho. A integração de recursos lexicais que estão disponíveis em diferentes formatos (os dados da FrameNet estão armazenados em um banco de dados enquanto as ontologias usadas estão descritas em OWL) é feita através de uma rede semântica cujos nós (representando conceitos) são estruturas de traço e a realização de inferências é feita através do uso da técnica de ativação propagada.

## **4. FRAMEWORK LUDI**

O trabalho descrito deu origem a uma aplicação computacional, denominada Framework LUDI, baseado em um recurso lexical (a FrameNet), apoiado em uma visão ontológica (a ontologia SIMPLE e a ontologia LUDI, desenvolvida pelos autores) e implementado como um modelo conexionista estruturado. Esta seção apresenta alguns aspectos relacionados à construção da rede semântica utilizada no framework.

## 4.1. ENTIDADES, EVENTOS E QUALIDADES

A ontologia LUDI faz a distinção entre três categorias básicas: Entidades, Eventos e Qualidades. Foi adotada a abordagem da ontologia DOLCE (MASOLO et al., 2003), em que eventos são perdurantes (*perdurants*), ou seja, ocorrem ou acontecem no tempo, e entidades são endurantes (*endurants*), ou seja, existem independente do tempo. As qualidades, por sua vez, estão associadas às características das entidades ou eventos.

Ruppenhofer et al. (2010, p.5) reconhece que nomes referentes a tipos artefatuais ou naturais possuem uma estrutura de *frame* mínima, por si próprios. Artefatos geralmente ocorrem com outras expressões, indicando seu subtipo, o material de que é feito, a maneira de produção ou seu propósito e/ou uso, ou seja, sua estrutura qualia. Assim, optamos, no trabalho, pela substituição dos *frames* de Entidades atualmente registrados pela FrameNet, pelas classes da ontologia SIMPLE.

Com relação aos Eventos, foi mantida a modelagem da FrameNet, com a organização dos Frames (e das relações entre eles) segundo uma perspectiva ontológica, utilizando a própria ontologia SIMPLE. É importante ressaltar que as classes de eventos não visam substituir os Frames da FrameNet, mas sim classificá-los.

Para organização do nível estrutural da ontologia LUDI, referente a Qualidades, adotamos uma abordagem baseada nos chamados espaços conceituais (*conceptual spaces*), apresentados em Gärdenfors (2000). Embora fosse necessária uma reestruturação dos *frames* para se trabalhar mais completamente com a ideia de espaços conceituais, optamos por manter os frames atuais, realizando apenas uma simples adaptação das relações entre eles.

## 4.2 ELEMENTOS DE FRAME

Dentro da hipótese que aumentar a densidade da rede pode melhorar a compreensão do significado, foram feitas duas adições à estrutura da FrameNet:

- a) associação dos Elementos de Frame com uma classe da ontologia SIMPLE (que funciona então como um Tipo Ontológico, na perspectiva da FrameNet);
- b) associação dos Elementos de Frame com outros Frames da própria FrameNet.

A associação (b) necessita de uma justificativa, uma vez que ela não existe na estrutura da FrameNet atual e não foi ainda proposta em trabalhos anteriores, configurando uma extensão de caráter conceitual. A ideia é registrar a referência que o elemento de uma *frame* pode fazer a um outro *frame*. Assim, além de atuarem em termos linguísticos como papéis microtemáticos, específicos para a situação descrita pelo *frame*, os elementos podem desempenhar uma função cognitiva, como evocadores de outros *frames*, a partir do *frame* a que pertencem.

### 4.3. Integração de Ontologias

A ontologia LUDI faz a integração de duas outras ontologias: uma Ontologia de Frames e a Ontologia SIMPLE. Cada uma atende a um determinado propósito. A Figura 1 ilustra esta integração.

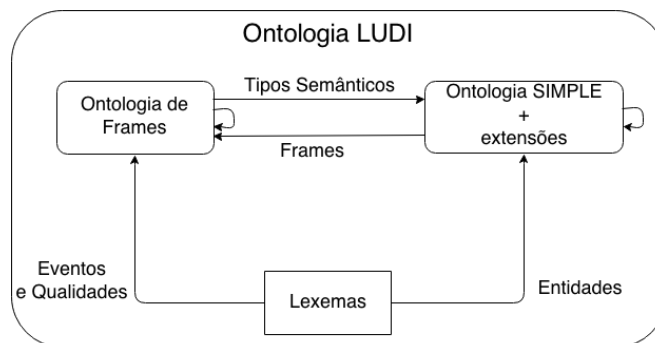


Figura 1 – LUDI - Integração de ontologias.

A Ontologia de Frames não é, na verdade, uma ontologia descrita formalmente, já que há dificuldades em formalizar a estrutura da FrameNet. Esta ontologia consiste na representação de alguns elementos da FrameNet como classes ou propriedades. Ela não contém a totalidade das informações disponibilizadas pela FrameNet, mas apenas aquelas utilizadas nas operações do framework LUDI: os Frames, os Elementos de Frame, as Unidades Lexicais, as Relações entre Frames e as Relações entre Elementos de Frame.

A Ontologia SIMPLE (com algumas extensões) é usada para registrar os tipos semânticos estruturados. Estes tipos possuem relações entre si e são usados basicamente para categorizar os *frames* e os elementos de *frames*. Os lexemas referentes a entidades (tipos naturais e artefatos) são associados a tipos da ontologia SIMPLE. Os tipos da ontologia podem ser associados também a *frames*, quando as relações indicam algum tipo de evento ou qualidade. Mecanismos de inferência são usados para a classificação dos lexemas em relação à estrutura qualia.

### 4.4. A REDE LUDI

#### 4.4.1. ESTRUTURA DE TRAÇOS

Para a implementação da rede LUDI é realizada uma operação intermediária, que consiste em representar os dados da FrameNet e da SIMPLE formatados como MVAs (Matriz de Valores de Atributos), representadas via uma Estrutura de Traços Tipados (*Typed Features Structure*). A redução a um formato comum facilita a integração das duas fontes de dados.

Em relação à FrameNet, cada Estrutura de Traços representa um *frame*. O tipo da estrutura é o nome do *frame*. Cada traço representa um elemento do *frame*. O valor do traço é uma lista de apontadores (i) para elementos de outros *frames* ou (ii) para as classes da Ontologia SIMPLE, que funcionam como Tipos Ontológicos ou (iii) para outros *frames*. Os lexemas, os lemas e as Unidades Lexicais também são representados por Estruturas de Traços, o que possibilita considerar as Unidades Lexicais como tipos

estruturados com traços próprios (que podem ser unificados com os traços do *frame*, por exemplo).

Em relação a ontologia SIMPLE, cada Estrutura de Traços representa uma classe (um tipo semântico da SIMPLE). O tipo da estrutura é o nome da classe. Cada traço representa uma relação qualia associada com o tipo semântico. O valor do traço é uma lista de apontadores para outros tipos semânticos ou para *frames*.

#### 4.4.2. VALÊNCIAS E FUNÇÕES GRAMATICAIS

Nesta implementação usamos o termo “valência” no mesmo sentido apresentado por Fillmore (2008): as valências registram as diversas possibilidades combinatórias da interface sintaxe- semântica. As valências indicam as várias configurações dos constituintes associados a uma determinada Unidade Lexical. É, portanto, uma estrutura complexa: ela está associada simultaneamente a uma Unidade Lexical e a vários elementos de *frame*, juntamente com as funções gramaticais atribuídas a cada um destes elementos. Cada valência é representada como uma unidade da rede. Assim, cada *frame* pode ser conectado a diversas valências e cada valência pode ser conectada a diversos elementos de *Frame*. O processo de análise consiste em determinar a valência mais provável para os dados fornecidos e, conseqüentemente, o *frame* cuja evocação é mais provável.

No trabalho foram consideradas as funções gramaticais descritas em Torrent e Ellsworth (2013): Argumento Externo (Ext), Objeto Direto (DObj), Objeto Indireto (IObj) e Dependente (Dep). As funções gramaticais são implementadas como traços dos links entre as valências e os elementos de *frame* e não como unidades da rede. Durante o processo de análise, a ativação ou não de um nó correspondente a uma valência vai depender da compatibilidade entre a função gramatical atribuída aos lexemas e a função gramatical atribuída ao link.

#### 4.4.3. UNIDADES E CONEXÕES DA REDE

A rede LUDI é ativada quando uma análise deve ser realizada. As estruturas de traço constituem as Unidades da rede. A rede é representada como um grafo dirigido, cujos nós recebem uma letra indicando o tipo de Unidade representada. A Tabela 1 apresenta a descrição de cada Unidade.

A Tabela 2 apresenta as conexões (*links*) entre as diversas unidades. O peso atribuído a cada link é sempre 1.0, exceto para os *links* em que o destino é um tipo semântico da ontologia SIMPLE. Neste caso, o peso do *link* é calculado em função da posição do tipo na hierarquia da ontologia.

Símbolo	Unidade	Descrição
W	Word	Palavra: uma das formas que uma palavra pode assumir
X	Lexema	A forma não flexionada da palavra
L	Lema	Um lema é composto de um ou mais lexemas
U	Unidade Lexical Unidade Semântica	Pareamento de um lema com um significado. Na FrameNet o pareamento de um lema com um <i>frame</i> é chamado Unidade Lexical. Na SIMPLE, o pareamento é chamado Unidade Semântica (USem).
F	Frame	Estrutura usada para representar situações, eventos e qualidades
E	Elemento de Frame	Usado para representar os participantes de um <i>frame</i> .
V	Valência	Representa a estrutura argumental de uma Unidade Lexical
T	Tipo Ontológico	Tipo semântico estruturado ontologicamente e baseado nas relações qualia

Tabela 1 - Unidades da Rede LUDI

Conexão	Descrição
XL	Conexão entre o Lexema (fornecido como entrada para a análise) e o Lema correspondente.
LF	Conexão entre o Lema e o Frame, representando o pareamento forma-significado.
FV	Conexão entre o Frame e as Valências consideradas para uma dada análise.
VE	Conexão entre uma Valência e os Elementos de Frame correspondentes.
EE	Conexão entre Elementos de Frame de Frames distintos.
ET	Conexão entre um Elemento de Frame e o Tipo Ontológico associado.
TE	Conexão entre um Tipo Ontológico e um Elemento de Frame (um elemento de um Frame esquemático, representando a conceptualização para um dado Tipo Ontológico).
LU	Conexão entre um Lema e uma Unidade Semântica (cada sentido do Lema, no caso de polissemia).
UT	Conexão entre uma Unidade Semântica e o Tipo Ontológico associado.
UF	Conexão entre uma Unidade Semântica e o Frame evocado por ela.
FT	Conexão entre um Frame e o Tipo Ontológico atribuído ao Frame.
UE	Conexão entre uma Unidade Semântica e um Elemento de Frame (quando a Unidade Semântica puder ser conceptualizada como participante em um Frame).
TT	Conexão entre dois Tipos Ontológicos.

Tabela 2 - Links da Rede LUDI

#### 4.4.4. CONSTRUÇÃO E ATIVAÇÃO DA REDE

O processo de análise é iniciado com a construção da rede completa, chamada **Rede LUDI**, que consiste de todos os nós e relações disponíveis na ontologia SIMPLE, mais os *frames* e os lexemas sob análise. A seguir, a partir da Rede LUDI, é construída uma rede intermediária, chamada **Rede Foco**, em que constam apenas os nós e relações associados aos lexemas analisados. A Ativação Propagada é realizada nesta rede. A rede resultante, composta pelos nós ativados, é chamada **Rede Resultado**.

### 5. RESULTADOS EXPERIMENTAIS

#### 5.1. O USO DE CORPORA

Uma das hipóteses do trabalho é o uso de corpora, por dois grandes motivos. Em primeiro lugar, ainda que a intuição do analista como usuário falante da língua seja rica, o uso feito pelos outros milhões de falantes é muito mais rico. Em segundo lugar, uma análise superficial de algumas centenas de sentenças é suficiente para concluir que qualquer categorização realizada *a priori* é bastante limitada. Isto não é surpresa, se adotamos como representação das conceptualizações o uso de esquemas

imagéticos (ao invés de uma solução meramente simbólica). A riqueza conceptual de cada esquema, acrescida das possibilidades de combinações e transformações, torna virtualmente impossível uma categorização completa. Assim, o uso de corpora para avaliar a categorização dos *frames* visa também testar a abrangência desta categorização, sem a ilusão de que uma cobertura de 100% possa ser alcançada.

## 5.2. O ESQUEMA BÁSICO DE <ABRIR.V>

Usando uma abordagem cognitiva, os muitos sentidos de <abrir.v> podem ser projetados a partir da junção do esquema imagético CONTAINER e da noção de dinâmica de forças (*force-dynamics*) (TALMY, 2000). Os sentidos são derivados através das operações de perfilamento, perspectivação e transformações no esquema imagético, bem como sua projeção em domínios mais abstratos, via metáforas. Assim, o *frame* esquemático para esta análise de <abrir.v> (Figura 2) é um CONTAINER sobre cujo portal podem ser aplicadas duas forças (esquema FORCE), através de um Agonista e de um Antagonista.

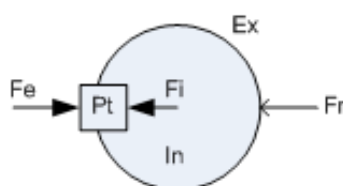


Figura 2 – Esquema para <abrir.v>.

Este é um esquema conceptual básico, servindo para a projeção dos vários sentidos de <abrir.v> apresentados no trabalho. Os elementos do esquema são detalhados na Tabela 3.

Sigla	Nome	Descrição
Ex	Exterior	representa a região externa ao container, incluindo os elementos que podem estar localizados nesta região.
In	Interior	representa a região interna ao container, incluindo os elementos que podem estar localizados nesta região.
Pt	Portal	representa um elemento que funciona como ligação entre as regiões exterior e interior, possibilitando o trânsito de elementos entre essas regiões (de "fora" para "dentro", ou vice-versa). O portal perfila o conceito de "possibilidade de acesso", e pode ser realizado, por exemplo, por uma abertura, uma passagem ou uma barreira.
Fr	Fronteira	representa o elemento que separa as regiões interior e exterior do container, servindo para delimitar o que é considerado interior ao container ou não.
Fe	Força externa	representa a ação de uma força atuando no portal "de fora para dentro" do container.
Fi	Força interna	representa a ação de uma força atuando no portal "de dentro para fora" do container.

Tabela 3 – Elementos do esquema para <abrir.v>

## 5.3. OS FRAMES DE <ABRIR.V>

A Figura 3 apresenta o *frame* não-lexical ABRIR\_ESQUEMA, usado para representar o esquema básico. Os elementos Agente, Causa, Instrumento são definidos mais especificamente como:

- a) Agente: representa uma entidade consciente - dotada de volição e que age com intenção - responsável pelo evento causador de <abrir.v>.

- b) Causa: representa um evento causador de <abrir.v>, que não esteja em relação com uma entidade consciente.
- c) Instrumento: representa o elemento em destaque no evento causador de <abrir.v>, usado pelo Agente ou relacionado a Causa.

I - ABRIR_ESQUEMA		
Elemento de Frame	Tipo	Descrição
Agente	CoreSet	Representa a força associada à ação de <abrir.v>.
Causa		
Instrumento		
Container	Coreset	Representa a entidade que sofre a ação da força e que muda seu estado.
Portal		
Conteúdo	Periférico	Representa os elementos na região interior do container.
Interior	Periférico	Representa a região interior do container.
Exterior	Periférico	Representa a região exterior ao container.
Fronteira	Periférico	Representa a separação entre as regiões interior e exterior do container.

Figura 3 - ABRIR\_ESQUEMA

Em relação ao conjunto (Container, Portal) é importante ressaltar a relação parte-todo existente. O Portal é “parte” do Container. Como o padrão metonímico “Parte-pelo-todo” é bastante comum, as sentenças podem lexicalizar tanto o Portal quanto o Container.

A partir do esquema básico, uma rede de *clusters* de sentidos é apresentada na Figura 4. O *cluster* Movimento\_corporal é definido para o sentido concreto e corporificado de <abrir.v>. O *cluster* Abrir\_físico também possui um sentido concreto. A partir destes *clusters*, os demais são obtidos através de interpretações que consideram o uso de metáforas.

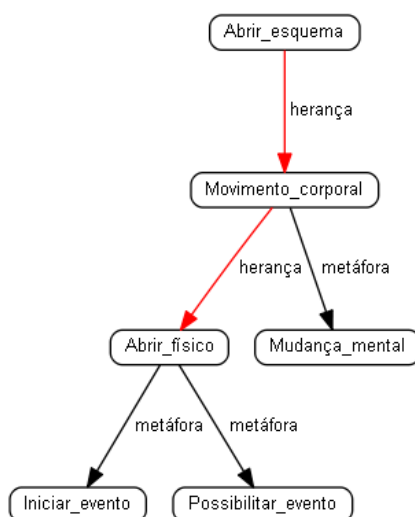


Figura 4 – Cluster de frames herdeiros de ABRIR\_ESQUEMA.

A rede de *clusters* é expandida na Figura 5. Basicamente é considerada a alternância causativo-incoativo e são adicionados alguns *frames* já definidos pela FrameNet. A fim de facilitar as referências, a Figura 6 apresenta a numeração dos *clusters* e dos *frames*.



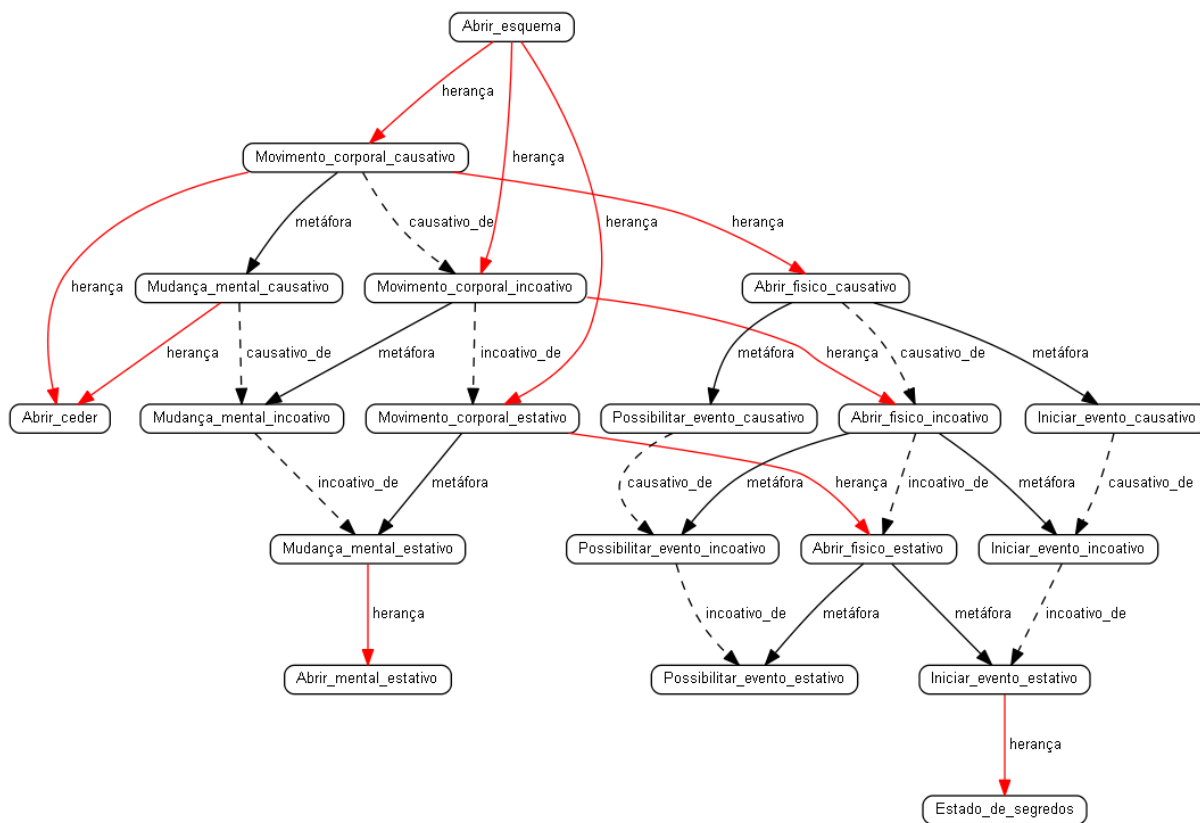


Figura 5 - Rede de frames herdeiros de ABRIR\_ESQUEMA

Domínio	Cluster	#Frame	Nome do Frame
Esquema básico	I – Esquema básico	1	Abrir_esquema
Corpo físico	II – Movimento_corporal	1	Movimento_corporal_causativo
		2	Movimento_corporal_incoativo
		3	Movimento_corporal_estativo
Entidades físicas	III – Abrir_fisico	1	Abrir_fisico_causativo
		2	Abrir_fisico_incoativo
		3	Abrir_fisico_estativo
Aspecto	IV – Iniciar_evento	1	Iniciar_evento_causativo
		2	Iniciar_evento_incoativo
		3	Iniciar_evento_estativo
		4	Estado_de_segredos
Modalidade	V – Possibilitar_evento	1	Possibilitar_evento_causativo
		2	Possibilitar_evento_incoativo
		3	Possibilitar_evento_estativo
Estado Mental, Emocional ou Psicológico	VI – Mudança_mental	1	Mudança_mental_causativo
		2	Mudança_mental_incoativo
		3	Mudança_mental_estativo
		4	Abrir_mental_estativo
		5	Abrir_ceder

Figura 6 - Frames por clusters

## 5.4. VALÊNCIAS

Usando a terminologia de Fillmore(2008), a “descrição de valência” de uma dada Unidade Lexical mostra o conjunto de associações dos elementos do *frame* com as Funções Gramaticais e com os

Tipos Sintagmáticos. A Figura 7 mostra algumas valências utilizadas no trabalho. Esta descrição é reconhecidamente incompleta, pois é derivada da análise de algumas poucas dezenas de sentenças. As valências que se diferenciam apenas pela variação dos elementos de *frame* foram agrupadas. Por exemplo, o conjunto (Agente, Causa, Instrumento) com a função “Argumento Externo” é representado como [Agente|Causa|Instrumento].ext. No entanto, cada valência é implementada individualmente.

<b>Frame II.1 – Movimento_corporal_causativo</b>	
<abrir.v>	V1: [Agente Causa Instrumento].ext, Parte_corporal.dobj, Experienciador.dep
<abertura.n>	V1: [Agente Causa Instrumento].ext, Parte_corporal.dep, Experienciador.dep
<b>Frame II.2 – Movimento_corporal_incoativo</b>	
<abrir.v>	V1: Parte_corporal.ext, Experienciador.dep
<b>Frame II.3 – Movimento_corporal_estativo</b>	
<aberto.a>	V1: Parte_corporal.ext, Experienciador.dep
<b>Frame III.1 – Abrir_físico_causativo</b>	
<abrir.v>	V1: [Agente Causa Instrumento].ext, [Container Portal].dobj, [Conteudo Interior Exterior Frenteira].dep V2: [Agente Causa Instrumento].ext, Portal.dobj, Container.dep, [Conteudo Interior Exterior Frenteira].dep
<abertura.n>	V1: [Agente Causa Instrumento].ext, [Container Portal].dep, [Conteudo Interior Exterior Frenteira].dep
<b>Frame III.2 – Abrir_físico_incoativo</b>	
<abrir.v>	V1: [Container Portal].ext, [Conteudo Interior Exterior Frenteira].dep V2: Container.ext, Portal.dep, [Conteudo Interior Exterior Frenteira].dep V3: Container.dep, Portal.ext, [Conteudo Interior Exterior Frenteira].dep
<b>Frame III.3 – Abrir_físico_estativo</b>	
<aberto.a>	V1: [Container Portal].ext, [Conteudo Interior Exterior Frenteira].dep V2: Container.ext, Portal.dep, [Conteudo Interior Exterior Frenteira].dep V3: Container.dep, Portal.ext, [Conteudo Interior Exterior Frenteira].dep
<b>Frame IV.1 – Iniciar_evento_causativo</b>	
<abrir.v>	V1: [Agente Causa Instrumento Evento_inicial].ext, [Evento Entidade].dobj
<abertura.n>	V1: [Agente Causa Instrumento Evento_inicial].ext, [Evento Entidade].dobj
<b>Frame IV.2 – Iniciar_evento_incoativo</b>	
<abrir.v>	V1: [Evento Entidade].ext, Tempo.dep
<b>Frame IV.3 – Iniciar_evento_estativo</b>	
<aberto.a>	V1: [Evento Entidade].ext
<b>Frame IV.4 – Estado_de_segredos</b>	
<aberto.a>	V1: Fenômeno.ext, Conhecedor.dep
<b>Frame V.1 – Possibilitar_evento_causativo</b>	
<abrir.v>	V1: [Agente Causa Instrumento Evento_inicial].ext, Meio.dobj, Evento.iobj
<b>Frame V.2 – Possibilitar_evento_incoativo</b>	
<abrir.v>	V1: Meio.ext, Evento.iobj
<b>Frame V.3 – Possibilitar_evento_estativo</b>	
<aberto.a>	V1: Meio.ext, Evento.iobj
<b>Frame VI.1 – Mudança_mental_causativo</b>	
<abrir.v>	V1: [Agente Causa Instrumento].ext, Parte_corporal.dobj, Evento.iobj, Experienciador.dep
<b>Frame VI.2 – Mudança_mental_incoativo</b>	
<abrir.v>	V1: Parte_corporal.ext, Evento.iobj, Experienciador.dep
<b>Frame VI.3 – Mudança_mental_estativo</b>	
<aberto.a>	V1: Parte_corporal.ext, Evento.iobj, Experienciador.dep
<b>Frame VI.4 – Abrir_ceder</b>	
<abrir_mão.v>	V1: Capitulador.ext, [Força Questão].iobj
<b>Frame VI.5 – Abrir_mental_estativo</b>	
<aberto.a>	V1: Experienciador.ext, Topico.dep

Figura 7 – Valências

## 5.5. EXEMPLOS DE APLICAÇÃO

Esta seção apresenta alguns exemplos de resultados da aplicação do *framework* LUDI para desambiguação de sentenças com o verbo <abrir.v> (o trabalho completo, em Matos (2014), apresenta

os resultados para as expressões abrir\_boca, abrir\_olho, abrir\_mão, abrir\_carteira, abrir\_porta, abrir\_jornal, abrir\_museu, abrir\_exposição, abrir\_shopping, abrir\_caminho, abrir\_cerveja, abrir\_marcador). Para cada experimento são apresentados: a motivação da escolha da expressão usada; a sentença e os segmentos usados como entrada para os algoritmos (juntamente com as funções gramaticais); a rede resultante, com a inferência final realizada e um comentário sobre o resultado obtido.

### 5.5.1. ABRIR\_PORTA

Motivação: Testar a inferência com entidades físicas conceptualizadas como containers.

Sentença: (...) *quem me abriu a porta do apartamento foi uma empregada* (...)

Segmentos: abrir.target, porta.dobj, empregada.ext

Comentário: A Figura 8 mostra que, apesar do processo de inferência ser direto, são ativadas mais de uma valência. Isto acontece porque <porta> é sempre conceptualizada como uma abertura para um container (uma construção 'building' neste caso).

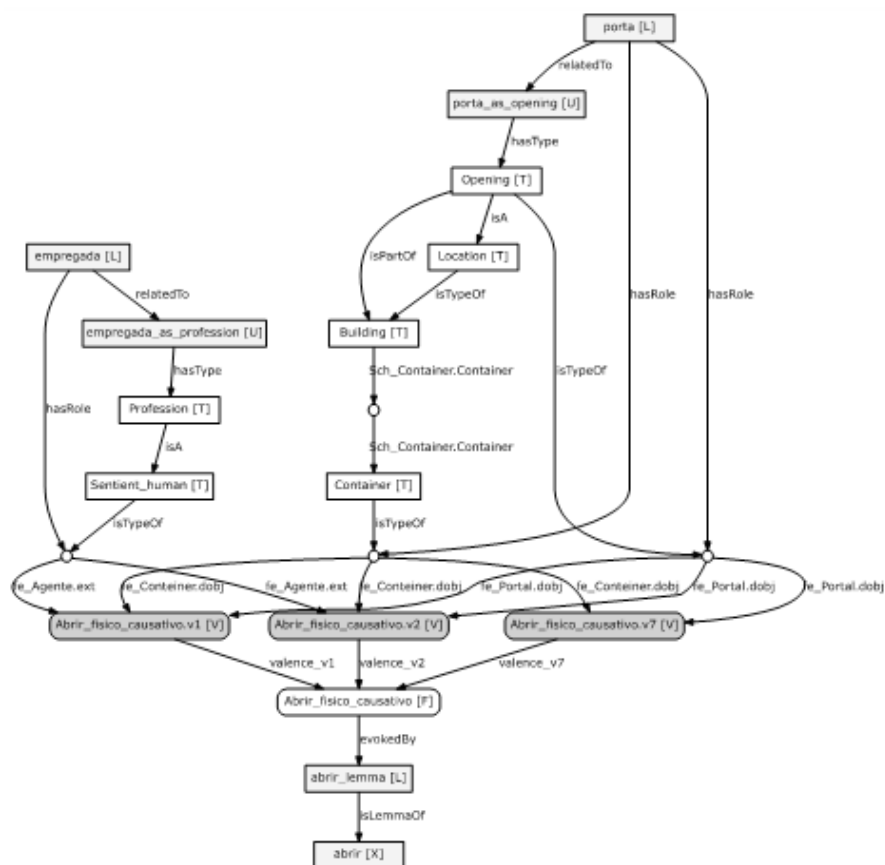


Figura 8 - Abrir\_porta

## 5.5.2. ABRIR\_OLHO

Motivação: Testar o cluster VI, onde partes corporais são usadas metaforicamente relacionadas a estados mentais.

Sentença: (...) *Esse curso abriu meus olhos para o lado da informática* (...)

Segmentos: abrir.target, olho.dobj, curso.ext, informatica.iobj

Comentário: Dois fatos se destacam na inferência apresentada na Figura 9:

- A associação de <curso> (uma Unidade Semântica) com o elemento de *frame* EDUCATION\_TEACHING.COURSE, mostrando a integração entre a FrameNet e a ontologia SIMPLE.
- Adistinção entre os *frames* VI.1.MUDANÇA\_MENTAL\_CAUSATIVO e II.1.MOVIMENTO\_CORPORAL\_CAUSATIVO só foi possível porque, na configuração da rede, as valências cadastradas para II.1.MOVIMENTO\_CORPORAL\_CAUSATIVO obrigam que o elemento de *frame* EXPERENCIADOR seja expresso, exceto quando o Externo é um AGENTE (que neste caso deve ser o próprio EXPERENCIADOR).

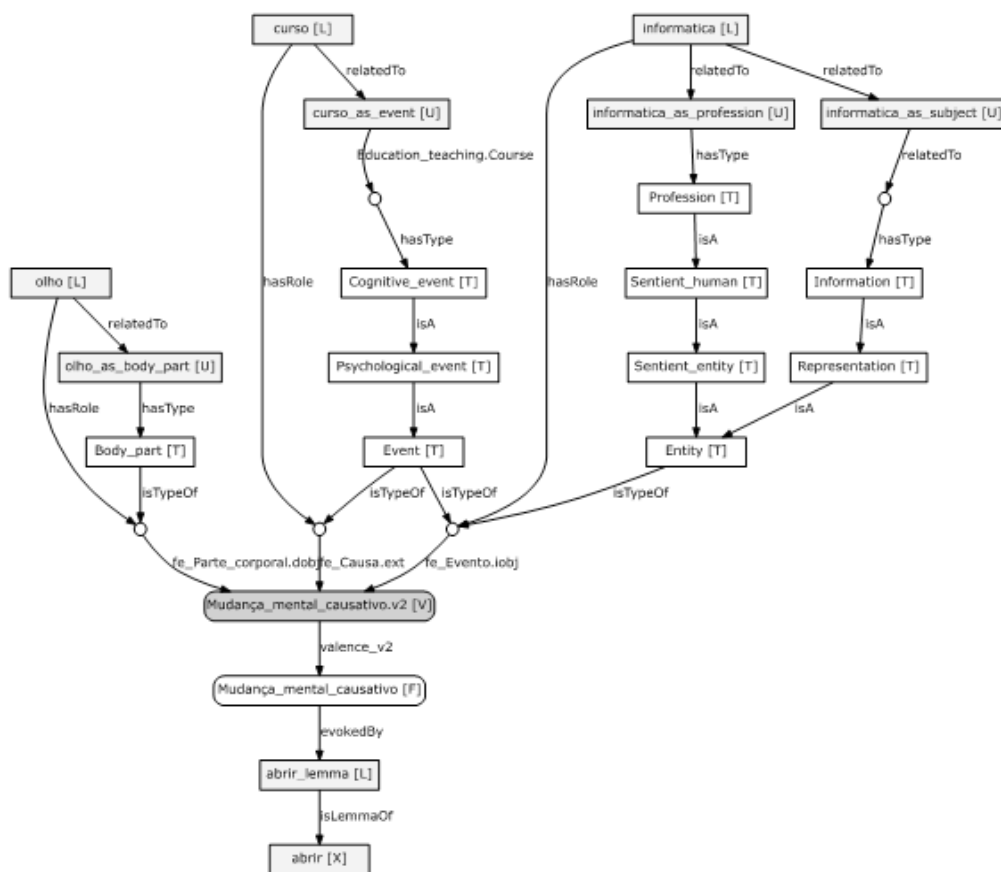


Figura 9 - Abrir\_olho

### 5.5.3. ABRIR\_JORNAL

Motivação: Testar o processo de inferência com entidades polissêmicas. No caso, <jornal> tem tanto o sentido de “informação” quanto de um “artefato”.

Sentença: (...) *Evidentemente está pior, basta a gente abrir o jornal e vê o que?* (...)

Segmentos: abrir.target, jornal.dobj, gente.ext

Comentários: Neste caso o processo de inferência (Figura 10) foi praticamente direto, por dois motivos:

- Existe uma relação qualia entre INFORMATION e SEMIOTIC\_ARTIFACT, estabelecendo que informações são geralmente contidas em artefatos semióticos. Esta relação permite ao processo de inferência reforçar a ativação de <jornal> como um artefato.
- Uma vez que geralmente contém informações, SEMIOTIC\_ARTIFACT é conceptualizado como um contêiner associado com o elemento de *frame* ABRIR\_ESQUEMA.CONTEINER. Este é mais um exemplo da integração entre a ontologia e o recurso lexical.

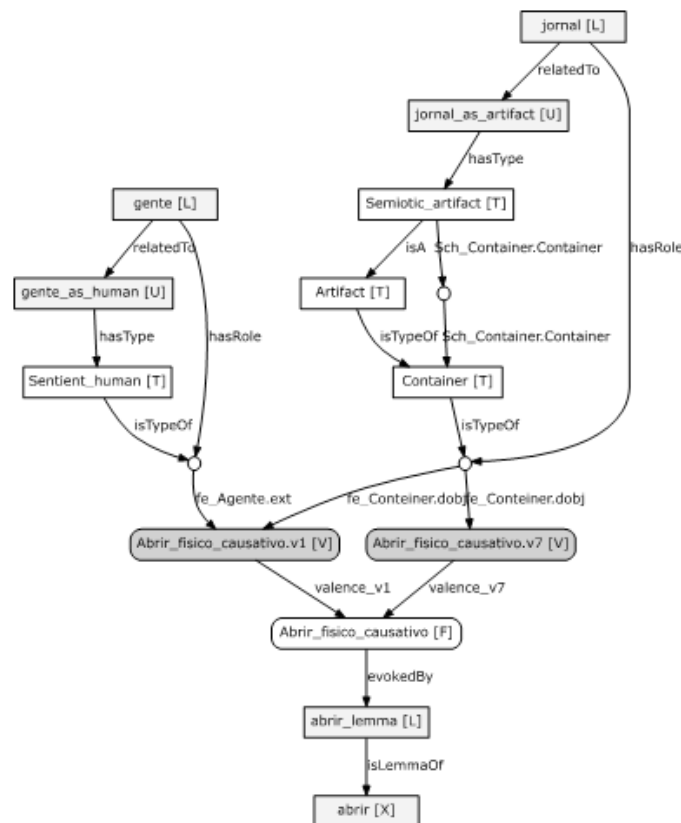


Figura 10 – Abrir\_jornal

#### 5.5.4. ABRIR\_CERVEJA

Motivação: Testar os algoritmos na resolução de uso metonímicos. Neste caso, o container é conceptualizado pelo seu conteúdo.

Sentença: *Estas são as melhores maneiras de abrir uma cerveja.*

Segmentos: abrir.target, cerveja.dobj

Comentários: A Figura 11 mostra que uma interpretação possível para a sentença é a consideração que <cerveja> é subentendida por um container (a garrafa de cerveja, por exemplo). Foi criada uma valência adicional (V8) e um elemento de *frame* CONTEUDO\_AS\_CONTEINER, específicos para esta situação, a fim de deixar mais claro o processo metonímico que está ocorrendo.

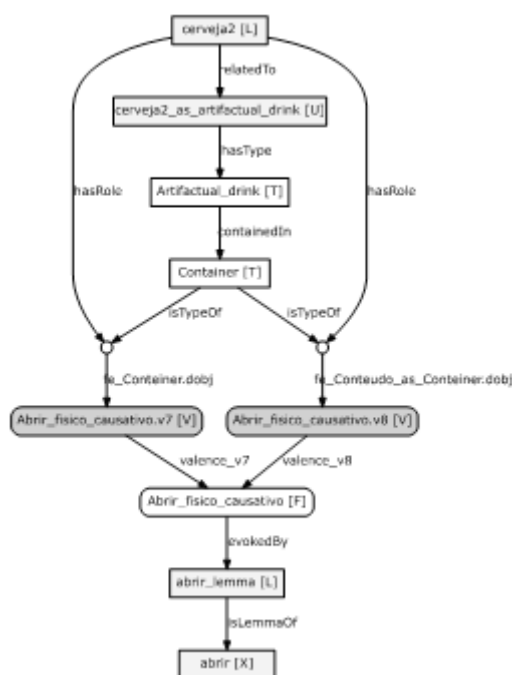


Figura 11 – Abrir\_cerveja

## 6. CONCLUSÃO

O objetivo do trabalho apresentado foi implementar um *framework* para a desambiguação de itens lexicais. O processo de desambiguação pressupõe que dois (ou mais) itens lexicais, que possuem a mesma forma, possuem significados diferentes. Este trabalho lidou, então, com a difícil questão do significado das palavras e da construção/compreensão do significado de sentenças.

O trabalho adotou uma abordagem híbrida, reunindo contribuições da Semântica de Frames (para representação do significado considerando aspectos cognitivos) e da Teoria do Léxico Gerativo (para representação do significado considerando aspectos mais formais). O resultado foi uma aplicação

computacional, chamada Framework LUDI (*Lexical Unit Discovery through Inference*), que pode ser enquadrada na área de Compreensão da Língua Natural.

As contribuições e impactos do trabalho realizado podem ser observados tanto na área da Linguística Cognitiva, quanto na área da Linguística Computacional. Para a Linguística Cognitiva, uma contribuição importante é a demonstração de como pode ser realizado o alinhamento dos papéis temáticos com as relações qualia. A formalização deste alinhamento permite que interpretações de sentenças possam ser geradas dinamicamente, simulando (de maneira ainda bastante limitada, naturalmente) o processo cognitivo humano. Uma vez que a associação entre os elementos de um *frame* e as Entidades que participam do *frame* não está restrita à atribuição de um Tipo Ontológico para o elemento, a construção de uma especificação semântica para o enunciado pode ser feita com base em um número maior de opções.

Para a Linguística Computacional, a contribuição é uma técnica de análise de sentenças baseada em um modelo conexionista estruturado, usando a técnica de Ativação Propagada. Os modelos encontrados na literatura, relativos à desambiguação de *frames*, utilizam predominantemente técnicas baseadas em probabilidade, seja através de Máxima Entropia, seja através de Redes Bayesianas. No framework LUDI usamos um modelo conexionista estruturado, neurologicamente plausível, e um processo de inferência baseado em evidência, não em probabilidade. Consideramos que este tipo de inferência é mais adequado cognitivamente, pelo menos para uma certa classe de situações, em que experiências anteriores possibilitam o estabelecimento de evidências.

A Tabela 4 sintetiza as abordagens escolhidas neste trabalho, em comparação com as escolhas mais comuns em PLN.

PLN	LUDI
Linguística Gerativa	Linguística Cognitiva
Semântica Formal	Semântica de Frames
Forma Lógica	Especificação Semântica
Máquinas de Estados, Gramáticas Livres de Contexto	Modelo conexionista estruturado
Modelos probabilísticos	Modelo Fuzzy

Tabela 4 – Abordagens usadas no framework LUDI

Estas contribuições possuem valor por si mesmas. No entanto, julgamos que a maior contribuição é justamente a demonstração de que conceitos e ideias caros à Linguística Cognitiva podem ser usados em uma implementação computacional. A literatura de viés cognitivista, relacionada aos processos de categorização, conceptualização através de espaços mentais e abordagens gramaticais usando construções, por exemplo, apresenta, via de regra, um sentimento de resistência (quando não de rejeição) a qualquer tipo de formalização. Acreditamos que pesquisas como as realizadas pelo grupo NTL (*Neural Theory of Language*) ou esta, que levou à construção do framework LUDI, mostram que há um caminho viável para uma Linguística Cognitiva Computacional (FELDMAN, 2004).

Este trabalho fornece uma ideia inicial que pode ser expandida em diversas direções. Como exemplo do que pode ser realizado a partir dos resultados alcançados até agora, citamos:

- a) Atualmente as funções gramaticais estão sendo fornecidas como dados de entrada; para uma automatização maior da análise, seria necessário o uso de um *parser* que atribuisse estas funções;
- b) Estamos trabalhando apenas com palavras que evocam Frames (*Frame Evoking Elements*) e Unidades Semânticas associadas a tipos da ontologia SIMPLE; para uma análise completa da sentença, é necessário o cadastro dos *frames* e/ou Esquemas associados às preposições, artigos, quantificadores, etc;
- c) Não foi feita nenhuma abordagem do ponto de vista construcional, mas apenas baseada em valências; o modelo, como está hoje, pode ser facilmente estendido para alguns caso simples de construções;
- d) A ontologia SIMPLE foi utilizada por ser uma ontologia já existente e baseada na TLG; no decorrer do trabalho ficou claro que, para uma aplicação mais genérica, ela precisa ser expandida e possivelmente associada ou alinhada com outras ontologias de topo, como a DOLCE (MASOLO et al., 2003);
- e) Apenas os itens lexicais necessários aos experimentos foram adicionados à ontologia; para uma aplicação mais genérica, seria necessário um léxico mais completo e abrangente.

À guisa de conclusão, acreditamos que a implementação do *framework* LUDI mostra que a Semântica de Frames, enriquecida com a contribuição da Teoria do Léxico Gerativo e o uso de Tipos Ontológicos, possibilita não só a desambiguação lexical, mas se constitui em uma promissora base de pesquisa na área da Compreensão da Língua Natural.

## REFERÊNCIAS

AMARO, R. *Computation of verbal predicates in portuguese: relational network, lexical-conceptual structure and context*. Tese (Doutorado), Universidade de Lisboa, 2009.

BOS, J. *A Survey of Computational Semantics: Representation, Inference and Knowledge in Wide-Coverage Text Understanding*. *Language and Linguistics Compass*, v. 5, n. 6, p. 336–366, jun. 2011. ISSN 1749818X.

BRUGMAN, C.; LAKOFF, G. *Cognitive topology and lexical networks*. In S. Small, G. Cottrill, & M. Tannenhaus (Eds.), *Lexical ambiguity resolution* (pp. 477-508), 1988. Palo Alto: Morgan Kaufman.



CRUSE, A. *Meaning in Language: An Introduction to Semantics and Pragmatics*. 3a. ed. Oxford: Oxford University Press, 2010.

DODGE, E. *Constructional and Conceptual Composition*. Tese (Doutorado), University of California at Berkeley, 2010.

FELDMAN, J. A. *Computational cognitive linguistics* in *Proc. 20th Intl. Conf. on Computational Linguistics (Coling 2004)*, Morristown, NJ: Association for Computational Linguistics, 2004, pp. 1114-1118.

FELDMAN, J. A.; BALLARD, D. *Connectionist models and their properties*. *Cognitive science*, v. 254, n. 1982, p. 205–254, 1982.

FILLMORE, C. J. *The need for a frame semantics in linguistics*. In: KARLGREN, H. (Ed.). *Statistical Methods in Linguistics*. Scriptor, 1977.

FILLMORE, C. J. *Frame semantics*. *Linguistics in the morning calm*, Hanshin Publishing Co., Seoul, South Korea, p. 111–137, 1982.

FILLMORE, C. J. *Frames and the semantics of understanding*. *Quaderni di Semantica*, v. 6, n. 2, p. 222–254, 1985.

FILLMORE, C. J. *Frame Semantics*. In: *Cognitive Linguistics: Basic Readings*. Berlin: Walter de Gruyter, 2006. cap. 10, p. 373–400.

FILLMORE, C. J. *Valence issues in Framenet*. In: GUTZ, T.; VOTTELER, K. (Ed.). *Trends in Linguistics, Studies and Monographs :Valency : Theoretical, Descriptive and Cognitive Issues*. Berlin: Mouton de Gruyter, 2008. p. 141–174.

FILLMORE, C. J.; BAKER, C. *A frames approach to semantic analysis*. *The Oxford Handbook of Linguistic Analysis*, n. 1, p. 313–340, 2010.

FILLMORE, C.; JOHNSON, C. R.; PETRUCK, M. R. *Background to Framenet*. *International Journal of Lexicography*, v. 16, n. 3, p. 235–250, 2003.

GÄRDENFORS, P. *Conceptual spaces - the geometry of thought*. MIT Press, 2000. I–X, 1–307 p. ISBN 978-0-262-07199-4.

GAWRON, J. M. *Frame Semantics*. In: MAIENBORN, C.; HEUSINGER, K. von; PORTNER, P. (Ed.). *Semantics: An International Handbook of Natural Language Meaning*. Berlin: Mouton de Gruyter, 2008, (HSK Handbooks of Linguistics and Communication Science Series, v. 23). p. 1–35.

- GEERAERTS, D. *Theories of Lexical Semantics*. Oxford: Oxford University Press, 2009. 368 p. ISBN 9780198700302.
- GRIES, S. *Corpus-based methods and cognitive semantics: The many senses of to run*. In *Linguistics Studies and Monographs*, 2006.
- LAKOFF, G. *Women, Fire, and Dangerous Things*. Chicago: University of Chicago Press, 1987. ISBN 9780226468044.
- LANGACKER, R. W. *Foundations of Cognitive Grammar, Vol. 1*. Stanford: Stanford University Press, 1987.
- LENCI, A. et al. *SIMPLE Work Package 2 - Linguistic Specifications - Deliverable D2.1*, 2000.
- LYONS, J. *Introdução à Linguística teórica*. São Paulo: Companhia Editorial Nacional, 1979.
- MASOLO, C. et al. *Wonderweb deliverabled18, ontology library (final)*.2003.
- MATOS, E. E. *LUDI: Um framework para desambiguação lexical com base no enriquecimento da Semântica de Frames*. 200 p. Tese (Doutorado), Universidade Federal de Juiz de Fora, 2014.
- PETRUCK, M. R. L. *Frame Semantics*. In: VERSCHUEREN, J. et al. (Ed.). *Handbook of Pragmatics*. John Benjamins, 1996.
- PUSTEJOVSKY, J. *The Generative Lexicon*. *Computational Linguistics*, The MIT Press, Cambridge (USA), v. 17, n. 4, p. 409–441, 1991.
- PUSTEJOVSKY, J. *The Generative Lexicon*. Cambridge, MA: MIT Press, 1995.
- RUPPENHOFER, J. et al. *FrameNet II: Extended Theory and Practice*. Berkeley, California: International Computer Science Institute, 2010.
- SALOMÃO, M. M. *FrameNet Brasil: um trabalho em progresso*. *Calidoscópico*, v. 7, n. 3, p. 171–182, dez. 2009. ISSN 16798740.
- SHASTRI, L.; AJJANAGADDE, V. *A connectionist system for rule based reasoning with multi-place predicates and variables*. 1989.
- SOWA, J. *Semantic Networks*. 2015. Disponível em: <http://www.jfsowa.com/pubs/semnet.htm>.

TALMY, L. *Toward a Cognitive Semantics, Vol 1: Concept Structuring Systems*. MIT Press, 2000.

TORAL, A.; MONACHINI, M. *SIMPLE-OWL: a generative lexicon ontology for nlp and the semantic web*. In: Workshop of Cooperative Construction of Linguistic Knowledge Bases, 2007.

TORRENT, T. et al. *Copa 2014 FrameNetBrasil: a frame-based trilingual electronic dictionary for the Football World Cup*. In: COLING (Demos). 2014. p. 10–14.

TORRENT, T. T.; ELLSWORTH, M. *Behind the Labels: Criteria for defining analytical categories in FrameNetBrasil*. Revista Veredas, v. 17, n. 1, p. 44–65, 2013.

ZIEM, A. *Frames of Understanding in Text and Discourse: Theoretical foundations and descriptive applications*. John Benjamins Publishing Company, 2014. (Human Cognitive Processing). ISBN 9789027269645.

**Recebido em 04/04/2016**

**Aceito em 15/05/2016**