

**Aplicação de Redes Bayesianas e de tipos ontológicos a partir de um lexema para  
estimar a pertinência a um Frame Semântico**

Alexandra Moreira (UFJF)

Maria Margarida Martins Salomão (FrameNet Brasil / UFJF)

Alcione de Paiva Oliveira (UFV)

**RESUMO:** O trabalho apresentado neste artigo está inserido em uma pesquisa mais ampla que busca definir um framework para inserir o uso de tipos ontológicos associados aos lexemas de enunciados de modo a auxiliar na detecção da ocorrência e no estabelecimento das fronteiras de um frame, buscando a elaboração de uma FrameNet com maior rigor metodológico. Neste artigo é apresentada, por meio de um estudo de caso, a aplicação de ferramentas de mineração de dados e inferência probabilística que, fazendo uso apenas de tipos ontológicos, demonstram ser possível estimar o grau de pertinência de uma sentença em um determinado frame. A execução da técnica sobre o conjunto de teste obteve uma taxa de acerto de 85,7%, o que indica que a técnica possui um bom potencial de utilização.

**PALAVRAS-CHAVE:** Semântica de Frames; FrameNet; Ontologia; Redes Bayesianas.

**ABSTRACT:** The work presented in this paper is part of a broader research project that seeks to define a framework to incorporate the use of ontological types associated with lexemes so as to assist in detecting the presence and establishing the boundaries of a frame, seeking the development of a FrameNet with higher methodological rigorousness. This article presents, through a case study, the application of data mining tools and probabilistic, which, using only the ontological types, point to the possibility of estimating the degree of relevance of a sentence in a particular frame. The implementation of the technique on the test set achieved a hit rate of 85.7%, which indicates that the technique has a good potential for use.

**KEYWORDS:** Frame Semantics; FrameNet Ontology; Bayesian Networks.

## **Introdução**

O significado de um lexema em um enunciado só pode ser estabelecido a partir do conhecimento da cena a que o enunciado está associado. Esta cena foi chamada de *Frame* por Charles Fillmore (1977), em sua teoria do significado lexical, denominada de *Semântica de Frames*. A Semântica de Frames inspirou a criação de uma base léxico-semântica, a FrameNet (BAKER; FILLMORE; LOWE, 1998). A FrameNet procura estabelecer o significado de um elemento lexical por meio da definição dos elementos que compõem o frame em que ele ocorre e por meio do posicionamento do frame em uma rede de relações semânticas com outros frames. A FrameNet foi desenvolvida para a língua inglesa pelo *International Computer Science Institute* em Berkeley, mas, atualmente, várias instituições estão criando uma “FrameNet” associada a outras línguas, como é o caso da Universidade Federal de Juiz de Fora, que está desenvolvendo uma FrameNet para o Português Brasileiro (SALOMÃO, 2009).

As ontologias desenvolvidas no âmbito da Ciência da Computação podem ser compreendidas como estruturas conceituais que descrevem os objetos de um domínio ou as tarefas executadas sobre os objetos dentro de um domínio. Estudos que unem estes sistemas de representação conceitual (frames e ontologias) podem ser observados tanto na Linguística como na Ciência da Computação (SCHEFFCZYK; PEASE; ELLSWORTH, 2006; CHOW;

WEBSTER, 2007; BURCHARDT; PENNACCHIOTTI, 2008; OVCHINNIKOVA ET AL., 2010). O objetivo global desses trabalhos é unir as vantagens de cada método de representação e minimizar suas deficiências para aumentar as potencialidades de sua aplicação.

Este artigo também aborda o uso em conjunto desses dois métodos de representação: a FrameNet e as ontologias da Ciência da Computação, no entanto, com motivações e objetivos distintos das pesquisas anteriores. A primeira motivação é a percepção de que o aspecto ontológico não foi tratado completamente na FrameNet. A Semântica de Frames, como uma teoria de conceptualização, trata de todos os aspectos inerentes a um conceito. É uma teoria sobre as estruturas conceituais que ocorrem na mente humana e, como tal, abarca todos os aspectos de um conceito, inclusive sobre sua natureza ontológica, não ocorrendo distinção entre níveis e separações que se apresentam como um todo indissociável (*gestalt*). No entanto, quando é criada uma linguagem de representação para expressar essa teoria de conceptualização, como é o caso da FrameNet em relação à Semântica de Frames, não é possível estabelecer uma relação isomórfica em função da própria natureza do objeto teorizado: o conceito. A FrameNet é uma linguagem de representação e, como qualquer sistema de símbolos, não é capaz de capturar todos os aspectos de um conceito. Sendo assim, dependendo do propósito da representação, ela deve ser enriquecida para cumprir adequadamente sua tarefa. Acreditamos que o aspecto ontológico deva ser abordado de forma mais explícita na FrameNet, uma vez que se trata de um aspecto importante para a compreensão de sentenças em linguagem natural. Muitas sentenças polissêmicas, que ocorrem constantemente nas conversações, são resolvidas pela compreensão da natureza dos objetos envolvidos no domínio de discussão. Resumindo, a primeira motivação deste trabalho é resgatar o papel das estruturas ontológicas na compreensão do significado.

Outra motivação partiu da observação de que a rede de frames expressa na FrameNet apresenta algumas falhas e inconsistências. Algumas dessas falhas e inconsistências já haviam sido observadas por alguns pesquisadores, tais como Ovchinnikova *et al.* (2010). Observamos também algumas inconsistências na cadeia hierárquica iniciada pelo frame MOTION. Estas falhas serão discutidas com mais detalhes mais adiante. Acreditamos que estas falhas possam ser evitadas, ou pelo menos minimizadas, com a inclusão de etapas que envolvam a análise ontológica na metodologia de criação de frames na FrameNet.

O trabalho apresentado neste artigo está inserido em uma pesquisa mais ampla que busca definir um framework para inserir o uso de tipos ontológicos associados aos lexemas de enunciados para auxiliar na detecção da ocorrência e para o estabelecimento das fronteiras de um frame, buscando a elaboração de uma FrameNet com maior rigor metodológico. Para este fim, foi proposta uma alteração nos passos metodológicos de desenvolvimento da FrameNet, uma formalização dos elementos da FrameNet e uma aplicação de ferramentas de mineração de dados e inferência probabilística para calcular o grau de pertinência de uma sentença a um determinado frame. A formalização dos elementos da FrameNet e a alteração dos passos metodológicos e sua aplicação foram descritos por Moreira (2012) e Moreira e Salomão (2012). Este artigo apresenta a aplicação de ferramentas de mineração de dados e inferência probabilística que, por meio de um estudo de caso, mostra que, fazendo uso apenas de tipos ontológicos, é possível estimar o grau de pertinência de uma sentença em um determinado frame. Para a realização da inferência probabilística, foi utilizado o formalismo denominado de redes bayesianas, cujo propósito é definir a probabilidade da ocorrência de uma determinada hipótese a partir de um conjunto de evidências.

## 1. A Semântica de Frames

A Semântica de Frames surgiu como resposta à incapacidade da semântica tradicional de dar subsídios para interpretações de diversas situações, tais como explicar porque não é adequado caracterizar o Papa como um homem solteiro (KATZ; FODOR, 1964). Para entender o conceito evocado pela unidade lexical *solteiro*, 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 solteiro/casado e o que é estar apto a exercer essas funções. Só assim é possível aplicar adequadamente item lexical “*solteiro*” a alguém. Existem vários itens lexicais como esses numa língua natural, itens lexicais cujo significado só pode ser compreendido pelo entendimento de todos os conceitos envolvidos (*gestalt*) e não por sua análise isolada.

A Semântica de Frames opera com 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 TERRORISM, possui como papéis temáticos centrais o *Terrorista, o Ato e a Vítima*, enquanto o enfoque anterior postularia os papéis de Agente, Instrumento e Paciente

## 2. A Rede Semântica da FrameNet

A FrameNet é uma base lexical semântica fundada na Semântica de Frames e apoiada por evidências obtidas em *corpora*. A iniciativa para o desenvolvimento de tal recurso coube ao *International Computer Science Institute* em Berkeley, sob a coordenação de Collin F. Baker, Charles J. Fillmore e John B. Lowe. (BAKER; FILLMORE; LOWE, 1998). O projeto tem por objetivo registrar as possibilidades combinatórias semânticas e sintáticas (valências) de cada palavra predicativa (nomes, adjetivos e verbos), em cada um dos seus sentidos. Os conceitos básicos que fundamentam o projeto FrameNet são os conceitos de frames, unidades lexicais (UL), elementos de frames (EF) e relações entre frames.

Uma unidade lexical (UL) é o pareamento de uma palavra com um significado <*palavra, significado*> (RUPPENHOFER ET AL., 2006, p.5). Ainda segundo os mesmos autores, cada sentido de uma palavra polissêmica pertence a um frame semântico diferente. As ULs evocam os frames. Por exemplo, a ocorrência da palavra “*buy*” em uma sentença evoca o evento de uma compra comercial, capturado pelo frame COMMERCE\_BUY.

Elementos de frame (EF) são papéis que ocorrem em um determinado frame. Por exemplo, o frame COMMERCE\_BUY descreve situações comuns envolvendo papéis, tais como *buyer (comprador), goods (bens), seller (vendedor), place (lugar) e money (dinheiro)*. Ao apresentar um determinado frame, o sistema exibe uma definição e uma lista de elementos de frames, sendo que, para cada EF, é apresentado um conjunto de sentenças anotadas, extraídas de um *corpus*. Alguns elementos de frames são associados a determinados tipos semânticos, que atribuem restrições às instâncias dos elementos de frames, de modo que todas devam ser consideradas instâncias dos tipos semânticos. A FrameNet não oferece nenhuma formalização para os tipos semânticos, nem discrimina o conjunto completo de tipos que possam ser atribuídos a todo elemento de frame.

Os frames são interligados, formando um sistema de frames. Eles se ligam por meio de relações semânticas, tais como *herança, uso, subframe e perspectiva sobre*. Detalhes sobre as relações semânticas podem ser encontrados em Ruppenhofer *et al.*(2006).

### 3. Ontologias

A palavra *ontologia* vem do grego *ontos* (ser) e *logos* (palavra). Apesar do estudo do ser ter suas raízes em Aristóteles e Platão, o uso do termo ontologia para designar este ramo da filosofia é muito mais recente, tendo sido introduzido entre os séculos XVII e XVIII por filósofos alemães. O termo “ontologia” começou a ser empregado na Ciência da Computação, dentro da subárea denominada de Inteligência Artificial (IA), segundo Moreira *et al.* (2004), no início dos anos 90, em projetos para organização de grandes bases de conhecimento, como CYC (LENAT; GUHA, 1990) e Ontolingua (GRUBER, 1993).

A definição mais aceita dentro da comunidade de representação de conhecimento e que estamos adotando é de Gruber (1993), que diz que uma ontologia é a “*especificação formal e explícita de uma conceptualização compartilhada*”, em que *formal* significa que a ontologia pode ser expressa em uma linguagem formal, *explícita* significa que é um objeto de nível simbólico e *compartilhada* indica que o conhecimento é aceito dentro da comunidade. Já uma *conceptualização* é uma visão abstrata do mundo que se deseja representar para algum propósito (GRUBER, 1993, p.1). A ontologia na Ciência da Computação, de acordo com Guarino (1998), tem o papel de limitar as interpretações de uma sentença. Por exemplo, quando alguém diz “o servidor caiu”, existem várias interpretações possíveis, sendo que duas são bastante frequentes: 1) o servidor é um trabalhador, geralmente ligado ao setor público, e ele caiu no chão; ou 2) o servidor é um computador usado para disponibilizar serviços e ele parou de funcionar. No primeiro caso, “servidor” é um papel social exercido por uma pessoa, e no segundo caso, “servidor” é uma função tecnológica exercida por uma máquina. A natureza do evento “caiu” também é distinta nos dois casos. Ao definir a natureza dos conceitos envolvidos no enunciado, estaremos também limitando as interpretações possíveis. Note que a limitação não impõe uma interpretação exata, uma vez que isso não é possível no caso geral.

#### 3.1. A ontologia SIMPLE-CLIPS

Nesta pesquisa, a ontologia tem a função de fornecer tipos ontológicos para a anotação de itens lexicais de enunciados. 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 onde está inserido o item. Por exemplo, um objeto, como um carro, pode ser visto como um objeto concreto, como um artefato, um objeto estruturado em suas partes ou 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. Esse é o caso da Teoria do Léxico Gerativo de Pustejovsky (1995). Na Teoria do Léxico Gerativo (TLG), os sentidos de um item lexical não são enumerados e tratados separadamente. 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. 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)

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*. Por exemplo, seja o nominal “livro”. O *quale* TÉLICO registra a função do livro (ser lido) e o *quale* AGENTIVO registra que sua criação envolveu um evento de escrever. Finalmente, a estrutura de herança lexical, captura como a estrutura lexical está relacionada a outras, em um reticulado de tipos.

Nesta pesquisa foi utilizada 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 foi a SIMPLE-CLIPS<sup>1</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 se baseia na estrutura *qualia* e consiste de 157 tipos semânticos organizados por meio de relações conceituais hierárquicas e não hierárquicas.

#### 4. Passos realizados para o experimento

Para evidenciar o papel do nível ontológico no entendimento de um enunciado, executamos um experimento que envolveu as seguintes etapas:

1) Seleção de um domínio para análise. Foi selecionado o domínio de **viagem** por ser um domínio com facilidade para encontrar enunciados em *corpora* e por ser um domínio que possui diversas variações de cenários: viagens de turismo, viagem de negócios, deslocamentos longos individuais e em grupos, deslocamentos de animais, deslocamentos com objetos, deslocamentos com auxílio de veículos ou por meios próprios, e evocações metafóricas de experiências (vida, amor, trajetória profissional, etc.).

2) A escolha de uma ontologia abrangente o suficiente para abordar tanto categorias de nível topo como categorias com um detalhamento suficiente para descrever os tipos dos lexemas empregados nos enunciados tratados. Foi adotada a ontologia SIMPLE-CLIPS (LENCI ET AL. 2000) que possui a característica de se basear na Teoria do Léxico Gerativo de Pustejovsky e, em função disso, analisar mais de um aspecto ontológico do lexema.

3) Aplicar uma técnica automática para avaliar a pertinência dos enunciados ao frame. Utilizamos as Redes Bayesianas para essa automação em função de sua adequação para estimar probabilidades com base em relações de dependência e também estar alinhada à hipótese de que o procedimento estatístico de melhor enquadramento (*best fit*) é um componente importante da cognição humana (FELDMAN, 2006, p.67). As Redes Bayesianas foram treinadas por meio de um *subcorpus* de 57 sentenças previamente selecionadas.

As evidências linguísticas apresentadas neste trabalho foram obtidas do *Corpus* do Português (DAVIES; FERREIRA, 2006-) de livre acesso do portal da BRIGHAM YOUNG UNIVERSITY (<http://corpus.byu.edu>). Esse *corpus* possui em torno de 45 milhões de palavras obtidas de textos que vão desde o século XIV até o século XX.

<sup>1</sup> <http://www.ilc.cnr.it/clips/Ontology.htm>.

A escolha deste *corpus* deveu-se a grande quantidade de sentenças associadas ao domínio em estudo: o domínio de **viagem**. Restringindo a seleção ao século XX e ao Português Brasileiro, o *corpus* apresenta 2375 sentenças com lexema *viagem* e 1587 sentenças com o lexema *viajar*.

Após a extração das sentenças, foi realizado um processo manual de filtragem para eliminação de sentenças de difícil entendimento, ou que a ocorrência do lexema não estava relacionada com um deslocamento físico ou metafórico, como é o caso de nomes próprios, como em “Festa de Nossa Senhora de Boa Viagem”. Após essa filtragem, restaram 2133 sentenças. Deste *corpus* foram selecionadas sentenças que caracterizassem o uso polissêmico dos lexemas para ilustrar o benefício da aplicação da metodologia proposta neste trabalho para distinção destes significados. Este *corpus* serviu para produzir um *subcorpus* com 57 sentenças que apoiou a criação de uma ontologia de domínio. Com este *subcorpus* foi possível o treinamento e o teste das redes bayesianas. O critério para a seleção destas sentenças foi a variação do uso no sentido típico, correlato e metafórico.

#### 4.1 A cena de *Viagem* na FrameNet

A cena típica que se deseja descrever é a cena de **viagem** realizada por uma entidade consciente ou grupo de entidades a uma determinada localidade, por meios próprios ou por um meio de transporte e com algum propósito. Lexemas candidatos a unidades léxicas nessa cena são: nominais - *viagem, excursão, safari, peregrinação, expedição*; verbos - *viajar, excursionar, peregrinar, visitar*. No entanto, em função das limitações de espaço, restringiremos nossa análise ao lexema *viagem* e ao lexema *viajar*. Analisaremos então, os enunciados onde ocorrem esses elementos, sob o ponto de vista ontológico. As sentenças selecionadas contêm o nominal *viagem* e as seguintes flexões do verbo *viajar*: *viajar, viajou, viaja, viajado, viajava, viajam, viajavam, viajado, viajei, viajaram e viajo*.

A FrameNet do português brasileiro (FrameNetBr)<sup>2</sup> (SALOMÃO, 2009) possui uma entrada para o frame VIAGEM com a seguinte definição:

*Neste frame um Viajante sai em uma viagem, uma atividade, geralmente planejada com antecedência. Nesta viagem o Viajante se desloca de um local Fonte para um Alvo por um Caminho ou dentro de uma Área. A viagem pode ser acompanhada por Coparticipantes e por Bagagem. A Duração ou a Distância da viagem, de modo geral longas, também podem ser descritas, assim como o Meio\_de\_Transporte. Palavras neste frame enfatizam todo o processo de deslocamento de um lugar para outro, ao invés de perfilar apenas o início ou o fim da viagem.*

Porém, diferentemente do frame TRAVEL, registrado na FrameNet de Berkeley, apenas o verbo **viajar** é associado como unidade lexical do frame. A razão disso é que, como a criação de um frame exige atestação em *corpus*, a estrutura do frame varia de língua para língua. Em uma análise simples, pode-se notar que alguns lexemas candidatos evocam cenas distintas da cena de viagem típica, como é o caso do lexema *peregrinar*. Neste caso, é um deslocamento associado a sofrimento e sacrifício, sendo o deslocamento um fim em si mesmo e não um meio para chegar a algum destino.

Apesar da FrameNetBr não atribuir a unidade lexical **viagem** ao frame VIAGEM, ela foi utilizada nesta pesquisa, uma vez que foram encontradas várias evidências de seu uso no *corpus* criado para o estudo de caso.

<sup>2</sup> <http://www.framenetbr.ufjf.br>

A partir da definição do frame VIAGEM, é possível notar que se trata de um deslocamento geralmente longo, tanto em distância quanto em duração. Deslocamentos curtos, em geral, não podem ser considerados como uma viagem. Mas como alguém caracteriza um deslocamento como uma viagem? Outro elemento que parece ser essencial é o deslocamento espacial. Ontologicamente é um processo, em que todas as etapas são percebidas.

Na FrameNet de Berkeley o frame TRAVEL herda diretamente do frame SELF-MOTION (movimento próprio), que, por sua vez, herda do frame MOTION (movimento) e do frame INTENTIONALLY\_ACT (ato intencional), sendo esse último um frame abstrato. Essa hierarquia é muito 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 sobre o próprio poder, mas, em vários exemplos de viagem na FrameNet de Berkeley, 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, não se pode dizer que a **Viagem** é um movimento próprio. Recentemente, o frame SELF-MOTION foi redefinido na FrameNet de Berkeley para acomodar a possibilidade de uso de veículos. No entanto, outros questionamentos surgem com essa redefinição. Se um veículo é permitido e se o frame SELF\_MOTION herda do frame INTENTIONALLY\_ACT, uma pergunta é necessária: uma pessoa ou animal que é transportado em um veículo, mas não realiza um ato intencional (não se desloca por sua vontade), está viajando? Essas dúvidas surgem justamente da falta de formalização da FrameNet.

## 4.2 Procedimento de anotação

Para realizar o enquadramento de uma sentença em uma cena, ou seja, para sua contextualização, é necessário realizar um procedimento de anotação nos elementos da sentença. Para os objetivos deste estudo havia a necessidade de se realizar uma anotação indicando os tipos ontológicos, de acordo com uma ontologia de domínio. Como o foco da pesquisa não era a criação de um procedimento automático de anotação, partes do processo foram feitas manualmente, notadamente, a criação da ontologia de domínio.

Após a seleção das sentenças, foi gerada uma lista de palavras com o auxílio da biblioteca de programas NLTK (BIRD; KLEIN; LOPER, 2009) voltada para o processamento de linguagem natural. Em seguida, a lista de palavras foi anotada com os tipos ontológicos presentes na ontologia de nível topo SIMPLE-CLIPS (LENCI ET AL., 2000), gerando uma ontologia de domínio. A ontologia de domínio foi complementada com uma lista de nomes de pessoas e de lugares. É importante destacar que o objetivo desta ontologia de domínio é gerar argumentos em favor das alterações metodológicas sugeridas neste trabalho. Portanto, não se trata de uma ontologia completa sobre um domínio e que sirva para ser aplicada em outros sistemas.

O *subcorpus* foi analisado sintaticamente pelo analisador MXPOST (RATNAPARKHI, 1996). A anotação sintática ajudou a guiar a anotação ontológica. Assim, se um determinado item não for reconhecido pelo anotador ontológico, em alguns casos seu tipo pode ser inferido pela sua classe. Por exemplo, pronomes na primeira pessoa podem indicar um humano. Posteriormente, as sentenças foram anotadas ontologicamente por meio de um programa desenvolvido com esse propósito. A anotação ontológica foi realizada a partir da ontologia de domínio previamente construída.

Após a anotação ontológica, o programa emitiu um arquivo na forma de tabela com indicação da ocorrência (denotada pela letra V) ou ausência (denotada pela letra F) do tipo ontológico em cada sentença. Este arquivo foi usado como entrada para o treinamento e teste

do módulo de rede bayesiana do sistema Weka<sup>3</sup> de mineração de dados. Detalhes sobre redes bayesianas serão apresentados na próxima seção. Antes de submeter este arquivo ao módulo do Weka, foi necessário registrar manualmente no arquivo qual sentença pertenceria a uma cena prototípica, ou quasi-prototípica, ou metafórica. Isso se faz necessário para que a rede possa treinar e verificar seu *score* de acerto no conjunto de teste.

### 4.3 Distinção ontológica por meio de Redes Bayesianas

Redes Bayesianas é a denominação de um conjunto de formalismos para a representação e inferência do conhecimento probabilístico. Esse conjunto de formalismos tem por base o teorema de Bayes, proposto pelo matemático e teólogo inglês Thomas Bayes, que viveu de 1702 a 1761 (COPPIN, 2010, p. 284). O teorema de Bayes procura estabelecer a probabilidade de uma hipótese  $H$ , dada uma ou mais evidências  $E$ , representada pela notação  $P(H|E)$ . Na maioria dos casos, há mais de uma evidência para a hipótese. Uma pessoa pode ter outros sintomas, além da febre, para corroborar com a hipótese de gripe. Nesse caso é preciso combinar as probabilidades. Por exemplo, dada uma evidência prévia  $E$  e uma nova evidência  $e$ , a conjunção das probabilidades é dada por:

$$P(H|E,e) = P(H|E) \cdot P(e|E,H)/P(e|E) \quad (1)$$

Pode-se observar que, se o número de evidências e hipóteses for muito grande, o número de probabilidades a serem fornecidas também será muito grande e pode tornar o uso do teorema de Bayes inviável. Para contornar esse problema, surgiram as *Redes Bayesianas*. Essas redes formam um grafo direcionado acíclico (DAG), onde o arco direcionado estabelece uma dependência probabilística entre os nós. Assim, apenas as probabilidades entre os nós relacionados precisam ser fornecidas, limitando o número de probabilidades que alimentam o sistema. As redes bayesianas têm sido aplicadas com sucesso na mineração de dados, classificação de documentos, em estudos de bioinformática, recuperação de informação etc. Meurs *et al.* (2009) apresentaram um sistema que derivava frames semânticos de *corpus* de fala por meio de redes bayesianas. Palmer, Alishahi e Sporleder (2011) apresentam um anotador de papéis semânticos baseados em redes bayesianas. Narayanan e Jurafsky (1998) utilizaram modelos bayesianos para simular o processamento de sentenças humanas.

Neste trabalho de pesquisa, em que se propõe um aperfeiçoamento na metodologia da elaboração da FrameNet, as redes bayesianas foram utilizadas para verificar o grau de pertinência de um enunciado em um frame, seguindo o que foi estabelecido na formalização dos conceitos e relacionamentos da FrameNet. A ideia básica é relacionar a probabilidade de um enunciado estar enquadrado em uma cena a partir dos tipos ontológicos que ocorrem no enunciado, mimetizando, em certo grau, a interpretação de melhor enquadramento (*best fit*) (FELDMAN, 2006, p.67) que o cérebro humano realiza quando tenta estabelecer o significado de uma sentença polissêmica. A nossa proposição é que os tipos ontológicos é que terão um papel fundamental na diferenciação do significado.

Agora mostraremos como esta técnica foi efetivamente aplicada no trabalho. O objetivo é usar a técnica para verificar o grau de pertinência de um enunciado em um frame baseando-se nos tipos ontológicos dos elementos de frames.

As redes bayesianas foram aplicadas sobre a saída produzida pelo programa de anotação. O *subcorpus* utilizado nos testes, com 57 sentenças, está apresentado na Figura 1.

<sup>3</sup> <http://www.cs.waikato.ac.nz/ml/weka>

1	Seu pensamento viajava por caminhos sem fim
2	As duas vítimas estavam viajando na caminhonete Saveiro
3	Fujimori viajou domingo pela manha para a Costa Rica
4	Che Guevara passou por Temuco na viagem que fez de moto pela América do Sul
5	Minha última viagem a Montevideú foi para ver Julián
6	A viúva vai viajar com os filhos
7	Como secretário do general Saint-Clair, viajou pela Europa em missão diplomática
8	por ser uma viagem mais rápida eu já fui até a Europa em navio
9	Naquela madrugada estavam de viagem marcada para a fazenda do Coronel Raimundo de Natuba.
10	A viagem até Ural forneceu-lhe material para a obra Doctor Zhivago.
11	A ave viajara espremida num cesto de vime
12	imaginava os porcos-voadores na sua viagem para o infinito a percorrerem signos e constelações com as suas asas dentadas.
13	O historiador também viajou muito pelo sul da Itália
14	As sondas espaciais viajam até aos planetas em órbitas de transferência.
15	Viajavam nele Paulo Turkiewicz e Issei Akiki, além de um terceiro passageiro ainda não identificado.
16	O presidente Itamar Franco viajou ontem a Juiz de Fora
17	eu viajei em todo tipo de avião desde o teco-teco
18	O presidente Antonio Pithon viaja hoje para São Paulo a fim de participar de reunião do Clube dos 13
19	o delegado e o prefeito Hermes Bonfim Nascimento viajaram para Salvador
20	cometas que viajam próximos àquela região
21	A bagagem viaja em carroças puxadas por juntas de bois
22	Sigmund Freud está viajando pra cá
23	os noivos partiram para uma longa viagem que devia durar três meses
24	Fez uma viagem à Europa, para estudar o mecanismo financeiro dos países do Velho Mundo.
25	Cassini e Huygens serão lançadas por um foguete Tita IVB/Centauro para uma viagem de 3,5 bilhões de quilômetros
26	A nossa viagem demorará cerca de hora e meia
27	Eu viajo para comprar coisas.
28	com semente de jaca e do feijão camaratu, que viajaram nas naves espaciais Columbia e Atlantis.
29	Viajei até Arajá, em Minas Gerais, com o Cadilac sem nenhum problema.
30	Inicialmente a espaçonave viajou com o 'nariz' apontado para a Lua
31	A análise, feita através das ondas sonoras produzidas pelo Sol, revelou que estas viajavam mais devagar na superfície
32	já seus olhos viajavam nessa casa velha
33	as formigas que viajavam sob aquele toldo tão suave
34	Nos primeiros tempos que se lhe vinham notícias da corte, essas viajavam com truncadas informações.
35	No ano passado, dois executivos japoneses já sentiram a sensação de viajar no espaço
36	Viajei ontem ao lado de um camelô
37	O estudante estava gostando de viajar e, como dizia, procurava saborear o tempero do sul
38	Para elas, essas relíquias são uma verdadeira viagem no tempo.
39	Um baseado leve, bom para sonhar e viajar, correr mundos e projetar fantasias além da imaginação.
40	Soldados viajaram de uniforme
41	os policiais viajaram para Itajaí
42	Pássaros negros viajavam sobre os sítios das emboscadas
43	os planetas, incluindo a Terra, viajam com órbitas circulares em torno do Sol
44	o radar emite ondas electromagnéticas que viajam a velocidade da luz
45	Yuri Gagarin torna-se o primeiro homem a viajar no espaço
46	Oscar nem viajou para o jogo de sexta-feira, em Santa Cruz do Sul
47	Na antiguidade, acreditava-se que a luz viajava a uma velocidade infinita
48	As duas sondas percorrerão uma trajetória particularmente complicada em sua viagem de quase sete anos
49	A primeira experiência brasileira a viajar na nave espacial Columbia na semana passada
50	albatrozes são aves migratórias e que ... podem viajar mais de cento e trinta milhas náuticas por dia.
51	Os fótons são partículas elementares que viajam com a velocidade da luz
52	crianças da minha idade, na Inglaterra, já viajavam com os pais em trem elétrico, de Londres para Liverpool
53	Viajou hoje no bonde um homem embriagado, meio dormindo
54	A imaginação de Verne viajou por todos os quadrantes do planeta
55	ela fez uma viagem de volta ao fim do século passado
56	Um arrepio viajou-lhe pelo corpo
57	Elesbão viajava de ônibus para São Paulo

Figura 1: *Subcorpus* utilizado na anotação e para o treinamento e testes das redes bayesianas.

As sentenças do *subcorpus* foram selecionadas de forma a possuírem variações no grau de pertinência na cena prototípica de **viagem**. A ideia era incluir sentenças típicas do uso e que tivessem a cena de viagem (prototípica, metafórica ou com significado relacionado) como a cena central do enunciado. Dado o pequeno número de sentenças do *subcorpus*, os resultados servem apenas para ilustrar as potencialidades do uso das anotações ontológicas e regras de distinção formalizadas na automação da criação de FrameNets. Não é a intenção aqui de apresentar um sistema automático pronto para ser disponibilizado para os usuários. Para isso, seria necessário um *subcorpus* maior, o acréscimo de informações sobre os elementos de frames e um tempo maior de treinamento da rede bayesiana.

A partir do *corpus* selecionado foi extraída a ontologia de domínio criada apenas com o propósito de se testar a anotação do *subcorpus* e apoiar a análise das sentenças, não devendo ser considerada como uma ontologia completa para o domínio de viagem. Buscou-se anotar os termos mais relacionados com os elementos de frames centrais, tais como viajante, local e veículo. A Figura 2 mostra alguns dos itens de domínio anotados. Estes elementos estariam na base da estrutura hierárquica que compõe a ontologia. O restante da estrutura hierárquica seriam os elementos da ontologia SIMPLE-CLIPS.

```

Alagoinhas/formal=local_geopolitico
América/formal=local_geopolitico
Arajá/formal=local_geopolitico
Arcebispo/formal=humano;constitutivo=estrutura;agenteivo=;telico=papel_social
Atlantis/formal=veiculo;constitutivo=estrutura;agenteivo=artefato;telico=transporte
Brasil/formal=local_geopolitico
Cadillac/formal=veiculo;constitutivo=estrutura;agenteivo=artefato;telico=transporte
Columbia/formal=veiculo;constitutivo=estrutura;agenteivo=artefato;telico=transporte
Comprei/formal=processo
Coronel/formal=humano;constitutivo=estrutura;agenteivo=;telico=papel_social
Embarcou/formal=evento
Eu/formal=humano;constitutivo=estrutura
Europa/formal=local_geopolitico
Fez/formal=evento
Foi/formal=evento
França/formal=local_geopolitico
Itajaí/formal=local_geopolitico
Japão/formal=local_geopolitico
Jerusalém/formal=local_geopolitico
Lisboa/formal=local_geopolitico
Manaus/formal=local_geopolitico
noivos/formal=humano;constitutivo=agrupamento;agenteivo=;telico=papel_social
Noruega/formal=local_geopolitico
Pernambuco/formal=local_geopolitico
Pássaros/formal=animal;constitutivo=agrupamento
Rodoviária/formal=local;constitutivo=estrutura;agenteivo=artefato
Soldados/formal=humano;constitutivo=agrupamento;agenteivo=;telico=papel_social
Universidade/formal=entidade_concreto;constitutivo=agrupamento;agenteivo=;telico=ensino
Velho/formal=propriedade;constitutivo=;agenteivo=;telico=
Viagem/formal=processo
Viajei/formal=processo
agosto/formal=entidade_abstrata
ano/formal=entidade_abstrata
asas/formal=entidade_concreto;constitutivo=parte-de;telico=movimento
atacante/telico=papel
automóvel/formal=veiculo;constitutivo=estrutura;agenteivo=artefato;telico=transporte
ave/formal=animal
avião/formal=veiculo;constitutivo=estrutura;agenteivo=artefato;telico=transporte
axilas/formal=entidade_concreto;constitutivo=parte-de
bagagem/formal=entidade_concreta;agenteivo=artefato
bailarina/formal=humano;constitutivo=agrupamento;agenteivo=;telico=papel_social
boate/formal=local;constitutivo=estrutura;agenteivo=artefato
bonde/formal=veiculo;constitutivo=estrutura;agenteivo=artefato;telico=transporte
bois/formal=animal;constitutivo=agrupamento
bom/formal=propriedade
brasileira/formal=propriedade
brasileiras/formal=propriedade
brasileiro/formal=propriedade
braços/formal=entidade_concreta;constitutivo=parte_corpo
cã/formal=local
calado/formal=propriedade
camelô/formal=humano;telico=papel_social
caminhoneiro/formal=veiculo;constitutivo=estrutura;agenteivo=artefato;telico=transporte

```

Figura 2: Segmento da Ontologia de domínio.

Os tipos ontológicos anotados geraram as variáveis independentes da rede bayesiana. Ou seja, as variáveis cuja ocorrência independe de outro elemento. A suposição de que os tipos ontológicos que ocorrem em uma sentença são independentes entre si é uma condição de contorno importante para viabilizar a aplicação da rede. No entanto, isso não é verdade no

caso geral. Certamente há uma influência mútua na ocorrência dos tipos ontológicos em uma sentença, mas isso não foi tratado neste trabalho.

A análise executada com as redes bayesianas relacionou a existência do tipo ontológico na sentença com a probabilidade da sentença estar associada a uma cena. Não foi levada em consideração a ocorrência do tipo ontológico juntamente com o papel semântico, ou seja, verificado se o tipo ontológico *humano* ocorria no papel de *viajante*. Acreditamos que essa associação aumentaria a precisão da inferência bayesiana, no entanto, seria preciso ter à mão um anotador automático de papel semântico para o português, o que não foi o caso no tempo desta pesquisa. Deixamos essa análise como um trabalho futuro. Contudo, é importante ressaltar que, apenas com a anotação dos tipos ontológicos, foi possível obter resultados promissores.

Outra limitação da análise é o fato de não termos utilizado hierarquia da ontologia SIMPLE-CLIPS na análise dos tipos ontológicos. Em função disso, escolhemos anotar os lexemas com os tipos ontológicos mais adequados aos nossos objetivos. Por exemplo, a palavra *feijão* foi anotada como sendo do tipo *entidade\_concreto*, uma vez que essa é a informação relevante para a análise da cena de *viagem*. Porém, a ontologia SIMPLE-CLIPS possui a classe *entidade\_vegetal* que é subclasse de *entidade\_concreto*, o que permite inferir que toda a instância de *entidade\_vegetal* é também uma instância de *entidade\_concreto*. Mas, para se beneficiar dessa inferência, seria necessário o uso de raciocinadores em lógica de descrições, o que não foi possível em função do tempo disponível para a realização da pesquisa. A inclusão desses raciocinadores é também uma proposta para trabalhos futuros.

#### 4.4 Análise com a Ferramenta Weka

As probabilidades de cada nó foram calculadas automaticamente pelo software Weka a partir do arquivo gerado pelo programa anotador. O Weka<sup>4</sup> é um conjunto de programas escritos na linguagem de programação Java voltados para tarefas de mineração de dados. O software foi desenvolvido pela Universidade de Waikato na Nova Zelândia e é uma ferramenta de código aberto. A ferramenta pode ser executada diretamente ou incorporada a outros programas e possui aplicações para pré-processamento, classificação, regressão, agrupamento e visualização de dados. Particularmente, para o problema em questão, decidiu-se usar as capacidades da ferramenta para geração de uma rede bayesiana, cálculo das probabilidades condicionais, treinamento da rede e inferência probabilística em um conjunto de teste.

Para se usar o Weka, é necessário gerar um arquivo texto com um formato específico: o formato *arff*. Nesse formato, os dados são tabulados, onde cada coluna representa um atributo, e o atributo na última coluna, em geral, representa o atributo cujo valor se deseja identificar. Os atributos ontológicos utilizados para a avaliação das sentenças são *humano*, *veiculo*, *animal*, *abstrato* e *local*. A Figura 3 mostra o arquivo para as 57 sentenças do arquivo de teste.

@relation FrameViagem	F,F,F,F,V,quasi V,F,F,F,F,quasi	V,F,F,F,F,quasi V,F,F,F,F,quasi
@attribute humano {V, F}	V,F,F,F,V,prototipico	F,F,F,V,F,metaforico
@attribute veiculo {V, F}	V,V,F,F,F,quasi	F,F,F,V,F,metaforico
@attribute animal {V, F}	V,F,F,F,V,prototipico	V,F,F,F,F,quasi
@attribute abstrato {V, F}	V,F,F,F,V,prototipico	V,F,F,F,V,prototipico
@attribute local {V, F}	F,F,F,F,F,quasi	F,F,V,F,F,quasi
@attribute frame {prototipico, quasi, metaforico}	F,V,V,F,F,quasi	F,F,F,F,V,quasi

<sup>4</sup> <http://www.cs.waikato.ac.nz/ml/weka>.

@data	V,F,F,F,V,prototipico	F,F,F,F,F,quasi
F,F,F,V,F,metaforico	V,F,F,F,F,prototipico	V,F,F,F,F,quasi
F,V,F,F,F,quasi	F,F,F,F,V,prototipico	V,F,F,F,V,prototipico
V,F,F,F,V,prototipico	F,V,F,F,F,quasi	F,F,F,F,F,quasi
V,V,F,F,V,prototipico	F,F,F,F,F,quasi	F,F,F,F,F,quasi
V,F,F,F,V,prototipico	V,F,F,F,F,prototipico	F,V,F,F,F,quasi
V,F,F,F,F,prototipico	F,V,F,F,F,quasi	F,F,V,F,F,quasi
V,F,F,F,V,prototipico	F,V,F,F,V,prototipico	F,F,F,F,F,quasi
V,V,F,F,V,prototipico	F,V,F,F,F,quasi	V,V,F,F,V,prototipico
V,F,F,F,V,prototipico	F,F,F,F,V,quasi	V,V,F,F,F,quasi
F,F,F,F,V,prototipico	F,F,F,F,F,metaforico	F,F,F,V,V,metaforico
F,F,F,F,V,prototipico	F,F,V,F,F,quasi	F,F,F,V,F,metaforico
F,F,V,F,F,quasi	F,F,F,V,F,metaforico	F,F,F,F,F,metaforico
F,F,F,V,V,metaforico	F,F,F,V,F,quasi	V,V,F,F,V,prototipico
V,F,F,F,V,prototipico		

Figura 3: Arquivo arff emitido pelo anotador ontológico.

Note que, antes da tabulação dos dados, são definidos os atributos e suas faixas de valores. É importante ressaltar que o valor do atributo de conclusão (*frame {prototipico, quasi, metaforico}*), para as sentenças foi definido manualmente. Isto é necessário para treinar a rede e para verificar seu desempenho no conjunto de teste. A melhor partição do *subcorpus* em conjunto de sentenças de treinamento e conjunto de sentenças de teste foi de 75% para o primeiro grupo e 25% para o segundo. Portanto, o conjunto de treinamento ficou constituído pelas primeiras 43 sentenças extraídas do *subcorpus* de 57 sentenças. As 14 sentenças restantes foram usadas para o teste de inferência probabilística.

As probabilidades condicionais também são geradas automaticamente. A Figura 4 mostra a influência probabilística do tipo ontológico *abstrato* sobre o valor do frame, extraída do conjunto de treinamento. Dessa forma, um valor “verdadeiro” para a ocorrência de um tipo *abstrato* indica a probabilidade de 92% de ser um uso metafórico do lexema.

frame	V	F
prototipico	0,031	0,969
quasi	0,039	0,961
metaforico	0,917	0,083

Figura 4: Probabilidades relacionadas com o tipo *abstrato*.

A execução da rede sobre o conjunto de teste obteve uma taxa de acerto de 85,7%, o que indica que a técnica possui um bom potencial, uma vez que esse número foi obtido com poucos exemplos de treinamento. A Figura 5 mostra o resultado da predição da rede.

Scheme:weka.classifiers.bayes.BayesNet weka.classifiers.bayes.net.estimate.SimpleEstimator -- -A 0.5 Relation: FrameViagem Instances: 57 Attributes: 6 humano veiculo animal abstrato local frame Test mode:split 75.0% train, remainder test === Classifier model (full training set) === Bayes Network Classifier not using ADTree #attributes=6 #classindex=5	Time taken to build model: 0 seconds === Predictions ontest split=== inst#, actual, predicted, error, probability distribution 1 2:quasi 2:quasi 0.069 *0.806 0.125 2 2:quase 1:prototip + *0.493 0.491 0.015 3 1:prototip 1:prototip *0.953 0.045 0.002 4 2:quasi 2:quasi 0.069 *0.806 0.125 5 2:quasi 2:quasi 0.069 *0.806 0.125 6 2:quasi 2:quasi 0.038 *0.939 0.023 7 2:quasi 2:quasi 0.008 *0.953 0.039 8 2:quasi 2:quasi 0.069 *0.806 0.125 9 1:prototip 1:prototip *0.907 0.092 0.001 10 2:quasi 2:quasi 0.319 *0.677 0.003 11 3:metafori 3:metafori 0.056 0.094 *0.85 12 3:metafori 3:metafori 0.004 0.124 *0.872 13 3:metafori 2:quasi + 0.069 *0.806 0.125 14 1:prototip 1:prototip *0.907 0.092 0.001
---	--

Network structure (nodes followed by parents) humano(2): frame veiculo(2): frame animal(2): frame abstrato(2): frame local(2): frame frame(3): LogScore Bayes: -185.33555315649258 LogScore BDeu: -206.00368309155624 LogScore MDL: -211.2683582674526 LogScore ENTROPY: -176.90242249085895 LogScore AIC: -193.90242249085895	=== Evaluation on test split === === Summary === Correctly Classified Instances      12      85.7143 % Incorrectly Classified Instances    2      14.2857 % Kappa statistic                        0.7544 Mean absolute error                  0.1478 Root mean squared error            0.2428 Relative absolute error              35.1572 % Root relative squared error        52.4091 % Total Number of Instances         14
---	---

Figura 5: Saída da rede para o conjunto de teste.

Na seção de *Predictions* da saída do programa, a coluna “*actual*” mostra qual é o valor que era esperado, e a coluna “*predicted*” indica qual foi o valor atribuído pela rede. É interessante notar que o uso prototípico foi sempre inferido corretamente. Isso se deve, possivelmente, ao fato de que o enquadramento na categoria prototípico deve-se em geral à presença de tipos ontológicos típicos (humano e local). Duas sentenças foram classificadas de forma diferente do que foi especificado. Uma do tipo metafórico e outra do tipo *quasi*. No caso da falha do uso metafórico podemos atribuir ao fato de o anotador ontológico não ter conseguido produzir uma anotação do tipo “abstrato” para algum elemento da sentença. Essas imprecisões são esperadas e o sistema pode melhorar a precisão com um maior número de sentenças, um número maior de atributos e com uma associação com os elementos de frames.

## Conclusões

A FrameNet é uma base lexical semântica fundada na Semântica de Frames. Ao tentar capturar toda a riqueza de aspectos significativos de uma rede conceitual em uma representação simbólica, como a FrameNet, parte do significado se perde. Parte do significado que não é capturado completamente pela FrameNet é o significado intuitivo sobre a natureza e os propósitos das entidades do mundo. Esse conhecimento corresponde ao que é denominado, na Ciência da Computação, de *ontologia*.

O conhecimento ontológico limita as interpretações de um enunciado, facilitando sua compreensão. Sendo assim, esse conhecimento possui uma grande utilidade na construção de uma base lexical, em particular para a FrameNet. O objetivo deste trabalho foi mostrar que a análise ontológica pode auxiliar na identificação do frame associado a um enunciado. Além disso, pode ajudar a explicar a ligação entre diferentes frames, mas com sentidos próximos, por meio da detecção de variação na propriedade ontológica. Todo o trabalho se baseia nas seguintes hipóteses: 1) A Semântica de Frames é, atualmente, a teoria que melhor descreve as estruturas conceituais presentes no aparato cognitivo humano; 2) A FrameNet é uma representação parcial dessas estruturas; e 3) A análise ontológica dos itens lexicais permite melhorar a análise e enquadramento dos enunciados.

A aplicação das redes bayesianas mostrou que essa é uma técnica viável para a automação parcial do processo de análise. Essa é, na verdade, uma consequência natural da proposta de Narayanan e Jurafsky (1998), que apresentam essa técnica como um modelo adequado para simular o processamento de sentenças da forma como é realizado por seres humanos. Ou seja, a decisão sobre a interpretação adequada de uma sentença polissêmica é baseada em uma série de evidências, inclusive sobre a natureza dos participantes, que resultam em uma interpretação mais provável. Claro que o sistema apresentado analisou apenas um tipo de evidência, a natureza ontológica, mas os resultados comprovam a importância desse aspecto na interpretação da sentença.

Para a Linguística Computacional, a contribuição é a apresentação de uma técnica de análise de sentenças baseada na Semântica de Frames e na análise ontológica e implementada

por uma rede bayesiana. A técnica permite verificar a probabilidade de um enunciado pertencer a uma determinada cena, com base na natureza ontológica de seus elementos, de uma maneira que, provavelmente, se assemelha à cognição humana. Se comprovada essa última hipótese, essa seria uma forte contribuição para a linguística cognitiva. Essa contribuição é inédita, uma vez que, na literatura, as contribuições da junção ontologia-FrameNet ocorrem após a criação dos frames e não na metodologia de criação.

Apesar do trabalho desenvolvido se apoiar em uma ontologia topo particular, a SIMPLE-CLIPS, os aspectos gerais do trabalho podem ser adaptados a qualquer ontologia de nível topo que aborde os principais aspectos de uma entidade expressos em enunciados da língua natural.

### Agradecimentos

Agradecemos ao apoio, no desenvolvimento deste trabalho, aos órgãos de fomento CAPES e FAPEMIG.

### Referências bibliográficas

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

BIRD, S.; KLEIN, E.; LOPER, E. *Natural Language Processing with Python*. O'Reilly Media, 2009.

BURCHARDT, A.; PENNACCHIOTTI, M. FATE: a FrameNet-Annotated Corpus for Textual Entailment. In: LREC'08, Marrakech Morocco, 2008. *Proceeding of LREC'08*, Marrakech Morocco, 2008.

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, Mexico. *Proceedings...* GELBUKH, Alexander (Ed.). LNCS 4394/2007, Mexico City, Mexico, 2007. p.1-11.

COPPIN, B. *Inteligência Artificial*. Rio de Janeiro : LTC editora, 2010.

DAVIES, M.; FERREIRA, M. *Corpus do Português: 45 million words, 1300s-1900s*. 2006-. Available from Internet:<<http://www.corpusdoportugues.org>>

FELDMAN, J. A. *From Molecule to Metaphor*. Cambridge. MA: Bradford MIT Books. 2006.

FILLMORE, C.J. Frame semantics and the nature of language. In: ANNALS OF THE NEW YORK ACADEMY OF SCIENCES: Conference on the Origin and Development of Language and Speech. 1976. *Proceedings...* 1976, v. 280, p. 20-32.

FILLMORE, C. J. Scenes-and-frames semantics, Linguistic Structures Processing. In: ZAMPOLLI, Antonio (Ed.). *Fundamental Studies in Computer Science*, n.59, North Holland Publishing, 1977. p. 55-88.

GRUBER, T. R. A translation approach to portable ontologies. *Knowledge Acquisition*, v.5, n.2, p.199-220, 1993.

GUARINO, N. Formal Ontology, conceptual analysis and knowledge representation. In: GUARINO, N., POLI, R. (Eds). *International Journal of Human-Computer Studies*, v. 43, n. 5/6, 1995. Número especial.

GUARINO, N. Formal ontology and information systems. In: FOIS, 1998, Trento, Italy. *Proceedings...* Amsterdam : IOS Press, 1998. p. 3-15.

KATZ, J. J.; FODOR, J. A. The structure of a semantic theory. *Language*, v. 39, n. 2, p. 170-210, 1964.

LENAT, D. B.; GUHA, R. V. *Building large knowledge-based systems: representation and inference in the CYC project*. Massachusetts: Addison-Wesley. 1990.

LENCI, A. ET AL. *SIMPLE* work package 2 - Linguistic specifications, Deliverable D2.1. , Pisa :ILC-CNR, March 2000.

MEURS, M. J.; LEFEVRE, F.; MORI, R. Spoken language interpretation: on the use of dynamic Bayesian networks for semantic composition. In: IACASSP/IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING, Taipei, 2009. *Proceedings...* Taipei, 2009. p. 4773-4776.

MOREIRA, A. *Proposta de um framework apoiado em ontologias para detecção de frames*. Tese de Doutorado. Programa de Pós-graduação em Linguística. Universidade Federal de Juiz de Fora , 2012.

MOREIRA, A.; SALOMÃO, M. M. M. Análise ontológica aplicada ao desenvolvimento de Frames. *Alfa*, São Paulo, vol.56, n.2, p.491-521, 2012.

MOREIRA, A.; ALVARENGA, L.; OLIVEIRA, A. P. Thesaurus and ontologies: a study over the definitions found in the computer and information science literature, by means of analytical-synthetic method. *Knowledge Organization.*, v.31, p.231-244, 2004.

NARAYANAN; JURAFSKY, D. Bayesian models of human sentence processing. In: COGNITIVE SCIENCE SOCIETY, 1998. *Proceeding...*, 1998. p. 752–758.

OVCHINNIKOVA, E. ET AL. Data-driven and ontological analysis of FrameNet for natural language processing. In: LREC - INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION, 7, Valletta, Malta, May 17-23 2010. *Proceeding...* Valletta, Malta, May 17-23 2010.

PALMER, A.; ALISHAHI, A.; SPORLEDER, C. Robust semantic analysis for unseen data in FrameNet. In: INTERNATIONAL CONFERENCE RECENT ADVANCES IN NATURAL LANGUAGE, 2011, Hissar, Bulgaria. *Processing...* Hissar, Bulgaria, 2011.

PETRUCK, M. R. L. Frame Semantics. In: VERSCHUEREN, J. ET AL (eds.). *Handbook of Pragmatics*. Philadelphia : John Benjamins,1996.

PUSTEJOVSKY, J. *The Generative Lexicon*. Cambridge: The MIT Press, 1995. 298 p.

RATNAPARKHI, A. A maximum entropy part-of-speech tagger. In: EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING CONFERENCE, May 17-18, 1996. *Proceedings...* May 17-18, 1996.

RUIMY, N.; TORAL, A. More semantic links in the SIMPLE-CLIPS database. In: LREC, 2008, Marrakech (Morocco). *Proceedings...* Marrakech (Morocco), 2008. (CD-ROM).

RUPPENHOFER, J. ET AL. *FrameNet II: extended theory and practice*. ICSI Technical Report. ICSI. 2006. (Technical Report).

SALOMÃO, M. M. M. FrameNet Brasil: um trabalho em progresso. *Calidoscópio*. v. 7, n. 3, p. 171-182, set/dez 2009.

SCHEFFCZYK, J.; PEASE, A.; ELLSWORTH, M. Linking FrameNet to the Suggested Upper Merged Ontology. In: FORMAL ONTOLOGY IN INFORMATION SYSTEMS, 2006. *Proceedings of Formal Ontology in Information Systems (FOIS-2006)*. BENNETT, Brandon; FELLBAUM, Christiane (Eds). IOS Press, p. 289–300, 2006.

*Received: 30/01/2013*

*Revised: 16/05/2013*

*Accepted: 20/05/2013*

*Published: 31/10/2013*