

UNIVERSIDADE FEDERAL DE JUIZ DE FORA  
PÓS-GRADUAÇÃO EM LINGUÍSTICA

ALEXANDRA MOREIRA

**PROPOSTA DE UM FRAMEWORK APOIADO EM ONTOLOGIAS  
PARA DETECÇÃO DE FRAMES**

JUIZ DE FORA

2012

ALEXANDRA MOREIRA

**PROPOSTA DE UM *FRAMEWORK* APOIADO EM ONTOLOGIAS  
PARA A DETECÇÃO DE FRAMES**

Tese apresentada ao Programa de Pós-Graduação em Linguística da Universidade Federal de Juiz de Fora, como requisito parcial para obtenção do título de Doutora em Linguística.

Orientadora: Profa. Dra. Maria Margarida M. Salomão

JUIZ DE FORA

2012

M838p

Moreira, Alexandra

Proposta de um *framework* apoiado em ontologias para a detecção de frames./ Alexandra Moreira - 2012.

194 f. il.

Orientador: Margarida Salomão.

Tese (Doutorado em Linguística) – Universidade Federal de Juiz de Fora, 2012.

1. Frames. 2. Ontologias. 3. Semântica de Frames. I. Título.

CDD:

CDU:

## AGRADECIMENTOS

É bom quando temos a oportunidade de agradecer àqueles que são parceiros em prol de um objetivo. Reconhecemos que estas parcerias ocorreram de diversas formas, seja financeira, seja pela transmissão de conhecimentos e, sobretudo, pelo apoio e carinho recebido durante esta longa jornada.

A parceria de pessoas e instituições é o que permitiu a realização desta pesquisa. A estes dedico o meu MUITO OBRIGADO!

Destaco o apoio financeiro concedido pela CAPES que permitiu dedicação exclusiva à pesquisa.

No decurso desta tarefa pude contar com a estrutura da Universidade Federal de Juiz de Fora e conhecer pessoas que ficarão para sempre em minha lembrança.

Destaco a participação de minha orientadora, a professora MARGARIDA SALOMÃO, que com competência e dedicação soube conduzir-me nesta tarefa e explorou minhas qualidades, adquiridas em outras áreas do conhecimento.

Os professores LUIZ FERNANDO, MARIA CRISTINA NAME e NEUSA SALIM, que introduziram e guiaram na aquisição de conhecimento sobre o estado da arte dos estudos sobre a linguagem e cognição.

À coordenadora LUCIANA TEIXEIRA e a secretária ROSÂNGELA MONTEIRO que com paciência e dedicação auxiliam os estudantes na resolução de suas dúvidas e problemas administrativos.

À revisora Sandra Del-Gaudio que em curto espaço de tempo dedicou-se no aperfeiçoamento da redação do texto.

Agradeço e espero manter a amizade adquirida com os colegas de classe ROBLEDÓ, FRANCINE, GABRIELA, IGOR, ANNA CARRARA, muitos outros que pude contar com esclarecimentos, ideias e informações.

Não posso deixar de mencionar aqueles que sempre estiveram comigo antes mesmo de iniciar o curso de Doutorado. Começo pelo meu MARIDO que me incentivou nesta tarefa e em tantas outras atividades profissionais exercidas.

A MINHA FAMÍLIA – pais e irmã pela torcida por meu sucesso pessoal e profissional, e por compreenderem as minhas ausências em várias ocasiões. Tenho certeza que eles sabem que meu pensamento estará sempre com eles.

E ao meu grande e maravilhoso DEUS, amigo inseparável, em todos os meus momentos.

*First there is Ontology, the treatment of abstract and quite general philosophic categories, such as Being (ÓL) and its being the One and Good.*

(Christian Wolff)

## **Proposta de um *framework* apoiado em ontologias para a detecção de frames**

### **RESUMO**

A Semântica de Frames é uma teoria que busca estabelecer o significado de um item lexical pela evocação da cena conceitual associada. A FrameNet é uma base léxico-semântica fundamentada na Semântica de Frames e é constituída de um conjunto de frames relacionados por ligações semânticas. A metodologia de descoberta e incorporação de frames à base obedece a critérios semi-formais e envolvem alguns critérios ontológicos. Nesta pesquisa buscamos o aprofundamento dessa análise ontológica de forma a permitir uma formalização mais abrangente do processo. Em linhas gerais, esta pesquisa visou estabelecer um arcabouço metodológico que relaciona ontologias e frames de modo a verificar a ocorrência e as fronteiras de um frame, segundo critérios ontológicos e uma análise mais fundamentada do enquadramento de um enunciado em um frame, permitindo a elaboração de uma FrameNet com maior precisão. Os resultados práticos desta pesquisa foram a formalização da FrameNet, a inserção da análise ontológica na metodologia da FrameNet e a sugestão de métodos de automação desta análise. A automação aplicada possibilitou a categorização de sentenças em frames de sentidos correlatos. Como resultado teórico essa pesquisa demonstrou que os aspectos ontológicos contribuem fortemente para a elucidação do significado de sentenças polissêmicas.

**Palavras-chave:** Frames; Ontologias; Semântica de Frames.

## **Proposal of a framework supported on ontologies for detecting frames**

### ***ABSTRACT***

*Frame semantics is a theory that aims to establish the semantics of a lexical item by the evocation of the associated conceptual scene. FrameNet is a lexical-semantic database developed based on the semantics frames and is constituted of a set of frames related by semantics links. The methodology for the discovery and incorporation of frames is based on semi-formal criteria and incorporate some ontological criteria. This thesis seeks to deepen this ontological analysis in order allow greater formalization of the process. Broadly speaking, this research aims to establish a methodological framework that relates to ontologies and frames to help identify the occurrence and the borders of a frame, according to ontological criteria and a more grounded analysis of the relationship of an utterance on a frame, allowing the development of a FrameNet more accurate and more useful. The practical results of this research were the formalization of FrameNet, the insertion of an ontological analysis step in the FrameNet methodology, and the suggestion of automation methods for this analysis. The automation applied enabled the categorization of sentences in frames of related meanings. As a theoretical result this research showed that the ontological aspects contribute greatly to the elucidation of the meaning of polysemic sentences.*

**Keywords:** Frames; Ontologies; Frame semantics.



## SUMÁRIO

1	INTRODUÇÃO .....	1
1.1	Objetivos .....	5
2	O TRATAMENTO COGNITIVISTA DA SIGNIFICAÇÃO .....	7
2.1	Os princípios da Linguística Cognitiva .....	7
2.2	A posição do Programa Minimalista .....	11
2.3	Dualismo versus corporificação .....	13
3	FUNDAMENTOS TEÓRICOS ADOTADOS DA LINGUÍSTICA COGNITIVA. .....	17
3.1	A questão da corporificação da cognição.....	17
3.1.1	A conceptualização baseada na corporificação da cognição .....	18
3.2	O Conceito de Frames .....	26
3.2.1	Frames Básicos e Frames Culturais .....	28
3.3	Uma nova concepção de categoria .....	32
3.3.1	O Conceito de prototipia.....	33
3.3.2	A noção de categoria básica.....	36
3.4	A Semântica de Frames .....	37
3.4.1	Elementos participantes do frame .....	40
3.4.2	Perfilamento de frame.....	40
3.4.3	As redes de frames .....	41
3.5	Relações de projeção de frames .....	42
3.5.1	Metáforas .....	42
3.5.2	Metonímias .....	46
4	FUNDAMENTOS COMPUTACIONAIS ADOTADOS.....	49

4.1	Tratamentos computacionais ligados a Linguística Cognitiva.....	49
4.2	A rede semântica da FrameNet .....	54
4.2.1	Tipos Semânticos .....	63
4.2.2	Anotação de relações metafóricas na FrameNet.....	65
4.3	Ontologias .....	66
4.3.1	Frames na Ciência da Computação.....	69
4.3.2	Ontologia na Ciência da Computação.....	73
4.3.3	Teoria, conceitos, relacionamentos e atributos de uma ontologia .....	75
4.3.4	Classificação das ontologias .....	80
4.4	Alinhamento entre Ontologias e a FrameNet.....	82
4.5	Limitações da FrameNet .....	88
4.6	Benefícios da junção de Ontologias com a FrameNet .....	90
5	METODOLOGIA .....	96
5.1	<i>Corpus</i> e campo de conhecimento.....	99
5.2	Procedimento de Anotação.....	100
5.3	Anotação baseada na Ontologia SIMPLE-CLIPS.....	103
5.4	Redes Bayesianas .....	111
6	DESCRIÇÃO DO EXPERIMENTO .....	114
6.1	O <i>FRAMEWORK</i> PROPOSTO .....	114
6.2	Formalização .....	118
6.3	Comparação com a Formalização de Ovchinnikova.....	125
6.4	Proposta de Alteração Metodológica .....	127
6.5	Aplicação da formalização .....	129

6.6	Experimento Corroborador.....	131
6.7	Distinção ontológica por meio de Redes Bayesianas .....	140
6.7.1	Análise com a Ferramenta MSBNx .....	146
6.7.2	Análise com a Ferramenta Weka .....	149
6.8	Discussão dos resultados .....	155
7	CONCLUSÕES .....	157
	REFERÊNCIAS .....	162
	ÍNDICE.....	175
	ANEXO – PLANILHAS DA ANOTAÇÃO DAS SENTENCAS.....	176

## LISTA DE FIGURAS

FIGURA 1 – Córtex pré-motor: áreas F1 a F7.....	21
FIGURA 2 – Generalização do Esquema-x. ....	24
FIGURA 3 – Diferença da aplicação do conceito complexo associado ao lexema inglês on. ....	30
FIGURA 4 - O Robô iCub.....	51
FIGURA 5 - O Robô CB2. ....	52
FIGURA 6 – Relações do frame COMMERCE_BUY com outros frames e informações sobre as unidades léxicas.....	56
FIGURA 7 – Relações do frame COMMERCE_BY com outros frames apresentadas de forma gráfica.....	60
FIGURA 8 – Editor de subcorpus com as camadas de anotações.....	62
FIGURA 9 - Hierarquia topo de tipos semânticos da FrameNet.....	64
FIGURA 10 – Sistema de frame envolvendo jogador de futebol.....	71
FIGURA 11 – Roteiro para um restaurante.....	73
FIGURA 12 - Relação entre conceptualização, linguagem e ontologia.....	76
FIGURA 13 - Exemplo de base de conhecimento em lógica de descrições. ....	78
FIGURA 14 - Exemplo de base de conhecimento em OWL. ....	79
FIGURA 15 – Tipos de ontologias.....	80
FIGURA 16 – Graduação de linguagens de representação de conhecimento em relação à categoria de ontologia.....	81
FIGURA 17 – Classificação de ontologias segundo o nível do conhecimento representado.....	82

FIGURA 18 – A rede relacional do mapeamento entre FrameNet, WordNet e SUMO.	84
FIGURA 19 – Cadeia de predicados de uma sentença.....	87
FIGURA 20 – Níveis de representação do conhecimento.....	91
FIGURA 21 – Relação entre os tipos ontológico e a FrameNet. ....	93
FIGURA 22 – Fases do sistema de análise.....	102
FIGURA 23 – Tela do anotador ontológico. ....	102
FIGURA 24 - Representação do lexema “livro” na TLG. ....	105
FIGURA 25 - Rede de herança para “livro” na TLG.....	106
FIGURA 26 - Mapeamento do item livro, segundo a TLG com os frames relacionados. .....	108
FIGURA 27 – Parte da ontologia SIMPLE-CLIPS. ....	110
FIGURA 28 - Rede bayesiana para sentido do lexema rede. ....	112
FIGURA 29 - Mapeamento de elementos de frames do COMMERCE_BUY.....	116
FIGURA 30 - Anotação de sentença baseada na FrameNet e na SIMPLE-CLIPS - caso 1. ....	117
FIGURA 31 - Anotação de sentença baseada na FrameNet e na SIMPLE-CLIPS - caso 2. ....	118
FIGURA 32 - Frame de VIAGEM - definição e elementos nucleares.....	133
FIGURA 33 – Frame de TRAVEL: relações semânticas.....	134
FIGURA 34 - Anotação ontológica de sentença - exemplo 1.....	135
FIGURA 35 - Anotação ontológica de sentença - exemplo 2.....	136
FIGURA 36 - Anotação ontológica de sentença - exemplo 3.....	136
FIGURA 37 - Anotação ontológica de sentença – exemplo 4. ....	137
FIGURA 38 - Anotação ontológica de sentença - exemplo 5.....	137

FIGURA 39 - Anotação ontológica de sentença - exemplo 6.....	138
FIGURA 40 - Anotação ontológica de sentença - exemplo 7.....	138
FIGURA 41 - Anotação ontológica de sentença - exemplo 8.....	138
FIGURA 42 – Subcorpus utilizado na anotação e para o treinamento e testes das redes bayesianas.....	141
FIGURA 43 – Ontologia de domínio.....	144
FIGURA 44 – Parte da anotação das sentenças do subcorpus.....	144
FIGURA 45 - Rede bayesiana para o Frame de VIAGEM.....	147
FIGURA 46 - Probabilidades de uma cena de viagem ser metáfora.....	148
FIGURA 47 - Avaliação de um enunciado na rede bayesiana para o Frame de VIAGEM.....	149
FIGURA 48 - Tela inicial da Weka.....	150
FIGURA 49 – Arquivo arff emitido pelo anotador ontológico.....	151
FIGURA 50 – Rede bayesiana gerada pelo Weka.....	152
FIGURA 51 – Probabilidades relacionadas com o tipo abstrato.....	153
FIGURA 52 – Saída da rede para o conjunto de teste.....	154
FIGURA 53 – Sugestão de rede para o frame VIAGEM.....	156

## **LISTA DE TABELAS**

Tabela 1	Comparação sobre inferência textual no domínio médico.	p.	86
----------	--	----	----

## 1 INTRODUÇÃO

Devido à sua complexidade, à sua natureza singular dentre os sistemas de comunicação dos animais e às suas interações com as diversas áreas do conhecimento, a linguagem natural humana é objeto de estudo de pesquisadores de diversas áreas da ciência.

Os resultados e elucidações obtidas em uma área de conhecimento repercutem nas outras, produzindo um saudável efeito em cadeia. Apesar do grande investimento em pesquisa nesse objeto de estudo, a linguagem natural permanece com diversos problemas em aberto, que vão desde seu surgimento na espécie humana, até seu processamento pelos seres humanos, produzindo diversas vertentes teóricas competidoras a respeito de seus variados aspectos, tais como behavioristas versus cognitivistas e gerativos versus construcionistas. A vertente cognitivista sobrepujou a behaviorista desde os anos sessenta (SALOMÃO, 2006, p.1), graças às contribuições teóricas lideradas por Noam Chomsky, sendo o paradigma predominante desde então. Essa vertente estuda a ligação entre linguagem e mente, algo que só foi possível a partir da evolução de outras áreas da ciência, como a Psicologia, a Computação e a Neurociência.

Não obstante a guinada para uma visão cognitivista proporcionada por Chomsky, o termo “Linguística Cognitiva” só foi empregado a partir de uma dissidência da linha gerativa no final dos anos setenta e início dos anos oitenta, tendo como expoentes George Lakoff, Ron Langacker, Len Talmy e Charles Fillmore. Um dos pontos principais de discordância entre a linha gerativa e a linha cognitivista é a relação da linguagem com a cognição humana. Para a linha gerativa, a linguagem é uma capacidade que depende de um módulo cognitivo distinto de outras capacidades cognitivas humanas. Este módulo é restrito à espécie humana e, em função de sua natureza recursiva, é o responsável pela capacidade de o ser humano emitir infinitos enunciados a partir de um conjunto finito de fonemas.

Já para a Linguística Cognitiva, a linguagem humana resulta da evolução das capacidades cognitivas e sensorio-motoras comuns a outros animais superiores, tais como os mamíferos, particularmente a ordem dos primatas. A linguagem possui uma



natureza corpórea e, mesmo quando lidamos com conceitos abstratos, lançamos mão do uso de metáforas que têm por base elementos físicos básicos e corpóreos. Resumidamente, a linguagem funciona como um tipo de gesto.

Na abordagem da Linguística Cognitiva, segundo SALOMÃO (2006, p.8), não existe distinção entre léxico e sintaxe, e a gramática é concebida como uma grande rede de unidades denominadas de construções. Além disso, o significado de um lexema só pode ser estabelecido a partir do conhecimento da cena onde ocorre. Esta cena foi denominada de Frame por Charles Fillmore. Charles Fillmore (1977) foi também o proponente da Semântica de Frames. Nas palavras de CROFT (2009, p. 1) a Semântica de Frames é “uma teoria de como os conceitos são organizados e representados na mente”. Na Semântica de Frames, o conceito fundamental é o conceito de frames: estruturas conceituais estabelecidas na memória permanente, fruto de nossa interação com o mundo e da consolidação de nossa experiência diária. A grande contribuição da Semântica de Frames para a Linguística é propor que o significado das expressões linguísticas seja interpretado de acordo com o frame de fundo que representa a cena no momento do enunciado.

Os resultados das pesquisas relacionadas à Semântica de Frames transcendem as fronteiras da Linguística tendo impacto em outras áreas do conhecimento, tais como o processamento da linguagem natural (PNL), a Psicologia e a Inteligência Artificial. Notadamente na área de PNL destacam-se as pesquisas para busca de informação com uso de *corpus* para análise de texto completo, tradução e inferência textual.

A Semântica de Frames inspirou a criação de uma base lexical semântica, a FrameNet (BAKER; FILLMORE; LOWE, 1998). A FrameNet procura estabelecer o significado de um elemento lexical definindo os elementos que compõem a cena onde ele ocorre e o posicionamento dessa cena em uma rede de ligações semânticas com outras cenas. A cena é definida por meio de evidência de *corpus*. A FrameNet foi desenvolvida para a língua inglesa pelo *International Computer Science Institute in Berkeley*, mas, atualmente, diversas instituições estão criando uma “FrameNet” para outras línguas, como é o caso da Universidade Federal de Juiz de Fora, que está desenvolvendo uma FrameNet para o Português Brasileiro.

A FrameNet está relacionada com a organização do conhecimento e, portanto, atua em um nível epistemológico. Mas alguns pesquisadores sugerem outros níveis de

conhecimento. Guarino (1995) propõe a existência de vários níveis de organização do conhecimento - lógico, epistemológico, ontológico, conceitual e linguístico. Segundo o mesmo pesquisador, o uso combinado dos vários níveis de conhecimento gera benefícios no entendimento e na construção de sistemas de representação de conhecimento.

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. As ontologias também são classificadas de acordo com o grau de generalidade das classes que englobam os objetos do domínio. Assim, as classes que são independentes de domínio são denominadas de categorias e ocorrem nas chamadas ontologias de nível topo. Já as classes específicas de determinados domínios ocorrem nas ontologias de domínio.

Trabalhos que unem estes sistemas de representação conceitual podem ser verificados tanto na Linguística como na Ciência da Computação (CHISHMAN; ALVES; BERTOLDI, 2006; BĀRZDIŅŠ *et al.*, 2008; 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.

Esta pesquisa 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 FrameNet é uma linguagem de representação e, como qualquer sistema de símbolos, não é capaz de capturar todos os aspectos de um conceito. 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. Sendo assim, dependendo do propósito da representação, ela deve ser enriquecida para cumprir adequadamente sua tarefa.

Apesar de ciente desta limitação, 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. O trabalho de Cruse (2002) destaca a importância do contexto na delimitação do sentido de um enunciado. 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 em um capítulo subsequente. 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, com a formalização dos elementos da FrameNet, e com a formalização das regras que envolvem os relacionamentos entre os frames.

Em linhas gerais, esta pesquisa visa estabelecer um arcabouço metodológico que relaciona ontologias e frames de forma que permita verificar a ocorrência e as fronteiras de um frame, segundo critérios ontológicos e uma análise mais fundamentada do enquadramento de um enunciado em um frame, permitindo a elaboração de uma FrameNet com maior rigor metodológico. Denominamos esse arcabouço de *framework* por tratar-se de um conjunto de medidas: a formalização dos elementos da FrameNet; uma alteração nos passos metodológicos para a criação de frames em um domínio levando em consideração os tipos ontológicos; e uma sugestão de automação.

O interesse para a Linguística no relacionamento dessas linguagens de representação insere-se nas pesquisas sobre semântica e sobre o desenvolvimento de bases de frames. Assim, acreditamos que a presente pesquisa, ao focar na determinação da interação entre esses dois níveis de conhecimento, se insere dentro dos limites do objeto de pesquisa da Linguística (a linguagem humana) e contribui para o avanço do campo de estudo. As contribuições obtidas auxiliariam no desenvolvimento de bases lexicais com maior precisão e de forma mais metodológica. Outro benefício da ligação dos Frames com ontologias formais é a ampliação do uso dos Frames em inferências, limitada atualmente pela falta de formalização dos Frames.

Esta pesquisa aborda como assuntos principais uma discussão sobre a noção de conceptualização e corporificação; sobre a Semântica de Frames; sobre a FrameNet; sobre a noção de ontologia e suas denominações, enfatizando a ontologia fundamental; os benefícios da junção entre ontologias e a FrameNet e a apresentação do *framework* que combina essas duas teorias para criação e a definição de frames.

## 1.1 Objetivos

### **Objetivo geral:**

O objetivo deste trabalho é estabelecer um framework teórico que integre frames e ontologias, de modo que os passos da metodologia para o desenvolvimento de uma FrameNet e sua respectiva rede de frames sejam esclarecidos de forma explícita e rigorosa.

### **Objetivos específicos:**

- a) Elaborar e justificar o *framework* a partir dos estudos semânticos considerados;
- b) justificar o modelo proposto a partir de evidências em corpora;

- c) formalizar as entidades da FrameNet visando à sistematização do uso do *framework*;
- d) aplicar uma camada ontológica aos elementos da sentenças para permitir um melhor ajuste na rede de frames;
- e) apresentar exemplos da aplicação do *framework*, focando no auxílio à identificação do enquadramento de sentenças em Frames e usos metafóricos de lexemas;
- f) aplicar um método semiautomático ao *framework* de forma a permitir o processamento de grandes volumes de enunciados.

## 2 O TRATAMENTO COGNITIVISTA DA SIGNIFICAÇÃO

A Linguística Cognitiva<sup>1</sup> (doravante LC) é uma área, dentro da Linguística, que abriga um conjunto de teorias que partilham alguns princípios fundamentais comuns. Esses princípios comuns constituem importante quebra de paradigma nos estudos sobre a cognição humana. Primeiramente, estabelece-se uma ruptura em nível filosófico, pelo desafio ao dualismo corpo-mente. Fundamentada nesta nova disposição, a LC propõe novas hipóteses epistemológicas sobre a cognição humana e busca evidências que as corroborem em elementos da Neurociência e da Ciência da Computação. Claramente, as teorias sobre a aquisição e uso da linguagem, como parte integrante dos processos cognitivos humanos, tanto quanto os estudos sobre memória e significação, foram amplamente influenciados por essas alterações paradigmáticas. Como a presente pesquisa está situada no âmbito da LC, neste capítulo apresentaremos a posição da área e seus princípios fundamentais.

### 2.1 Os princípios da Linguística Cognitiva

Geeraerts e Cuyckens (2007, p. 5) apresentam a seguinte definição para a LC: “*Cognitive Linguistics is the study of language in its cognitive function, where cognitive refers to the crucial role of intermediate informational structures in our encounters with the world*”<sup>2</sup>. Essa definição pressupõe que nossas interações com o mundo são mediadas por estruturas de conhecimento. A linguagem é concebida como um meio para organizar, processar e expressar informação como campo de estruturação do conhecimento sobre o mundo (GEERAERTS; CUYCKENS, 2007).

---

<sup>1</sup> A Linguística Cognitiva é escrita com as iniciais em maiúsculas para diferenciá-la de outras teorias linguísticas, que também enquadram a linguagem humana como uma habilidade cognitiva, mas a tratam como habilidade distinta das demais, como é o caso da Teoria Gerativa. Já a LC trata a cognição humana como criticamente integrada às outras habilidades cognitivas humanas.

<sup>2</sup> LC é o estudo da linguagem em sua função cognitiva, onde cognitiva se refere ao papel crucial de intermediar estruturas informacionais em nossos encontros com o mundo.

A partir dessa definição, é possível motivar três outros princípios que caracterizam todas as teorias relacionadas com o empreendimento cognitivista: *a primazia da semântica* na análise linguística, *a natureza enciclopédica do significado linguístico*, e o fato de que *o significado linguístico é perspectivizado*. Segundo Geeraerts (2006, p.3), a afirmação de que a linguagem é basicamente significação representa uma guinada radical em relação à abordagem vigente até então, a saber, a visão de que a linguagem é um objeto formal sintático, defendida pelo programa gerativo de Chomsky.

A tese sobre a natureza enciclopédica do significado linguístico tem raízes no fato de a linguagem ser um sistema para a categorização do mundo, fundamentado e moldado nas estruturas do conhecimento que lhe são contínuas. Como destaca Geeraerts (2006, p. 5), o significado linguístico não está separado de outras formas de conhecimento que temos do mundo e, nesse sentido, é enciclopédico, e não autônomo: envolve o conhecimento do mundo, que é integrado às nossas outras capacidades cognitivas. A natureza enciclopédica da linguagem implica que temos que levar em conta a familiaridade real que as pessoas têm com os objetos, tais como *pássaros*: não é apenas a definição geral de pássaro que conta, mas também o que sabemos sobre os pardais e os canários e sabiás etc. E essas experiências variam de cultura para cultura: os pássaros típicos em uma cultura serão diferentes daqueles típicos em outra cultura, influenciando o conhecimento que as pessoas associam com a categoria *pássaro* (GEERAERTS, 2006, p. 5).

O princípio da natureza perspectivizada da linguagem se opõe à visão objetivista, que determina a existência de uma e única visão correta do mundo. Esse princípio alega que a linguagem reflete uma visão do mundo influenciada pela cultura, vivência e necessidades de um indivíduo e não constitui um espelho da realidade. Segundo Geeraerts (2006, p. 4), o modo mais fácil para entender este princípio é pensar em perspectivas espaciais que aparecem em expressões linguísticas e em como a mesma situação objetiva pode ser interpretada linguisticamente de formas diferentes. Pense em uma situação em que você está no seu quintal, *atrás de sua casa*, e você quer contar onde você deixou sua bicicleta que está *na porta da frente da casa, do lado de fora*. Você poderia, então, dizer tanto que *ela está atrás da casa*, assim como que *ela está na frente da casa*. Apesar de parecerem contraditórias, as expressões apenas corporificam perspectivas diferentes quanto ao posicionamento do narrador frente ao marco (*a casa*).

Na primeira expressão, a perspectiva é da pessoa e, na segunda, a perspectiva é a da casa.

Lakoff (1987, p. 373), argumenta contra a visão objetivista prevalente nas ciências cognitivas tradicionais, demonstrando as insuficiências dessa concepção. Defende, pois, a posição experiencialista da LC, segundo a qual nossas experiências estruturam nosso raciocínio. Por exemplo, os objetivistas citam como evidência da existência de uma realidade objetiva as categorias denominadas *tipos naturais*, encontradas em áreas tais como a biologia. Exemplos dessas categorias seriam peixes, aves, gatos etc. No entanto, segundo Lakoff, as categorias clássicas e espécies naturais são resquícios da filosofia pré-darwiniana. Elas se encaixam nos estudos da Grécia antiga, e até mesmo na biologia de naturalistas como Lineu, mas estão em desacordo com descobertas modernas sobre fenômenos centrais da evolução natural - variação dentro de espécies, adaptação ao ambiente, mudança evolucionária gradual, reserva genética: todos esses fenômenos estão em conflito com a ideia clássica de tipos naturais, definidos segundo condições necessárias e suficientes, compartilhadas por todos os membros da categoria.

Salomão, em entrevista (ABRAÇADO, 2010) defende que a abordagem cognitivista é caracterizada pelo *idiomatismo constitutivo das formas linguísticas* e pela importância do *uso na gramatogênese*, tanto em termos ontogenéticos, como em termos históricos. O idiomatismo constitutivo das formas linguísticas rompe com a visão compartimentada da linguagem e com a noção composicional tradicional, segundo a qual o significado do todo é função do significado das partes. Na perspectiva cognitiva, a linguagem é formada por blocos (construções) que transcendem as barreiras léxico-sintática-semântica, e podem ser combinados de forma não determinística nem puramente sintática, como preconiza o tratamento gerativo.

A gramatogênese não é um princípio, mas uma área de pesquisa, cujos resultados evidenciam o caráter idiomático e corporificado da linguagem. Segundo Givón (2002, p. 39), à semelhança de outros fenômenos biológicos, a linguagem não pode ser plenamente compreendida sem referência aos fatos (hipotéticos ou comprovados) da evolução da espécie humana. Os estudos realizados por pesquisadores como Michael Tomasello mostram que a linguagem surgiu dos aprimoramentos das capacidades cognitivas dos primatas, em função da pressão evolucionária exercida pela



necessidade de atuação em grupo e pela participação dos atores em complexas relações sociais.

Em termos epistemológicos, Lakoff (1990) propõe dois compromissos que são subjacentes ao empreendimento da LC: o *compromisso da generalização* e o *compromisso cognitivista*.

O compromisso da generalização fundamenta a investigação dos princípios gerais que governam todos os aspectos da linguagem humana, numa arquitetura não-modularista. A LC lida com a necessidade metodológica da divisão do tratamento da linguagem, em disciplinas tais como a morfologia, a sintaxe e a semântica, mas entende que essa divisão apenas atende às necessidades práticas de análise e não é um reflexo da existência de módulos distintos de processamento para esses aspectos na cognição. Dessa forma, os princípios que motivam e explicam a linguagem instanciam-se em todas as suas dimensões estruturantes.

A *categorização* é considerada como dimensão constitutiva da estrutura da linguagem e, no âmbito da LC, o conceito de categoria é visto não como um agrupamento homogêneo de instâncias equivalentes, todas com o mesmo *status* de pertinência, mas como um agrupamento de elementos com diferentes *stati*, sendo alguns mais salientes que outros (ROSH, 1978). Os elementos são agrupados por familiaridade ou por semelhanças, mas não existe uma definição de categorias em termos de condições mínimas e necessárias de pertinência.

Sendo assim, considerando uma categoria que agrupe objetos usados pelas pessoas para sentar, uma cadeira seria mais prototípica como assento do que um caixote de madeira. Constata-se que esta mesma estrutura de categorização ocorre também para os elementos da linguagem em todos os seus aspectos expressivos, sejam morfemas, palavras ou construções mais complexas. Considere as palavras que estão enquadradas na classe de verbos. Existem algumas instâncias de verbos que apresentam um comportamento mais típico que é esperado de um elemento dessa categoria, podendo ocorrer tanto na voz ativa quanto na passiva (ex. *chutar*). Por outro lado, existem outros verbos que não apresentam todas as características típicas, como é o caso de verbos cuja forma passiva não é gramatical (ex. *gostar*) (EVANS; GREEN, 2006, p.32).

O compromisso cognitivo refere-se à ideia de que os princípios da estrutura linguística devem refletir o que se sabe da cognição humana a partir de disciplinas tais como as da Neurociência, da Psicologia, da Filosofia e da Inteligência Artificial. Isso significa que a linguagem e a organização linguística devem expressar princípios gerais que não são exclusivos da linguagem, ilustrando, deste modo, o quanto a linguagem é constitutiva da cognição humana.

## 2.2 A posição do Programa Minimalista

Para estabelecer um contraponto e para situar os princípios da LC no debate da Linguística contemporânea, apresentamos aqui a visão da cognição humana em relação à linguagem defendida pelo programa minimalista, cujo desenvolvimento segue-se das ideias de Chomsky (1995).

As hipóteses da LC representam uma base teórica oposta à visão chomskiana. Chomsky vê a linguagem como faculdade inata e autônoma, independente de outras faculdades cognitivas do ser humano, como defendido expressamente em artigo que escreve em cooperação com Hauser e Fitch (HAUSER *et al.*, 2002). Nas palavras do próprio Chomsky (1995, p. 167), “podemos distinguir a linguagem propriamente de um sistema conceitual e de um sistema pragmático de competência”.

Sob o enfoque do programa minimalista, a gramática é vista como um conjunto de estruturas que atuam na interface de dois sistemas, o *conceptual-intencional* (C-I) e o *articulatório-perceptual* (A-P), fazendo a ligação entre eles e atendendo as restrições impostas pelas interfaces dos sistemas. Cada um desses sistemas possui estruturas próprias, que impõem restrições à linguagem devido às suas inerentes limitações, tais como capacidade auditiva e articulatória, como salienta Borges Neto (2007). Uma sentença é dita “gramatical” ou “aceitável”, se satisfizer inteiramente as condições dessas interfaces.

Além disso, como salienta Smith (1999), a distinção entre a estrutura profunda e a estrutura de superfície não só foi abolida, como a representação desses níveis simplesmente desapareceu.

Os níveis de representação da estrutura da sentença na gramática, no programa minimalista, passam a ser o LF (do inglês *logical form*) e o PF (do inglês *phonetic form*). O PF faz interface com o sistema articulatório-perceptual e o LF faz interface com o sistema conceitual-intencional. Esses níveis de representação fazem interface com os sistemas cognitivos, sendo que LF faz interface com o sistema conceitual-intencional, e PF faz interface com o sistema sensorio-motor. A faculdade da linguagem no sentido amplo (FLB – do inglês *faculty of language in the broad sense*) englobaria esses sistemas cognitivos, juntamente com o módulo responsável pela geração das formas linguísticas que obedecem às restrições impostas pelas interfaces com os sistemas cognitivos. Esse último módulo, denominado por Hauser *et al.* (2002) como faculdade da linguagem no sentido estrito (FLN – do inglês *faculty of language in the narrow sense*), possui uma condição restrita à capacidade linguística dos seres humanos, que é a condição de recursão, ou seja a capacidade de gerar infinitas formas partindo de um conjunto finito inicial, em função da aplicação repetida de alguma regra de composição. Essa capacidade é a que permitiria o encaixamento de frases como “Maria disse que Márcia pensa que João é bonito”.

Nas palavras de Hauser *et al.* (2002), a FLN “é o sistema de computação abstrato isolado, independente de outros sistemas com os quais interage e faz interface”. A FLN seria, então, o componente que nos distingue de outros animais, capaz de gerar representações internas da linguagem e mapeá-las para as interfaces com os sistemas de desempenho conceitual-intencional e articulatório-perceptual, por meio dos sistemas LF e PF respectivamente. A FLN inclui apenas o mecanismo recursão e é o único componente especificamente humano da linguagem, mas não surgiu primariamente para a comunicação e é improvável que tenha se originado como adaptação evolucionária.

A informação disponibilizada para as interfaces deve conter apenas elementos legíveis nesses níveis – impõe-se a eliminação de traços formais que não sejam interpretáveis (legíveis para os sistemas de interface). Augusto (2007, p.117) destaca que “a informação disponibilizada para as interfaces deve conter apenas elementos legíveis nesses níveis”, ou seja, o “PF só interpreta traços fonológicos e o LF só interpreta traços semânticos”. Segundo Hauser *et al.* (2002), as expressões formadas pela FLN e autorizadas pelas interfaces que geram pares de som e significado é o que se denomina como linguagem: “um sistema de conexões som-significado”, potencialmente infinito.

Essa posição a favor da modularização é totalmente descartada na LC, que preconiza que faculdade da linguagem emerge das mesmas estruturas usadas em outras modalidades cognitivas, não existindo um módulo separado para a gramática e outro para a semântica. É essa a hipótese langackeriana sobre a gramática, como salienta Salomão, em entrevista (ABRAÇADO, 2010, p.17), segundo a qual “a sintaxe não é senão semântica congelada”.

Para a LC, a linguagem é uma manifestação derivada da pressão evolucionária operando sobre capacidades cognitivas corporificadas, previamente existentes.

Outro elemento destacado da teoria de Chomsky é a restrição da semântica linguística à forma lógica, limitando a significação à verificação dos valores verdades dos enunciados.

No entanto, se a gramática é uma estrutura explicável parcialmente por sua motivação conceitual, então é impossível reduzir a significação a um modelo objetivo do mundo “real”.

Finalmente, a ideia de que o conhecimento da linguagem emerge do uso da linguagem rejeita a proposição de que diversos fenômenos linguísticos, ligados ao uso da linguagem, possam ser relegados à periferia dos estudos linguísticos, permanecendo no foco da atenção disciplinar apenas as representações sintáticas e semânticas mais abstratas.

### **2.3 Dualismo versus corporificação**

Uma das principais áreas de debate na filosofia é o tipo de ligação que se postula entre mente e corpo. Existe a posição dualista, que preconiza que a mente e o corpo são entidades radicalmente distintas. Nesse caso, a mente possui uma natureza imaterial. Segundo Robinson (2009), essa visão da separação entre mente e o corpo tem origem no pensamento de Platão, que acreditava que as verdadeiras substâncias não são corpos físicos, os quais são efêmeros, mas formas eternas, das quais os corpos não são senão cópias imperfeitas. Ainda segundo Robinson (2009), as formas mais modernas de

dualismo originam-se dos escritos de Descartes, que acreditava em dois tipos de substâncias: a matéria e a mente. A matéria possui a propriedade essencial de ter extensão espacial. Já a mente possui a propriedade essencial de ser capaz de pensar.

A ideia de que a mente independe de um suporte material específico está em sintonia com a linha filosófica funcionalista (MALONEY, 1999). Na abordagem funcionalista, o objeto de estudo é descrito por suas funções, sendo irrelevante o aspecto físico. Particularmente, no caso da mente humana, o funcionalismo estabelece que o que define um estado mental, tal como um pensamento ou um desejo, é sua relação causal com estímulos sensoriais, com outros estados mentais e com o comportamento (LEVIN, 2010). No entanto, diferentemente da visão dualista, a abordagem funcionalista não propõe que a mente possua uma substância distinta. A mente precisa de um suporte físico, mas sua estrutura independe da natureza desse suporte.

As teorias cognitivas de primeira geração foram influenciadas por essas bases filosóficas e foram projetadas tendo por premissa básica a ideia de que os conceitos possuem uma natureza abstrata e amodal (independente da forma como são percebidos) e são representados em uma linguagem de pensamento (o mentalês), como proposto por Fodor (1975). Essa abordagem desconecta a percepção e a ação motora da cognição, tornando essa última um processo de manipulação simbólica.

Nas teorias cognitivas de geração mais recente, desponta a abordagem do *inatismo*, adversária da vertente empirista e racionalista da compreensão entre corpo e vida mental. O *inatismo* adota a visão de que uma parte significativa de nosso conhecimento está codificada em nosso organismo, ou seja, é inatamente pré-definida. Do lado oposto, encontra-se o *empirismo*, que apregoa que nosso conhecimento é adquirido por meio de nossa experiência cotidiana. O *inatismo* tem suas raízes na filosofia de Platão e foi defendido por filósofos, como Descartes e Leibniz, e combatido por filósofos empiristas, como John Locke e David Hume (SAMET, 1999).

O *inatismo* foi superado pelo empirismo até meados do século XX. Esse domínio do empirismo sobre o nativismo até esse período deveu-se à falta de base teórica para apoiar os argumentos dos nativistas e à fundamentação filosófica oferecida pelo empirismo para o behaviorismo, que era a teoria de aprendizado vigente. Noam Chomsky encerrou o domínio do empirismo e provocou o ressurgimento do nativismo, com seus trabalhos seminais na área da Linguística. Chomsky argumentou em favor da

hipótese de que a linguagem é uma capacidade inata, uma vez que, segundo ele, a simples interação com o mundo não é suficiente para explicar a rápida aquisição da linguagem nativa pelas crianças. É o argumento da *pobreza de estímulo*. Para Chomsky, as crianças não são expostas a estímulos em volume suficiente para explicar a capacidade que elas possuem de gerar novos padrões de sentenças, uma vez que elas são basicamente expostas a exemplos positivos e não a exemplos de estruturas incorretas. Segundo Chomsky (1965), a exposição a sentenças gramaticalmente incorretas é fundamental para o entendimento das regras gramaticais.

É importante salientar que a premissa da pobreza de estímulo é fortemente contestada pelos pesquisadores da Linguística Cognitiva. A rede neural cerebral e seu comportamento estatístico e de casamento de padrões conseguiria induzir padrões a partir de exemplos positivos. Pullum e Scholz (2002) realizaram um experimento que, utilizando-se de redes neurais e outros métodos estatísticos, induziu algumas regras de linguagem (referentes à depreensão de padrões) usando apenas as evidências positivas.

Apesar da tradição filosófica mencionada anteriormente e das teorias cognitivas de primeira geração, a Linguística Cognitiva adota uma posição oposta: a de que a cognição está ligada estruturalmente ao corpo, ou seja, a tese da **cognição corporificada**. A posição da Linguística Cognitiva é contrária à visão inatista da Linguística Gerativa e reconhece o papel da interação cérebro/corpo/ambiente na realização da cognição. Essa é a visão denominada de enatista ou enativista (do inglês *enactive*), que propõe o conceito de Cognição Situada. Segundo Rossa e Rossa (2011, p.19), a ideia em comum às abordagens enatistas é a de que o organismo e seu ambiente constituem uma unidade indissociável e interagem de forma dinâmica, contínua e simultânea.

Para o enativismo, a cognição é um processo emergente das interações cérebro/corpo/ambiente, endossando a tese da corporificação da cognição e da gramática. A mente é abordada como um sistema dinâmico corporificado no mundo, antes que como uma rede neural “na cabeça”. A cognição é a reorganização material do sistema, provocada por alterações em seu acoplamento com o ambiente. A relação mente-mundo não é de “espelhamento”, mas de interferência recíproca. A experiência não é uma dimensão epifenomênica da vida mental, mas uma dimensão fundante da mente e de sua fenomenologia (SALOMÃO, 2009b).

Essa visão alterou o desenvolvimento das ciências cognitivas, tais como a Inteligência Artificial, cujo redirecionamento será comentado mais adiante.

### 3 FUNDAMENTOS TEÓRICOS ADOTADOS DA LINGUÍSTICA COGNITIVA

Neste capítulo iremos descrever os fundamentos teóricos da LC que foram adotados nesta pesquisa: a corporificação da cognição, a nova noção de categorização, e a Semântica de Frames. Esses fundamentos teóricos são importantes para o entendimento da conceptualização e seu relacionamento com a Semântica de Frames, que é a base teórica para a FrameNet, foco desta pesquisa.

#### 3.1 A questão da corporificação da cognição

A cognição é corporificada, segundo Gallese e Lakoff (2005), porque é estruturada pela nossa interação constante com o mundo, através de nosso corpo e de nosso cérebro, de tal modo que nossa conceptualização está relacionada com ações corporificadas e, possivelmente, com suas extensões metafóricas, envolvendo sua simulação no sistema sensorio-motor cerebral. Segundo Wilson e Foglia (2011), alguns dos principais fenômenos que evidenciam a influência mútua entre a mente e o corpo são:

1) As pessoas gesticulam quando falam umas com as outras e os gestos não só facilitam a comunicação, como também o próprio processamento da linguagem.

2) Existem certos tipos de neurônios, os *neurônios espelho* (DI PELLEGRINO *et al.*, 1992) que disparam, não somente quando executamos uma ação, mas também quando assistimos outras pessoas realizando as mesmas ações.

3) A visão é, com frequência, um guia para ação, sendo que o movimento do corpo e o *feedback* que ele gera estão intimamente integrados a algum processamento visual.

4) O uso de partes do nosso corpo e mesmo de objetos a nossa volta, nos permite realizar tarefas cognitivas (tais como lembrar) de forma mais eficiente, simplificando o processamento cognitivo e permitindo o acesso às memórias armazenadas.



### 3.1.1 A conceptualização baseada na corporificação da cognição

Os estudos da percepção e conceptualização das cores fornecem uma visão clara de como os conceitos são formados tendo por base nossos recursos perceptivos, nosso aparato neural e de como são influenciados pela cultura na qual estamos imersos. Como afirma Lakoff (1989), o conceito de cores não reflete uma realidade objetiva. As cores não existem no mundo. Não existe algo como “a cor azul”. O que existem são ondas eletromagnéticas cuja reflexão na superfície de determinados objetos podem sensibilizar o aparato neurosensorial dos seres vivos, causando a sensação de visualização dessa cor. Esse aparato neurosensorial varia muito entre os seres vivos, acarretando diferentes capacidades visuais e de reconhecimento de cor.

Restringindo a análise aos seres humanos, sabe-se que o olho humano é capaz de perceber comprimentos de onda que vão de 400 nm a 700 nm (faixa da luz em nanômetros), tendo picos de percepção nos comprimentos de onda associados a azul, verde e vermelho. Esses picos de estimulação correspondem à variação de sensibilidade dos tipos de células sensíveis à luz que existem no olho, denominadas de cones e bastonetes. A combinação da estimulação dessas células resulta em picos na área das cores vermelha, azul, verde, amarela, branco e preto (cores focais).

As outras cores, como laranja e roxa, são “computadas” pelo aparato cognitivo, a partir da entrada neurofisiológica (LAKOFF, 1989). Assim, a cor laranja é computada pela composição nebulosa (*fuzzy*) das cores vermelha e amarela.

Já a influência cultural, a partir dessa entrada, determina um sistema de conceitos de cores pela promoção de deslocamentos de cores centrais e outras alterações nas propriedades das cores (LAKOFF, 1989). Portanto, os conceitos associados a cores são balizados pelas nossas condições orgânicas e, a partir dessa base, moldados pela nossa cultura. Não possuem correspondência com uma realidade objetiva, a não ser indiretamente, pelas propriedades de reflexão dos objetos do mundo real. A base neurofisiológica é evidenciada pelo caráter universal das cores focais. Segundo Feldman (2006, p.98), as pessoas em todo o mundo concordam que o vermelho prototípico se assemelha mais à cor de um carro de bombeiro do que a outros tons de vermelho.

Outra evidência do relacionamento entre a cognição e o corpo pode ser obtida a partir da verificação de que o substrato neural usado quando uma pessoa realiza uma ação é também ativado quando uma pessoa imagina realizar a mesma ação ou quando uma pessoa vê alguém realizar a ação. Essa descoberta da Neurociência, reportada em Gallese e Lakoff (2005), fornece uma forte evidência da ligação entre a cognição e o corpo. Os autores lançam a hipótese “de que o mesmo substrato neural usado para imaginar algo é usado para entendê-lo” (GALLESE; LAKOFF, 2005, p.2) e exemplificam com a sentença “*Harry pegou um copo*”. Se você não pode imaginar-se pegando um copo ou vendo alguém pegar um copo, então você não consegue entender essa sentença. A hipótese dos autores dá um passo adiante e diz que o entendimento procede pela simulação da situação linguisticamente evocada, e que o que você imagina ao interpretar uma sentença em um contexto é o significado da sentença nesse contexto.

A proposta não é uma teoria internalista do significado, **mas uma teoria interacionista da significação**, uma vez que a imaginação, da mesma forma que a percepção e a ação, é corpórea, isso é, estruturada por nosso encontro e constante interação com o mundo através de nossos corpos e cérebros.

Segundo os autores, um aspecto chave da cognição humana é o aproveitamento produtivo de estruturas neurais pré-existentes para novas funções: a adaptação dos mecanismos sensório-motores do cérebro, para servirem a novos propósitos relacionados com a cognição, enquanto preservam sua utilização nas funções originais.

As ideias básicas subjacentes à teoria da cognição corporificada são:

- **Multimodalidade (vs. modularidade) dos conceitos:** os circuitos neurais conectam entre si as diversas experiências modais (visão, audição, tato, movimento...) e a linguagem explora esta condição.
- **Agrupamentos neurais (nós ou *clusters*):** as unidades conceptuais correspondem a *clusters* funcionais, que podem ser *clusters* locais ou estruturados em redes neurais.
- **Simulação:** condição semântica que articula percepção, ação e imaginação e permite a experiência da significação: *imaginar* e *fazer* usam o mesmo substrato neural. A imaginação é simulação mental.

- **Parâmetros:** os elementos esquemáticos (ex. “funções temáticas”, “fases aspectuais”), que governam a simulação, articulam as inferências e emergem na linguagem.

Para ilustrar como essas ideias básicas atuam no modelo proposto, Gallese e Lakoff (2005) discutem o conceito *grasp* (pegar), mostrando como ele atua em nossa cognição.

Os agrupamentos neurais, também denominados de *clusters funcionais*, para a ação de pegar são multimodais, ou seja, disparam quando a ação de pegar é praticada, vista, ouvida, imaginada ou inferida, da mesma forma que disparam quando a ação de pegar é realizada por qualquer agente, envolvendo qualquer objeto, praticada de qualquer maneira ou em qualquer lugar. Por conta dessa multimodalidade da rede neural ativada para a ação de pegar, podemos dizer que eles asseguram uma condição universal à conceptualização de pegar. Em outras palavras, assim como a conceptualização das cores é indissociável da experimentação perceptual do espectro cromático, a conceptualização da ação de pegar é indissociável da experimentação sensório-motora desta ação (SALOMÃO, 2009c).

Para se entender como se processa a ação de pegar, assim como sua simulação, é preciso entender que áreas do cérebro são responsáveis por esse processamento e como elas se integram. Uma das áreas de fundamental importância que participa na execução da ação é o córtex pré-motor que está localizado no lobo frontal. O córtex pré-motor, tradicionalmente considerado homogêneo, é na verdade um mosaico de áreas funcionalmente distintas (F1 a F7, como mostra FIG. 1) (SALOMÃO, 2009c).

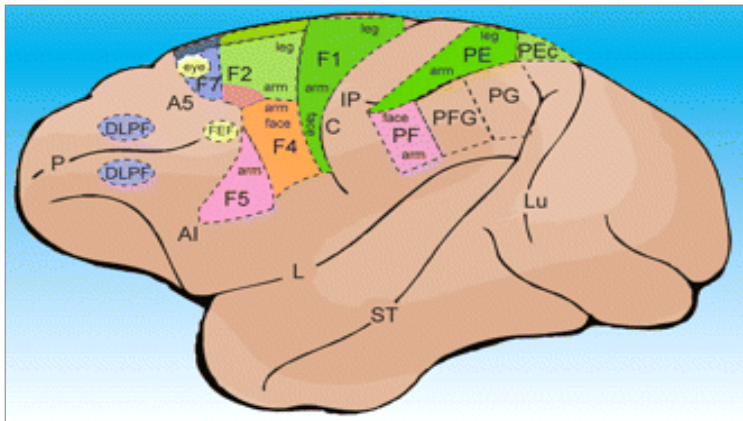


FIGURA 1 – Córtex pré-motor: áreas F1 a F7.

FONTE – HARI & KUJALA, 2009. p. 462.

O córtex pré-motor é, na verdade, uma série de *clusters* funcionais processados em paralelo. Suas regiões são dotadas de propriedades sensoriais e respondem a estímulos visuais, auditivos e somatosensórios. Por exemplo, a área F4 contém neurônios que integram as modalidades motoras, visuais e somatosensórias, com propósito de controlar ações no espaço e na percepção do espaço peripessoal (a área alcançável pelos nossos membros) (GALLESE; LAKOFF, 2005, p.5).

Além disso, segundo os mesmos autores, resultados de pesquisas recentes mostram que os neurônios da área F4 integram também informação auditiva sobre a localização de objetos dentro do espaço peripessoal. Portanto, o que temos aqui é a caracterização da multimodalidade neuronal: os mesmos neurônios que controlam a ação também respondem à informação visual, auditiva e somatosensória sobre objetos que participam da ação. Esse fato rejeita a noção anterior de que as modalidades possuíam áreas neurais separadas, que necessitavam de uma integração em um nível superior.

Participando desse processo existem três classes relevantes de neurônios: neurônios genéricos, neurônios espelho e neurônios canônicos. Os **neurônios genéricos** ativam grandes programas motores para ações “genéricas” (ex. pegar, segurar, manipular...). Eles asseguram uma universalidade parcial à conceptualização, já que disparam correlacionados com qualquer ação orientada para uma finalidade, pouco importando o modo da ação ou as partes do corpo envolvidas (ex. pegar com dois dedos, ou com a mão inteira, ou com a boca etc.) (SALOMÃO, 2009c).

Os **neurônios espelho** operam por simulação: quando o sujeito observa um indivíduo realizando alguma ação, ele a compreende porque ele simula realizá-la. Como ação e simulação empregam a mesma base neural, isso explica porque os mesmos neurônios disparam tanto na execução da ação como na observação da ação.

Os **neurônios canônicos** disparam durante a execução de uma ação como segurar, assim como pela observação de objetos que podem ser submetidos à ação. Porém eles não disparam durante a observação da ação realizada por terceiros (GALLESE; LAKOFF, 2005, p. 461).

Gallese e Lakoff (2005) mostram que essas funções multimodais foram descritas em três importantes circuitos corticais pré-motores: **F4-VIP**, **F5ab-AIP** e **F5c-PF**. Esses circuitos são denominados de *clusters funcionais*, porque funcionam como uma unidade com relação a computações neurais relevantes.

O *cluster* **F4-VIP** conecta a área pré-motora **F4** e a área parietal **VIP**. Ele transforma a posição espacial de um objeto situado no espaço peripessoal em programas motores para a interação com o objeto. Virar a cabeça porque alguém chamou, esticar a mão para pegar um objeto, afastar uma parte específica do corpo do ataque de um inseto ou de um desconforto. As propriedades do objeto são menos importantes do que sua posição espacial em relação a quem percebe.

O *cluster* **F5ab-AIP** conecta a área pré-motora **F5ab** com a área parietal **AIP**. Contém os neurônios canônicos, que transformam as propriedades físicas do objeto (formato, tamanho) em programas motores para ações praticadas com as mãos (manipular, pegar, segurar, rasgar etc.). Nesse *cluster* as propriedades dos objetos são bem mais importantes do que sua localização espacial.

O *cluster* **F5c-PF** conecta a área pré-motora **F5c** com a área parietal **PF**. Contém os neurônios espelho, que disparam quando o sujeito (humano ou macaco) pratica com as mãos ações orientadas para uma finalidade, ou observa outro indivíduo realizando ações dessa mesma natureza, ou quando infere que ações desta natureza tenham sido realizadas.

Agora que foram descritas as grandes áreas do cérebro que participam na ação e na simulação da ação e os tipos dos neurônios que atuam no processo, é preciso descrever como os conceitos, como o de PEGAR, são representados. Em seu trabalho

de doutoramento, Narayanan (1997) descobriu a estrutura do circuito geral que executa a programação motora: o *Esquema-x*. Os Esquemas-x são responsáveis por “coreografar” e estabelecer a sequência apropriada de movimentos simples para o córtex motor.

Essas estruturas modelam as atividades exercidas pelos neurônios espelho, neurônios canônicos e neurônios localização-ação. Esses esquemas são capazes de executar simulações imaginativas que, por sua vez, permitem o estabelecimento do raciocínio conceptual abstrato, resultando em uma teoria neural para a metáfora conceitual.

Narayanan usou esse modelo neural para realizar mapeamentos entre domínios abstratos e domínios físicos e mostrou que seria possível realizar inferências nos dois domínios, por meio do mesmo circuito neural. Além disso, as conexões dos Esquemas-x com o córtex motor podem ser tanto ativadas como inibidas, assim oferecendo uma estrutura inferencial para o raciocínio com conceitos abstratos.

Continuando a análise sobre o conceito de PEGAR o esquema neural que corresponde à ação de pegar é um Esquema de Execução (Esquema-x), que pode ser ativado para realizar a ação ou para simulá-la mentalmente. Segundo Johnson (2008, p. 159), o esquema associado (*grasping schema*) possui a seguinte sequência de fases de execução:

- **Estado Inicial**- Localizar o objeto: dentro do espaço peripessoal.
- **Fase Inicial** – Alcançar com direção: mover em direção à localização do objeto; abrindo efetuador.
- **Fase Central de Transição** – Fechar o efetuador com força: é uma função parametrizada pela fragilidade e massa do objeto.
- **Condição Objetivo** - Efetuador envolve o objeto com modo: um aperto determinado por valores de parâmetros e condições situacionais.
- **Estado Final** – Agente em controle do objeto.

Feldman (2006, p. 231) ilustra o Esquema-x por meio de uma FIG.2, aqui adaptada.

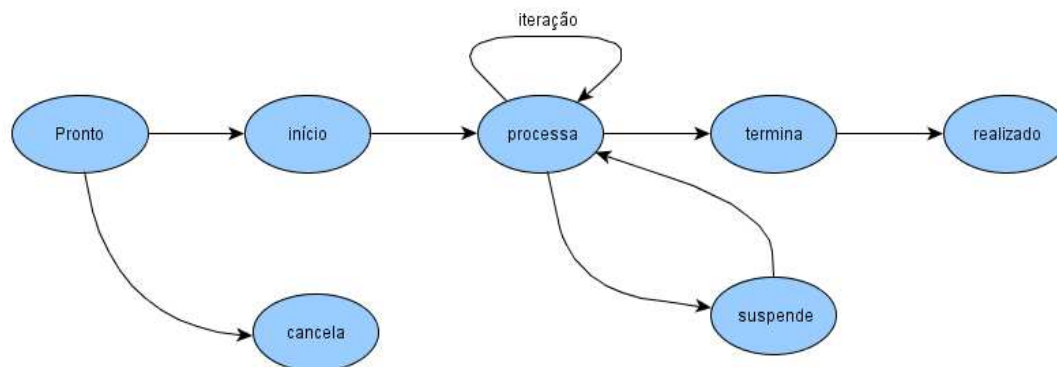


FIGURA 2 – Generalização do Esquema-x.

FONTE – FELDMAN, 2006. p. 231.

Relacionada com os Esquemas-x, existe a ideia de *cogs*. Os *cogs* surgiram a partir do trabalho de doutoramento de Narayanan (1997), ele propunha um modelo computacional para a semântica do verbo, que fornecesse uma base semântica para interpretar expressões aspectuais. Expressões aspectuais são usadas para caracterizar a estrutura de eventos e são usadas para raciocinar sobre eventos. Salomão (2006, p.8) apresenta o seguinte exemplo de marcadores aspectuais em português, baseados em metáforas convencionalizadas TEMPO É ESPAÇO E MUDANÇA É MOVIMENTO:

- (a) Entrei no Mestrado ano passado. (Inceptivo)
- (b) Estou no Mestrado desde o ano passado. (Estativo)
- (c) Passei o Mestrado todo trabalhando. (Progressivo)
- (d) Saí do Mestrado com novas ideias. (Terminativo)

Segundo Feldman (2006), a estrutura de controle dinâmico neural da FIG. 2 possui a organização necessária para definir a semântica das expressões aspectuais. O aspecto inerente a um verbo pode ser capturado pela marcação apropriada do nó do esquema de controle para a ação denotada pelo verbo. Os esquemas aspectuais identificados por Narayanan estão localizados em área secundária do córtex pré-motor e foram denominados por Lakoff (GALLESE; LAKOFF, 2005, p.17) de *cogs* ou *cogs estruturantes*, uma vez que são usados para estruturar eventos.

Os Esquemas-x e os *cogs* são elementos importantes para a representação conceitual uma vez que perspectivam e impõem uma estrutura cronológica linear ao conceito (que no âmbito da LC, corresponde ao *frame*), determinando uma sequência de passos para a execução de uma ação.

Dentro da ideia de que a cognição é corporificada, a linguagem se enquadra como um tipo de ação ou gesto, governado pelos mesmos mecanismos descritos. Conceitos são o que as palavras, morfemas e construções gramaticais expressam, sendo esse o propósito da linguagem. Gallese e Lakoff acreditam que, se estiverem certos, então:

- A linguagem faz uso direto das mesmas estruturas do cérebro utilizadas na percepção e ação.
- A linguagem não é uma inovação estritamente humana.
- Não há tal coisa como um "módulo de linguagem".
- A gramática reside nas conexões neurais entre os conceitos e a expressão dos conceitos através da fonologia. Ou seja, a gramática é constituída pelas conexões entre os esquemas conceituais e esquemas fonológicos. A estrutura gramatical hierárquica é a estrutura conceitual. A estrutura gramatical linear é fonológica, determinada pela necessidade de linearizar a estrutura conceitual para comunicar pela fala.
- A semântica da gramática é constituída por circuitos *cogs*-estruturantes usados no sistema sensorio-motor.
- A semântica e a gramática não são modalmente neutras.
- A semântica e a gramática não são simbólicas no sentido da teoria dos sistemas formais, que consiste em regras para manipulação de símbolos descorporificados e sem sentido.

Concluindo, a computação de abstrações pelo sistema sensorio-motor é possível em função do mapeamento metafórico desses elementos em correspondentes físicos, o que permite a realização de inferências análogas àquelas que seriam realizadas sobre os elementos físicos. Evidências Psicolinguísticas e os resultados das pesquisas na Neurociência apoiam essa hipótese. A hipótese da cognição corporificada, além de



alterar a base filosófica tradicional dualista, produz consequências profundas nos estudos da linguagem, pois altera a noção da formação dos conceitos e de sua transmissão por meio da linguagem humana. Mostra também que os conceitos possuem uma estrutura e que se relacionam em uma rede de conceitos, tornando a noção de conceitos equivalente à noção de frames semânticos.

### 3.2 O Conceito de Frames

Central ao estudo da cognição humana é a noção de **Frame** ou **Estrutura Conceitual**. A definição do que seja um conceito tem sido alvo de discussão de filósofos, psicólogos e cientistas da área cognitiva e outras áreas correlatas, sem que uma definição de consenso tenha sido obtida. Dentro da LC existe a definição fornecida por Gallese e Lakoff (2005, p. 455), segundo a qual “concepts are the elementary units of reason and linguistic meaning”<sup>3</sup>. Essa definição é suficientemente genérica para não ser controversa. No entanto, questões tais como: qual a natureza dessas unidades? como são formadas? qual é seu papel no raciocínio e na linguagem? continuam a ser fonte de amplos e acalorados debates.

Na concepção da primeira geração da ciência cognitiva, conceitos possuem uma natureza simbólica e descorporificada como um sistema formal que, apenas por circunstâncias, são processados no cérebro. Deste modo, o mesmo sistema poderia ser transferido para outro suporte físico, sem alterações em seus efeitos.

A visão da LC é diametralmente oposta. Ou seja, conceitos são indissociáveis de nosso aparato sensorio-motor e de seu substrato neural. Gallese e Lakoff (2005, p. 456) argumentam que o sistema sensorio-motor não só fornece a estrutura para o conteúdo conceitual, mas também caracteriza o conteúdo semântico dos conceitos, em termos do modo com que atuamos com os nossos corpos no mundo. Uma forma de compreender o que são os conceitos é o que os conceitos não são, sob a perspectiva da LC. Lakoff (1989) apresenta uma lista dessas negativas, das quais retiramos um subconjunto que, acreditamos, seja o mais relevante:

---

<sup>3</sup> conceitos são as unidades elementares de raciocínio e de significado lingüístico.

- Conceitos não refletem a natureza: não são representações internas de uma realidade externa.
- Conceitos, em sua maior parte, não são definidos por condições suficientes e necessárias.
- Conceitos não são entidades puramente mentais. Não são entidades puramente formais e não estão dissociados do corpo.
- Aparentemente, não existem primitivos conceituais, ou seja, conceitos que existam de forma independente, sem estrutura interna e que sejam cognitivamente autônomos.
- Conceitos não são, em geral, “literais”, no sentido de possuírem significado independente de outros domínios semânticos. Muitos conceitos são estruturados metaforicamente.
- Os conceitos, em sua maior parte, não são universais. Tampouco são apenas produtos da cultura.

Uma vez estabelecida as características que os conceitos não possuem, é necessário agora definir, no âmbito da LC e deste trabalho, quais características os conceitos possuem. **Conceitos ou Estruturas Conceituais são Frames.** Estruturas conceituais são corporificadas e motivadas pelo uso. Segundo Gallese e Lakoff (2005), conceitos/frames são:

- universais: no sentido de que caracterizam todas as instâncias específicas. Ou seja, no sentido de generalidade e não no sentido de ser algo independente da cultura. Por exemplo, o conceito de "pegar" é o mesmo, não importando quem seja o agente ou o que seria o paciente ou como a operação é processada;
- estáveis;
- internamente estruturados;
- composicionais;
- inferenciais;
- relacionais por equivalência, inclusão, herança etc.;
- significativos;

- não vinculados às formas específicas de palavras usadas para expressá-los.

Além disso, como observado por Rosch *et al.* (1976), alguns conceitos são mais imediatos à cognição humana que outros. Por exemplo, o conceito de “cadeira” é mais acessível à cognição do que o conceito de *mobília* (mais amplo) ou do que o conceito de “cadeira de balanço” (mais específico).

É importante notar que esses conceitos estão situados no meio da cadeia de generalização/especialização. Ou seja, não são os mais específicos e nem os mais genéricos. São os que aprendemos a nomear primeiro e com os quais interagimos no mundo com nossos corpos. Outros conceitos são mais complexos e envolvem uma rede rica de relacionamentos e influências culturais. Tome-se o caso do conceito de “feriadão”. Para entendê-lo, é preciso entender o conceito de semana, de fim de semana, de dia útil e do posicionamento de um dia útil em relação ao final de semana, todos eles baseados na cultura de tempo do mundo moderno.

### 3.2.1 Frames Básicos e Frames Culturais

Segundo Johnson (1987), a interação com o ambiente por meio do aparato sensório-motor humano e a repetição dessas experiências levam ao surgimento, no sistema conceitual, de **esquemas imagéticos**. Esquemas imagéticos correspondem a um subtipo de frame que parece ser fundamental e universal, possuindo um papel em todas as linguagens e que lidam, em grande parte, com relações físicas, movimento e percepção. São estruturas conceituais primitivas, ou seja, frames, denominados também de **frames básicos**.

Esses frames básicos resultam de experiências pré-linguísticas corporificadas. A partir de nossa experiência cotidiana com objetos que ocorrem dentro de outros objetos, surge na mente humana o esquema de imagem CONTAINER (RECIPIENTE). Os frames básicos surgem pela repetição da experiência e, uma vez gerados, são registrados

no sistema conceitual como um padrão. Uma vez que surgem a partir de nossa interação corporificada, formam a nossa base de conceitos mais primários. Eles são tão fundamentais que não estamos conscientes de possuí-los. Em geral, somos conscientes de conhecer conceitos tais como *ÁRVORE*, porém, não estamos conscientes de dispor da noção de *RECIPIENTE*. Os frames básicos geram conceitos mais específicos. O conceito de *RECIPIENTE* leva ao surgimento dos conceitos de *DENTRO* e *FORA*, elementos deste frame.

A observação do comportamento de crianças muito pequenas brincando é uma evidência forte de que elas aprendem os frames básicos, como *SUPORTE* e *RECIPIENTE*, antes de aprenderem as palavras associadas.

Segundo Feldman (2006, p. 136), o primeiro grande *insight* sobre esquemas imagéticos surgiu em 1983, com os trabalhos de Len Talmy (1983). A partir da análise de uma grande variedade de línguas, ele observou que os construtos usados para descrever o espaço podiam ser reduzidos a relações espaciais primitivas, que eram as mesmas em todas as linguagens. Essas estruturas primitivas são organizadas de forma diferente em cada língua para a geração dos construtos.

Em inglês, o sentido central da palavra *on* usa as primitivas conceituais “acima”, “contato” e “suporte”. Apesar de todas as línguas possuírem maneiras de expressar essas relações, nem todas possuem um conceito complexo estruturado da mesma forma que o lexema *on*. Em japonês, por exemplo, o lexema com o significado mais próximo não exige o contato, mas é mandatório o objeto estar localizado acima de outro. Ou seja, em japonês a relação se aplica apenas a superfícies posicionadas horizontalmente e em inglês, é possível aplicar a superfícies verticais. A Fig. 3 mostra a diferença de aplicação do conceito nas duas linguagens.

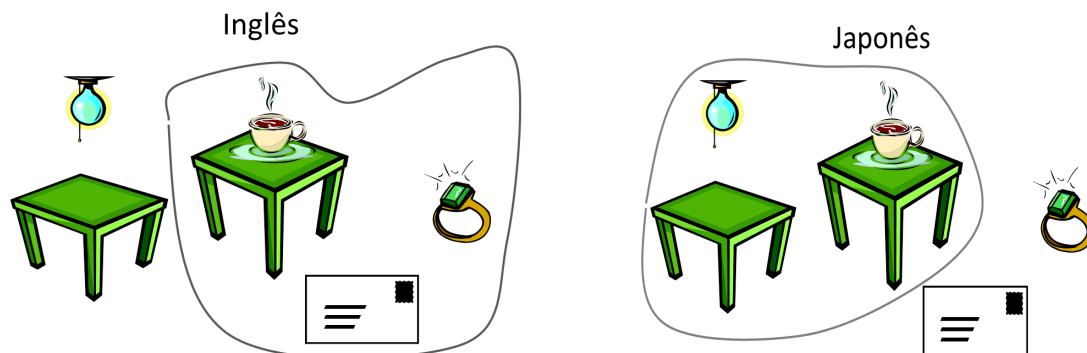


FIGURA 3 – Diferença da aplicação do conceito complexo associado ao lexema inglês *on*.

FONTE – FELDMAN, 2008.

Mas quantos esquemas imagéticos ou frames básicos existem? Segundo Hampe (2005, p. 2), a lista de esquemas imagéticos nunca foi uma lista fechada. Além disso, cada autor apresenta uma lista, sendo que, em alguns casos, parte dos elementos que não estão totalmente alinhados com o espírito da teoria original. Hampe (2005) apresenta uma lista de esquemas imagéticos que aparece em Lakoff (1987) e que constitui o seu inventário básico:

- a) RECIPIENTE/CONTEÚDO, CAMINHO/ORIGEM- CAMINHO-DESTINO, LIGAÇÃO, PARTE-TODO, CENTRO-PERIFERIA, EQUILÍBRIO.
- b) Esquemas de forças: DESBLOQUEIO, BLOQUEIO, CONTRA FORÇA, ATRAÇÃO, COMPULSÃO, RESTRIÇÃO, REMOÇÃO, DESVIO.

Interessante, também é mencionar os esquemas orientacionais, definidos em relação à orientação corporal, propostos por Talmy *apud* Feldman (2006, p. 136): ACIMA-ABAIXO, FRENTE-ATRÁS, ESQUERDA-DIREITA etc.

Pode-se notar que alguns dos esquemas estão diretamente relacionados com a nossa experiência do mundo, como é o caso dos conceitos ACIMA e ABAIXO, que estão relacionados com a gravidade. A palavra **acima**, em português, é baseada nesse esquema.

Feldman (2006) argumenta que a palavra *in*, em inglês, envolve uma situação complexa. Envolve a existência de um recipiente esquemático ou região delimitada no espaço composta de um *interior*, um *exterior* e *limites*. É preciso existir também um ponto de referência no interior e um elemento, denominado de *trajetor*, que está, pelo menos parcialmente no interior do recipiente. Além disso, como ressalta Feldman, alguns usos da palavra *in* implicam em esquemas de dinâmica de forças, em que o recipiente exerce força sobre o *trajetor*, como é o caso de garrafa, que impede que a água em seu interior derrame.

Grady (2005, p. 38) destaca que, apesar da ênfase frequente no caráter sensório-motor dos frames básicos, existem esquemas imagéticos propostos na literatura que não estão ligados a qualquer aspecto da experiência sensorial. Considere-se o caso de frames básicos como CICLO, PROCESSO e ESCALA (JOHNSON, 1987). Esses esquemas referem-se a uma dimensão que está relacionada a diferentes tipos de experiências sensoriais e não sensoriais.

Os frames básicos são usados para descrever unidades básicas e universais de conhecimento. Porém, grande parte do conhecimento humano é específica para uma determinada cultura ou ambiente social ou para um determinado contexto. As estruturas conceituais que armazenam esse tipo de conhecimento são os modelos culturais. Um modelo cultural é um *frame complexo*, estruturado por muitos frames básicos. Esses frames complexos são estruturados a partir dos básicos e moldados pela cultura. Por exemplo, um frame complexo como é o do JOGO DE FUTEBOL envolverá diversos frames mais básicos, tais como CORRER, SEGURAR, CONTATO e OBJETIVO, que por sua vez, se baseiam em esquemas imagéticos

Além de uma estrutura complexa, o frame delimita um conjunto de cenas ou objetos de forma não homogênea. Ou seja, o enquadramento de uma cena percebida em um conceito pode ter um nível de gradação. Esta forma de enquadramento não exato está relacionada com a concepção de categoria da LC que será discutida a seguir.

### 3.3 Uma nova concepção de categoria

Categorizar é um ato essencial para todo ser vivo. Feldman (2006) afirma que mesmo uma simples ameba, um ser vivo unicelular, realiza, por meio de processos químicos, a atividade da categorização: a ameba categoriza o que deve ser fagocitado e o que não deve ser fagocitado.

No caso dos seres humanos, em função do complexo aparato neural, o processo de categorização é bem mais elaborado. As categorias geradas pelos processos cognitivos dos seres humanos possuem uma estrutura interna e são influenciadas pela experiência individual.

Apesar de a linguagem não ser imprescindível à categorização, uma vez que todos os seres vivos categorizam, a categorização está intrinsicamente ligada à linguagem. Segundo Geeraerts e Cuyckens (2007, p. 5), a função primária da linguagem é categorização. Dessa forma, quando alguém diz “*Maria é bonita*”, está enquadrando *Maria* na categoria *bonita*. Mesmo em enunciados complexos tais como “*Câmara de Campinas cassa mandato do prefeito*” (Jornal Paraná Online de 20/08/2011), está sendo realizada uma categorização em que se caracteriza uma situação em detrimento de outras possibilidades. Na verdade, isso não é privilégio somente da linguagem natural. Toda linguagem, mesmo as artificiais, descrevem um estado de coisas, excluindo outras possibilidades; contudo a categorização realizada na linguagem não possui a delimitação rígida das linguagens artificiais.

A categorização expressa na linguagem humana está relacionada com a estrutura da realidade conceptualizada na mente. Isso não significa que a estrutura da linguagem é isomórfica com a estrutura da conceptualização que ocorre na mente. Existem evidências de que a mente possui estruturas conceituais que não são expressas na linguagem. Em um experimento conduzido por Rosch e Olivier (1972), foi mostrado que o povo *Dani*, da Papua Nova Guiné, apesar de possuir palavras apenas para duas cores, era capaz de categorizar os objetos a partir de suas cores, mesmo para aquelas para as quais não possuía palavras.

O ser humano, por meio de seu aparato sensorial e de suas capacidades introspectivas, percebe características proeminentes de objetos e eventos e realiza um

processo de categorização. Até recentemente, a teoria de categorização predominante era a teoria clássica, que estabelecia que a categorização era realizada por meio de atendimento de condições mínimas e necessárias. Apesar de ser a posição adequada em algumas situações, ela falha em diversas outras situações, como mostrado por Rosch (1975, 1977 e 1978). Em uma série de experimentos e argumentações, Rosch propõe que a categorização humana ocorre por meio da verificação do grau de familiaridade com um exemplar típico da categoria: o **protótipo**.

### 3.3.1 O Conceito de prototipia

O protótipo é um exemplar que reúne as características esperadas em um representante ideal abstrato de uma categoria. São os membros mais proeminentes de uma categoria e são os primeiros que vêm à mente, quando pensamos em uma categoria. O grau de semelhança com o protótipo gera uma estruturação dentro da categoria, fazendo com que alguns exemplares sejam mais centrais e outros sejam colocados em uma posição mais próxima à “fronteira psicológica” dessa categoria, ou seja, as categorias possuem uma gradação. As próprias fronteiras de uma categoria podem não ser bem delimitadas, adquirindo um caráter difuso (ou *fuzzy*, que é o termo utilizado na lógica *fuzzy* de ZADEH, 1975). Rosch (1975 apud EVANS; GREEN, 2006, p. 254) menciona a categoria FURNITURE (móvel), na qual uma pessoa teria dificuldade de indicar se um tapete pertence ou não à categoria, por não ser um elemento prototípico. Mesmo em categorias com fronteiras claras e nas quais os elementos podem ter sua pertinência verificada por condições mínimas e necessárias, ocorre uma gradação na prototipicidade. A categoria dos números naturais pares, por exemplo, possui alguns exemplares que são mais típicos (os números 2, 4 e 6) que outros (os números 108 ou 762), como foi mostrado por Armstrong *et al.* (1983).

O efeito prototípico, apesar de atestado experimentalmente, foi questionado como teoria para explicar o processo de categorização realizada na mente humana e de representação do conhecimento na mente, de forma geral. Como observado em Laurence e Margolis (1999), os principais questionamentos mencionavam a dificuldade de explicar como é possível que as pessoas possuam um conceito sem conhecer ou estar



enganado sobre suas propriedades, denominado **problema da ignorância e erro**. Um exemplo desse problema é o fato muito comum de as pessoas saberem identificar uma baleia, mas acharem que ela é um peixe. Segundo os autores, esse problema ocorre devido à necessidade da teoria de Rosch requerer uma forma de fixar a extensão de um conceito, mas ela não é capaz de determinar corretamente sua extensão, como é o caso da relação BALEIA-PEIXE.

Outra crítica é o problema denominado de **protótipo ausente**. Esse problema surge devido à existência de categorias que não possuem nenhuma instância e, portanto, não possuem protótipo, como é o caso da categoria “os reis dos EUA”. Também não é possível detectar um protótipo em categorias heterogêneas, tais como “sapo ou lâmpada” (LAURENCE; MARGOLIS, 1999, p.36).

Outro problema descrito por Fodor e Lepore (1996), é que os protótipos não se combinam composicionalmente, denominado de **problema da composicionalidade**. A categoria de PET FISH (peixe de estimação) (FODOR; LEPORE, 1996, p. 264) não pode ser inferida das características prototípicas das categorias de PET e de FISH. Em geral, um PET (animal de estimação) é um animal meigo e carinhoso e FISH (peixe) é um animal, em geral, cinza e indiferente. Fodor e Lepore argumentam que a razão da impossibilidade dessa derivação é que os tipos de peixes que as pessoas adotam como animal de estimação é um fato sobre o mundo e não sobre conceitos e linguagem.

Essas argumentações se baseiam na ideia de que as pessoas armazenam conceitos como protótipos e que comparam as novas instâncias com o protótipo. Porém, isso é uma interpretação equivocada do trabalho de Rosch. Rosch declara (1978, p. 41): “prototypes only constrain but do not specify representation and process models.”<sup>4</sup> Ela defende que existe uma diferença entre reconhecer o efeito de tipicidade (efeito prototípico) e propor uma teoria de como os conceitos são armazenados e aplicados.

Lakoff (1987, p.45) vai mais além e sustenta que o efeito prototípico é superficial e resulta de vários fatores. Mas a ideia principal é que as estruturas conceituais são frames complexos e interligados e o efeito prototípico surge a partir dos vários processos de enquadramento estatístico/*fuzzy* de nossas percepções nos frames que nosso sistema sensorio-motor realiza.

---

<sup>4</sup> Protótipos apenas restringem mas não especificam representações e modelos de processamento.

Evans e Green (2006, p. 270) citam o caso da categoria *ad-hoc* *O QUE RETIRAR DE CASA DURANTE UM INCÊNDIO*, sugerida por Barsalou (1983). Segundo Barsalou, essa categoria, criada por demanda, também exibe o efeito de prototipicidade, uma vez que a maioria das pessoas selecionaria as mesmas entidades para resgatar primeiro. Nesse caso, em particular, os itens que se enquadram na categoria são selecionados a partir de conhecimentos pré-existentes relacionados com valores sentimentais e monetários levando em consideração conhecimentos sobre sua localização na casa e o tempo disponível para resgatá-los. Para Lakoff (*apud* EVANS; GREEN, 2006, p. 270), essas categorias são construídas *online*, com o apoio de estruturas conceituais pré-existentes, para atender à necessidade de raciocínio local.

O exemplo clássico apresentado por Lakoff (1987) para ilustrar a aplicação dos frames é o do conceito associado à palavra inglesa *bachelor* (solteiro). Uma definição nos moldes clássicos do termo poderia ser “um homem adulto não casado”. No entanto, como apontado por Fillmore (1982 *apud* CIENKI 2007, p.177), o entendimento sobre o conceito de *solteiro* só pode ser feito a partir de sua relação com a cena onde ocorre, determinando expectativas sobre casamento e requerimentos de elegibilidade para ele. Isso não é capturado em uma definição com condições mínimas e necessárias. Na sociedade moderna, existem vários exemplos de homens adultos e não casados que não se enquadrariam na definição de solteiro: o papa (exemplo mencionado por FILLMORE); homens em uma relação estável não formalizada, casais homossexuais etc. Isso mostra também que o significado de um conceito é estabelecido por meio de suas relações com outros conceitos. O conceito de *solteiro* está inter-relacionado com o significado que é atribuído ao conceito de casamento. Para Lakoff, esse frame contém o conhecimento idealizado do que é uma relação de casamento típica e abstrai vários aspectos que podem ocorrer no mundo real. É com respeito a esse frame que a palavra *bachelor* é interpretada. Mas, na verdade, para uma categoria podem existir diversos frames, cada um dos quais licenciando uma interpretação distinta para a categoria. Isso contorna o problema da ignorância e erro, citado anteriormente.

Laurence e Margolis (1999) mencionam o caso do enquadramento de uma pessoa na categoria de AVÓ, na qual pessoas podem ser enquadradas erroneamente apenas por serem idosas e terem cabelos brancos, enquanto outras podem não ser enquadradas por não se assemelharem com o elemento prototípico. Para Lakoff, o exemplo típico é apenas um entre vários para a categoria AVÓ, sendo que o critério do

enquadramento correto ou incorreto dependerá do frame evocado. Da mesma forma, é possível possuir o conceito de BALEIA acreditando que é um PEIXE e não um MAMÍFERO, uma vez que possui propriedades típicas de peixe, como possuir nadadeiras e viver na água (EVANS; GREEN, 2006, p. 279).

O outro problema apontado por Laurence e Margolis sobre a teoria dos protótipos relatava sobre a dificuldade de se descreverem os protótipos para categorias sem extensão, como é o caso de “os reis dos EUA” ou para categorias heterogêneas, tais como “sapo ou lâmpada”. Mas esses problemas são contornados da mesma forma que o anterior, uma vez que essas categorias são construídas online a partir de frames pré-existentes.

Finalmente, o problema da composicionalidade mencionado por Fodor e Lepore é explicado da seguinte forma: a estrutura do conceito é independente das categorias com as quais se relaciona. Desse modo, apesar do peixe de estimação ser tanto um tipo de peixe como um animal de estimação, a experiência com peixe de estimação leva à criação de um frame, sendo que, nesse frame, o elemento prototípico é o peixinho dourado (EVANS; GREEN, 2006, p. 279).

### **3.3.2 A noção de categoria básica**

Outra descoberta interessante a respeito do processo cognitivo de categorização que contraria a ideia clássica de categorização é a observação de que as categorias que são mais imediatas à cognição são as que se encontram no “meio” da hierarquia de classes.

A visão clássica é a de que as categorias são formadas de “baixo para cima”, não existindo nenhuma diferença qualitativa entre os níveis superiores. Em uma hierarquia de categorias, tal como *mamífero*->*gato*-> *angorá*, a categoria gato é a que mais se destaca cognitivamente. Rosch denominou essas categorias de Categorias de Nível Básico e demonstrou, através de experimentos (ROSCH *et al.*, 1976), que essas categorias são as mais econômicas cognitivamente, pois agrupam elementos cujas

características comuns são facilmente identificáveis. Além disso, são as categorias que aprendemos a nomear primeiro (aprendemos, em geral, a falar *gato*, antes de falar *mamífero* ou *angorá*) e com as quais interagimos no mundo com nossos corpos, evidenciando o caráter corpóreo da categorização (GALLESE; LAKOFF, 2005).

### 3.4 A Semântica de Frames

No caso específico desta pesquisa, estamos interessados em estabelecer a conexão entre conceitos e Semântica de Frames, para destacar o papel fundamental dos frames na cognição humana. Os conceitos não ocorrem isoladamente, mas no âmbito de uma rede de relações com outros conceitos de tal modo que para compreendê-lo e usá-lo é preciso entender seu posicionamento na rede.

Além disso, cada conceito consiste de uma estrutura interna, composta por outros e por suas relações. Segundo Lakoff (1989, p.12), as estruturas imaginativas requeridas para definição de conceitos tais como FERIADÃO são chamadas de “frames” ou “esquemas”. Frames, como proposto por Fillmore (1976), são estruturas conceituais estabelecidas na memória permanente, fruto de nossa interação com o mundo e da consolidação de nossa experiência diária. Nos termos de Fillmore (2006): “By the term ‘frame’ I have in mind any system of concepts related in such a way that to understand any one of them you have to understand the whole structure in which it fits.”<sup>5</sup> Trata-se de conhecimentos construídos pela vivência e experimentação, formando um corpo de conhecimento estereotipado.

A ideia de frames propõe um modelo alternativo de semântica para os elementos sintáticos ocorrentes em um enunciado. Na Semântica de Frames, o significado das expressões linguísticas é interpretado de acordo com o frame de fundo, que representa a cena no momento do enunciado. Por exemplo, a palavra “risco” pode significar

---

<sup>5</sup> Pelo termo ‘frame’ o que eu tenho em mente é qualquer sistema de conceitos relacionados de tal forma que, para entender qualquer um deles, você tem que entender toda a estrutura na qual ele se encaixa.

“assumir um risco” ou “estar em risco”, dependendo do frame de fundo onde é aplicada (FILLMORE, 2008):

(a) *Vou correr o risco novamente [de comentar sobre o Kaká].* (ousar).  
(publicado em 15/01/2009 no Último segundo Online).

(b) *Patrimônio sacro de Jerusalém corre o risco de destruição.* (estar\_em\_risco).  
(publicado em 02/03/2004 na História Viva).

O que define o significado efetivo do termo é o frame evocado pelo termo e pelos outros elementos linguísticos que simultaneamente ocorrem no mesmo enunciado. No primeiro caso, a cena evocada é a de se ARRISCAR, na qual um agente realiza uma ação que pode lhe trazer um benefício, mas existe a possibilidade de que, caso não ocorra conforme desejado, sofra algum tipo de prejuízo. No segundo caso, a cena evocada é a de se ESTAR\_EM\_RISCO, na qual um recurso corre o risco de degradação ou destruição. Ou seja, são duas cenas com elementos participantes distintos. Assim é possível estabelecer que o sentido de *risco* em (a) difere do sentido em (b).

Segundo Croft e Cruse (2004), Fillmore descreve o modelo de Semântica de Frames como um modelo semântico do entendimento em oposição à semântica formal, que é baseada na ocorrência de condições que tornam uma sentença verdadeira.

A Semântica de Frames surgiu como resposta à incapacidade da semântica tradicional de dar subsídios para interpretações para diversas situações, tais como explicar porque não é adequado caracterizar o Papa como um homem solteiro (KATZ; FODOR, 1964). Esse é um exemplo clássico, usado em várias argumentações (LAKOFF, 1990; PETRUCK, 1996; FILLMORE, 1977) sobre a insuficiência da abordagem semanticista composicional, que define os conceitos segundo propriedades mínimas e necessárias.

No entanto, para entender o conceito evocado pela unidade lexical *casado*, é necessário entender uma série de estruturas conceituais inter-relacionadas, como a instituição do casamento no mundo ocidental, a noção das funções típicas de um homem casado e o que é estar apto a exercer essas funções. Só assim é possível aplicar adequadamente o termo “*solteiro*” a alguém. Existem vários termos como esses na língua natural, termos cujo significado só pode ser compreendido pelo entendimento de todos os conceitos envolvidos (*gestalt*) e não por sua análise isolada.

A semântica baseada em frames proposta por Fillmore tem se mostrado útil na interpretação de diversas figuras de linguagem tais como metáforas. Como salientado por Croft e Cruse (2004, p. 106), a abordagem cognitiva não só abre novos horizontes para a análise da linguagem, como também aborda preocupações tradicionais dos gramáticos e semanticistas de uma forma mais satisfatória.

Existem elementos linguísticos (doravante lemas) que estão mais naturalmente associados a certos cenários. No entanto, o dicionário tradicional não é capaz de retratar a cena de forma adequada, focando basicamente em uma definição sucinta para os sentidos mais comuns do lema. O dicionário Michaelis possui a seguinte entrada para a palavra *pedestre*:

adj m+f (lat pedestre) 1 Que anda ou está a pé. 2 Designativo da estátua que representa um homem a pé. 3 Humilde. s m+f Pessoa que anda a pé. sm Antigo soldado de polícia, no Rio de Janeiro.

A menção não trata da estrutura de relacionamento entre os elementos da cena, que é fundamental para a compreensão do significado da palavra pedestre.

Note-se que, além de não caracterizar a cena prototípica que dá o sentido típico da palavra *pedestre*, o dicionário apresenta alguns sentidos pouco comuns do lema. Os dicionários permitem apenas um acesso indireto à estrutura conceitual subjacente ao significado das palavras. Tipicamente, o lema *pedestre* está associado a uma cena onde existem pessoas andando a pé e veículos de transporte (FILLMORE, 2009, p. 319). Portanto, só existe o *pedestre*, quando relativizamos seu significado com respeito a outros elementos da cena, nesse caso em particular, automóveis se deslocando (automóveis parados em um estacionamento não compõem a cena).

Se alguém disser “eu desviei de um pedestre”, o mais provável é que o ouvinte imagine que quem pronunciou a sentença estava dirigindo, quando evitou uma pessoa que andava a pé. É menos provável imaginar que alguém, caminhando, afastou-se de outra que também caminhava. Isso ocorre porque ser pedestre é um papel que alguém exerce em uma cena prototípica e que esse papel possui uma relação de dependência com outros elementos da cena.

### 3.4.1 Elementos participantes do frame

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 postulava os papéis de Agente, Instrumento e Paciente.

Segundo Salomão (2009a), essa abordagem foi adotada na Semântica de Frames, em virtude da incapacidade das abordagens anteriores apresentarem um conjunto finito, consensual e preciso de papéis temáticos. Um conjunto fechado de casos semânticos não descreveria adequadamente o exemplo apresentado em Salomão (2009a, p. 4):

(a) *Dunga* (como técnico) *substituiu o atacante Robinho* no último jogo da seleção.

(b) *Nilmar* (como jogador) *substituiu o atacante Robinho* no último jogo da seleção.

No primeiro caso, temos como participantes o *substituidor* (o técnico) e o *substituído* (o jogador). Já no segundo caso, temos o que *substitui* (o jogador 1) e o *substituído* (o jogador 2). Isso mostra a necessidade de haver papéis semânticos adequados à cena, uma vez que os papéis tradicionais não seriam suficientes para capturar a efetiva participação dos envolvidos.

### 3.4.2 Perfilamento de frame

Um frame não é analisado isoladamente e, sim, a partir de um sistema de frames, no qual se relaciona, de modo que, ao acessarmos um, o conjunto de frames associados é trazido à mente.

Essa ideia pode ser mais bem ilustrada com o exemplo (extraído de Petruck, 1996) do frame para TRANSAÇÃO COMERCIAL, cujos elementos incluem um *comprador*, um *vendedor*, um *bem* a ser adquirido e algum *valor monetário*.

Existe um grande número de verbos ligados semanticamente a esse frame, tais como *comprar*, *vender*, *pagar* e *cobrar*, sendo que cada um referencia aspectos diferentes da cena. O *verbo* comprar perfila (LANGACKER, 1987, p.118) o *comprador* e o *bem* que está sendo adquirido, colocando como fundo o *vendedor* e o *dinheiro*. O *verbo* *vender* tem como elementos principais o *vendedor* e o *bem*. Já o *verbo* *pagar* focaliza o *dinheiro*, o *vendedor* e o *comprador*, e assim por diante.

O fato é, como ressalta Petruck, que para conhecer o significado de qualquer um desses verbos, é necessário saber o que ocorre em uma transação comercial e que saber o significado de qualquer verbo implica em, de alguma forma, saber o significado de todos os outros verbos relacionados com uma transação comercial. Essa questão do foco (*profile*) já havia sido analisada por Langacker, em seu exemplo sobre a semântica do lexema *hipotenusa*. Segundo Langacker (1987), é impossível compreender esse lexema, sem evocar o frame visual do TRIÂNGULO RETÂNGULO. No entanto, a perspectiva imposta pelo lexema é distinta da imposta pelo lexema *cateto*.

A Semântica de Frames postula que atribuir sentido a um lema é considerar tanto o que a palavra perfila, mais o que é assumido como o frame de fundo (GOLDBERG, 2009, p. 2). Por exemplo, a palavra *divórcio* perfila a separação legal de um casamento, mas tem como frame de fundo a pressuposição de um casamento realizado anteriormente.

### 3.4.3 As redes de frames

Já mencionamos que as estruturas conceituais não ocorrem isoladamente. Elas se relacionam semanticamente formando uma rede e o posicionamento relativo dentro de um frame, nessa rede, contribui para seu significado. A evocação de um frame pode ativar, em determinado grau, os frames próximos na rede. Segundo Feldman (2006, p. 5), estruturas conceituais nunca trabalham isoladamente: “Thought and language are the



result of complex interactions among schemas; recall that schemas are just our way of writing down basic neural structures.” Um exemplo fornecido por Feldman (2006) discute o caso da palavra *into*, que combina *in* e *to*, da seguinte forma: *in* é definido em relação um frame básico de RECIPIENTE. *To* é definido em associação a um frame básico de FONTE-CAMINHO-OBJETIVO. A palavra *into* combina ambos os frames RECIPIENTE e FONTE-CAMINHO-OBJETIVO, de forma que o objetivo é o interior do recipiente e a fonte é exterior do recipiente. Um frame pode estabelecer uma ligação de subtipo com outro, como é o caso entre a cena de CRIME e ASSALTO. Um frame pode impor uma perspectiva sobre outro, como é o caso entre TRANSAÇÃO\_COMERCIAL e VENDA. Um frame pode estar ligado a outro por uma necessidade definicional, como é o caso do frame DIVÓRCIO em relação ao frame de CASAMENTO.

### 3.5 Relações de projeção de frames

Um conceito complexo pode ser ligado a outro ou mapeado em outro, com o objetivo de auxiliar no raciocínio, tornar mais clara uma explicação, organizar ou economizar na fala. Essas ligações e mapeamentos entre frames resultam em fenômenos que são reconhecidos, na interface de comunicação, como metonímias e metáforas.

#### 3.5.1 Metáforas

O marco da inclusão da metáfora como operação fundamental do processamento cognitivo foi a publicação do livro de Lakoff e Johnson (1980), intitulado *Metaphors We Live By*, no qual propunham a ocorrência das metáforas no nível conceitual e não apenas como um fenômeno estilístico. Daí a denominação de **metáfora conceitual** proposta pelos autores. Segundo essa teoria, a forma como as pessoas pensam, o que experienciam e o que fazem diariamente estão relacionados com metáforas (LAKOFF; JOHNSON, 1980. p. 3). Os autores definem a metáfora como o “entendimento e a

experimentação de um tipo de coisa em termos de outra” (LAKOFF; JOHNSON, 1980, p. 5).

Para evidenciar a teoria da metáfora conceitual e para mostrar como os conceitos metafóricos permeiam as atividades cotidianas, os autores analisaram a metáfora conceitual ARGUMENTO É GUERRA, por meio de sentenças como as listadas a seguir:

- *Your claims are **indefensible**.* (Suas reivindicações são indefensáveis.)
- *He **attacked every weak point in** my argument. His criticisms were **right on target**.* (Ele atacou cada ponto fraco na minha argumentação. Suas críticas foram bem no alvo.)
- *I **demolished** his argument.* (Eu demoli seus argumentos.)
- *I've never **won** an argument with him.* (Eu nunca venci uma discussão com ele.)
- *You disagree? Okay, **shoot!*** (Você discorda? Ok, atire!)
- *If you use that **strategy**, he'll **wipe you out**. He **shot** down all of my arguments.* (Se você usar essa estratégia, ele vai acabar com você. Ele derrubou todos os meus argumentos.)

Em uma argumentação, o oponente é encarado como inimigo a ser derrotado e o processo de argumentação se desenvolve como em uma guerra, envolvendo estratégias, ataques, defesas, batalhas e terminando com um vencedor e um derrotado. Ou seja, quase todos os elementos de uma guerra se fazem presentes, metaforicamente, em uma argumentação. A estruturação dos dois domínios ocorre de forma semelhante, quase isomórfica, sendo que o domínio da argumentação é estruturado a partir do domínio da guerra. Esse mapeamento entre domínios conceituais é denominado de **Projeção Conceitual** (EVANS; GREEN, 2006, p. 286). O domínio de onde parte o mapeamento, fornecendo as características para a compreensão do segundo domínio, é denominado de **domínio fonte**. O outro domínio, que é interpretado sob a estrutura fornecida pelo domínio fonte, é chamado de **domínio alvo**.

A teoria da metáfora conceitual se fundamenta na teoria da cognição corporificada, uma vez que os domínios básicos, que são mapeados para domínios mais abstratos, são gerados a partir da experiência corporificada vivenciada pelo aparato sensório-motor.

A metáfora permeia quase todos os enunciados e processos cognitivos. A ideia de que existe uma distinção clara entre linguagem literal e linguagem metafórica não resiste a uma análise mais aprofundada. Em Gibbs (1994), é apresentada a dificuldade de se estabelecer uma linha divisória entre o sentido puramente literal e o figurativo. E qual é a relação da metáfora conceitual com o processo de categorização? Segundo Glucksberg (2008), metáforas são declarações categoriais sobre pertinência em classes: “Quando alguém diz que o seu cirurgião é um açougueiro, é isso mesmo que ele quer dizer, que o cirurgião pertence a uma categoria de pessoas que são açougueiros de uma forma ou de outra.” (GLUCKSBERG, 2008. p. 68).

A categoria é criada quando a metáfora é cunhada pela primeira vez. A teoria da metáfora conceitual proposta por Lakoff e Johnson (1980) continha os seguintes elementos:

- Unidirecionalidade – ou seja, o mapeamento ocorre em apenas uma direção, do domínio fonte para o domínio alvo e não ao contrário. Por exemplo, se o AMOR pode ser conceitualizado como uma JORNADA, uma JORNADA não pode ser parcialmente mapeada para os elementos do domínio do AMOR. Viajantes não são vistos como amantes e acidentes não são encarados como decepções amorosas. Isso não quer dizer que não possa existir outro mapeamento no sentido inverso, porém os elementos mapeados são distintos.
- Metáforas permitem inferências – em função do mapeamento entre os domínios, é possível inferir que situações que ocorrem no domínio alvo também podem ocorrer no domínio fonte.
- Motivações para Domínio Alvo e Fonte – segundo a teoria, existem elementos que habilitam um domínio a ser candidato a domínio alvo ou fonte. Os domínios alvo tendem a ser mais abstratos, enquanto os domínios fonte mais concretos. Isso evidenciaria o caráter corporificado da cognição. Em função disso, Lakoff e

Johnson propuseram que os esquemas imagéticos poderiam funcionar como domínio fonte para o mapeamento metafórico.

- Princípio da invariância – o mapeamento metafórico preserva a estrutura do esquema imagético. Ou seja, o mapeamento não viola a estrutura do domínio alvo, restringindo a forma como o mapeamento deve ser realizado.

Um ponto de potencial inconsistência na teoria da metáfora conceitual é o princípio que estabelece que o mapeamento não viole a estrutura do domínio alvo. Se o domínio alvo, que é mais abstrato, já possui uma estrutura, então qual é a necessidade do mapeamento? Posteriormente, alguns aspectos da teoria conceitual da metáfora foram modificados pela teoria que distingue **metáforas primárias e metáforas complexas**, proposta por Grady (1997), em sua tese de doutorado. Essa teoria elimina a inconsistência potencial da teoria da metáfora conceitual, dividindo as metáforas em primárias e em complexas. As metáforas primárias são fundacionais, enquanto as complexas são construídas a partir da unificação das metáforas primárias (EVANS; GREEN, 2006, p. 304). O mapeamento que ocorre nas metáforas primárias é realizado entre conceitos que são igualmente básicos, pois ambos são frutos de experiência direta e de percepção.

Essa proposta está em desacordo com a ideia de que o mapeamento seja entre conceitos com diferentes níveis de abstração. Para Grady, o que distingue o domínio fonte do alvo é que, apesar de ambos serem igualmente fundamentais, os conceitos do domínio alvo são diretamente percebidos (GRADY, 2007, p. 192). Uma das metáforas primárias fornecidas por Grady é “QUANTIDADE É ELEVAÇÃO VERTICAL”, como no enunciado “*The price of shares has gone up*” (Os preços das ações subiram). Nesse caso, a questão não é o grau de abstração que é considerado, mas o fato de o conceito de ELEVAÇÃO VERTICAL ser diretamente percebido pelos nossos sentidos, enquanto o conceito de QUANTIDADE tem a possibilidade de ser um parâmetro em qualquer domínio.

As metáforas complexas são estruturadas a partir das metáforas primárias, incorporando crenças culturais e, por essa razão, tendendo a ser mais restritas a certas culturas. As metáforas complexas envolvem domínios mais complexos e não exibem evidências de terem surgido a partir de uma base experimental. Como mostrado em Lakoff (2008, p.25), a metáfora “amor é uma jornada” é composta pelas seguintes

metáforas primárias: *propósitos são destinações; dificuldades são impedimentos ao movimento; relacionamentos são recipientes e intimidade é proximidade.*

A essas metáforas é acrescentado o seguinte conhecimento literal enciclopédico de senso comum:

*Um veículo é um instrumento para viagem; um veículo é um recipiente onde os viajantes ficam próximos; espera-se que as pessoas possuam objetivos de vida e, idealmente, amantes possuem objetivos de vida compatíveis.* (Lakoff, 2008, p.25).

Esses aspectos são combinados da seguinte forma: *os objetivos de vida são destinações; as pessoas que se amam são viajantes que buscam alcançar esses destinos; os relacionamentos são veículos, de modo que os amantes estão dentro da relação; eles estão próximos; o relacionamento os ajuda a atingir seus objetivos de vida; as dificuldades do relacionamento são impedimentos para o movimento.*

Concluindo, metáforas fazem parte do processo cognitivo e são fundamentais para o raciocínio abstrato. Elas são estruturadas, primariamente, pelos frames básicos e são combinadas para formarem metáforas mais complexas, que são influenciadas pelo ambiente cultural no qual o indivíduo está inserido. Feldman (2006, p.194), sugere que todos os frames culturais derivam do mapeamento metafórico de nossas experiências corpóreas representadas nos frames básicos, estabelecendo relações complexas entre eles.

### 3.5.2 Metonímias

Segundo Panther e Thornburg (2007, p. 237), metonímia (em grego μετωνυμία, e latim *denominatio*) é uma das figuras de linguagem principais na retórica clássica. Dentro da tradição clássica, a metonímia é entendida como uma relação de substituição ou representação, em que o nome de algo é usado para referenciar outro elemento. Panther e Thornburg mencionam que esta visão é chamada de *teoria da metonímia*. Assim, a sentença (x1), “Google” e “Apple” substituem as

pessoas responsáveis por essas empresas e “jornal” substitui o jornalista que publicou a notícia.

(x1) *Google monitorava usuários de Safari, da Apple, diz jornal.* (publicado em 17/02/2012 no “O Globo online” <<http://oglobo.globo.com/>>).

No entanto, existem falhas na teoria clássica, como não explicar os casos em que a metonímia não é usada apenas como substituição, mas também de forma predicativa. Por exemplo, na sentença “*ela não é apenas um rostinho bonito*” a expressão “*rostinho bonito*” não é usada como substituição, mas de forma predicativa.

Já no caso da LC, segundo Lakoff (1987), metonímia é um fenômeno conceitual caracterizado por deslocamento referencial dentro de um mesmo frame. Isso difere da metáfora conceitual que envolve dois frames. Na sentença (x1), “Google” e as pessoas responsáveis por sua gerência estão no mesmo frame relacionado com empresas e sua condução.

Segundo Croft (2006) o deslocamento realizado na metonímia é bastante natural e produtivo. No caso da sentença (x1), a empresa é o elemento proeminente do domínio, sendo natural que seja associada às ações de sua diretoria. Nesse caso, a metonímia ativa ou destaca certo aspecto do domínio. Segundo Evans e Green (2006, p. 312), a metonímia fornece uma “rota” de acesso para um elemento particular dentro de um mesmo frame. Sendo assim, para a LC as metonímias conceituais são elementos cognitivos, motivados por necessidades comunicacionais e referenciais, produzindo “atalhos” mentais a elementos do domínio.

Existem vários tipos de metonímia. Seguem apenas alguns exemplos, adaptados de Evans e Green (2006):

**PRODUTOR PELO PRODUTO**

- Acabei de comprar um Fiat novo.
- Pegue o Jorge Amado na prateleira.

**LOCAL PELO EVENTO**

- A Opinião pública americana teme outro Vietnã.

**LOCAL PELA INSTITUIÇÃO**

- Brasília não comentou o assunto.
- Europa elevou as apostas na guerra comercial com os Estados Unidos.

**PARTE PELO TODO**

- Ela não é apenas um rostinho bonito.
- Várias pernas passavam apressadamente.

**TUDO PELA PARTE**

- A União Europeia acaba de aprovar nova legislação de direitos humanos.
- Meu carro possui uma falha mecânica.

**EFEITO POR CAUSA**

- Sócrates bebeu a morte.
- Seu rosto está radiante.

Segundo Lakoff (2008, p.21) as metonímias ocorrem pelo estabelecimento, em um mesmo frame, de uma ligação entre papéis semânticos, de tal modo que um papel semântico deflagra, por esta relação de substituição interna (entre papéis semânticos) a evocação de todo o frame.

Metonímias diferenciam-se, portanto, de metáforas porque as últimas requerem o processamento de relações entre frames, enquanto as primeiras requerem processamento de relações semânticas dentro de um mesmo frame.

## 4 FUNDAMENTOS COMPUTACIONAIS ADOTADOS

Neste capítulo iremos descrever os fundamentos computacionais adotados nesta pesquisa: as hipóteses fundamentais da Inteligência Artificial e suas relações com a LC, as ontologias e o recurso lexical da FrameNet, cujo aperfeiçoamento metodológico é o foco deste trabalho. Em função deste foco, nas seções finais deste capítulo iremos abordar as limitações da FrameNet e os possíveis benefícios de inserir em sua metodologia o apoio do conceito de **tipo ontológico**.

### 4.1 Tratamentos computacionais ligados a Linguística Cognitiva

A Inteligência Artificial possui várias vertentes e linhas de pesquisa. A vertente mais tradicional é a linha simbólica, que se baseia na *hipótese dos sistemas físicos simbólicos* proposta por Newell e Simon (1963). Essa hipótese supõe que um sistema de símbolos composto por operadores capazes de gerar novos símbolos a partir de símbolos anteriores é condição suficiente para a construção de um sistema inteligente.

A partir dessa hipótese, conclui-se que o pensamento humano é um sistema simbólico, uma vez que essa é condição suficiente para a inteligência. Essa conjectura segue a mesma linha formal da Teoria Gerativa.

Existe também a linha conexionista, que entende que não é possível ter um dispositivo computacional que exiba inteligência análoga à humana, se esse dispositivo não operar fisicamente de forma análoga ao “*hardware*” humano. No caso da linha conexionista, o “*hardware*” se restringe basicamente ao cérebro humano.

Existe também a linha em que o conexionismo procura se inspirar na natureza, para elaborar dispositivos computacionais capazes de solucionar problemas de forma semelhante à realizada pela natureza. É a linha dos sistemas inteligentes bioinspirados, da qual também fazem parte a programação evolutiva, os algoritmos genéticos e os sistemas multiagentes.



Apesar do sucesso no desenvolvimento de diversos sistemas inteligentes, tais como provadores de teorema e sistemas especialistas, a linha simbólica tradicional apresenta limitações na construção de sistemas inteligentes em áreas tais como raciocínio de senso comum e interação com o meio ambiente. Em função disso, a hipótese dos sistemas físicos simbólicos passou a ser questionada.

A ênfase atual é a combinação de raciocinadores dotados de corpos e sensores artificiais que possam atuar e ser moldados pela interação com o mundo (DHARMENDRA *et al.*, 2011).

Um dos laboratórios envolvido diretamente nesse tipo de pesquisa é o *Laboratory of Artificial Life and Robotics*<sup>6</sup> do *Institute of Cognitive Sciences and Technologies*, em Roma. O laboratório realiza pesquisas sobre organismos artificiais corporificados. Os organismos são dotados de uma rede neural artificial, interagem com o ambiente externo onde estão situados e são adaptativos, variando ontogeneticamente e/ou filogeneticamente. Segundo informações do site, o laboratório foca tanto no comportamento individual (i.e. coordenação sensório-motora, motivação e determinação de comportamento), quanto no comportamento coletivo (i.e. coordenação, cooperação e comunicação) e cognição (categorização, antecipação e integração da informação sensório-motora ao longo do tempo).

Outro projeto que possui o objetivo de apoiar os estudos de robótica com base na cognição corporificada foi desenvolvido por um consórcio dentro de um projeto financiado pela Comissão Europeia. Neste projeto foi desenvolvido um robô humanoide com a forma de uma criança de quatro anos. O robô, denominado de *iCub*, é disponibilizado sob uma licença aberta e é capaz de engatinhar, sentar, manipular objetos. Possui capacidades visuais, auditivas e sensórias. O robô *iCub* já foi adotado por mais de 20 laboratórios em todo o mundo. A FIG. 4 mostra o robô *iCub*.

---

<sup>6</sup> <http://gal.istc.cnr.it/>

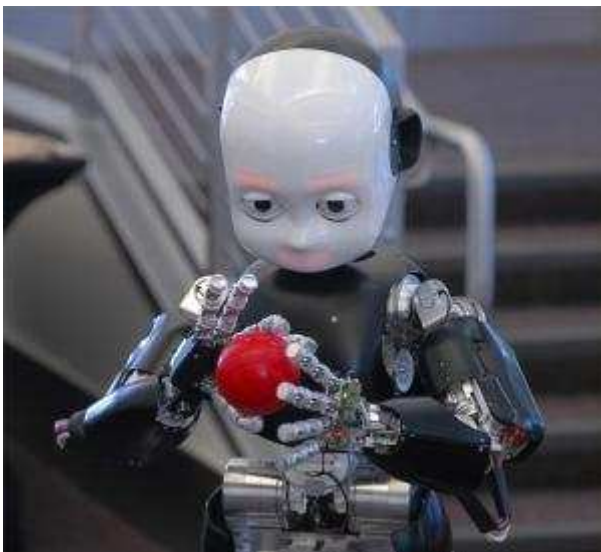


FIGURA 4 - O Robô *iCub*.<sup>7</sup>

Outro trabalho marcante nesse aspecto é o desenvolvido por Minato *et al.* (2007). Os pesquisadores desenvolveram uma “criança robô”, o CB2 (*Child Robot with Biomimetic Body*), para permitir o estudo do desenvolvimento cognitivo e inteligência social-sinérgica, por meio da robótica. CB2 é capaz de aprender por meio de observação e interação. Ainda não possui capacidade de comunicação verbal. Além de observar expressões faciais, o robô está aprendendo a andar. Em 2009, ele já era capaz de andar de modo lento, porém de modo mais coordenado do que no início de seu desenvolvimento, mas ainda requer auxílio humano (NARIOKA; HOSODA, 2011). Futuramente, os pesquisadores esperam que ele tenha a inteligência de uma criança de dois anos e seja capaz de falar sentenças básicas. A FIG. 5 mostra o robô CB2.

---

<sup>7</sup> <http://www.icub.org/>

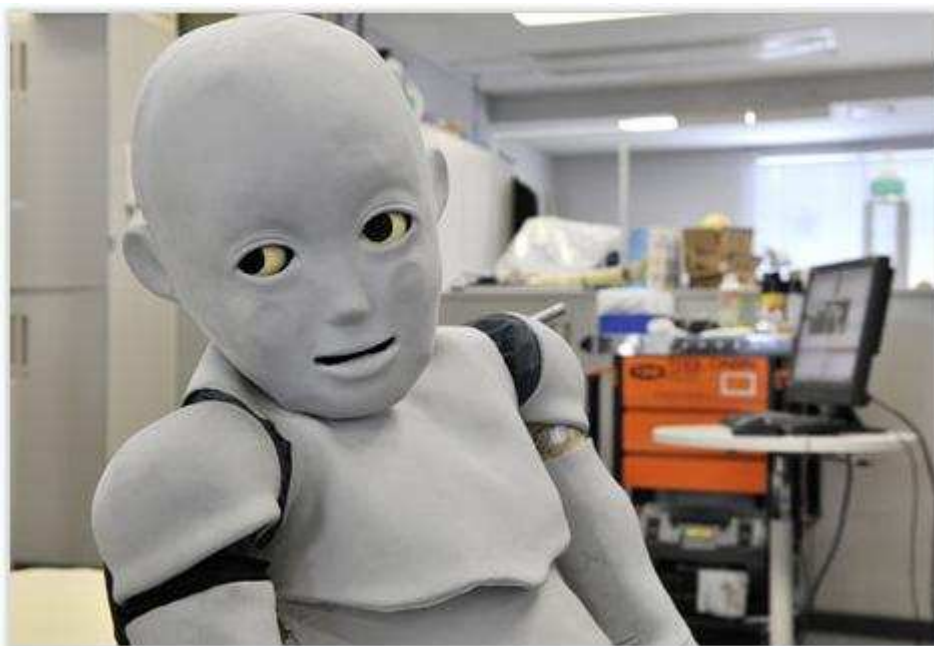


FIGURA 5 - O Robô CB2.<sup>8</sup>

O projeto **MirrorBot** (WERMTER; PALM; ELSHAW, 2005) é um projeto europeu que foca no desenvolvimento de robôs com aprendizado biomimético multimodal, inspirado nos neurônios espelho. Esse projeto obteve resultados importantes em relação modelos neurais baseados nos neurônios espelho e na organização modular do córtex cerebral associado a ações. Esses modelos foram utilizados em robôs inteligentes, em cenários complexos, em operações do tipo “pegar e colocar”. Segundo Wermter *et al.* (2005) a hipótese sob investigação no projeto é se os modelos baseados nos neurônios espelho podem produzir sistemas capazes de ações similares aos sistemas vivos.

Recentemente, a IBM (DHARMENDRA *et al.*, 2011) apresentou o primeiro processador cognitivo que, segundo nota da empresa, é projetado “para emular as capacidades do cérebro de percepção, sensação, ação, interação e cognição”. O processador, desenvolvido dentro do projeto Synapse (*Systems of Neuromorphic Adaptive Plastic Scalable Electronics*), ainda está em sua versão inicial, mas o objetivo é desenvolver um dispositivo que, inspirado nas novas teorias cognitivas e nas descobertas da Neurociência, seja capaz de analisar informações em paralelo e

---

<sup>8</sup> <http://crave.cnet.co.uk/gadgets/photos-freakish-cb2-child-robot-knows-if-youre-sad-49301884/>

reconfigurar-se dinamicamente em tempo de execução, aprendendo com a sua interação com o mundo.

Com base nos estudos sobre o papel da interação social no desenvolvimento da linguagem, tal como proposto por Tomasello (2008), Schulz *et al.* (2011) criaram experimentos com robôs que desenvolveram uma linguagem para permitir a comunicação entre eles e para permitir a resolução de tarefas. Como relatado por Schulz *et al.*, o experimento mostra que os robôs formaram um modelo interno independente do mundo, com base na experiência pessoal e geraram uma linguagem toponímica compartilhada, baseada na experiência compartilhada.

Esses relatos mostram a comprovação experimental dos resultados dos estudos abordados pela Linguística Cognitiva sobre a relação corpo/mente e sobre o surgimento da linguagem. No sentido oposto, é observado que as teorias formuladas pelos estudos sobre a cognição na área da linguística cognitiva ajudam na elaboração de sistemas computacionais e robôs mais adaptados a lidarem com problemas complexos.

Outra área da Computação que está ligada às pesquisas na área da linguagem humana é a de processamento em linguagem natural. O processamento em linguagem natural é uma subárea da Inteligência Artificial, que tem por objetivo o desenvolvimento de sistemas computacionais que comuniquem com o ser humano por meio da língua natural. Dentro dessa área, temos como sublinhas de pesquisa a extração de informação, a recuperação de informação e a tradução automática. Para a execução dessas tarefas, é necessário dispor de recursos léxicos bem desenvolvidos e de fácil manipulação por sistemas automatizados. Um recurso lexical é uma base de dados contendo os lexemas de uma ou mais línguas. Além disso, deve conter informações adicionais, tais como estrutura argumental e relações sintáticas e semânticas entre os lexemas. Os lexemas são organizados em uma rede, de forma que o significado de um lexema não pode ser entendido isoladamente, mas sim a partir de seu posicionamento na rede. As bases lexicais auxiliam, principalmente, na resolução de ambiguidades no sentido das palavras.

Existem vários recursos lexicais disponíveis, tais como o VERBNET (KIPPER *et al.*, 2006), o WordNet (FELLBAUM, 1998), e o PROBANK (PALMER; KINGSBURY; GILDEA, 2005). Neste trabalho, foi examinada a FrameNet que trata

os itens lexicais sob uma estrutura organizacional de cenários. Nas próximas seções, discutiremos a FrameNet, destacando suas características.

## 4.2 A rede semântica da FrameNet

Uma vez adotada a Semântica de Frames como uma teoria plausível para a organização da conceptualização humana, surge uma questão pertinente: é possível desenvolver uma representação do conhecimento que reflita essa organização? Caso seja possível, tal representação seria mais facilmente compreendida e manipulada, uma vez que naturalmente se encaixa na forma como as pessoas percebem o mundo.

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. Este recurso foi denominado FrameNet. A FrameNet (BAKER; FILLMORE; LOWE, 1998) é um recurso lexical *online* para a língua inglesa, implementado com base na Semântica de Frames e apoiado por evidências obtidas em *corpora*.

Segundo Salomão (2009a, p.174), o “projeto FrameNet surge do cruzamento da Semântica de Frames com a Lexicografia através da cooperação entre Fillmore e Atkins no princípio da década de noventa...”.

O projeto tem por objetivo registrar as possibilidades combinatórias semânticas e sintáticas (valências) de cada palavra predicativa (nomes, adjetivos e verbos), em cada um dos seus sentidos. O projeto utiliza sistemas computacionais para auxiliar na anotação das sentenças, para tabular os resultados e para exibir os resultados da anotação. Os conceitos básicos que fundamentam o projeto FrameNet são os conceitos de frames, relações entre frames, unidades lexicais (UL) e elementos de frames (EF). O conceito de frames já foi abordado. Definiremos aqui os outros conceitos.

Uma unidade lexical (UL) é o pareamento de uma palavra com um significado <palavra, significado> (RUPPENHOFER *et al.*, 2006, p.5). Ainda segundo o mesmo autor, cada sentido de uma palavra polissêmica pertence a um frame semântico diferente. O termo *palavra* deve ser interpretado, nesse contexto, como cada sentido de

um lema. Cada significado pertence a uma estrutura de informação, projetada de acordo com a visão de frames, que descreve uma determinada situação, um objeto ou um evento e os elementos participantes. Essa estrutura de informação é o frame.

Atualmente, a FrameNet possui mais de 10.000 unidades léxicas (UL) anotadas, sendo que 5249 foram finalizadas durante o projeto FrameNet II, 1276 criadas durante o projeto FrameNet I, 4273 foram criadas, mas ainda não foram associadas com evidência de *corpus*. As ULs evocam os frames. Por exemplo, a ocorrência 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*). A FIG. 6 mostra parte do frame COMMERCE\_BUY, obtida em uma consulta ao site da FrameNet<sup>9</sup>, as relações do frame COMMERCE\_BUY com outros frames e informações sobre as unidades léxicas, obtidas na mesma consulta.

---

<sup>9</sup> <http://framenet.icsi.berkeley.edu/>

## Commerce\_buy

### Definition:

These are words describing a basic commercial transaction involving a buyer and a seller exchanging money and goods, taking the perspective of the buyer. The words vary individually in the patterns of frame element realization they allow. For example, the typical pattern for the verb BUY: BUYER buys GOODS from SELLER for MONEY. Abby bought a car from Robin for \$5,000.

### FEs:

#### Core:

##### Buyer [Byr]

The **Buyer** wants the **Goods** and offers **Money** to a **Seller** in exchange for them.

**Jess BOUGHT** a coat.

Lee **BOUGHT** a textbook **from Abby**.

##### Goods [Gds]

The FE **Goods** is anything (including labor or time, for example) which is exchanged for **Money** in a transaction.

Only one winner **PURCHASED the paintings**

#### Non-Core:

##### Duration [Dur]

##### Semantic Type

The length of time that the **Goods** are (or have been) in the **Buyer**'s possession. (This FE is generally only relevant for the LU's *rent.v* and *lease.v* which imply a temporary change of possession.)

Our neighbors have been **RENTING** their house **for the last ten years**.

Inherits From: **Getting**  
 Is Inherited By: **Renting**  
 Subframe of:  
 Has Subframes:  
 Precedes:  
 Is Preceded by:  
 Uses:  
 Is Used By: **Importing, Shopping**  
 Perspective on: **Commerce\_goods-transfer**  
 Is perspectivized in:  
 Is Causative of:  
 See Also:

## Lexical Units

*buy.v, purchase.v, purchase\_((act)).n*

Created by miriamp on Thu Jul 12 12:38:02 PDT 2001

### Lexical Unit Information

Lexical Unit	LU Status	Lexical Entry Report	Annotation Report
<b>buy.v</b>	Finished_Initial	LE	Anno
<b>purchase.v</b>	Finished_Initial	LE	Anno
<b>purchase_((act)).n</b>	Finished_Initial	LE	Anno

FIGURA 6 – Relações do frame COMMERCE\_BUY com outros frames e informações sobre as unidades léxicas.

FONTE – FrameNet de Berkeley.

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*. As anotações destacam as ocorrências dos EFs nas sentenças. Note, no destaque com cores, a ocorrência nas palavras associadas aos EFs. Cada cor destaca um EF diferente. É possível observar também uma divisão entre os EFs, que separa os elementos centrais ao frame (*core*) e os não centrais (*non-core*).

Na verdade, segundo Ruppenhofer *et al.* (2006), os EFs são divididos em três grupos: centrais (*core*), periféricos e extra-temáticos. Esse último grupo denota os EFs que se acredita não pertencerem conceitualmente aos frames com os quais coocorrem e aparecem. As palavras que aparecem nas sentenças com o fundo preto são as unidades léxicas que ativam o 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*. Isso os diferencia de outras bases lexicais, como os tesouros. Os tesouros se baseiam em relações lexicais, tais como sinonímia, antonímia, hiponímia e meronímia, relações essas que ocorrem apenas entre os lexemas. Já as relações da FrameNet ligam frames, que são estruturas de conhecimento. Portanto, as relações introduzidas na FrameNet são relações semânticas.

As relações semânticas entre frames são assimétricas formando um grafo direcionado. Na relação de **herança**, um frame possui o papel de pai e outro de filho. É a relação mais forte entre frames. O frame filho é uma especialização do frame pai e herda os elementos do frame pai, sendo que tudo que é estritamente verdadeiro no frame pai, também deve ser verdadeiro no frame filho (RUPPENHOFER *et al.*, 2006). Por exemplo, o frame RENTING herda do frame COMMERCE\_BUY e, portanto, tudo que é verdadeiro para o frame COMMERCE\_BUY é verdadeiro para o frame RENTING, como a transferência de propriedade de um bem. Mas o frame filho acrescenta algumas propriedades à cena, como o fato dessa transferência ser temporária. Ainda segundo Ruppenhofer *et al.* (2006), o que não é estritamente semântico por natureza, como, por exemplo, a relação “ver também” (*see also*), não é herdado.

A relação **perspectiva\_sobre** indica a existência de pelo menos dois diferentes pontos de vista sobre um frame neutro. A sentença “*Carla bought the computer from Sally for \$100*” apresentada por Petruck (1996), evoca o frame de transação comercial de bens, sob a perspectiva do comprador, o que resulta no frame de compra comercial.



Da mesma forma, a sentença “*Sally sold the computer to Carla for \$100*” evoca o mesmo frame, sob a perspectiva do vendedor. Os frames que impõem uma perspectiva são ligados por meio do frame neutro de perspectiva, sendo que esse último, geralmente, não é lexicalizado. A relação de perspectiva\_sobre é considerada um refinamento da relação semântica de uso.

A relação de **subframe** ocorre quando existe um frame complexo composto por uma sequência de estados, sendo que cada estado pode ser um frame distinto (RUPPENHOFER *et al.*, 2006). Cada frame da sequência é ligado ao frame complexo por meio da relação de subframe. A denominação subframe pode ser confusa, uma vez que se aproxima da denominação usada em outros formalismos para a herança (classe/subclasse). Acreditamos que uma denominação melhor seria obtida da relação parte-todo da merologia. O processo criminal é um frame adequado para ilustrar esse caso. Para cada etapa do processo (a prisão, o indiciamento, o julgamento, o sentenciamento e o aprisionamento), existe um frame distinto que está relacionado com o processo criminal pela relação de subframe. Além disso, existe uma relação de ordem entre os subframes. Segundo Ruppenhofer *et al.* (2006), participantes de um subframe não necessitam ocorrer em outro subframe. Nesse caso, o juiz não aparece no aprisionamento.

A relação **precede** ocorre entre dois frames concomitantemente com a relação de subframe, como indicação de que um frame precede outro na sequência de eventos. É importante ressaltar que, de acordo com Ruppenhofer *et al.* (2006), essa é a única relação que permite ciclos, sendo que todas as outras são acíclicas. Por exemplo, o frame ESTAR ACORDADO precede o frame ADORMECER, que, por sua vez, precede o frame ESTAR ACORDADO.

Outra importante relação é a que se estabelece entre frames estativos, causativos e incoativos. Essa distinção é importante, porque, segundo Ruppenhofer *et al.* (2006), comparações *cross-linguísticas* mostraram que existem línguas que diferenciam os lexemas por derivação morfológica e que existem alguns lexemas que só se aplicam em um dos frames. Um exemplo fornecido pelos autores é o relacionado com uma mudança em escala:

(1) (*raise.v*) *Billie Blount raised taxes on farmers 18 times in 2002!* (causativo)

(2) (*rise.v*) *During the Elizabethan age, there was an increased emphasis on genealogy in the heralds' work as the gentry class **rose** in importance.* (incoativo)

(3) (*high.a*) *Most fish from lakes is too **high** in mercury.* (estativo)

As diferenças nos enunciados justificam a criação de uma rede de frames relacionados por relações semânticas do tipo *causativo\_de* e *incoativo\_de*. A relação *causativo\_de* relaciona um frame causativo com um estativo ou incoativo, enquanto a relação *incoativo\_de* relaciona um frame incoativo com um estativo.

A relação de **uso** ocorre quando parte de uma cena evocada por um frame faz uma referência a outro frame (RUPPENHOFER *et al.*, 2006), podendo usá-lo como fundo. A cena relacionada com **PROFISSIONAIS MÉDICOS** usa a cena de **CURA**, sem ser um subframe e sem ter uma relação de herança com a última.

Assim sendo, ao apresentar um frame, a FrameNet apresenta também as relações do frame com outros frames, assim como informações sobre as unidades léxicas associadas. Alternativamente, as relações entre os frames podem ser vistas graficamente, por meio de visualizador gráfico existente no próprio site, o FrameGrapher. A FIG. 7 mostra graficamente as relações entre frames, partindo do frame **COMMERCE\_BUY**.

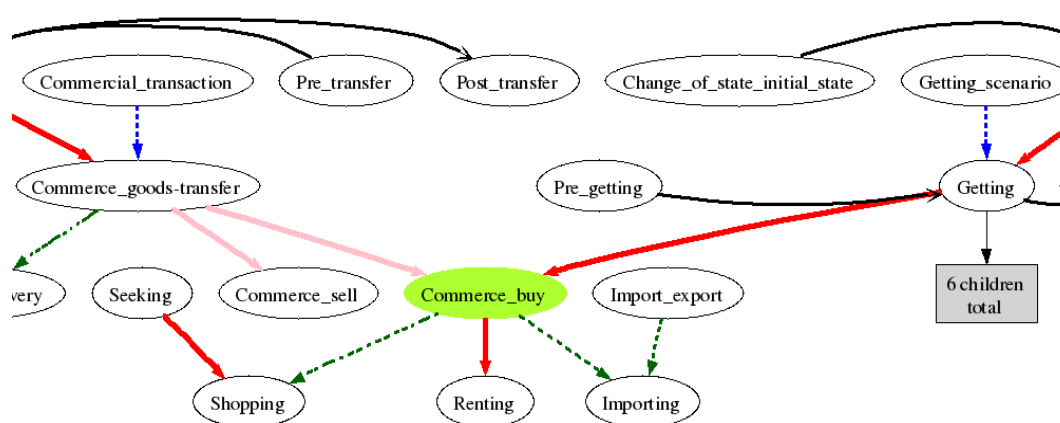


FIGURA 7 – Relações do frame COMMERCE\_BY com outros frames apresentadas de forma gráfica.

FONTE – FrameNet de Berkeley.

As setas em vermelho indicam herança (**GETTING** <--> **COMMERCE\_BUY**). As setas verdes indicam relação de uso (**IMPORTING** <--> **COMMERCE\_BUY**). As setas em rosa indicam relação de perspectiva (**COMMERCE\_GOODS-TRANSFER** <--> **COMMERCE\_BUY**). As setas em azul indicam relação de subframe (**GETTING\_SCENARIO** <--> **GETTING**). As setas em preto indicam relação de precedência (**PRE\_GETTING** <--> **GETTING**).

A construção de um léxico semântico como a FrameNet é uma tarefa árdua e complexa. O projeto FrameNet de Berkeley tem 13 anos e já cadastrou aproximadamente 1014 frames, mas não possui uma perspectiva de quantos ainda serão cadastrados. Várias etapas estão envolvidas no processo e muitas delas são subjetivas e sem apoio de automatização.

O desenvolvimento de um frame na FrameNet, segundo informações do sítio na web e de Ruppenhofer *et al.* (2006), começa a partir de uma proposição, sustentada por observações de evidências no *corpus* subjacente ao projeto (o projeto FrameNet II utiliza atualmente o *British National Corpus*, com 100 milhões de palavras e está adicionando o *LDC North American Newswire corpora*). Juntamente com a proposição do frame, algumas palavras são sugeridas para ilustrá-lo. Em seguida, alguns elementos de frames que, se acredita sejam comuns a essas palavras, são sugeridos. Os elementos

de frame são obtidos de forma indutiva, pela observação dos elementos que ocorrem nas sentenças do *corpus* relacionadas com a UL.

Durante esse processo, que atualmente segue uma metodologia mais formal do que no início do projeto, um frame proposto originalmente pode se dividir em dois ou mais. Isso ocorre em função das exigências de agrupamento, estabelecidas pelo projeto. Dentre os critérios de agrupamento pode-se citar a exigência de que todas as unidades léxicas devem possuir o mesmo número e o mesmo tipo de elementos de frames em todos os contextos, implícitos e explícitos. Os desenvolvedores da FrameNet, têm sistematicamente separado frames causativos dos incoativos. O processo de análise das LU e dos EF no estabelecimento de um novo frame está explicado detalhadamente em Ruppenhofer *et al.* (2006).

O próximo passo é o processo de anotação das sentenças, adicionando informações sintáticas e semânticas. Esse processo, atualmente, é essencialmente manual, mas existem pesquisas no sentido da criação de sistemas que automatizam essa tarefa, partindo de exemplos anotados manualmente (GILDEA; JURAFSKY, 2002; LITKOWSKI, 2004). De forma geral, segundo o sítio da FrameNet, as etapas para anotação são as seguintes:

- As sentenças onde ocorrem uma determinada palavra são extraídas do *corpus* e disponibilizadas para a realização das anotações.
- Os anotadores selecionam as frases que identificam um papel semântico determinado e anotam essas frases com os elementos de frame.
- Em seguida, por meio de um processo semiautomático, são adicionadas às sentenças anotadas camadas de informações correspondentes à função gramatical (GF) e à informação sintagmática (PT).

A FIG. 8 mostra um editor de *subcorpus* exibindo as anotações de um conjunto de sentenças relacionadas com a UL *want.v*, sendo que a sentença 10 é colocada em destaque. Podem-se notar as três camadas de informação associadas à sentença: FE, GF e PT. Essas camadas são sempre apresentadas pela ferramenta, mas pode-se acrescentar um número ilimitado de outras camadas, cuja apresentação pode ser selecionada manualmente.

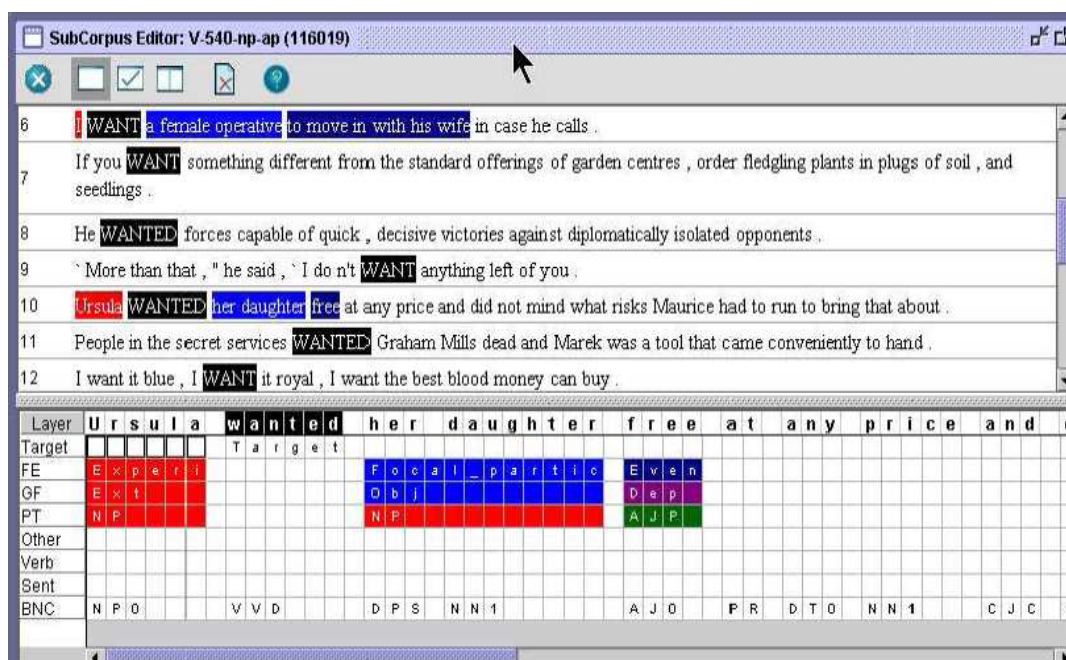


FIGURA 8 – Editor de *subcorpus* com as camadas de anotações.

FONTE – RUPPENHOFER *et al.*, 2006. p. 20.

As aplicações para um léxico com essa estrutura são diversas. Segundo informações do site da FrameNet<sup>10</sup>, pesquisadores da área de processamento da linguagem natural (PLN) têm usado a FrameNet para:

- Eliminação de ambiguidades do significado de um termo;
- tradução por máquina;
- extração de informações;
- sistemas automáticos de resposta.

<sup>10</sup> <http://framenet.icsi.berkeley.edu/>

### 4.2.1 Tipos Semânticos

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. O objetivo dos tipos semânticos é adicionar informação tipológica aos elementos de frames. O tipo semântico *sentient* (ser consciente) é atribuído para o EF Perpetrador no frame de PIRATARIA. Desta forma, limita-se o conjunto de possíveis “perpetradores” às entidades conscientes, mais particularmente, um humano e eliminam-se algumas sentenças com significado correlato, porém, distinto.

A FIG. 9 apresenta a hierarquia de tipos semânticos da FrameNet. Os tipos semânticos foram portados pelos autores dessa pesquisa para a notação OWL (*Web Ontology Language*) por meio do software de edição de ontologia Protégé (<http://protege.stanford.edu>), o que permitiu a exibição da hierarquia de tipos graficamente.

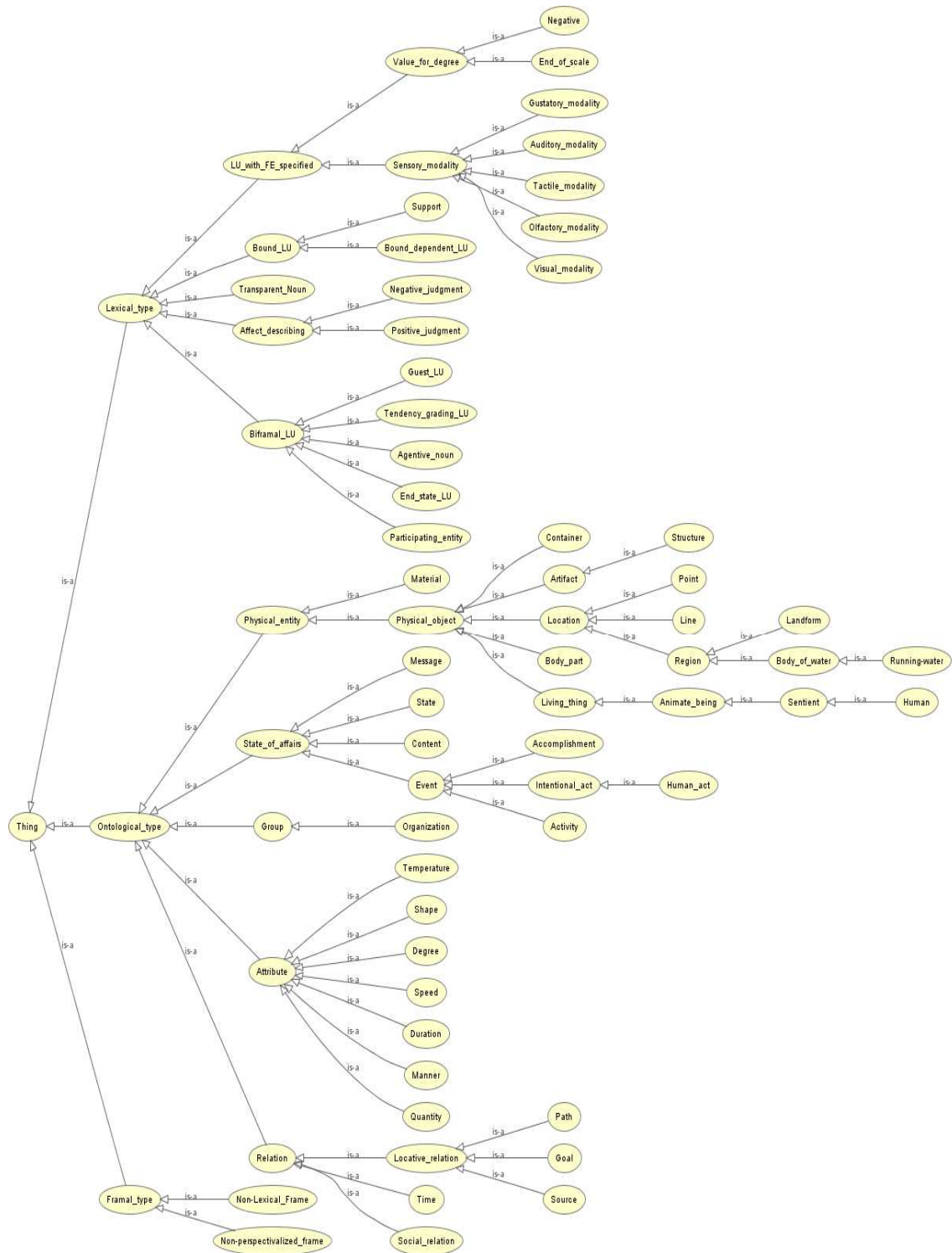


FIGURA 9 - Hierarquia topo de tipos semânticos da FrameNet.

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, UL, ou mesmo a um frame. No entanto, o conjunto dos tipos semânticos pode ser

considerado uma ontologia inicial de nível topo e pode ser, futuramente, objeto de formalização.

#### 4.2.2 Anotação de relações metafóricas na FrameNet

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

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

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

Nos casos desses critérios terem sido satisfeitos, então uma UL seria anotada no domínio alvo da metáfora e deveria ser criada uma relação Frame-a-Frame metafórica entre os domínios, mas, segundo Ruppenhofer *et al.* (2010, p. 91), esses *links* ainda não



foram adicionados à FrameNet. Apenas foi criado, experimentalmente um frame metafórico, denominado de “EMOTION\_HEAT” que representa o *blend* semântico da metáfora EMOÇÃO É CALOR. Portanto, conclui-se que existe uma proposta de tratamento de metáfora na FrameNet, porém essa proposta ainda não foi implementada completamente. Mesmo essa proposta não considera a anotação simultânea dos domínios alvo e fonte, uma vez que, segundo Ruppenhofer *et al.* (2010, p. 91), isso mereceria, por si só, todo um programa de pesquisa.

### 4.3 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. Segundo Welty e Guarino (2001), o termo foi cunhado em 1613, por Rudolf Goclenius e, aparentemente de forma independente, por Jacob Lorhard. Segundo Mora (1963), o termo ontologia é mencionado brevemente por Goclenius, na página 16 do *Lexicon philosophicum, quo tanquam clave philosophiae fores aperiuntur, Informatum opera studio Rodolphi Goclenii*<sup>11</sup>, da seguinte forma: “*ontologia, philosophia de ente*”, sendo esta a única menção ao termo. Coube a Christian Wolff, ainda segundo Mora, a popularização da palavra ontologia nos círculos filosóficos: “A palavra aparece no título de seu *Philosophia prima sive ontologia methodo scientifica pertractata, qua omnes cognitionis humanae principia continentur*<sup>12</sup>, publicado em 1730.”

Inspirados pelos estudos deste ramo da filosofia, os pesquisadores da área de representação do conhecimento se apropriaram do termo *ontologia* para designar representações que descrevem formalmente os objetos de um domínio de interesse e seus relacionamentos. Para se entender como ocorreu essa apropriação na área da

---

<sup>11</sup> Uma introdução à filosofia, como uma chave que abriu a porta da filosofia, informada pela obra de Rudolph Goclenii.

<sup>12</sup> Filosofia primeira ou método científico de tratamento do ser, que contém todos os princípios do conhecimento humano.

representação do conhecimento, descreveremos a seguir a evolução desta área ligada à Inteligência Artificial.

A representação do conhecimento é um ramo da pesquisa científica que abrange diversas áreas de investigação, como a filosofia, a ciência da informação, as ciências cognitivas e a Ciência da Computação. Nessa última área, a representação do conhecimento ganhou grande importância, a partir do surgimento da Inteligência Artificial, marcada pela publicação, em 1950, do artigo escrito por Alan Turing “*Computing Machinery and Intelligence*”, em que ele define o teste de Turing, que é usado ainda hoje como critério para decidir se um dispositivo computacional exibe um comportamento inteligente.

Outro evento importante, considerado por muitos como o que deu início a essa área de pesquisa, foi a conferência que aconteceu no verão de 1956, no Dartmouth College, organizada por John McCarthy e Marvin Minsky. A conferência reuniu pesquisadores interessados no desenvolvimento de sistemas computacionais que resolvessem problemas para os quais não existissem soluções algorítmicas. Na chamada para a conferência, John McCarthy introduziu o termo *inteligência artificial* (IA) que passou a caracterizar a área:

*“We propose that a 2 month, 10 man study of **artificial intelligence** be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves.”* (MCCARTHY *et al.*, 1955).

Nos primeiros anos da IA, o foco era no desenvolvimento de algoritmos de busca de soluções, por meio apenas de manipulações simbólicas, mas logo ficou claro que, para tornar essa busca eficiente, era necessário adicionar ao sistema uma grande quantidade de conhecimento sobre o domínio do problema. A partir de 1969 (RUSSELL; NORVIG, 2010), os pesquisadores passaram a estudar formas de inserir e manipular o conhecimento nos sistemas computacionais.

No princípio, essa representação do conhecimento era feita por meio da sintaxe da lógica matemática, tendo por base a proposta por Alfred Tarski, que interpreta as

sentenças da lógica por meio de um mapeamento para as estruturas da teoria de conjuntos. No entanto, à medida que a implementação de sistemas mais complexos era perseguida, as fragilidades dessa forma de representação se faziam mais evidentes. O mais óbvio foi a ineficiência na execução de inferências, à medida que a base de conhecimento crescia, em função da explosão combinatória<sup>13</sup>.

Outro problema se refere à dificuldade de expressar certos tipos de conhecimentos no formalismo lógico, como o conhecimento probabilístico, o conhecimento *fuzzy*, o raciocínio não monotônico<sup>14</sup> e o conhecimento estruturado hierárquico.

Um terceiro problema é a dificuldade de representar os efeitos de uma ação em lógica, sem representar explicitamente um grande número de condições que não são afetadas pela ação. Esse problema é conhecido como *frame problem* (SHANAHAN, 2009).

Esses problemas motivaram os pesquisadores de IA a buscarem formalismos de representação de conhecimento que contornassem essas deficiências. Houve, então, o período de ascensão das linguagens de representação de conhecimento estruturado: notações gráficas de nós e relações. Ross Quillian em sua tese de Doutorado em 1966 propôs o uso da notação gráfica, denominada de *redes semânticas*, inspirado na memória associativa humana. As redes semânticas são compostas por nós ligados por arestas, onde os nós representam os conceitos e as arestas representam as relações. Em 1975, Marvin Minsky propôs a notação de *frames* que possui um grau maior de estruturação. Os *frames* serão discutidos mais detalhadamente na próxima subseção.

Esses e outros formalismos estruturados começaram a ser criticados pela sua falta de semântica bem definida. Em 1975 Willian Woods publicou um artigo cujo título continha uma pergunta “What’s in a Link?” (O que existe em uma ligação?). Neste artigo o autor questiona a falta de uma semântica clara para as notações estruturadas. Em 1983 Ronald J. Brachman publica uma resposta no artigo intitulado “*What IS-A is and isn't. An Analysis of Taxonomic Links in Semantic Networks*”, em que ele propõe

---

<sup>13</sup> Na verdade esse não é um problema inerente à lógica, mas de todo formalismo com a mesma capacidade de representação da lógica clássica.

<sup>14</sup> O raciocínio não monotônico ocorre quando necessitamos eliminar conhecimentos prévios em função da aquisição de novos conhecimentos.

uma formalização para as representações gráficas: é o nascimento da *lógica de descrições*.

A lógica de descrições é um subconjunto da lógica clássica, voltada para formalização, principalmente de taxonomias, e cuja execução em dispositivos computacionais é bem mais eficiente do que a lógica clássica. É preciso destacar que, como é um subconjunto da lógica clássica, a de descrições não possui o mesmo poder de representação que a clássica. Ou seja, existem conhecimentos que podem ser expressos na lógica clássica que não são passíveis de serem representados na lógica de descrições.

A partir dos anos 90, as representações estruturadas, formalizadas pela lógica de descrições, passaram a ser usadas para representar os conceitos e terminologias que ocorrem no mundo, ou seja, em um domínio específico: é o surgimento das ontologias na Ciência da Computação, assunto a que retomaremos mais adiante.

Sumarizando, as representações estruturadas, particularmente os frames de Minsky, quando formalizadas pela lógica das descrições, são usadas para representar conceitos gerais, dando origem às ontologias de nível topo, e conceitos específicos, dando origem às ontologias de domínio. Outras variações notacionais surgiram, como a OWL (*Web Ontology Language*), mas a teoria subjacente a essas proposições permanece a mesma. Apesar de o termo *frames* ser comumente usado também na área da LC, na computação ele designa um objeto distinto, porém com algumas semelhanças com o da LC. Para caracterizá-lo mais claramente, a próxima subseção o descreve com mais detalhes.

### **4.3.1 Frames na Ciência da Computação**

O termo “frames” ocorre em outras áreas da ciência, como a da Inteligência Artificial (IA), em que possui um significado semelhante ao proposto por Fillmore. Para Minsky (1975), pesquisador na área de IA e proponente do termo frame para a área, um frame é uma representação de conhecimento estruturado, que captura o conhecimento

estereotipado de um determinado conceito. Cada frame possui um conjunto de elementos denominados de escaninhos (*slots*). Os escaninhos possuem valores que podem ser, inclusive, outros frames, formando um sistema de frames.

A representação de conhecimento através de frames está baseada na suposição de que as pessoas não analisam situações novas a partir do zero e depois constroem novas estruturas de conhecimento para descrever essas situações. Pelo contrário, elas possuem disponível na memória uma grande coleção de estruturas representando sua experiência anterior com objetos, localizações, situações e pessoas. Para analisar uma nova experiência, elas evocam as estruturas apropriadas e depois as preenchem com os detalhes do evento atual.

A principal característica desse método de representação é a capacidade de organização e estruturação do conhecimento.

Tipicamente, o frame descreve uma classe de objetos, tais como PESSOA ou AUTOMÓVEL, consistindo em uma coleção de escaninhos que descrevem os aspectos dos mesmos. Eles são preenchidos por outros quadros descrevendo outros objetos. Associado a cada escaninho poderá haver um conjunto de condições que deverá ser satisfeito por qualquer objeto a ele destinado. Cada escaninho também poderá ser preenchido com um valor padrão, para que, na ausência de informação específica, seja assumido esse valor padrão que representa o caso prototípico. A FIG. 10 mostra um exemplo simplificado de um sistema de frame relacionado com o termo *jogador de futebol*.

```

frame pessoa
    é-um: mamífero
    cardinalidade: 6.8 bilhões
    locomoção: bípede terrestre

frame adulto masculino
    é-um: pessoa
    cardinalidade: 2.5 bilhões
    altura: entre 1,5 a 1,9 metros

frame jogador de futebol
    é-um: pessoa
    cardinalidade: 50000
    time:

frame Zico
    instância: jogador de futebol
    altura: 1,7 metros
    time: flamengo

```

FIGURA 10 – Sistema de frame envolvendo jogador de futebol.

Apesar das semelhanças, os frames propostos por Minsky tratam conceitos estereotipados, mas não cenas. Os *Scripts* (roteiros) propostos por Schank e Abelson (1977) guardam uma maior semelhança com os frames de Fillmore. Os roteiros representam uma cena estereotipada, sendo o exemplo clássico o roteiro do restaurante.

Na verdade, o roteiro é uma especialização dos frames de Minsky. Enquanto os frames são estruturas de finalidade geral para representação de aglomerados comuns de fatos, os roteiros são estruturas com capacidade de explorar propriedades específicas de um domínio particular. A FIG. 11 mostra o roteiro para restaurante. O roteiro utiliza primitivas conceituais propostas por Schank (1975).

Roteiro: RESTAURANTE

Trilha: Cantina

Acessórios:

Mesas

Cardápio

F=Comida

Cheque

Dinheiro

Papéis:

S= Cliente

G=Garçon

C=Cozinheiro

M=Caixa

D=Dono

### Cena 1: Entrada

S PTRANS S no restaurante

S ATTEND olhos na mesa

S MBUILD onde sentar

S PTRANS S para mesa

S MOVE S para posição

### Cena 2: Pedido

(Cardápio na mesa)

S PTRANS cardápio para S

(S pede cardápio)

S MTRANS sinalizar para G

G PTRANS G para mesa

S MTRANS 'precisar

cardápio' para G

G PTRANS G para cardápio

Condições de Entrada:

S está com fome.

S tem dinheiro

G PTRANS G para a mesa

G ATRANS para cardápio S

Resultados:

S tem menos dinheiro.

D tem mais dinheiro.

S não está com fome.

S está satisfeito (opcional).

S MTRANS G para a mesa

\*S MBUILD escolha de F

S MTRANS sinalizar para G

G PTRANS G para a mesa

S MTRANS 'Eu quero F' para G

G PTRANS G para C

G MTRANS (ATRANS F) para C

C MTRANS 'não há F' para G

G PTRANS F para S

(voltar para \*) ou

(voltar para cena 4 na trilha não

pagar)

C DO (preparar roteiro F)

para cena 3

### Cena 3: Refeição

C ATRANS F para G

G ATRANS F para S

S INGEST F

### Cena 4: Saída

G MOVE (preencher cheque)

G PTRANS G para S

G ATRANS cheque para S

G ATRANS gorjeta para G

S PTRANS S para M

S ATRANS dinheiro para M

S PTRANS S para fora do restaurante (trilha não pagar)

**Legenda:**

ATTEND- enfocar o órgão de sentido em um estímulo  
 ATRANS - transferência de uma relação abstrata.  
 INGEST - ingestão de um objeto  
 MBUILD - construir nova informação da antiga  
 MTRANS - transferência de informação mental  
 MOVE - movimento da parte de um corpo de alguém  
 PTRANS - transferência de uma relação física de um objeto

FIGURA 11 – Roteiro para um restaurante.

FONTE – SCHANK e ABELSON, 1977.

Os eventos descritos num roteiro formam uma gigantesca cadeia causal. O início da cadeia é o conjunto de condições de entrada que permite a ocorrência dos primeiros eventos do roteiro. O final da cadeia é o conjunto de resultados, em consequência da execução do roteiro e que poderão levar à ocorrência de eventos posteriores. Dentro da cadeia, os eventos estão ligados tanto aos anteriores que os tornam possíveis, como aos posteriores que eles, por sua vez, possibilitam. Se um roteiro em particular for apropriado a uma dada situação, pode ser muito útil para prever a ocorrência de eventos que não foram explicitamente citados. Os roteiros também podem ser úteis para indicar como os eventos que foram citados se relacionam entre si. Qual é a ligação entre alguém pedir um filé e alguém comer um filé?

Os roteiros descrevem cenas de modo a serem úteis à atuação de dispositivos computacionais, mas não buscam o entendimento da cena sob o ponto de vista linguístico. Eles descrevem os elementos da cena, mas não os elementos linguísticos que são utilizados pelas pessoas nas cenas.

### **4.3.2 Ontologia na Ciência da Computação**

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, 1992).



Desde os anos 70 (RUSSELL; NORVIG, 2010), a IA já se preocupava com a organização e manipulação de bases de conhecimento, mas, a partir da década de 90, houve um impulso para a criação de bases de conhecimento compartilháveis e reutilizáveis. Esse impulso deveu-se à percepção de que os problemas complexos deveriam ser atacados por diferentes sistemas, atuando de forma cooperativa/competitiva em uma rede de multiagentes<sup>15</sup>. Além disso, o esforço para a criação de uma base de conhecimento pode ser muito caro e a reutilização e o compartilhamento das bases podem baratear os custos. No entanto, o compartilhamento de bases de conhecimento só pode ocorrer se houver um claro entendimento dos “*comprometimentos ontológicos*” associados às bases. Por comprometimentos ontológicos entendem-se as escolhas que levaram uma pessoa a selecionar um determinado conjunto de conceitos em vez de outro (VALENTE, 1995, p.34).

Ou seja, os comprometimentos ontológicos determinam o que é relevante em um determinado domínio para que seja representado em uma base de conhecimento. Ao escolher representar o objeto “livro” por meio de um predicado em uma linguagem lógica, o desenvolvedor está se “comprometendo” com a existência desta propriedade no domínio, ou seja, a de que existem objetos que possuem a propriedade de serem livros. A própria escolha da linguagem de representação revela alguns comprometimentos ontológicos. O uso da lógica de primeira ordem revela a crença com a existência de fatos, objetos e relações (RUSSELL; NORVIG, 2010, p.166). O compartilhamento dos mesmos comprometimentos ontológicos é que torna possível a comunicação entre dois agentes (humanos ou não), uma vez que estabelece que os mesmos objetos foram compreendidos pelos agentes envolvidos.

O registro explícito e formal dos comprometimentos ontológicos é o que tem sido denominado, na maioria das vezes, de ontologia no âmbito da Inteligência Artificial. Enquanto as bases de conhecimento tradicionais acumulavam o conhecimento necessário para atender uma aplicação específica, uma ontologia deveria ter a propriedade de ser usada em várias aplicações e em aplicações distribuídas, como

---

<sup>15</sup> Agentes na Inteligência Artificial é uma forma de classificar os componentes de um sistema que reagem às percepções recebidas do ambiente e que possuem autonomia e objetivos próprios. Dentro dessa visão, várias entidades, naturais ou artificiais podem ser classificadas como agentes, indo desde certos tipos de programas até pessoas.

sistemas multiagentes, fornecendo o suporte necessário para a troca de informação entre agentes.

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 "*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, segundo Gruber,

*“... os objetos, conceitos, e outras entidades que se assume existir em alguma área de interesse e as relações que ocorrem entre eles. Uma conceptualização é uma visão abstrata e simplificada do mundo que nós desejamos representar para algum propósito. Toda base de conhecimento, sistema baseado em conhecimento, ou agente atuando no nível do conhecimento são comprometidos com alguma conceptualização, explícita ou implicitamente”. (tradução da autora) (GRUBER, 1993, p.1).*

As ontologias tiveram um grande impulso nos últimos anos e sofreram diversas ampliações, em função da aplicação em inúmeros sistemas computacionais. No entanto, os fundamentos teóricos, que possuem raízes na lógica de descrições permanecem inalterados. Esses fundamentos serão descritos na próxima seção.

### **4.3.3 Teoria, conceitos, relacionamentos e atributos de uma ontologia**

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. A FIG. 12 ilustra o papel da ontologia. Note que a limitação não impõe uma interpretação exata, uma vez que isso não é possível no caso geral.

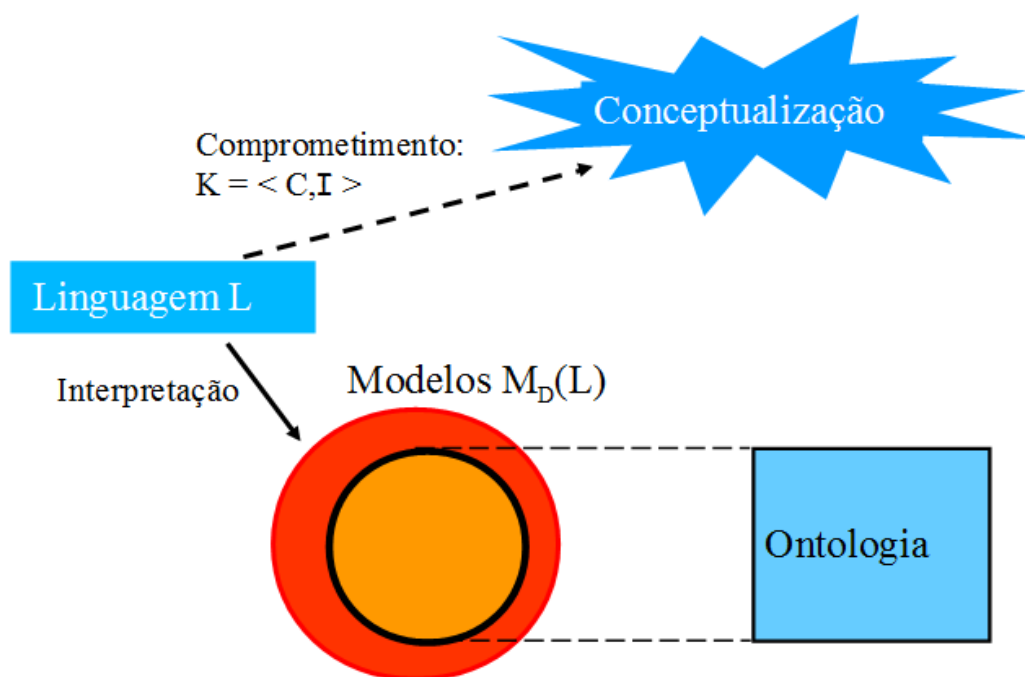


FIGURA 12 - Relação entre conceptualização, linguagem e ontologia.

FONTE – GUARINO, 1998.

O entendimento sobre conceitos na ontologia é equivalente à ideia clássica de categoria, ou seja, um conceito é uma categoria cuja pertinência de um elemento é definida por condições mínimas e necessárias. Uma ontologia é uma rede de conceitos ligados por relações.

Uma ontologia não está ligada a nenhuma linguagem de representação específica, mas algumas linguagens e notações lógicas foram desenvolvidas com o intuito de representar ontologias. A notação mais utilizada é a lógica de descrições, derivada da linguagem KL-ONE (BRACHMAN; SCHMOLZE, 1985). KL-ONE foi desenvolvida para formalizar as linguagens de representação estruturadas como os frames de Minsky. A lógica das descrições é um subconjunto computável da lógica. Ou seja, é sempre possível verificar se uma sentença pode ser inferida de outro conjunto de sentenças. Os elementos que são expressos na lógica de descrições são os seguintes:

- **Indivíduos** – representados por constantes. Exemplo: Alexandra
- **Conceitos** - representados por predicados unários. Exemplo: Estudante(Alexandra)
- **Papéis** - representados por predicados binários. Exemplo: Orienta(Margarida, Alexandra)

Existem também os **construtores**, que são usados para formar novos conceitos e papéis, a partir de conceitos e papéis prévios. Eles devem obedecer à restrição de serem decidíveis (passíveis de serem verificados computacionalmente) em relação à veracidade e à subjugação (taxonomia). Assim, podemos usar os conceitos de “estudante” e “pesquisador” para definir o conceito de “estudante\_doutorado”, usando o operador de interseção de conceitos ( $\cap$ ):  $\text{estudante\_doutorado} \subseteq \text{estudante} \cap \text{pesquisador}$ .

A lógica de descrições é, na verdade, uma família de lógicas. Cada lógica particular recebe um nome, de acordo com os operadores que possui. Por exemplo, a lógica de descrições mais simples é a *ALC*, que possui os operadores para conceitos de união, interseção e de negação restrito a alguns casos. A lógica de descrições que é a base da Web semântica e da OWL é a *SHOIN*, que possui operadores para papéis transitivos (S), para papéis hierárquicos (H), para nominais (O), para papéis inversos (I, exemplo:  $\text{temPai} \times \text{temFilho}$ ), e para restrições de número (N, exemplo:  $\text{tem uma mãe}$ ).

Uma base de conhecimento expressa em lógica de descrições é separada em duas seções: a *TBOX* e a *ABOX*. A seção *TBOX* (*Terminological BOX*) define a terminologia que será usada, ou seja, os conceitos e papéis do domínio em questão. Já

ABOX (*Assertion BOX*) registra as declarações sobre os indivíduos do domínio, usando o vocabulário estabelecido pela TBOX. A FIG. 13 mostra um exemplo de base de conhecimento em lógica de descrições.

TBOX	ABOX
Estudante $\subseteq$ Pessoa Empregado $\subseteq$ Pessoa Docente $\subseteq$ Empregado Pesquisador $\subseteq$ Pessoa Estudante_doutorado $\subseteq$ Estudante $\cap$ Pesquisador ORIENTA $\subseteq$ Docente $\times$ Estudante_doutorado Estudante_doutorado $\subseteq$ Estudante $\cap$ $\exists$ ORIENTA.Docente	Estudante(Alexandra) Pesquisador(Alexandra) Docente(Margarida) ORIENTA(Alexandra, Margarida)

FIGURA 13 - Exemplo de base de conhecimento em lógica de descrições.

Note-se que não basta afirmar que um estudante de doutorado é um estudante e pesquisador. É preciso adicionar a restrição de que deve existir um orientador que o orienta. A interpretação das sentenças (semântica) é estabelecida ao modo da teoria dos modelos da lógica clássica. Deste modo, “Estudante(Alexandra)” é uma sentença verdadeira se e somente se a interpretação de “Alexandra”, denotada por *Alexandra*<sup>I</sup>, pertencer ao conjunto resultante da interpretação de “Estudante” denotada por *Estudante*<sup>I</sup>.

A notação OWL é basicamente a expressão da lógica de descrições (tipicamente a SHOIN) na notação XML. Por exemplo, a base de conhecimento da FIG. 13, traduzida para OWL, ficaria como mostrado na FIG. 14.

```

<Ontology xmlns="http://www.w3.org/2002/07/owl#"
...
  <Declaration> <Class IRI="#Docente"/> </Declaration>
  <Declaration> <Class IRI="#Empregado"/> </Declaration>
  <Declaration> <Class IRI="#Estudante"/> </Declaration>
  <Declaration> <Class IRI="#Estudante_doutorado"/> </Declaration>
  <Declaration> <Class IRI="#Pesquisador"/> </Declaration>
  <Declaration> <Class IRI="#Pessoa"/> </Declaration>
  <Declaration> <ObjectProperty IRI="#ORIENTA"/> </Declaration>
  <Declaration> <ObjectProperty IRI="#ORIENTADA_POR"/> </Declaration>
  <Declaration> <NamedIndividual IRI="#Alexandra"/> </Declaration>
  <Declaration> <NamedIndividual IRI="#Margarida"/> </Declaration>
  <EquivalentClasses>
    <Class IRI="#Estudante_doutorado"/>
    <ObjectIntersectionOf>
      <Class IRI="#Estudante"/>
      <Class IRI="#Pesquisador"/>
    </ObjectIntersectionOf>
  </EquivalentClasses>
  <EquivalentClasses>
    <Class IRI="#Estudante_doutorado"/>
    <ObjectSomeValuesFrom>
      <ObjectProperty IRI="#ORIENTADA_POR"/>

```

```

    <Class IRI="#Docente"/>
  </ObjectSomeValuesFrom>
</EquivalentClasses>
<SubClassOf>
  <Class IRI="#Docente"/>
  <Class IRI="#Empregado"/>
</SubClassOf>
<SubClassOf>
  <Class IRI="#Empregado"/>
  <Class IRI="#Pessoa"/>
</SubClassOf>
<SubClassOf>
  <Class IRI="#Estudante"/>
  <Class IRI="#Pessoa"/>
</SubClassOf>
<SubClassOf>
  <Class IRI="#Estudante_doutorado"/>
  <Class IRI="#Pessoa"/>
</SubClassOf>
<SubClassOf>
  <Class IRI="#Pesquisador"/>
  <Class IRI="#Pessoa"/>
</SubClassOf>
<ClassAssertion>
  <Class IRI="#Estudante_doutorado"/>
  <NamedIndividual IRI="#Alexandra"/>
</ClassAssertion>
<ClassAssertion>
  <Class IRI="#Docente"/>
  <NamedIndividual IRI="#Margarida"/>
</ClassAssertion>
<ObjectPropertyAssertion>
  <ObjectProperty IRI="#ORIENTADA_POR"/>
  <NamedIndividual IRI="#Alexandra"/>
  <NamedIndividual IRI="#Margarida"/>
</ObjectPropertyAssertion>
<InverseObjectProperties>
  <ObjectProperty IRI="#ORIENTADA_POR"/>
  <ObjectProperty IRI="#ORIENTA"/>
</InverseObjectProperties>
<InverseFunctionalObjectProperty>
  <ObjectProperty IRI="#ORIENTADA_POR"/>
</InverseFunctionalObjectProperty>
<ObjectPropertyDomain>
  <ObjectProperty IRI="#ORIENTA"/>
  <Class IRI="#Docente"/>
</ObjectPropertyDomain>
<ObjectPropertyDomain>
  <ObjectProperty IRI="#ORIENTADA_POR"/>
  <Class IRI="#Estudante_doutorado"/>
</ObjectPropertyDomain>
<ObjectPropertyRange>
  <ObjectProperty IRI="#ORIENTA"/>
  <Class IRI="#Estudante_doutorado"/>
</ObjectPropertyRange>
<ObjectPropertyRange>
  <ObjectProperty IRI="#ORIENTADA_POR"/>
  <Class IRI="#Docente"/>
</ObjectPropertyRange>
</Ontology>

```

FIGURA 14 - Exemplo de base de conhecimento em OWL.

### 4.3.4 Classificação das ontologias

Dependendo do domínio e da forma como é especificada, a ontologia recebe várias denominações. Assim, se uma ontologia captura conceitos gerais e independentes de domínio, é denominada de ontologia fundamental ou de nível topo, como é o caso da DOLCE, GUM, SUMO, UFO etc; se descreve os elementos de uma linguagem, é chamada de ontologia linguística (FrameNet); se descreve as ações em um domínio, é chamada de ontologia de tarefas; se descreve um domínio particular, é chamada de ontologia de domínio; se é formalizada em uma linguagem de primeira ordem é chamada de ontologia formal; se é registrada em uma linguagem menos expressiva, que captura apenas as relações principais, é chamada de ontologia light. A FIG. 15 mostra algumas das denominações.

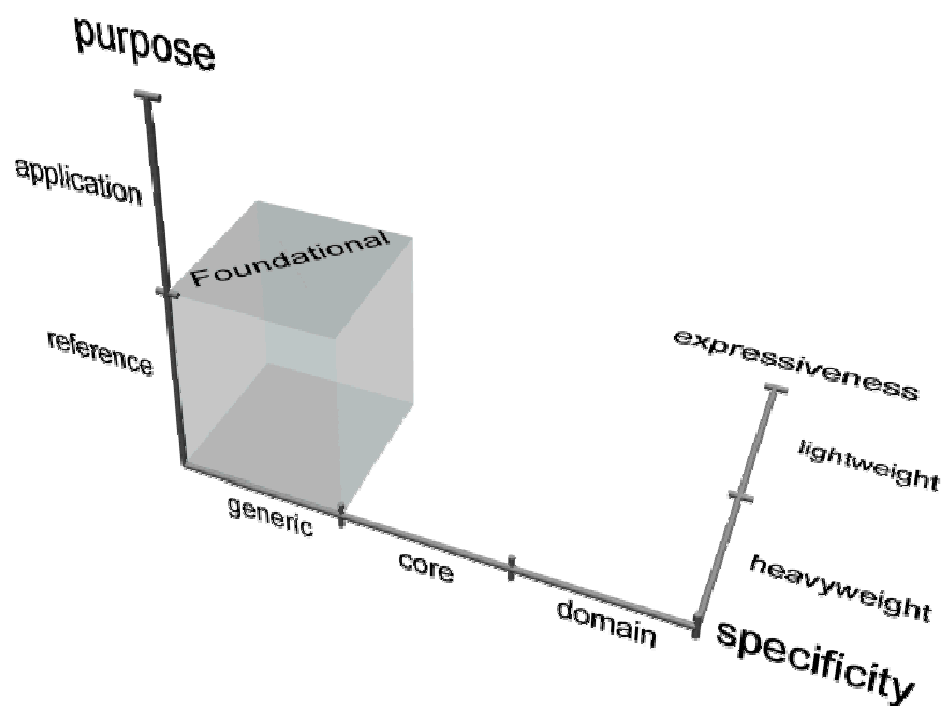


FIGURA 15 – Tipos de ontologias.<sup>16</sup>

---

<sup>16</sup> <http://cos.ontoware.org/>.

A definição de uma ontologia em termos de condições necessárias e suficientes pode ser uma tarefa difícil de ser realizada, uma vez que o termo denota conjunto que não possui fronteiras bem delimitadas, se comportando mais como uma categoria *fuzzy*. A FIG. 16 classifica as ontologias segundo a expressividade da linguagem de representação, interpretando o termo ontologia como uma categoria em que os membros possuem uma gradação quanto ao grau de pertinência à categoria. Na FIG. 16, os elementos posicionados mais à direita seriam mais prototípicos do que os elementos mais à esquerda. A linha vermelha que separa a gradação estabeleceria um limite para o que pode ser denominado de ontologia.

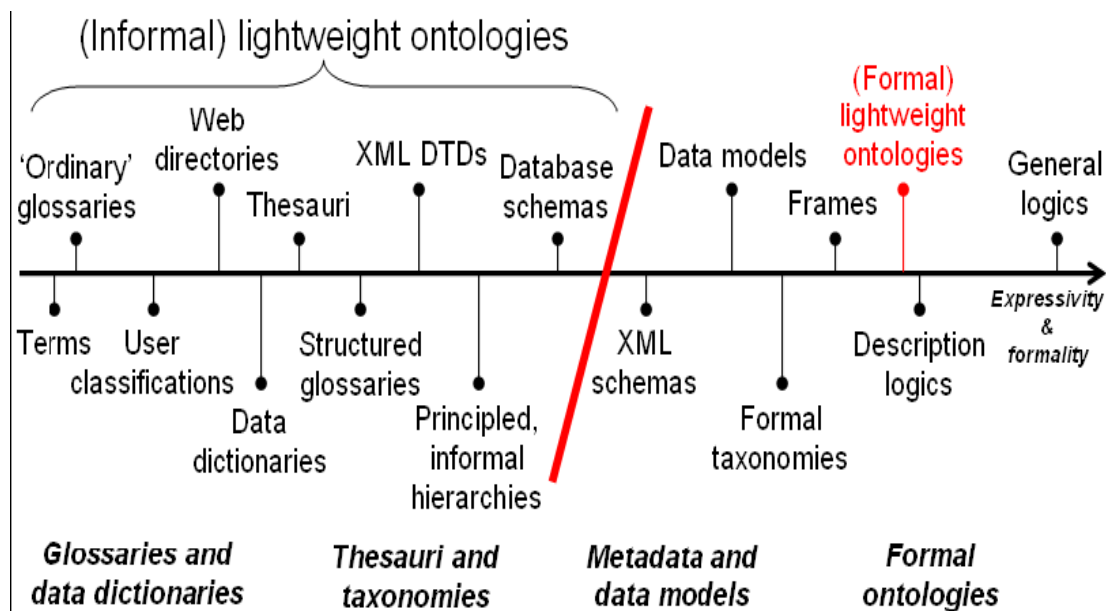


FIGURA 16 – Graduação de linguagens de representação de conhecimento em relação à categoria de ontologia.<sup>17</sup>

A FIG. 17 classifica ontologias segundo o nível do conhecimento representado. Os níveis mais altos representam o conhecimento mais genérico e independente de domínio. Ontologias filosóficas, como as dez categorias de Aristóteles, se enquadrariam neste nível. Os níveis mais inferiores estão associados a conhecimentos mais específicos e dependentes de domínio.

<sup>17</sup> [www.weblab.isti.cnr.it/.../2009/iccu/slides.html](http://www.weblab.isti.cnr.it/.../2009/iccu/slides.html).



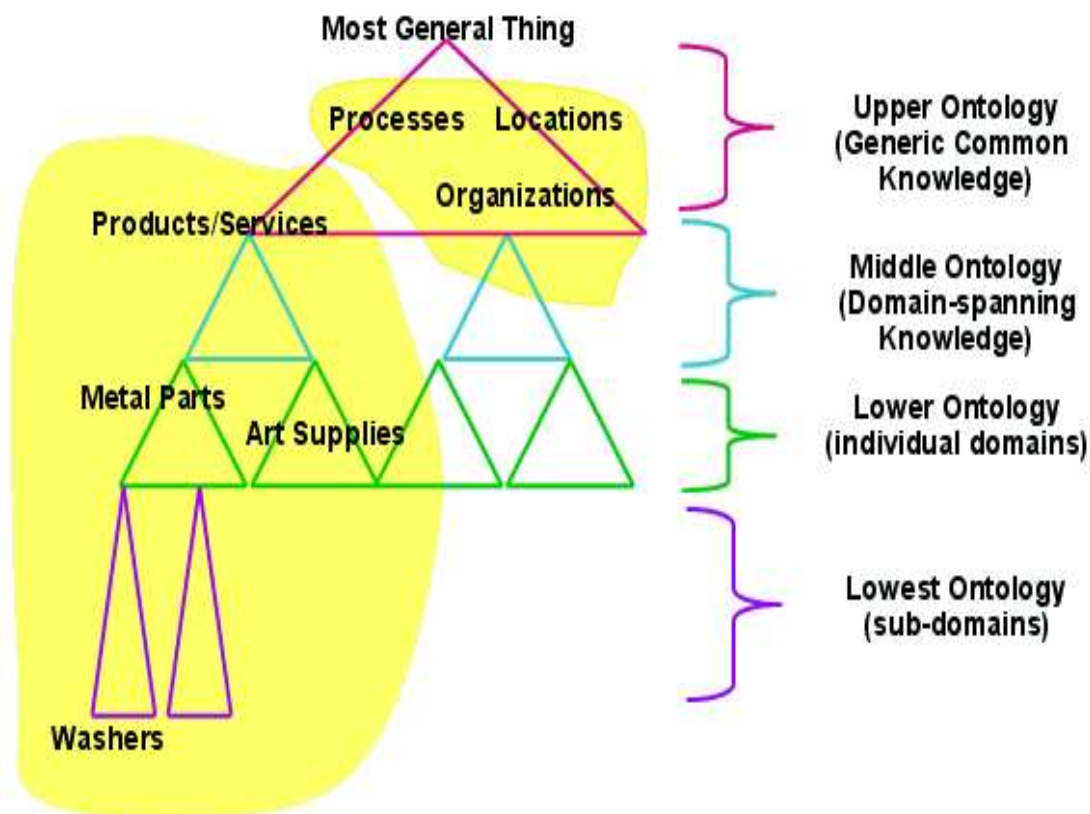


FIGURA 17 – Classificação de ontologias segundo o nível do conhecimento representado.<sup>18</sup>

#### 4.4 Alinhamento entre Ontologias e a FrameNet

A combinação da FrameNet e ontologias não é nova. Existe um projeto que combina a ontologia de nível topo SUMO com a FrameNet para fins de inferência linguística (SCHEFFCZYK; PEASE; ELLSWORTH, 2006). Nesse caso, os tipos semânticos da FrameNet são mapeados para os conceitos da Ontologia. Esta combinação é importante porque combina a capacidade inferencial das ontologias (que falta na FrameNet) com os recursos linguísticos da FrameNet (que faltam nas ontologias).

<sup>18</sup> <http://www.mkbergman.com/category/structured-web/page/9/>.

Em Chishman (2009) é proposta uma integração entre léxicos semânticos e ontologias para o domínio jurídico tendo como base teórica os estudos de Lenci *et al.* (2002) e Guarino (1998). Em sua proposta, são apresentadas duas iniciativas do grupo de pesquisa SEMANTEC da Unisinos, cujo objetivo é desenvolver estudos semântico-computacionais voltados para a construção de léxicos e ontologias.

Uma das iniciativas trata do desenvolvimento de uma ontologia de verbos para o domínio jurídico, tendo por base um *corpus* composto de seis acórdãos judiciais relacionados com o tema “acidentes rodoviários”.

A segunda iniciativa trata da análise, sob o ponto de vista ontológico, dos adjetivos que ocorrem em acórdãos do Instituto das Tecnologias de Informação na Justiça de Portugal, totalizando 40 textos. No primeiro caso, os verbos foram sistematizados segundo os aspectos da semântica verbal, focando principalmente nas relações lógico-semânticas (antonímia, hiponímia etc.), papéis semânticos (agente, paciente, instrumento etc.) e elementos de frames. A organização na forma de ontologia permitiu a generalização de consultas, tais como a capacidade de incluir motocicletas, carros, ônibus e conceitos correlatos em respostas às perguntas como: *que veículos automotores mais se acidentam?* (CHISHMAN, 2009).

A especificação dos frames onde ocorrem os verbos permite que se observe a necessidade da ocorrência de determinados argumentos, para compor a cena relacionada com o verbo. Além disso, a ontologia determina restrições em relação à natureza dos argumentos. No caso dos adjetivos, o foco foi a análise das várias categorias de adjetivos (classificadores, qualificadores, intensionais remissivos, valenciais, modalizadores etc.) e, conforme cada caso, relacionando-os com os elementos da ontologia. Por exemplo, os adjetivos classificadores do domínio jurídico (público, administrativo, jurídico etc.) são de grande importância para uma taxonomia estrutural do domínio, explicitando relações tais como hiponímia/hiperonímia (CHISHMAN, 2009).

Chow e Webster (2007) propuseram a integração da FrameNet com a WordNet e a ontologia SUMO, com o propósito de realizar a classificação de verbos de acordo com a análise da metafunção Ideacional da Gramática Sistêmico-Funcional (GSF) (HALLIDAY, 1994). Nessa pesquisa, as unidades léxicas da classe de verbos da FrameNet foram mapeadas para os synsets de verbos da WordNet. Após isso, foi

utilizado o mapeamento WordNet/SUMO realizado por Niles e Pease (2003), para realizar o mapeamento FrameNet/SUMO. A FIG. 18 esquematiza o mapeamento realizado.

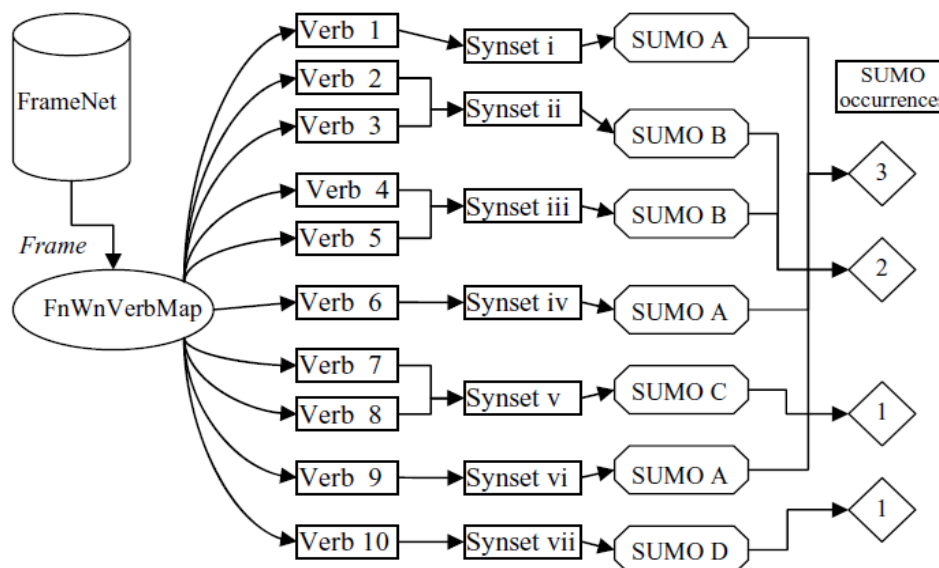


FIGURA 18 – A rede relacional do mapeamento entre FrameNet, WordNet e SUMO.

FONTE – CHOW & WEBSTER, 2007. p.6.

López *et al.* (2008) utilizaram a FrameNet alinhada com a ontologia DOLCE para descrever emoções. O sistema está sendo usado no desenvolvimento de interfaces computacionais sensíveis à expressão de emoções.

Burchardt e Pennacchiotti (2008) investigaram as razões do baixo desempenho de sistemas que utilizam a estrutura argumental na tarefa de reconhecer inferência textual. A tarefa de reconhecer inferência textual (*Recognizing Textual Entailment Track*, RTE) é um projeto patrocinado pela agência NIST (*National Institute of Standards and Technology*) do governo americano. Consiste do reconhecimento automático da relação inferencial entre pares de sentenças T e H, onde T representa “Texto” e H representa “Hipótese”. O RTE disponibiliza um conjunto desses pares para a análise, sendo que o critério para existência da relação inferencial é o seu

reconhecimento por falantes típicos da língua. Por exemplo, nas sentenças a seguir, um falante da língua inglesa reconheceria que a sentença T implica na sentença H:

T – *Kiesbauer was target of a letter bomb in June 1995. Her secretary got injured.*

H – *A letter bomb was sent to Kiesbauer.*

Em tarefas como essa é importante entender o papel dos participantes em cada cena, de modo a relacioná-las. Em T, *Kiesbauer* é o recipiente da carta bomba e em H ele é o destinatário da carta. Portanto, recursos lexicais que levam em conta a estrutura argumental e identificam os papéis semânticos possuem um potencial para contribuir com sistemas para RTE. No entanto, segundo os autores, o único sistema baseado na FrameNet obteve um desempenho apenas mediano. Os autores especulam que as razões podem estar entre as seguintes: a) baixa cobertura lexical; b) confiabilidade limitada dos *parsers* semânticos de frames; c) utilização não otimizada da informação semântica dos frames no módulo de raciocínio.

O primeiro passo da investigação de Burchardt e Pennacchiotti foi a anotação da base do RTE, composta por 800 pares de sentenças, com a informação existente na FrameNet. Após a realização dessa etapa, foi verificado que apenas 8% dos elementos evocadores de frames não possuíam frames na base da FrameNet. Portanto, a baixa cobertura lexical foi descartada como sendo a responsável pelo baixo desempenho e os autores concluíram que outras análises deveriam ser realizadas, para identificar os gargalos da utilização da FrameNet na RTE.

Ovchinnikova *et al.* (2010) propuseram melhorar a estrutura da FrameNet por meio de sua integração com a ontologia DOLCE, com o objetivo de ampliar seu uso em tarefas intensivas de processamento de linguagem natural. Segundo os autores, isso é necessário devido a algumas limitações atuais da FrameNet, tais como baixa cobertura lexical, incompletude das relações, inconsistências na herança de propriedades e falta de axiomatização (ver seção sobre limitações da FrameNet). Para contornar as limitações, a FrameNet foi mapeada para as categorias da DOLCE e suas relações foram formalizadas. Por exemplo, a relação de subframe para eventos foi formalizada da seguinte forma:

$$\forall p_1 p_2 (sub\_ev(p_1, p_2) \rightarrow (strict\_temp\_inc(p_2, p_1) \wedge spatially\_includes(p_2, p_1)))$$

Parafrazeando: *um evento é subevento de outro se está incluído temporal e espacialmente no outro.*

Após o alinhamento, axiomatização e ajuste das relações da FrameNet, foi feito um estudo de caso em um conjunto de pares (t,h) de textos relacionados com o domínio médico. A TAB. 1 mostra o resultado do estudo de caso:

Tabela 1 – Comparação sobre inferência textual no domínio médico.

	<b>Sem anotação de frames</b>	<b>de Com anotação de frames</b>	<b>de Com anotação de frames axiomatização</b>	<b>de +</b>
Provas corretas	1	4	7	
Provas erradas	1	1	1	
Precisão geral	0,56	0,5	0,61	

FONTE – Adaptado de OVCHINNIKOVA *et al.*, 2010. p.7.

Pode-se observar que a axiomatização permitiu um número maior de inferências corretas e um aumento na precisão geral.

Em Bārzdiņš *et al.* (2008, p.1) é sugerido que a FrameNet seria beneficiada pela formalização de seus tipos ontológicos, de modo a permitir um melhor uso por dispositivos computacionais. Com esse objetivo eles geraram, a partir das sentenças do *corpus* anotado da FrameNet, uma estrutura multidimensional: a dimensão dos frames, a dimensão das unidades léxicas, a dimensão dos elementos de frames e a dimensão das valências. Segundo os autores, essa estrutura multidimensional, juntamente com todos os elementos do *corpus* anotado, formam a ontologia da FrameNet. É importante ressaltar que isso não é o que é considerado como ontologia neste trabalho.

Ofoghi *et al.* (2007) e Ofoghi (2009) propuseram estender a FrameNet por meio de uma ontologia, para permitir a implementação de um sistema de perguntas e respostas que fosse capaz de lidar com perguntas das quais as respostas envolvessem o tratamento de cadeias de predicados. Cadeias de predicados, como definido por Morris e Hirst (1991), são cadeias de unidades lexicais que são sequências de termos relacionados semanticamente. Por exemplo, a sentença apresentada em Ofoghi (2009), “In 1974, using beams of electrons and antielectrones, or positrons, Richter discovered a particle that came to be called Psi/J. It contained two quarks possessing a previously unknown flavor called charm...”, forma a cadeia de predicados ilustrada na FIG. 19:

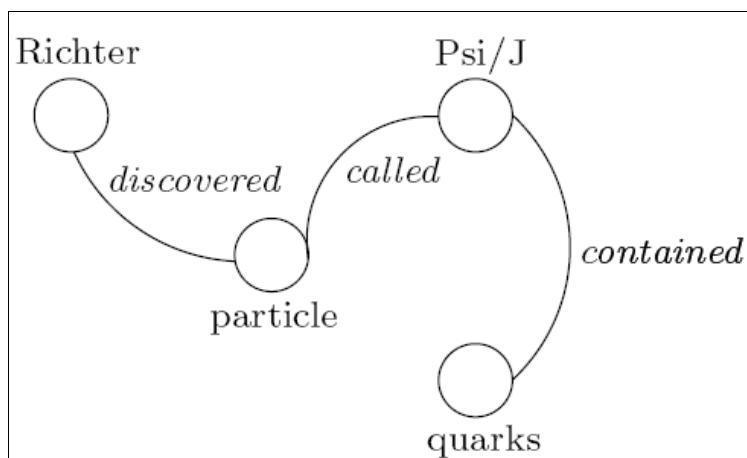


FIGURA 19 – Cadeia de predicados de uma sentença.

FONTE – OFOGHI, 2009. p.120.

Outra área onde a combinação da FrameNet com ontologias é útil é na interpretação de metáforas. A FrameNet não faz distinções entre uso metafórico e literal de unidades lexicais. Cada uso da unidade é um sentido distinto.

Essa abordagem é adequada para o entendimento de cenas, mas limita o uso em inferências, já que existe um isomorfismo entre estruturas literais e metafóricas e esse isomorfismo pode ser utilizado para realização de inferências e entendimento de um texto. Como já foi mencionado anteriormente, o uso metafórico de ULs tem sido anotado na FrameNet, mas ligações metafóricas entre os frames ainda não são aplicadas.

Gedigian *et al.* (2006) mostraram que a combinação da FrameNet com outros recursos lexicais que distinguem o uso metafórico do literal pode ser utilizada para ajudar na realização de inferências textuais. O uso combinado das ontologias com a

FrameNet também pode ser empregado para este fim. Por exemplo, no domínio fonte (mais concreto) algum elemento participante da estrutura teve alguma propriedade ontológica flexibilizada, para permitir a ocorrência de elementos mais abstratos na estrutura de alvo (mais abstrata). Seja o verbo *pegar* em seu uso literal. Neste caso, um ente animado tomou posse de um objeto concreto. Já no uso metafórico do verbo *pegar*, como na frase “pegar uma ideia”, o objeto que foi “pegado” passou de concreto particular para um objeto abstrato. Especulamos que as alterações das propriedades ontológicas não são grandes, de forma a permitirem que o ouvinte perceba a intenção do falante.

Finalmente, o uso combinado de ontologias e frames permite o estabelecimento de novas ligações entre os frames (relações de semelhanças de propriedades ontológicas), permitindo uma melhor verificação da adequação de determinados frames e indicação da necessidade de criação de novos. Um benefício adicional dessa ligação é o fornecimento de uma explicação plausível para o processo de criação de metáforas. Além disso, torna-se possível estabelecer o contexto para a identificação de frames, o que facilita a extração automática de informações. Estas aplicações serão mais exploradas na próxima seção.

## 4.5 Limitações da FrameNet

A FrameNet possui o mérito de promover o entendimento de um item lexical a partir das cenas nas quais é empregado. É um rico repositório de informações semânticas com um nível fino de granularidade. Porém a FrameNet possui algumas limitações, que dificultam seu uso em certas aplicações. Uma das limitações da FrameNet é a sua falta de formalização.

A FrameNet é apresentada como uma base textual em HTML, o que dificulta seu uso por sistemas computacionais. Existe uma formalização em OWL DL<sup>19</sup> (SCHEFFCZYK; PEASE; ELLSWORTH, 2006), mas sem dispor de uma

---

<sup>19</sup> Ontology Web Language — description logic (<http://www.w3c.br/>).

axiomatização, como é o caso das ontologias formais como SUMO e DOLCE. Outra queixa comum a respeito da FrameNet é a sua baixa cobertura lexical.

Ovchinnikova *et al.* (2010) também levantaram algumas limitações da FrameNet: baixa cobertura lexical, incompletude das relações, inconsistências na herança de propriedades e falta de axiomatização. Como exemplo de incompletude encontrado na FrameNet, os autores alegaram que existem frames que deveriam possuir ligação, mas não é postulada nenhuma relação entre eles na base. Por exemplo, os autores acreditam que os frames SURVIVING e RECOVERY deveriam estar conectados, uma vez que o primeiro implica no segundo:

(t) .[people]<sub>SURVIVOR</sub> who [survive]<sub>SURVIVING</sub> [Sars]<sub>DANGEROUS SITUATION</sub> ...

(h) [Those]<sub>PATIENT</sub> who [recovered]<sub>RECOVERY</sub> [from Sars]<sub>AFFLICTION</sub> ...

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

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

(h) [Three leaders ...]<sub>SUSPECT</sub> are [suspected]<sub>SUSPICION</sub> [of stealing money]<sub>INCIDENT</sub> .

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



Em nossa opinião, a principal limitação da FrameNet é a inconsistência das relações interframes e a precariedade da informação sobre a natureza dos elementos da FrameNet. Essas limitações levam a interpretações errôneas dos frames e limitam seu uso em sistemas computacionais. Para contornar essas limitações é necessário formalizar os elementos da FrameNet e adicionar passos na metodologia associados à análise ontológica. Uma formalização dificilmente será unanime, mas estabelecerá uma base sobre a qual será possível discutir e propor aperfeiçoamentos. O que não ocorre sobre uma base informal.

#### **4.6 Benefícios da junção de Ontologias com a FrameNet**

A semântica de frames e sua correspondente implementação na forma da FrameNet é uma importante contribuição para o entendimento da linguagem natural.

Existem, entretanto, alguns espaços para avanços. No uso da FrameNet para extração de informação, como determinar qual frame deve ser evocado? Como gerar novos frames, a partir de frames previamente estabelecidos? Como usar os frames para a descrição dos mapeamentos entre domínios, como no caso das metáforas? Que outras ligações semânticas existem entre os elementos de frames?

Um frame descreve uma cena conforme o nível conceitual segundo o qual se pode entender o significado de um item lexical, mas existem conhecimentos que podem ser agregados à FrameNet para ampliar sua aplicação. Para definirmos que conhecimento pode ser agregado à FrameNet, podemos utilizar as classificações de Brachman (1979) e Guarino (1995) para os tipos de conhecimento, organizados em níveis. Um desses níveis é o nível ontológico que estabelece o que existe no mundo, ou pelo menos como o percebemos, independentemente da linguagem. A FIG. 20 mostra os níveis de conhecimento propostos por Guarino (1995).

<i>Nível</i>	<i>Primitivas</i>
Lógico	Predicados, funções
Epistemológico	Relações de estruturação
Ontológico	Relações Ontológicas
Conceitual	Relações Conceituais
Linguístico	Termos Linguísticos

FIGURA 20 – Níveis de representação do conhecimento.

FONTE – Guarino, 1995.

Dentro desta visão, os frames semânticos operam no nível linguístico, por ser um léxico; no epistemológico, ao oferecerem uma organização estruturada do conhecimento; e, no nível conceitual, como atestado por GAWRON (2008), que declara:

*Frames are conceptual structures that provide context for elements of interpretation; their primary role in an account of text understanding is to explain how our text interpretations can (validly) leap far beyond what the text literally says.*

(GAWRON 2008, p.4)

O nível ontológico é responsável pela análise da natureza do que existe e as relações que podem ser observadas no nível individual, tais como relações parte-todo (*merologia*).

O que é proposto aqui é que o uso combinado do nível ontológico com o nível conceitual expresso na FrameNet permita não só ampliar as aplicações da FrameNet, mas também auxiliar no próprio desenvolvimento da rede. Como esses conhecimentos interagem? Suponha o seguinte segmento de sentença:

a) *Ele possui um livro ...*

O trecho (a) é semanticamente ambíguo, como pode ser observado pelo contexto fornecido pelas sentenças abaixo:

b) *Jorge Amado é um bom escritor. Ele possui um livro sobre meninos de rua.*

c) *Carlos gosta de estudar os astros. Ele possui um livro sobre astronomia.*

Claramente trata-se de dois frames distintos. Um sobre a produção de obras intelectuais e o outro sobre a posse de objetos materiais. Mas esta distinção pode ser verificada no nível ontológico, que detectaria que a natureza do livro em (b) é de um objeto intelectual abstrato, enquanto em (c), trata-se de um objeto material.

A mesma análise pode ser feita para a diferença de sentido do verbo “*pegar*” em *pegar a xícara de café*, que evoca o frame de MANIPULAÇÃO, e o verbo “*pegar*”, em *peguei a ideia que você me passou* ou *peguei a matéria da aula passada*, que evocam o frame de COMPREENSÃO”. Por exemplo, a ligação entre os frames GRASP e MANIPULATION, que não é estabelecida na base da FrameNet, a não ser pelo fato de compartilharem a unidade lexical *grasp*, é que existe um relaxamento em uma propriedade ontológica de um dos elementos do frame MANIPULATION, permitindo que sejam “manipulados” elementos abstratos, como conceitos. Sugere-se que o frame GRASP possua uma ligação com o frame MANIPULATION, seguindo a tese da criação de metáforas a partir do conhecimento corporificado. A FIG.21 resume o que foi dito. Note-se que é possível entender as ligações no nível ontológico entre os dois frames, o que não é possível observando-se a FrameNet isoladamente.

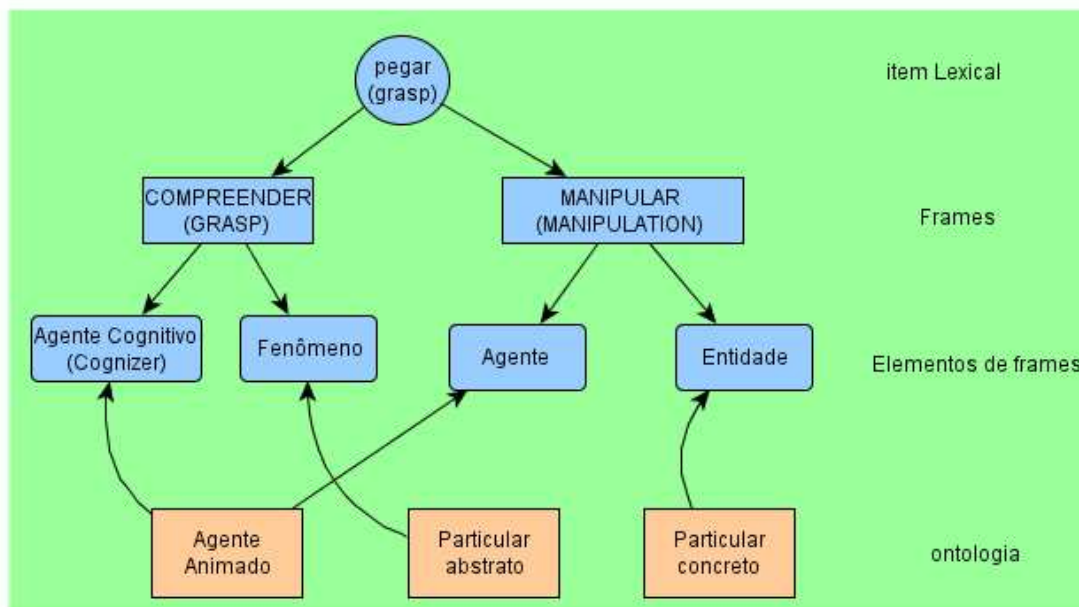


FIGURA 21 – Relação entre os tipos ontológico e a FrameNet.

As ontologias podem ser usadas, também, para definir novos frames, a partir de frames já definidos. Suponha uma cena onde existe um casamento parcial com um frame pré-existente. As diferenças das propriedades ontológicas servem para indicar o caminho que deve ser tomado para a criação de um novo frame. Assim, um elemento de frame que passa a ter um caráter mais abstrato pode indicar que o frame a ser definido deva atuar em um nível mais mental e menos no nível sensório-motor. O exemplo anterior também serve para ilustrar esta ideia. O frame GRASP trata da captura mental de um conhecimento, fazendo analogia com a captura física de um objeto descrito no frame MANIPULATION.

Outra aplicação possível é o estabelecimento de relações entre os elementos de frames. Guarino e Welty (2000) apresentaram um conjunto de meta-propriedades ontológicas (identidade, rigidez, unidade e dependência) que impõem certas restrições no relacionamento entre as entidades. Informalmente, identidade é a capacidade de uma entidade ser identificada; a rigidez é definida como algo que é essencial a uma entidade; unidade é definida como a capacidade de reconhecer todas as partes de uma entidade; e dependência ocorre quando a existência de uma entidade depende da existência de outra.

Desse modo, uma entidade que é um papel material na ontologia DOLCE, como, por exemplo, o conceito de “estudante”, possui as meta-propriedades identidade (um estudante é passível de ser identificado); não rigidez (uma vez que uma pessoa pode deixar de ser estudante sem deixar de existir); unidade (é possível reconhecer todas as partes do estudante) e possui a propriedade de dependência (uma vez que um estudante para existir precisa estar ligado a uma instituição). Portanto, ao encontrarmos um elemento de frame que pode ser enquadrado como papel material, é provável que participe da mesma cena outro elemento de frame com o qual possui uma relação de dependência. De fato, o frame EDUCATION\_TEACHING inclui os elementos de frame *student* e *institution*, como elementos nucleares. Assim, a análise ontológica de uma cena pode sugerir a existência de elementos de frames que poderiam passar despercebidos.

Atualmente a FrameNet não faz distinção explícita entre papéis e tipos. Assim as sentenças “*a person*<sub>[People]</sub> *bought*<sub>[Commerce\_buy]</sub> *a car*<sub>[Vehicle]</sub>” e “*a student*<sub>[Education\_teaching]</sub> *bought*<sub>[Commerce\_buy]</sub> *a car*<sub>[Vehicle]</sub>” não possuem distinção explícita na FrameNet; no entanto, a distinção é feita pelos frames evocados pelos elementos de frames. Assim, *person* evoca o frame PEOPLE e *student* evoca o frame EDUCATION\_TEACHING, que revelam características dos objetos. Porém, não é explicitado que *person* é um tipo e que *student* é um papel, apesar desta natureza carregar implicações sobre o significado, como, por exemplo, focalizar em uma determinada característica do EF. A falta de distinção entre papel e tipos ou a ligação rígida entre eles (por exemplo, a ligação fixa entre estudante e pessoa) engessa um relacionamento que, ontologicamente é dinâmico e mutável: quem pode afirmar que sempre todo estudante é uma pessoa?

Sumarizando, os problemas da FrameNet derivam de duas falhas: a falta de formalização e a inclusão parcial de informação ontológica. A formalização dos elementos da FrameNet permitiria uma compreensão mais precisa dos mesmos e eliminaria equívocos na sua utilização. A inclusão mais ostensiva de informação ontológica permitiria melhor distinção dos frames, seus relacionamentos e enquadramento dos enunciados que evocam o frame. Propomos aqui, avançar no sentido de eliminar essa lacuna, apresentando uma formalização e uma forma de utilizar a informação ontológica para a distinção de frames.

Uma decisão importante é sobre qual ontologia deve ser usada para a anotação dos lexemas de um domínio. Existem diversas propostas de ontologias em diferentes níveis de atuação. Desde as dez categorias de Aristóteles, em seu tratado denominado *Categorias*, vários conjuntos de categorias de nível topo têm sido sugeridas para organizar o que é percebido do mundo. Em se tratando de conceitos associados a um lexema, a questão é complexa, principalmente devido à polissemia. Taylor (1992) escreveu um artigo intitulado “How many meanings does a word have?”. Nele o autor reúne as discussões sobre a palavra “window” (janela) realizadas por George Lakoff e as considerações de John Searle para o verbo “open”. No caso da palavra “window”, ela é empregada para denotar diferentes categorias de conceitos. Por exemplo, segundo Lakoff (1987, p. 417), a palavra pode ser usada tanto no sentido de uma abertura em uma parede como para referenciar o retângulo de vidro que preenche a abertura. A abordagem estatística, associada à Semântica de Frames, pode fornecer uma solução para esse dilema, indicando o sentido mais provável de um lexema dentro de um enunciado. Essa solução se assemelha à que é empregada pelos seres humanos, segundo as descobertas recentes da Neurociência, tais como a natureza *best-fit* (melhor ajuste) de nossa rede neural (LAKOFF, 2008; FELDMAN, 2006). A natureza *best-fit* da nossa rede neural é usada para decidir qual interpretação é mais adequada dentre várias possíveis. É um “casamento” baseado em evidências perceptivas e estatísticas reunidas ao longo da vivência. Feldman (2006, p. 235) descreve um modelo que tem por base uma rede de crenças (*belief networks*<sup>20</sup>) para simular a inferência neural.

Além da dificuldade de se associar um lexema a uma categoria em função de sua natureza polissêmica, ainda existe a dificuldade de, uma vez definido seu significado, se decidir qual aspecto de seu sentido deve ser levado em consideração. Por exemplo, um automóvel pode ser visto como um objeto concreto, como um artefato, como um objeto de transporte, como um objeto de desejo, etc. Cada um desses aspectos é proeminente em uma situação específica. Os aspectos ou facetas são inúmeros, mas é preciso um critério para organizar e decidir qual aspecto deve ser privilegiado. Uma abordagem possível é proposta por James Pustejovsky (1995), que utiliza a estrutura *qualia* criada por Aristóteles, para definir qual aspecto do sentido de um lexema é proeminente.

---

<sup>20</sup> Redes de crenças é uma técnica de estatística bayesiana usada para inferência estatística.

## 5 METODOLOGIA

A pesquisa em questão tem como meta a análise das interações entre dois objetos de representação de conhecimento, a FrameNet e as ontologias, com intuito de gerar um *framework* que funcionará como um guia na tarefa de definir e ajustar os frames a serem disponibilizados em uma base léxico-semântica a partir do enquadramento dos enunciados em uma cena.

Para esta tarefa, foi realizado um levantamento bibliográfico sobre ambos os objetos de estudo para caracterizá-los no escopo deste trabalho que envolve as áreas da LC e da Ciência da Computação. Na linguística, o enfoque foram os frames e a Semântica de Frames proposta por Fillmore. Na computação, focamos nas ontologias, no sentido do termo nessa área, verificando possíveis contribuições para uma melhor formalização e aperfeiçoamento metodológico para a geração de frames.

Os passos executados durante a pesquisa foram os seguintes:

- Para compreender o estado da arte nas áreas pertinentes foi realizado um levantamento sobre a bibliografia relacionada com a Linguística Cognitiva, Semântica de Frames, FrameNet e ontologias, sendo que essas últimas serão pesquisadas dentro do âmbito da ciência da computação.
- Formalização da FrameNet em uma linguagem com semântica formal, tendo como objetivo a eliminação de ambiguidades na compreensão dos elementos da FrameNet. Partimos do princípio de que, como objeto de pesquisa, a FrameNet esteja sujeita a divergências e interpretações diversas, porém, para o uso em sistemas automatizados é necessário um entendimento não ambíguo de seus elementos. Essa formalização foi utilizada para servir de guia para verificar a pertinência de uma sentença a um frame a partir da obediência da sentença às definições constantes na formalização.
- Análise do extrato de evidências empíricas a partir da escolha de um campo de conhecimento ou campo temático para determinação dos frames que ocorrem no domínio. 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.).

- 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. Nesse caso, as candidatas naturais são as ontologias SIMPLE-CLIPS (LENCI *et al.* 2000) e ontologia DOLCE, ambas de inspiração linguística. A ontologia SIMPLE-CLIPS possui a vantagem de se basear na Teoria do Léxico Gerativo de Pustejovsky e, em função disso, analisar mais de um aspecto ontológico do lexema.
- Em seguida foram analisadas as unidades léxicas e elementos de frames que ocorrem no domínio sob o ponto de vista ontológico. Esta análise foi realizada em um *subcorpus* do *corpus* selecionado e foi feita manualmente, gerando uma ontologia de domínio que foi usada posteriormente para o aprendizado de sistemas de anotação automatizados. O objetivo da anotação ontológica foi resgatar o significado da natureza do objeto relacionado ao lexema que é capturado na estrutura conceitual, mas parcialmente perdido na FrameNet. Esta natureza é fundamental para determinar mais claramente o seu sentido nas sentenças.
- Aplicação de regras da formalização, definidas anteriormente, e a ontologia selecionada em um conjunto de sentenças do *corpus* para ilustrar o uso destes recursos na metodologia de desenvolvimento de frames. Procuramos mostrar que esta ação permite avaliar adequação de uma sentença à cena padrão e para a identificação de variações do uso e o uso em sentidos metafóricos.
- Sugerir uma forma de aplicação automática das regras. Utilizamos as Redes Bayesianas para essa automação. As Redes Bayesianas inferiram com um grau aceitável de acerto (acima de 80%) a pertinência de uma sentença a um frame tendo por base apenas a ocorrência dos tipos ontológicos. Isso demonstra a importância desse aspecto do conhecimento. Também reforça a hipótese de que o procedimento estatístico de melhor enquadramento (*best fit*) é um componente importante da cognição humana. A formalização ocorre, neste caso,



implicitamente pelas probabilidades condicionais capturadas automaticamente durante o aprendizado das Redes Bayesianas. As Redes Bayesianas foram treinadas por meio de um *subcorpus* de 57 sentenças previamente selecionadas.

- Sumariar as contribuições da pesquisa.

Esta pesquisa é caracterizada por um método de abordagem indutivo uma vez que partimos de uma cena particular descrita por um frame e induzimos os resultados no sentido de generalizar a aplicação do *framework* para detecção, criação e pertinência de sentenças aos frames registrados na FrameNet. Ainda entendemos que o *framework* gerado pela pesquisa pode ser aplicado à FrameNet de qualquer língua. Além disso, a ontologia utilizada nesta pesquisa, a SIMPLE-CLIPS pode ser substituída por outra ontologia, caso seja conveniente.

Especificamente como indicado por Lakatos e Marconi (2001, p.87) para toda indução são necessárias três fases: a observação, a descoberta e a generalização. No caso, em particular a observação ocorreu ao percebermos, durante a pesquisa, que os tipos ontológicos tratados e limitados na SUMO poderiam ser usados mais amplamente e com um novo objetivo: auxiliar na detecção e distinção de frames para a base FrameNet. A fase da descoberta ocorreu quando propusemos a forma de aplicar os tipos ontológicos para adicionar informações adicionais aos elementos de frames da FrameNet. Os elementos de frames receberiam uma camada ontológica relacionada a natureza destes elementos. Além disso, o *framework* contempla a formalização aos elementos e as relações dos frames dando maior rigor e precisão ao entendimento do recurso lexical FrameNet. A generalização se apresenta ao indicar que a metodologia proposta e testada poderá ser aplicada em outras situações quando da criação de frames.

Os métodos de procedimento utilizados foram o monográfico e o estatístico. É monográfico uma vez que se examinou a base FrameNet e a contribuição de uma ontologia para a FrameNet, avaliando e analisando seus aspectos com a finalidade de obter uma generalização para a detecção e distinção de frames para a base FrameNet. O método estatístico ocorre por conta do uso das Redes Bayesianas que conseguiram classificar o conjunto de sentenças nas categorias prototípica, *quasi* prototípica (quando não é uma ocorrência típica, tampouco metafórica) e metafórica.

As técnicas abordadas para coleta dos dados foram a documentação indireta e a direta. A documentação indireta foi útil para o preparo da fundamentação teórica para as duas áreas, Linguística e Computação. Enquanto a documentação direta relacionada à coleta de dados foi voltada para o campo experimental para selecionar a amostra e aplicar técnicas probabilísticas. Neste caso, os dados foram obtidos do *corpus* selecionado, pesquisando o lexema *viagem* e sua ocorrência nas sentenças, visando a determinação dos sentidos. Uma amostra, especificamente um *minicorpus*, recebeu um tratamento para análise de seu conteúdo seguida da técnica probabilística visando a generalização das descobertas.

Detalharemos a seguir a metodologia de anotação adotada e a forma de aplicação das redes bayesianas.

## 5.1 *Corpus* e campo de conhecimento

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.

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 nesta tese 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, itens que serão descritos mais adiante. O critério para a seleção destas sentenças foi a variação do uso no sentido típico, correlato e metafórico.

## 5.2 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 da pesquisa abordada nesta tese, foi necessário realizar uma anotação indicando os tipos ontológicos, de acordo com uma ontologia de domínio. Como o foco da pesquisa não é a criação de um procedimento automático de anotação, partes do processo foram feitas manualmente, notadamente, a criação da ontologia de domínio. Existem pesquisas, como a realizada por Zhang Rui-ling e Xu Hong-sheng (2009), que mostram como realizar essa etapa de forma automática, embora este ainda seja um tópico de pesquisa.

A ontologia de domínio foi criada a partir das 57 sentenças mencionado na seção anterior. 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 nesta tese. 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 na linguagem de programação Java. A anotação ontológica foi realizada a partir da ontologia de domínio previamente construída.

Após a anotação ontológica, o programa emitiu dois tipos de arquivo. Um na forma de tabela com indicação da ocorrência (denotada pela letra V) ou ausência (denotada pela letra F) do tipo ontológico em cada sentença. Esse arquivo foi usado como entrada para o treinamento e teste do módulo de rede bayesiana do sistema Weka<sup>21</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.

O segundo arquivo gerado pelo anotador é uma planilha, contendo as sentenças e suas camadas de anotação *Part of speech* (POS) e ontológica, uma em cada linha distinta da tabela. Esse arquivo serve para examinar o processo de anotação e também para realizar os cálculos de probabilidades usados na rede bayesiana projetada com o programa MSBNx (*Microsoft Bayesian Network Editor*) (KADIE; HOVEL; HORVITZ, 2001) da Microsoft<sup>®</sup>. Este último programa foi utilizado em função de sua saída gráfica ser mais fácil de ser compreendida do que a saída do Weka. No entanto, o Weka possui maiores facilidades para a automação e para o processamento de dados volumosos.

A FIG. 22 explica os passos executados e a FIG. 23 mostra a tela principal do programa anotador desenvolvido pelos autores da pesquisa.

---

<sup>21</sup> <http://www.cs.waikato.ac.nz/ml/weka>

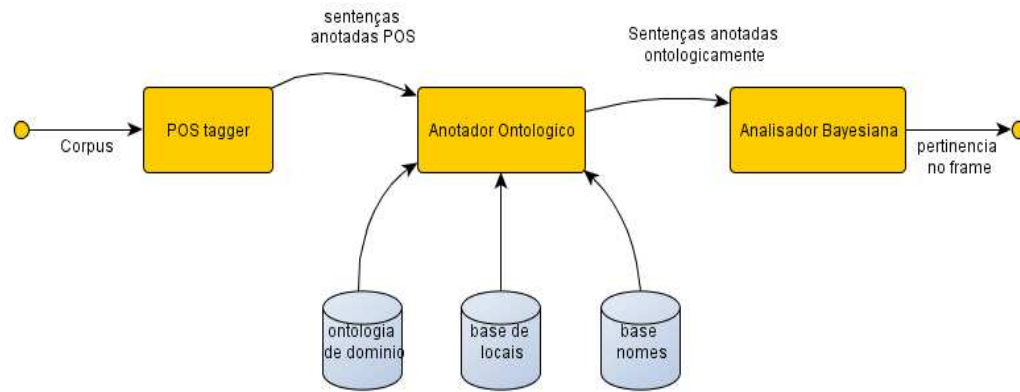


FIGURA 22 – Fases do sistema de análise.

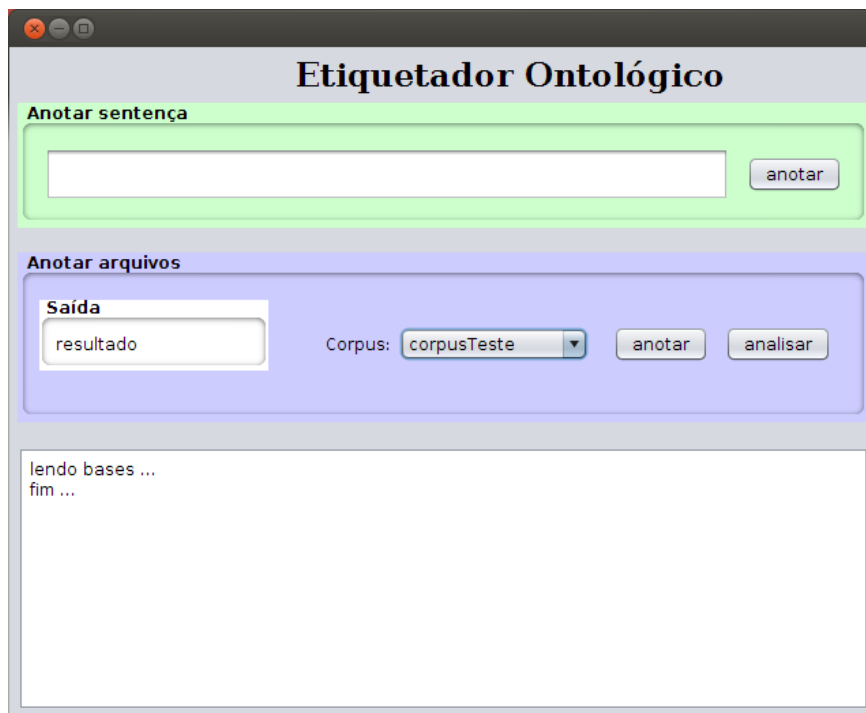


FIGURA 23 – Tela do anotador ontológico.

### 5.3 Anotação baseada na Ontologia SIMPLE-CLIPS

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, nas sentenças a seguir, o lexema *viagem* pode assumir diferentes naturezas.

a) O zagueiro Antônio Carlos desembarcou às 5h30, cansado, após uma **viagem** de 23 horas desde o Japão. (19N:Br:SCat). (**processo**)

b) A **viagem** custou uns R\$ 15,00/R\$ 20,00. (site: [www.mochileiros.com/rio-de-janeiro-em-um-fds-t43496.html](http://www.mochileiros.com/rio-de-janeiro-em-um-fds-t43496.html)). (**evento**)

Na sentença (a), **viagem** denota um processo com etapas que podem ser mencionadas e descritas. Já na sentença (b), **viagem** é vista como um evento atômico, sem etapas. Da mesma forma, um objeto, como um carro, pode ser visto 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. Uma possibilidade é o uso 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.

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

Assim, segundo Chishman (2003), a TLG propõe uma forma distinta de composicionalidade, onde a semântica do predicado é definida em conjunto com seus argumentos.

Um exemplo típico apresentado por Pustejovsky (1995, p. 91) trata sobre os sentidos do lexema “livro”:

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

Segundo Pustejovsky, a habilidade de um item lexical agrupar diversos sentidos é o que ele e Anick (1988) chamam de *Paradigma Lexical Conceitual* (lcp), que preconiza que existe algo inerente à semântica de um nome que o capacita a projetar qualquer um dos sentidos do nome em ambientes sintáticos e semânticos distintos.

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

A estrutura *qualia* é o nível mais proeminente da TLG e a mais utilizada em pesquisa de processamento de linguagem natural (ZAVAGLIA; GREGHI, 2003; GONZALEZ; LIMA, 2004; PARADIS, 2005; CIMIANO; WENDEROTH, 2005; MENDE; CHAVES, 2001; AMARO; MENDES; MARRAFA, 2010). Para a proposição da estrutura *qualia*, Pustejovsky inspirou-se na interpretação de Moravcsik (1973 *apud* PUSTEJOVSKY, 1995, p. 76) dos modos de explicação de Aristóteles. As definições dos componentes da estrutura *qualia* são as seguintes:

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

Para ilustrar como é feita a descrição de um item lexical, nesses níveis de representação, é mostrada a seguir (FIG. 24), a descrição da palavra “livro”.

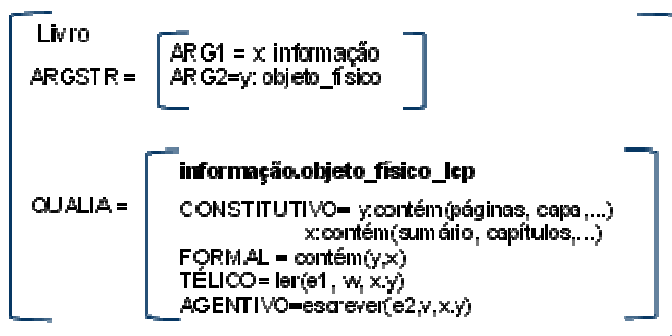


FIGURA 24 - Representação do lexema “livro” na TLG.

FONTE - Adaptada de PUSTEJOVSKY, 1995. p.101.

A FIG. 24 mostra os níveis *qualia* e estrutura argumental do nominal “livro”. A estrutura argumental define quais os tipos que o objeto pode assumir: um objeto físico ou uma informação. Sendo considerado um tipo complexo, esses sentidos não são tratados separadamente e o *quale* FORMAL define como um argumento se relaciona com o outro (o objeto físico contém informação). A etiqueta “**informação.objeto\_físico\_lcp**” indica tratar-se da representação de um tipo complexo (lcp), cuja referência da instância complexa é representada por “x.y”. O *quale* CONSTITUTIVO descreve as partes que formam os sentidos do lexema. O *quale*



TÉLICO registra, na forma de um predicado de evento<sup>22</sup>, a função do livro (ser lido em um evento e1). Finalmente, o *quale* AGENTIVO registra que sua criação envolveu um evento de escrever. A FIG. 25 mostra o relacionamento dessa estrutura lexical com outras estruturas. As etiquetas T, A e F junto aos relacionamentos indicam relacionamentos télico, agentivo e formal, respectivamente.

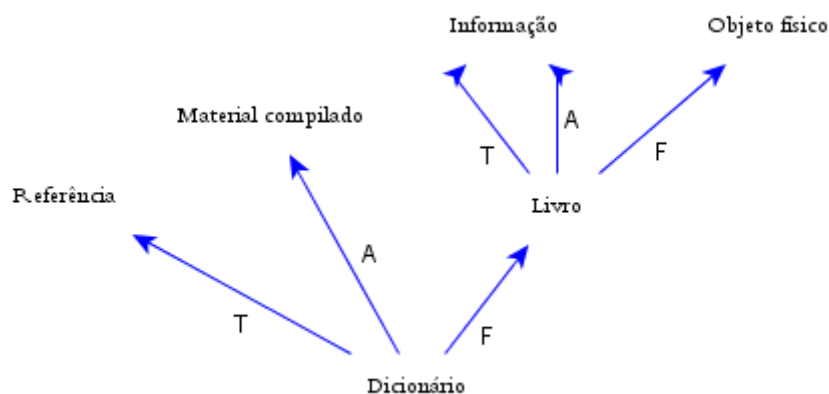


FIGURA 25 - Rede de herança para “livro” na TLG.

FONTE - Adaptada de PUSTEJOVSKY, 1995. p.145.

A possibilidade de agrupar os diversos sentidos de um lexema em uma estrutura formal que pode ser manipulada, posteriormente, de forma automática, é bastante atrativa e tem sido aplicada na construção de léxicos, ontologias e na interpretação de textos. A estrutura *qualia* foi utilizada no desenvolvimento da EuroWordNet (VOSSSEN, 1998), no desenvolvimento da WordNet para o português europeu (MENDES; CHAVES, 2001), no desenvolvimento da WordNet para o dinamarquês (PEDERSEN *et al.*, 2006) e no projeto SIMPLE-CLIPS (LENCI *et al.*, 2000). Paradis (2005) aplicou em *framework* para interpretação de semântica de itens lexicais. Kasama *et al.* (2010) utilizou a estrutura *qualia* na representação ontológica do domínio da nanociência e nanotecnologia. Chishman (2009) aplicou a estrutura *qualia* para integrar léxicos semânticos e ontologias.

<sup>22</sup> Predicado de evento é usado no Cálculo de eventos, que é um formalismo para raciocinar sobre situações que mudam com o tempo. Neste tipo de predicado, o primeiro argumento denota o evento.

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

A diferença de abordagem entre a TLG e a Semântica de Frames, se combinada, pode ser benéfica. Segundo Ruppenhofer *et al.* (2010, p. 46), em vários frames que possuem nominais denotando artefatos, alguns elementos de frames foram definidos, refletindo um tipo de estrutura *qualia*, como por exemplo, modificadores que denotam o material do artefato (*qualia* constitutivo). Ruppenhofer *et al.* (2010, p. 10) também menciona que a divisão de frames em função de inconsistências no tipo semântico facilita a busca de dados para estudar as coerções pustejovskianas. A ideia de que significado de um item lexical é função do relacionamento dos demais itens lexicais da sentença com determinadas dimensões da estrutura *qualia* do item, pode permitir o estabelecimento de novos relacionamentos na estrutura de frames (uma das deficiências apontadas na FrameNet). Assim, TLG fornece uma base teórica para a extensão dos relacionamentos da FrameNet. Se esses relacionamentos são suficientes para fornecer uma explicação completa para a polissemia de um item lexical, é algo que merece uma análise mais detalhada. A FIG. 26 mostra o mapeamento de um item lexical, segundo a TLG com os frames relacionados.

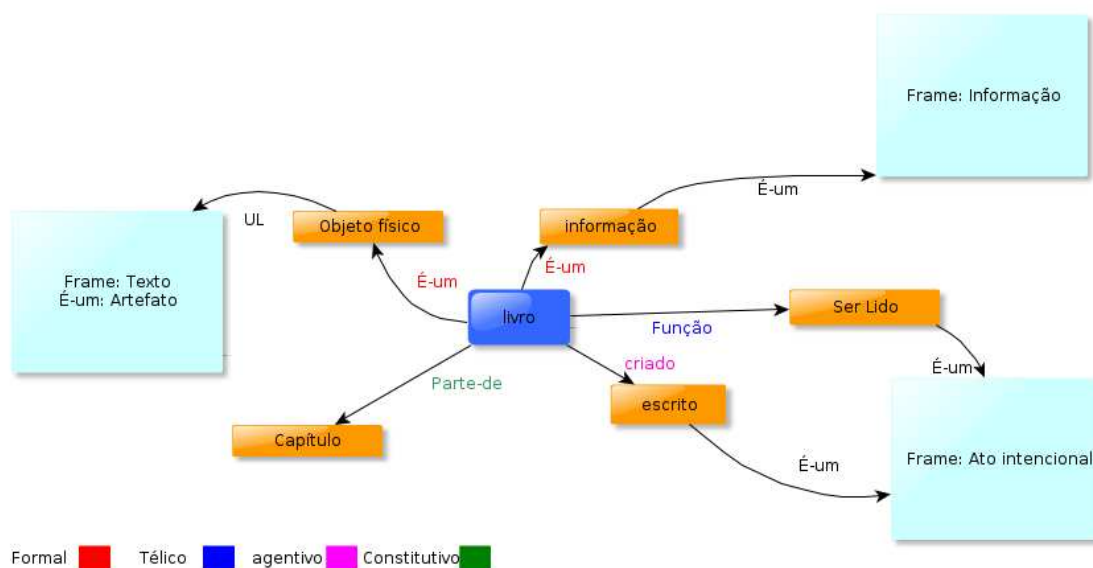


FIGURA 26 - Mapeamento do item *livro*, segundo a TLG com os frames relacionados.

Na FIG. 26 podemos observar que o lexema *livro*, pode possuir diversos aspectos que são explicitados pela estrutura *qualia*. Esses aspectos são perfilados em cenas distintas, descritas pelos frames. Dessa forma, podemos ter um entendimento claro de como a TLG pode se relacionar com os *frames*.

Neste trabalho utilizamos 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 ontologia SIMPLE-CLIPS<sup>23</sup>, (*Semantic Information for Multifunctional Plurilingual Lexica-Corpora e Lessici dell'Italiano Parlato e Scritto*) (LENCI *et al.*, 2000; RUIIMY; TORAL, 2008 consiste de 157 tipos semânticos organizados por meio de relações conceituais hierárquicas e não hierárquicas.

A base teórica para o desenvolvimento da SIMPLE-CLIPS é a Teoria do Léxico Gerativo (PUSTEJOVSKY, 1995). A ontologia SIMPLE-CLIPS divide os conceitos em três categorias de objetos (de primeira, segunda e terceira ordem), de acordo com o preconizado por Lyons (1977).

<sup>23</sup> <http://www.ilc.cnr.it/clips/Ontology.htm>.

Segundo Paradis (2005), de um modo geral, as entidades de primeira ordem são objetos físicos tais como: animais, pessoas, plantas, veículos. Essas entidades são relativamente estáveis do ponto de vista perceptivo. Existem no espaço tridimensional, em qualquer ponto do tempo, e são publicamente observáveis.

O status ontológico das entidades de segunda e de terceira ordem é vago, uma vez que não está associado ao mesmo número de propriedades estáveis, como as entidades de primeira ordem. Essas entidades são mais variáveis e, portanto, também mais difíceis de serem definidas e, conseqüentemente, mais controversas. Entidades de segunda ordem são acontecimentos, processos e estados, tais como 'vitória', 'discussão' e 'felicidade', respectivamente. Essas entidades estão localizadas no tempo e se diz que ocorrem em oposição às que existem.

Finalmente, as entidades de terceira ordem são entidades abstratas que estão fora do espaço e do tempo. São entidades como "conceitos", "ideias", "possibilidades" e "proposições", "dia" e "ano", referidas por Schmid (2000) como SHELLS (concha). A denominação de SHELLS ocorre, segundo Paradis (2005), em função das entidades de terceira ordem não possuírem um conjunto estável de propriedades, sendo que as propriedades são incorporadas à medida que se enquadram em determinada noção abstrata ou pelas necessidades pragmáticas das pessoas. Um exemplo de SHELL apresentado por Schmid (2000, p.3), seria o lexema *problem* (problema), na sentença “*the (big) problem was that I had no money*”, que serve de SHELL para “*I had no money*”.

Parte da estrutura pode ser vista na FIG. 27. Para que fosse possível capturar parte da estrutura da ontologia SIMPLE-CLIPS, realizamos a migração desta para o software de edição de ontologias Protégé (<http://protege.stanford.edu>).

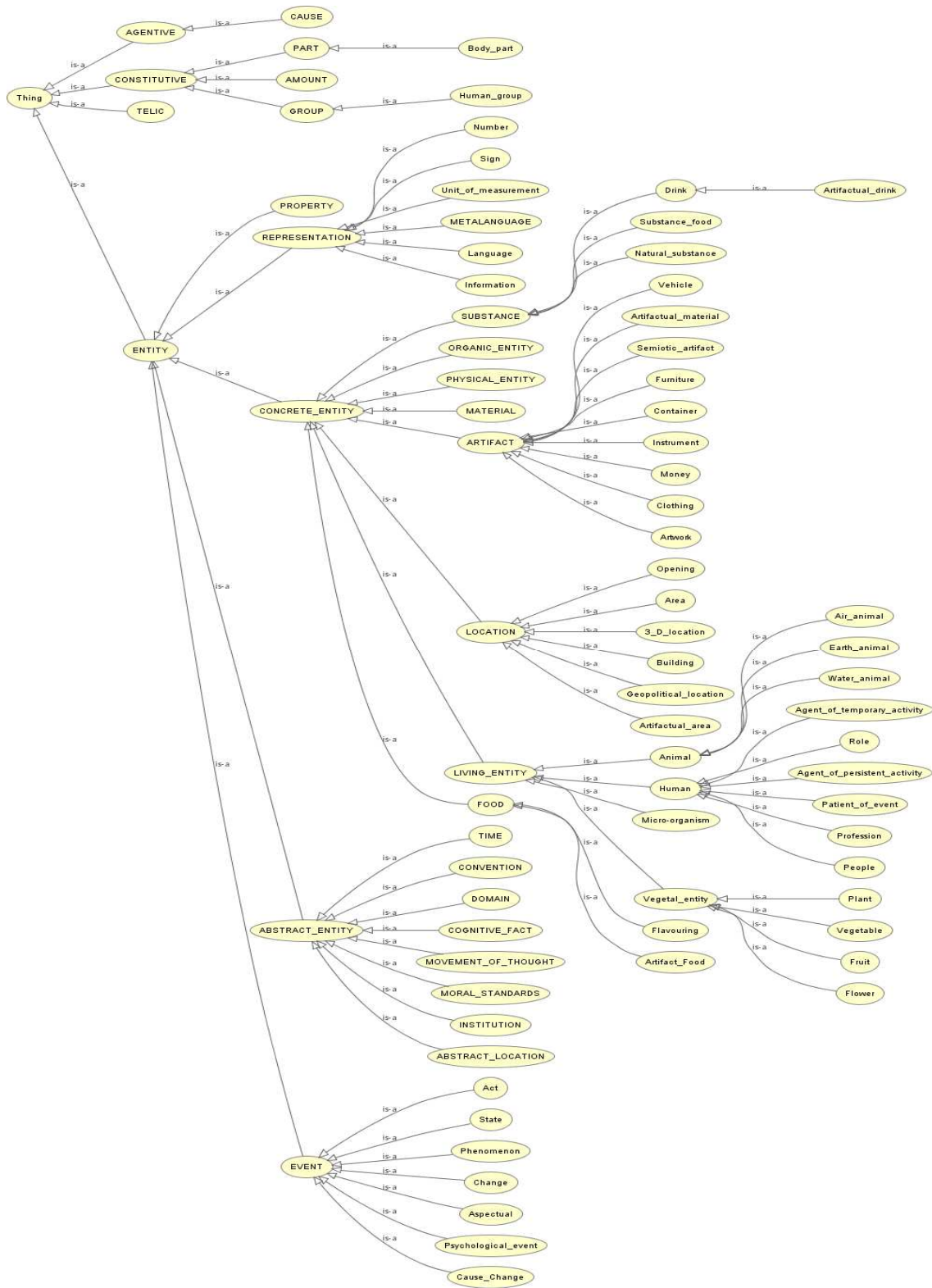


FIGURA 27 – Parte da ontologia SIMPLE-CLIPS.

## 5.4 Redes Bayesianas

Redes Bayesianas são 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  $P(H/E)$ . A fórmula para o cálculo dessa probabilidade é dada pela fórmula (RICH; KNIGHT, 1993, p. 268):

$$P(H_i|E) = \frac{P(E|H_i)P(H_i)}{\sum_{n=1}^k P(E|H_n)P(H_n)} \quad (1)$$

Onde,  $P(H_i/E)$  é a probabilidade de hipótese  $H_i$  dada a evidência  $E$ . Por exemplo, a probabilidade de alguém ter gripe ( $H_i$ ) dado que essa pessoa tem febre ( $E$ ).  $P(E/H_i)$  é probabilidade da evidência  $E$  dada a hipótese  $H_i$ . Por exemplo, a probabilidade de alguém ter febre ( $E$ ) dado que essa pessoa tem gripe ( $H_i$ ).  $P(H_i)$  é probabilidade *a priori* de  $H_i$  ocorrer na ausência de evidências.  $P(E/H_n)$  é probabilidade da evidência  $E$  dada outra hipótese  $H_n$ , por exemplo, estar com dengue.  $P(H_n)$  é probabilidade *a priori* de  $H_n$  ocorrer na ausência de evidências. Finalmente,  $k$  é o número de hipóteses.

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:

$$P(H|E,e) = P(H|E) \cdot P(e|E,H)/P(e|E) \quad (2)$$

Pode-se observar que, se o número de evidências e hipóteses for muito grande, o número de probabilidade 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. A FIG. 28 mostra uma rede bayesiana simples, para verificar o sentido do lexema *rede*, dada a coocorrência dos lexemas *computador* e *peixe*. Os valores que aparecem junto aos nós são as probabilidades do sentido dado às ocorrências dos lexemas, sendo que os valores servem apenas para exemplificar, não tendo qualquer base real. Para um sistema mais realista, os lexemas e as probabilidades de coocorrência devem ser levantados de um grande *corpus* textual.

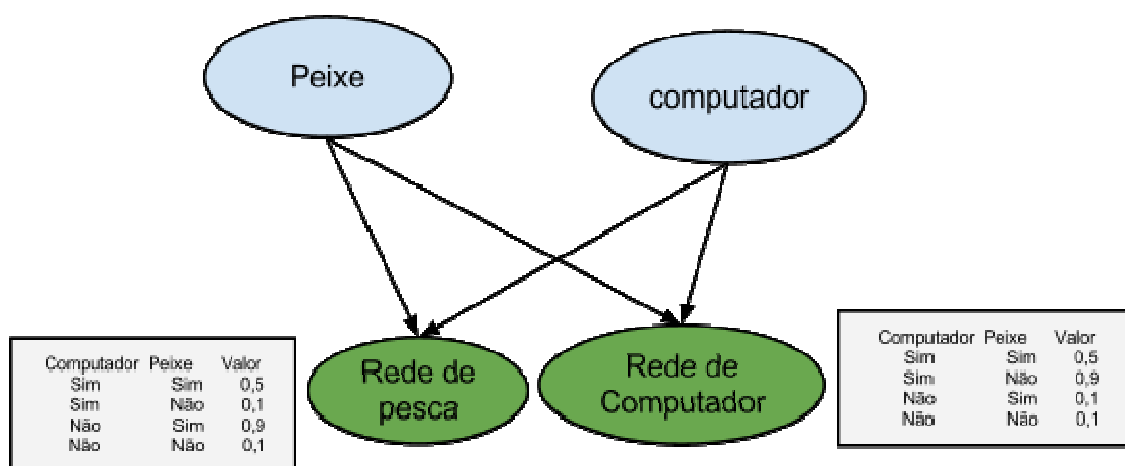


FIGURA 28 - Rede bayesiana para sentido do lexema *rede*.

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.

No caso da aplicação das redes bayesianas neste trabalho de pesquisa, em que se propõe um aperfeiçoamento na metodologia da elaboração da FrameNet, elas 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. Neste caso, os papéis semânticos não são tão relevantes, uma vez que as sentenças possuem basicamente os mesmos papéis. A nossa proposição é que os tipos ontológicos é que terão um papel fundamental na diferenciação do significado.



## 6 DESCRIÇÃO DO EXPERIMENTO

Neste capítulo será descrita a nossa efetiva contribuição científica, resgatando o aporte semântico do aspecto ontológico na compreensão de um enunciado. Iremos primeiramente delinear o *Framework* proposto. Em seguida será apresentada uma formalização dos elementos da FrameNet e das regras a serem obedecidas durante o trabalho lexicográfico. Após essa etapa, será sugerida uma alteração na metodologia de desenvolvimento, inserindo etapas para a análise ontológica. Para demonstrar a viabilidade da proposta será apresentado um estudo de caso envolvendo a cena de VIAGEM com a aplicação manual do *FrameWork*. Finalmente, apresentaremos a aplicação automática de uma técnica probabilística para verificação do enquadramento de sentenças na cena de VIAGEM, com o objetivo de comprovar o peso da contribuição ontológica. No final do capítulo, serão discutidos os resultados obtidos.

### 6.1 O FRAMEWORK PROPOSTO

Pode-se notar que existem vários espaços para contribuição mútua entre ontologias e a FrameNet: formalização, ampliação da base lexical, fornecimento de estrutura argumental, ampliação de capacidade de inferência e a própria reestruturação da rede. Mas falta o estabelecimento de uma teoria ou arcabouço que una e explique todos esses relacionamentos e que forneça uma base para novas conexões entre esses níveis. Essa teoria deve mostrar a correspondência que existe entre ontologia e FrameNet.

Segundo Vossen e Fellbaum (2009), a FrameNet não faz distinção entre significado linguístico e conhecimento do mundo, enquanto uma ontologia pode ter conceitos sem correspondência linguística:

*“Frames and frame elements are inspired by the vocabularies of natural language, and FrameNet does not attempt to draw a distinction between linguistic meaning and world knowledge. There are no knowledge constructs independent of linguistic evidence. By contrast, an ontology may contain concepts not directly motivated by linguistics.”* (VOSSEN; FELLBAUM, 2009, p. 341)

No entanto, o conhecimento do mundo determina a expressão linguística. Concretamente, o *framework* contém os seguintes elementos:

- Uma formalização das entidades da FrameNet para caracterizar sem ambiguidades seu significado e aplicação. A formalização inclui considerações sobre a influência dos tipos ontológicos no estabelecimento dos frames.
- Uma alteração na metodologia da FrameNet para aplicar a análise ontológica.
- Um conjunto de exemplos que ilustram a aplicação da metodologia na análise do grau de associação de uma sentença em uma cena.
- Uma sugestão de automação da análise ontológica por meio de redes bayesianas.

Para um melhor entendimento da contribuição desta proposta, apresentamos, a seguir, a inclusão das informações sobre os tipos ontológicos em um frame comumente citado nos artigos sobre a FrameNet, o frame `COMMERCE_BUY`. Nele mapeamos os elementos de frame para as categorias ontológicas da ontologia `SIMPLE-CLIPS`, sem sermos categóricos com relação a esse mapeamento. Mapeamos os elementos de frames nucleares e selecionamos alguns não nucleares, para exemplificar (FIG.29).

## COMMERCE\_BUY

		<b>Tipo semântico</b>	<b>Categoria ontológica</b>
<b>Unidade lexical</b>	<i>buy</i>		<i>Transaction</i>
<b>Nuclear</b>	<i>Buyer</i>		<i>Role</i>
	<i>Goods</i>		<i>Concrete-entity</i>
<b>Não Nuclear</b>	<i>Duration</i>	<i>Duration</i>	<i>Time</i>
	<i>Manner</i>	<i>Manner</i>	<i>Property</i>
	<i>Money</i>		<i>Money</i> [Concrete_entity] <i>ArtifactAgentive   Telic</i>
	<i>Means</i>	<i>State-of-affairs</i>	<i>abstract-entity</i>
	<i>Place</i>	<i>Locative-relation</i>	<i>Location</i>
	<i>Purpose</i>	<i>State-of-affairs</i>	<i>abstract-entity</i>

FIGURA 29 - Mapeamento de elementos de frames do COMMERCE\_BUY.

Em uma análise preliminar, percebem-se algumas semelhanças, assim como algumas diferenças na comparação efetuada. Não existe um tipo semântico associado ao EF *buyer*, isso porque não existe na FrameNet nenhum tipo que o enquadre adequadamente. Pode-se pensar em *sentient*, mas um comprador pode ser uma organização, o que não é exatamente indivíduo consciente. Um comprador se enquadra mais como um papel exercido por uma entidade. O enquadramento em uma categoria ontológica auxilia no reconhecimento de sentenças que, de fato, estão associadas ao frame. Por exemplo, “comprar uma ideia” no sentido de “concordar”, seria descartada, uma vez que a ideia é uma entidade abstrata e não uma entidade concreta, caracterizando o emprego metafórico do lexema *comprar*.

Mas é preciso observar que a ontologia não permite realizar todas as distinções semânticas. Ela atua para restringir e distinguir, a partir de propriedades ontológicas. A semântica permeia vários níveis e algumas distinções estão além do alcance da ontologia. Por exemplo, a forma de pagamento pode ser feita por dinheiro ou outros meios, mas, no nível ontológico, não existe distinção, uma vez que o que está envolvido é o valor abstrato inerente ao objeto. Já no nível linguístico existe uma distinção que deve ser levada em consideração e é importante.

Mas para se beneficiar da ontologia no tratamento de frames é preciso diretrizes claras para a categorização, tais como testes para identificar a natureza de um elemento.

Complementando a exemplificação do que esperamos obter com a proposta, apresentaremos algumas análises de sentenças em português, anotadas manualmente tanto com base na FrameNet em inglês quanto na ontologia SIMPLE-CLIPS.

Caso 1: *Todos no mundo entenderam o crime contra a humanidade cometido por Israel.* Jornal Correio Braziliense, 5 de junho de 2010.

*Todos no mundo entenderam o crime contra a humanidade* [*Crime*] *cometido* [*Committing\_crime*] *por Israel* [*Perpetrator*].

	O crime contra a humanidade	Cometido	por Israel
Frame		<b>Committing_crime</b>	
Elementos de Frame	crime		Perpetrator
Tipo semântico			sentient
Nível ontológico	event		Geopolitical_location

FIGURA 30 - Anotação de sentença baseada na FrameNet e na SIMPLE-CLIPS - caso 1.

Na análise da sentença, pode-se observar uma discrepância entre o nível ontológico e o tipo semântico da FrameNet. No nível ontológico, "*Israel*" é enquadrado na categoria *Geopolitical\_location*, enquanto a FrameNet atribui ao perpetrador do frame COMMITTING\_CRIME o tipo semântico *sentient*, ou seja, uma entidade consciente. Esta discrepância aponta para um uso metonímico ou metafórico de um elemento de frame.

A FrameNet não aponta um tipo semântico para o elemento de frame *crime*, mas a ontologia SIMPLE-CLIPS permite enquadrar o elemento como sendo um *event*, coleção de elementos que é um tipo de perdurante: uma entidade que possui diferentes partes em diferentes instantes, como um processo. Já endurantes são entidades que podem ser percebidas como um conceito completo, ou seja, todas as partes estão presentes em qualquer instante de tempo.

Caso 2: *Europa quer mais recursos do mercado para combate à crise*. Folha on line, 9 de maio de 2010. (www.folha.uol.com.br).

Europa [Experiencer] quer [Desiring] mais recursos do mercado [Event] para combate à crise [Purpose\_of\_Event].

	<i>Europa</i>	<i>Quer</i>	<i>mais recursos do mercado</i>	<i>para combate à crise</i>
Frame		Desiring		
Elementos de Frame	Experiencer		Event	Purpose_of_event
Tipo semântico	sentient		State_of_affairs	
Nível ontológico	Geopolitical_location	Modal_event	Money	act

FIGURA 31 - Anotação de sentença baseada na FrameNet e na SIMPLE-CLIPS - caso 2.

Na análise da sentença pode-se observar uma discrepância entre o nível ontológico e o tipo semântico da FrameNet. No nível ontológico a "*Europa*" é enquadrada na categoria *Geopolitical\_location*, enquanto a FrameNet atribui ao *Experiencer* do frame DESIRING o tipo semântico *sentient*. Esta discrepância aponta para um uso metonímico do elemento de frame *Experiencer*.

A seguir apresentaremos cada componente do *Framework* e sua aplicação.

## 6.2 Formalização

A Semântica de Frames é uma teoria abrangente sobre a organização do conhecimento humano no cérebro. Essa teoria tem sido atestada por meio de experimentos na área da Neurociência. Sendo a Semântica de Frames uma visão adequada da conceptualização na mente humana, ela abarca todos os aspectos relacionados ao estabelecimento de um conceito na mente. Isso inclui os aspectos ontológicos percebidos pelo ser humano em sua interação com o mundo. Ou seja, a Semântica de Frames, como uma teoria da conceptualização, incorpora a noção de ontologia.

Já a FrameNet é uma base de conhecimento lexical fundamentada na Semântica de Frames. Sendo assim, é um esquema de representação de conhecimento, com as limitações e potencialidades inerentes a qualquer linguagem de representação do conhecimento. Em decorrência desse fato, a FrameNet é passível de se beneficiar da combinação com outras representações de conhecimento, para ampliar sua expressividade. Mas para que essa combinação seja feita de forma adequada, com definição clara das atribuições de cada linguagem de representação, é necessário que exista uma definição precisa de cada elemento de todas as linguagens envolvidas.

Para a linguística, a formalização é importante para estabelecer uma compreensão da Semântica de Frames e sua relação com o aspecto ontológico dos conceitos.

Nesta seção, é proposta uma formalização dos elementos da FrameNet, tendo por base princípios ontológicos. Essa formalização foi utilizada para relacionar a FrameNet com a representação ontológica e na análise de sentenças para verificação de sua inserção em determinado frame. Por formalização entendemos a descrição dos elementos do objeto a ser formalizado em uma linguagem artificial cuja sintaxe e semântica sejam claras e sem ambiguidades. A linguagem escolhida deve possuir a capacidade de representar todos os aspectos formalizáveis do objeto em estudo. A formalização de qualquer entidade apresenta alguns benefícios, tais como a eliminação de ambiguidades, a verificação formal das propriedades do objeto, a adição da capacidade de inferir propriedades ainda não explicitadas e permitir a aplicação em sistemas computacionais. Particularmente, na presente pesquisa, busca-se obter uma metodologia mais formal para estabelecer a distinção entre frames correlatos e a detecção de novos frames.

A formalização da FrameNet parte de alguns **pressupostos** que autorizam a formalização:

- A linguagem expressa e estrutura os processos cognitivos de uma conceptualização do mundo e não representa o mundo diretamente.
- A comunicação ocorre quando existe uma conceptualização compartilhada.
- A FrameNet é uma base de conhecimento prototípico, ou seja, bem definida e, portanto, passível de ser formalizada.

- A FrameNet se baseia na Semântica de Frames, mas como qualquer linguagem, possui limitações de expressividade e, portanto, pode se beneficiar da combinação com representações auxiliares, como, por exemplo, ontologias.

Para uma formalização adequada é de suma importância a escolha da linguagem formal apropriada. Ela deve possuir recursos para expressar todos os aspectos relevantes do objeto, de maneira clara e simples. Alguns dos formalismos usados para representar aspectos da linguagem natural são a *Head-Driven Phrase Structure Grammar* (HPSG) (POLLARD; SAG, 1987), uma linguagem formal baseada em unificação para descrever fenômenos linguísticos; Gramáticas Categóricas (WOOD, 1993); ou a lógica matemática e suas extensões, como a lógica modal e o cálculo de situações. Neste trabalho foi adotada a lógica matemática clássica, por ter uma notação simples e uma semântica bem conhecida, baseada na teoria dos modelos de Tarski (HODGES, 2009). Além disso, é um resultado conhecido de que nenhum formalismo possui um poder de representação maior que a lógica clássica:

“... as versões modernas de lógica são capazes de representar qualquer informação factual que pode ser declarada com precisão em qualquer linguagem, natural ou artificial. As linguagens naturais podem expressar o mais amplo conhecimento que pode ser expresso, e a lógica permite que o subconjunto formulado com precisão seja expresso em uma forma computacional. Talvez existam alguns tipos de conhecimento que não possam ser expressos em lógica. Porém, se tal conhecimento existe, ele não pode ser representado ou manipulado em qualquer dispositivo computacional em qualquer linguagem. O poder expressivo da lógica inclui todo o tipo de informação que pode ser armazenado e ou programado em qualquer computador digital.” (SOWA, 2000, p. 12) (tradução da autora)

Sendo assim, tudo o que é representável pode ser representado nela. Já os adereços notacionais utilizados nos outros formalismos, tais como operadores modais, não foram necessários para a formalização realizada. É importante destacar que não estamos propondo que a semântica da linguagem natural possa ser descrita pela lógica clássica. O que estamos propondo é a formalização da FrameNet.

Apresentamos a seguir as definições que formam a base axiomática da formalização.

**Definição 1 (unidade lexical):** Uma unidade lexical ( $ul$ ) é um par forma-significado, onde a “forma” é um conjunto de formas que diferem apenas por inflexão mas não no significado nuclear:  $ul = \langle form, s \rangle$ . Para acessar o aspecto da forma da unidade lexical foi usada a notação  $form(u)$ . Para acessar o aspecto do significado da unidade lexical, foi usada a notação  $s(u)$ .

**Definição 2 (elemento de frame):** Um elemento de frame ( $ef$ ) representa um papel semântico em um frame específico. A extensão dos elementos de frames que ocorrem nas sentenças associadas ao frame é denotada por  $inst(ef)$ . A função  $tipo(ef)$  retorna o tipo ontológico de um elemento de frame  $ef$ .

**Definição 3 (frame):** Um frame ( $f$ ) é uma tupla  $\langle UL, EL, R \rangle$ , onde  $UL$  é um conjunto de unidades léxicas,  $EL$  é um conjunto de elementos de frames e  $R$  é um conjunto de relações semânticas, onde o primeiro elemento da relação é o frame  $f$ . Os frames podem ser divididos ontologicamente, como mostrado em Ovchinnikova (2010): evento, objeto, qualidade e relação.

**Definição 4 (tipos de elementos nucleares):** Os elementos de frames de um frame  $f$  são agrupados em dois conjuntos: nucleares e não nucleares. Os elementos nucleares são denotados pela relação  $C$ . Assim,  $C(ef, f)$  denota que o elemento  $ef$  é nuclear em relação ao frame  $f$ .

**Definição 5 (relação entre frames):** Uma relação semântica  $r$  é uma relação binária entre dois frames:  $r(f1, f2)$ . As relações semânticas são irreflexivas e assimétricas.

**Definição 6 (valência de UL):** Toda unidade lexical de um frame possui o mesmo número de argumentos:



$$\forall f \forall ul_1, ul_2 | ul_1, ul_2 \in f \Rightarrow nargs(ul_1) = nargs(ul_2)$$

**Definição 7 (instâncias de EF):** Todas as instâncias de elementos de frames de um mesmo frame devem possuir o mesmo tipo:

$$\forall f, ef, e1, e2 | ef \in f \wedge \{e1, e2\} \in inst(ef) \Rightarrow tipo(e1) = tipo(e2)$$

**Definição 8 (linha herança ontológica):** Um elemento de frame  $ef1$  é subtipo de outro  $ef2$  se for subtipo direto ou se existir um outro elemento de frame  $ef3$  que é subtipo direto de  $ef2$  e  $ef1$  for subtipo de  $ef3$  :

$$\forall ef1, ef2 \text{ subtipo}(ef1, ef2) \Rightarrow (\text{subtipodireto}(ef1, ef2) \vee (\exists ef3 \text{ subtipodireto}(ef1, ef3) \wedge \text{subtipo}(ef3, ef2)))$$

**Definição 9 (herança1):** Tudo o que for verdadeiro em termos semânticos para um frame é verdadeiro para o frame filho. Seja  $P$  uma predicação sobre um frame  $f1$  e seja  $f2$  um frame filho de  $f1$ , expressa pela relação  $pai(f1, f2)$ , então a meta regra de segunda ordem a seguir expressa a regra para a herança:

$$\forall P, f1, f2 \text{ pai}(f1, f2) \wedge P(f1) \Rightarrow P(f2)$$

Por exemplo, na definição frame MOTION, é declarado que uma entidade inicia em uma localização e termina em outra. Isso elimina a possibilidade de existir movimento cíclico. Pela definição 9, esta propriedade deve ser verdadeira para todo filho.

**Definição 10 (herança2):** Os elementos de frames de um frame  $f$  devem ser compatíveis ontologicamente com os elementos de frames herdados do frame pai. Ontologicamente compatível significa que são do mesmo tipo ontológico ou de um tipo subjugado pelo tipo do elemento de frame do frame pai.

$$\forall f1, f2 \text{ pai } (f1, f2) \Rightarrow (\forall ef \in f1 \wedge C(ef, f1) \Rightarrow (\exists ef' \in f2 \wedge C(ef', f2) \wedge (\text{subtipo}(ef', ef) \vee \text{tipo}(ef') = \text{tipo}(ef))))$$

**Definição 11 (Metáfora em UL):** Se duas unidades léxicas possuem a mesma forma ( $\text{form}(u1) = \text{form}(u2)$ ), mesma valência e pertencem a frames distintos, pode ser o caso que um dos usos da *UL* seja metafórico. Em caso positivo, a ocorrência no frame com elementos de frames mais abstratos corresponderia ao uso metafórico. A distinção entre tipos abstratos e concretos é feita a partir dos tipos ontológicos em que, invariavelmente, existe esta distinção.

**Definição 12 (perspectiva):** Se um frame  $f1$  possui uma relação de perspectiva sobre outro frame  $f2$ , então cada elemento de frame nuclear de  $f1$  é um elemento de frame de  $f2$  ou subtipo de um elemento de frame de  $f2$ .

$$\forall f1, f2 \text{ persp}(f1, f2) \Rightarrow (\forall ef \in f1 \wedge C(ef, f1) \Rightarrow ((ef \in f2 \wedge C(ef, f2)) \vee (\exists ef' \in f2 \wedge C(ef', f2) \wedge (\text{subtipo}(ef, ef')))))$$

**Definição 13 (precedência):** A relação de precedência temporal entre dois frames é representada pelo predicado  $Pre(f1, f2)$  e possui a relação inversa de sucessão denotada pelo predicado  $Suc(f2, f1)$ .

$$\forall f1, f2 \text{ Pre}(f1, f2) \Leftrightarrow \text{Suc}(f2, f1)$$

**Definição 14 (subframe):** Um frame  $f1$  possui uma relação de subframe com outro frame  $f2$ , então existem um ou mais elementos de frame nucleares de  $f1$  que ocorrem em  $f2$ .

$$\forall f1, f2 \text{ subframe}(f1, f2) \Rightarrow (\exists ef \in f1 \wedge C(ef, f1) \Rightarrow (ef \in f2 \wedge C(ef, f2)))$$

**Definição 15 (subframe e precedência):** Se a relação de subframe ocorre entre frames do tipo eventos  $(f1, f2, \dots, fn)$ , e onde  $(f2, \dots, fn)$  são subframes de  $f1$ , então existe uma relação de precedência entre  $(f2, \dots, fn)$ .

Para frames do tipo eventos, é possível definir pré-condições que devem ser verdadeiras para a ocorrência do evento, e pós-condições que são verdadeiras após a ocorrência dos eventos. Essas pré e pós condições são usadas pelas pessoas para realização de inferências, como nas implicaturas de Grice (DAVIS, 2010) e para o encadeamento de raciocínio para a realização de planejamento, como sugerido por Bratman (1987).

Para fins de automação e verificação, propomos o uso de um cálculo semelhante ao usado na execução de planos, com pré e pós-condições a serem testadas. A base para essa opção de formalização é o fato da FrameNet tratar, em sua maioria, de eventos. Sendo assim, uma formalização do tipo pré e pós-condições é adequada e permite a validação automática da cadeia hierárquica. A condição mínima para a caracterização de MOTION seria:

$$\{local(e, [x, y, z, t])\} MOTION \{ \neg local(e, [x, y, z, t']) \wedge t' > t \}$$

onde *local* é um predicado que declara o posicionamento de uma entidade  $e$  no espaço  $(x, y, z)$  e tempo  $t$ . Os elementos posicionados à esquerda do frame denotam a pré-condição para o evento e os elementos posicionados à direita denotam as condições que devem ser verdadeiras, após a ocorrência do evento. No exemplo, a pré-condição estabelece que a entidade deva estar em um determinado local e que, após a ocorrência do evento, a pós-condição estabelece que a entidade não esteja mais neste local e que tenha decorrido algum tempo.

### 6.3 Comparação com a Formalização de Ovchinnikova

Ovchinnikova *et al.* (2010) propõem uma formalização e análise ontológica das relações da FrameNet. Tanto a formalização como a análise ontológica são feitas tomando-se por base a ontologia fundamental DOLCE e suas bases filosóficas. As distinções entre a nossa proposta e a apresentada no trabalho de Ovchinnikova *et al.* (2010) são as seguintes:

- No trabalho de Ovchinnikova apenas as relações são formalizadas, sendo que os conceitos e elementos intraframes não são formalizados. Neste trabalho, apresentamos uma formalização para os elementos que compõem o frame, uma vez que postulamos a necessidade desta formalização para melhor formalizar as relações e possibilitar uma análise automatizada da rede de frames.
- Em função da ausência da formalização dos elementos que compõem um frame, a proposta de Ovchinnikova *et al.* é vaga, servindo apenas como diretrizes gerais a serem seguidas pelos analistas da FrameNet. Por exemplo, a formalização da herança é expressa da por meio da seguinte sentença lógica, onde  $s$  é uma situação que é uma instância do frame:

$$\forall s (f1(s) \Rightarrow f2(s)).$$

Esta formalização reflete a declaração que consta no “The book” (RUPPENHOFER *et al.*, 2010), que estabelece que tudo que é verdade para o frame pai deve ser verdade para o frame filho. No entanto, ela não fornece elementos para se verificar se esse é o caso. No caso deste trabalho, a herança é formalizada por mais de uma fórmula (definições 9 e 10) e partindo do fato de que a grande maioria dos frames tratam de eventos e que eventos causam mudanças no estado do mundo, lançamos mão de uma formalização com pré e pós condições, ao estilo de STRIPS (FIKES; NILSSON, 1971). Dessa forma, é

possível determinar, de forma automatizada, se as condições estabelecidas pelo frame pai são verificadas nos frames filhos.

- A formalização proposta por Ovchinnikova *et al.* não aborda a relação de metáfora entre frames. Grande parte da cognição humana está relacionada com o mapeamento metafórico. Apesar de não ter sido explicitado nas primeiras versões da FrameNet, a ligação metafórica entre frames é discutida na versão 2010 do “The Book” (RUPPENHOFER *et al.*, 2010). Apesar de reconhecer a necessidade do estabelecimento da relação metafórica entre frames Ruppenhofer declara que, na prática, essa ligação não é feita na FrameNet. A formalização proposta neste trabalho aborda a relação metafórica, ampliando a aplicação da FrameNet na inferência textual, uma vez que a relação metafórica permite a inferência por analogia.
- A formalização proposta por Ovchinnikova *et al.* se baseia na ontologia DOLCE e em seus princípios filosóficos. No presente trabalho, além dos princípios filosóficos subjacentes à uma ontologia propomos a aplicação dos princípios ontológicos da Teoria do Léxico Gerativo, de James Pustejovsky. Com isso, espera-se ser capaz de explicar os diferentes sentidos dos lexemas nas sentenças, para prover uma separação mais adequada das situações descritas pelas sentenças, fornecendo um encaixe mais preciso nos frames associados. Esse passo pode permitir a detecção automática de frames relacionados com as sentenças, o que não é previsto pela formalização do trabalho citado.
- A formalização proposta por Ovchinnikova *et al.* tem por objetivo “arrumar” as relações da FrameNet, tornando-as ontologicamente mais coerentes, visando o processamento da linguagem natural. Nosso objetivo é obter uma base de frames mais completa, segundo evidências linguísticas.

## 6.4 Proposta de Alteração Metodológica

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

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

Segundo Ruppenhofer *et al.* (2010) etapas típicas do desenvolvimento de um frame são as seguintes: buscar atestação de *corpus* para um conjunto de lexemas que se supõe que tenham alguma sobreposição semântica. Estas atestações são divididas em grupos. Após isso, os grupos são reagrupados em grupos maiores, com tamanho suficiente para desenvolver os frames.

A etapas utilizadas pelo grupo que desenvolve a FrameNet espanhol<sup>24</sup> são mais claramente explicitadas:

- 1. identificar os elementos pertencentes ao frame subjacente a cada lexema,**
- 2. encontrar os sintagmas que instanciam os EFs nas sentenças que devem ser anotadas,**

---

<sup>24</sup> Fonte: <http://gemini.uab.es:9080/SFNsite/spanish-framenet-process>

3. **anotar os EF com nomes apropriados, e**
4. **reunir as sentenças (descritas semântica e sintaticamente) e exibi-las com a descrição construcional do lexema.**

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

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

Na metodologia predominante na elaboração de frames, parte-se de um sentido para a obtenção dos itens lexicais associados ao significado. Claro que nesse processo, alguns sentidos relacionados são detectados e frames distintos, porém relacionados, são criados.

As evidências da necessidade para o particionamento de um frame são mencionadas em Ruppenhofer *et al.* (2010) e incluem: alteração no número de argumentos, perfilamento diferente dos EFs, relações distintas entre os EFs, alteração no tipo semântico dos EFs. Este último critério é o embrião de nossa proposta que aprofunda esta análise do tipo semântico para uma análise ontológica mais ampla, uma vez que os tipos semânticos apresentados na FrameNet não são aplicados a todos os

elementos. Isto ocorre porque sua destinação não é a realização de uma análise ontológica, mas sim, como ressalta Ruppenhofer *et al.* (2010, p. 79), registrar as informações que não são representadas na hierarquia de frames e elementos de frames. Sendo assim, as várias naturezas de um item lexical correlacionadas semanticamente, obtidas por meio de evidência de *corpus*, auxiliam no estabelecimento de uma rede de frames associados de forma mais abrangente que a metodologia corrente.

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

## 6.5 Aplicação da formalização

A formalização e a anotação ontológica propostas aqui podem ser utilizadas manualmente ou implementadas em um sistema computacional, para auxiliar no tratamento de *corpus* linguístico volumoso. Usada manualmente, os elementos de frames são anotados com os tipos ontológicos e as sentenças recebem mais uma camada de anotação. A partir dessa anotação e dos tipos ontológicos esperados para os elementos de frame, a sentença é analisada segundo as restrições definidas na formalização.

Para que o sistema seja útil em larga escala é preciso primeiro implementar um léxico anotado com tipos ontológicos presentes em uma ontologia de nível topo. Essa anotação pode ser feita manualmente, o que é um projeto complexo e longo, ou pode-se usar alguma técnica de anotação automática por meio de inferência estatística através de relações de coocorrência.

A construção de uma anotação ontológica de um léxico com base em inferência estatística é tópico promissor e existem pesquisadores estudando esse tema, como Zhang Rui-ling e Xu Hong-sheng (2009). Porém esse tópico está fora do escopo desta



pesquisa. Para procedermos às análises exigidas de forma a validarmos nossa proposta metodológica, utilizaremos uma anotação manual dos lexemas relevantes para um domínio de estudo. Isto foi suficiente para este trabalho, uma vez que nosso foco são os passos necessários para associar as sentenças a frames correlacionados, a partir de uma anotação existente previamente.

Além da aplicação manual da formalização e da anotação ontológica, é possível aplicar as regras de forma automática. Neste trabalho, são sugeridos dois caminhos para essa automatização, sendo que um foi o aplicado em um estudo de caso.

O primeiro caminho é o desenvolvimento de anotadores automáticos para cada camada (*Part of speech* ou POS, ontológica, papel semântico) que anotem precisamente cada lexema. A partir desta anotação e com base em uma ampla ontologia associada a um léxico é possível aplicar as regras para separação das sentenças, segundo frames relacionados. Essa abordagem exige anotadores precisos, principalmente sobre a natureza ontológica do termo, no contexto em que está sendo empregado. Isto pode ser feito a partir de uma análise probabilística condicional (análise Bayesiana<sup>25</sup>) realizada em uma grande base textual. Porém, essa análise é um amplo programa de pesquisa por si só e está fora do escopo deste trabalho. Uma análise das técnicas disponíveis para criação de ontologias a partir de bases textuais pode ser encontrada em CIMIANO (2006).

Uma forma menos direta de aplicar as regras da formalização é a utilização de um mapeamento *fuzzy* que embutiria em suas regras as normas definidas na formalização e emitiria para cada sentença o grau de pertinência para um determinado frame. Nesse caso, não é necessária uma anotação completa da sentença, mas apenas é necessário um subconjunto de evidências, o suficiente para estabelecer certo grau de pertinência ao frame. Por exemplo, no caso do frame VIAGEM, a ocorrência de lexemas que pertençam aos tipos ontológicos *consciente*, *veículo/transporte* (téllico) e *local geográfico* fortalecem a possibilidade de pertinência ao frame. Esse mapeamento reflete a noção codificada nas definições 6, 7, 11 e 12 da formalização. Ainda assim, a

---

<sup>25</sup> A inferência bayesiana é uma análise probabilística que infere o grau de um evento ocorrer, dada a ocorrência de uma ou mais evidências. Assim, pode ser empregada para se inferir a natureza ontológica de um lexema, dada a coocorrência de outros elementos linguísticos.

atribuição do grau de pertinência é apenas um indicador e seu grau de acerto deve ser verificado posteriormente por um analista humano.

## 6.6 Experimento Corroborador

Exemplificaremos, agora, a aplicação da formalização em sentenças retiradas de um *corpus*. Para isso é preciso estabelecer uma cena alvo, expressa por um frame, para teste. 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*. Seguem algumas sentenças do Corpus do Português com esses lexemas e que evocam a cena de *viagem*.

- *Entre os dias 5 e 7, a equipe do programa participou de um **safari** em Mala Mala, reserva que fica na província do Transvaal (leste do país).* <CF8688-4>
- *Quem quiser, vai em **excursão** até o resort local, Casa de Campo.* <CF944-5>
- *Pedro Francisco de Moraes retornou de **peregrinação** a Roma e participou de Missa Solene, na Basílica de São Pedro.* (19N:Br:PA: Sociedade/E. Conill)

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, usando a formalização definida. A ontologia usada na anotação é a ontologia SIMPLE-CLIPS. 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>26</sup> (SALOMÃO, 2009a) possui uma entrada para o frame VIAGEM, como mostra a FIG. 32. 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 Boas (2009), é discutido com profundidade, os desafios para criação de FrameNets Multilínguas. 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 neste trabalho, uma vez que foram encontradas várias evidências de seu uso no *corpus* criado para o estudo de caso.

---

<sup>26</sup> <http://www.framenetbr.ufjf.br>

## . Viagem [Travel]

### 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.

### Elementos do Frame

### Nucleares (Core)

**Área [Area]**

Este EF corresponde à **Área** em que a viagem ocorre. Descreve uma área em meio à qual a viagem ocorre, não especificando o **Caminho**, a **Fonte** ou o **Alvo**.

**Tipo Semântico**

Localização

**Direção [Direction]**

A **Direção** em que o **Viajante** vai.

**Alvo [Goal]**

O alvo é a localização aonde o **Viajante** chega.

**Tipo Semântico**

Alvo

**Meio de transporte [Mode of Transportation]**

O **Meio de transporte** expressa o modo como o **Viajante** se locomove.

**Caminho [Path]**

O **Caminho** é o curso que o **Viajante** percorre.

**Tipo Semântico**

Caminho

**Fonte [Source]**

Ponto de partida da viagem.

**Tipo Semântico**

Fonte

**Viajante [Traveler]**

Este é o ser vivo que viaja. Normalmente é expresso como argumento externo.

**Tipo Semântico**

Consciente

FIGURA 32 - Frame de VIAGEM - definição e elementos nucleares.

FONTE – FrameNet de Berkeley.

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.

A FrameNetBr não apresenta a rede de ligações semânticas do frame VIAGEM com outros frames da base. Nesse caso, vamos apresentar a rede de ligações que consta na FrameNet de Berkeley. A FIG. 33 mostra as relações.

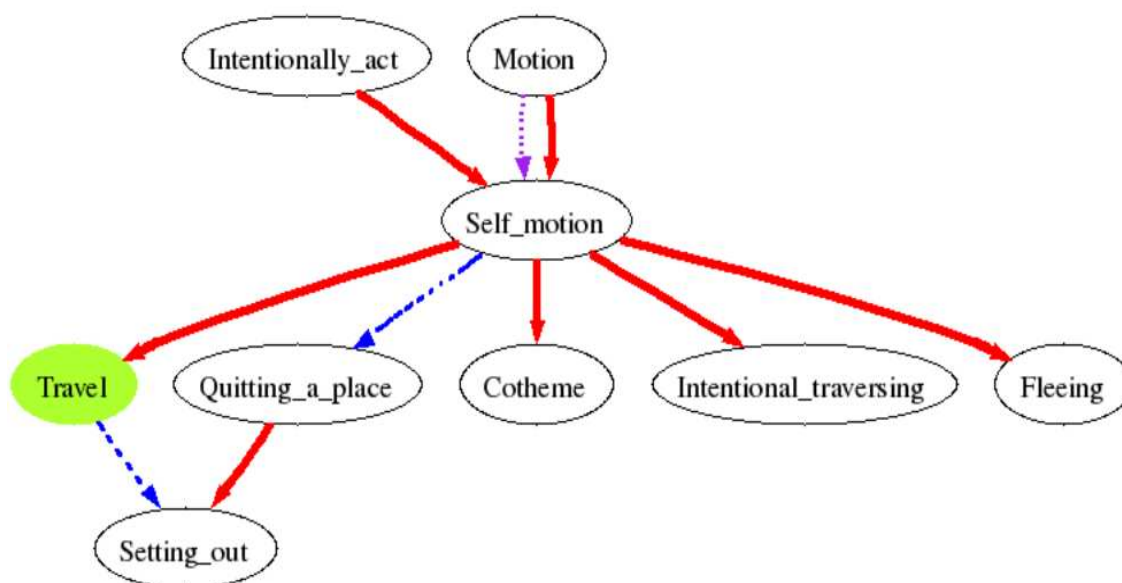


FIGURA 33 – Frame de TRAVEL: relações semânticas.

FONTE – 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. Essas relações podem ser observadas graficamente com o uso da ferramenta FrameGrapher<sup>27</sup>. 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

<sup>27</sup> <https://framenet.icsi.berkeley.edu/fndrupal/FrameGrapher>.

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, formalizada na **definição 9**, que estabelece que tudo que é verdade para o frame pai deve ser verdade para o frame filho, não se pode dizer que a **Viagem** é um movimento próprio. Essa inconsistência foi reportada no fórum da FrameNet de Berkeley e, recentemente, o frame SELF-MOTION foi redefinido para acomodar a possibilidade de uso de veículos:

The Self\_mover, a living being, moves under its own direction along a Path. Many of the lexical units in this frame can also describe the motion of vehicles (ex. as external arguments). We treat these as belonging in this frame. (FrameNet, SELF\_MOTION)

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 e impactam na **definição 9**.

O uso mais prototípico da palavra **viagem** ocorre em deslocamentos de longa distância realizados por pessoas, como mostrado na sentença a seguir (FIG. 34):

<i>Camada</i>	Eu uma vez fiz uma <b>viagem</b> a Mato Grosso (19Or:Br:lf:sp)
<i>EF (FrameNet)</i>	 <i>viajante UL alvo</i>
<i>ontológico (frame)</i>	Evento
<i>ontológico (elementos)</i>	humano Localização(constitutivo)

FIGURA 34 - Anotação ontológica de sentença - exemplo 1.

O uso prototípico evoca uma cena onde o viajante é um ser humano. Ao evocar a cena prototípica, pela **definição 6**, a valência tem que ser respeitada de forma implícita ou explicitamente. A verificação de que realmente a cena é evocada pode ser feita pela análise de que seria natural para um ouvinte, após ouvir a frase acima, realizar as seguintes perguntas: de onde você partiu (*fonte*)? Por onde você passou (*caminho*)?

Como você viajou (*meio\_de\_transporte*)? Outros elementos nucleares podem ser inferidos a partir do enunciado: a *área* é o Brasil, deduzido a partir do alvo. A *direção* pode ser inferida a partir do conhecimento da *fonte* e do *alvo*.

Existem casos em que o item lexical é empregado em cenas onde o viajante é uma entidade consciente não humano, como mostrado no exemplo a seguir (FIG. 35). Nos exemplos subsequentes, não será mostrada a camada *ontológico* (*frame*), uma vez que ela permanece inalterada.

<i>Camada</i>	No final do inverno, e primavera, os bandos iniciam sua <b>viagem</b> de volta para os pólos a fim de alimentar os adultos e filhotes (19Ac:Br:Enc:Baleia)									
<i>EF</i> ( <i>FrameNet</i> )	<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 50%; border: none;"> _____ </td> <td style="width: 20%; border: none;"> _____ </td> <td style="width: 30%; border: none;"> _____ </td> </tr> <tr> <td style="border: none;"> _____ </td> <td style="border: none; text-align: center;"><i>viajante</i></td> <td style="border: none; text-align: center;">UL</td> </tr> <tr> <td style="border: none;"><i>alvo</i></td> <td></td> <td></td> </tr> </table>	_____	_____	_____	_____	<i>viajante</i>	UL	<i>alvo</i>		
_____	_____	_____								
_____	<i>viajante</i>	UL								
<i>alvo</i>										
<i>Ontológico</i>	Localização(constitutivo)                      animal									

FIGURA 35 - Anotação ontológica de sentença - exemplo 2.

Este exemplo, a princípio, não viola as restrições ontológicas, uma vez que, pela definição, o viajante precisa ser, no mínimo, um ser consciente. Sendo assim, as **definições formais 7 e 10** são respeitadas. No entanto, o caso prototípico de viagem envolve seres humanos, sendo que viagens envolvendo animais teriam um significado relacionado com migração. Sendo assim, talvez fosse apropriado criar mais um frame com esse sentido correlato.

Em algumas sentenças, o lexema **viagem** possui um sentido de deslocamento no tempo e não no espaço, como mostrado na sentença a seguir. Apesar de qualquer deslocamento espacial implicar em um deslocamento temporal, o deslocamento mencionado é metafórico, uma vez que não houve um deslocamento físico pelo tempo (FIG. 36).

<i>Camada</i>	O passeio é uma <b>viagem</b> em o tempo e através de o Pacífico.                      (19N:Br:folha)								
<i>EF</i> ( <i>FrameNet</i> )	<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 25%; border: none;"> _____ </td> <td style="width: 25%; border: none;"> _____ </td> <td style="width: 25%; border: none;"> _____ </td> <td style="width: 25%; border: none;"> _____ </td> </tr> <tr> <td style="border: none;"></td> <td style="border: none; text-align: center;"><i>UL</i></td> <td style="border: none; text-align: center;"><i>area</i></td> <td style="border: none; text-align: center;"><i>caminho</i></td> </tr> </table>	_____	_____	_____	_____		<i>UL</i>	<i>area</i>	<i>caminho</i>
_____	_____	_____	_____						
	<i>UL</i>	<i>area</i>	<i>caminho</i>						
<i>Ontológico</i>	evento                      tempo                      Localização(constitutivo) (entidade_abstrata)								

FIGURA 36 - Anotação ontológica de sentença - exemplo 3.

A **definição 11** aborda este caso, sugerindo que se trata de um uso metafórico do lexema, uma vez que o elemento de frame *área* possui um tipo ontológico mais abstrato. A cena genérica que é evocada pelo enunciado está relacionada com o frame de RELEMBRAR\_EXPERIÊNCIA (*Remembering\_experience*). Esse frame consta na FrameNet de Berkeley.

A palavra **viagem** é empregada também em cenas em que o viajante não é um ente animado (FIG. 37).

<i>Camada</i>	As duas sondas percorrerão uma trajetória particularmente complicada em sua <b>viagem</b> de quase sete anos. (19N:Br:Recf)			
<i>EF</i> ( <i>FrameNet</i> )	_____  <i>viajante</i>	_____  <i>caminho</i>	_____  <i>UL</i>	_____  <i>duração</i>
<i>ontológico</i>	artefato(agentivo)	localização		tempo

FIGURA 37 - Anotação ontológica de sentença – exemplo 4.

Isso pode ser explicado pelo fato de que qualquer entidade que apresente um movimento, próprio ou causado, pode ser encarada como um viajante, desde que o deslocamento seja relativamente longo. No entanto, isso não é um sentido prototípico da palavra. Também não é um uso metafórico, como delimitado pela **definição 11**, uma vez que não envolve um elemento de frame de tipo mais abstrato, mas apenas de tipo diferente. Isso não quer dizer que não é um uso metafórico, mas que sua detecção não é contemplada pela **definição 11**. Ainda assim, existe a violação da **definição 7** e, em decorrência disso, deve-se analisar a necessidade da criação de um frame relacionado.

Ainda mais distante do uso prototípico é o emprego de itens lexicais associados à cena de VIAGEM, para designar deslocamento de entidades abstratas, como exemplificado a seguir (FIG. 38):

<i>Camada</i>	A imaginação de Verne <b>viagrou</b> por todos os quadrantes do planeta e até para bem longe dele. (19N:Br:SP)		
<i>EF</i> ( <i>FrameNet</i> )	_____  <i>viajante</i>	_____  <i>UL</i>	_____  <i>área</i>
<i>Ontológico</i>	fato_cognitivo		localização

FIGURA 38 - Anotação ontológica de sentença - exemplo 5.





As análises das sentenças por meio das definições formais mostram-se úteis para a diferenciação das cenas associadas aos enunciados. A aplicação no caso em estudo mostra que os enunciados estabelecem uma graduação em relação ao seu enquadramento na cena. O enunciado prototípico envolve uma pessoa realizando um deslocamento longo auxiliado por um veículo. Se afastando um pouco do significado central temos o deslocamento de animais utilizando-se de meios próprios, caracterizando mais uma migração. Menos prototipicamente, temos enunciados envolvendo objetos não conscientes. Temos também o uso do lexema em metáforas, reportando viagens através de espaços abstratos, como o tempo e a imaginação. As discrepâncias apontadas pelas definições podem ser utilizadas para agrupar os enunciados e indicar a necessidade de criação de novos frames. Por exemplo, os lexemas *viagem* e *viajar*, quando aplicados a animais, evocam a mesma cena quando aplicados a humanos? A análise mostra que não. A análise mostrou que o lexema *viagem* é usado em português no sentido de uma cena de VIAGEM prototípica (deslocamento de pessoa), no sentido metafórico (deslocamento numa dimensão abstrata ou de um viajante abstrato), no sentido de migração (em se tratando de animais, não possui frame correspondente na FrameNet) e no sentido de transporte (sem frame correspondente na FrameNet) e, por isso, não evocaria a cena de VIAGEM. Este sentido pode ser evidenciado pela distinção do tipo ontológico do agente, que, neste caso, exerce o papel de transportador. A ligação entre os dois frames (VIAGEM e TRANSPORTE) por meio de um *link* de distinção ontológica, do agente (consciente *versus* artefato) pode ajudar na compreensão dos fenômenos relacionados à polissemia. Como também, quando o lexema é usado no sentido metafórico, evocando uma cena distinta, novamente, as distinções ontológicas se fazem presentes.

As definições funcionam como guias e justificam as decisões. No entanto, para grandes volumes de sentenças, o uso das definições pode ser lento e tedioso. Por isso, o ideal é que as definições sejam codificadas em algum sistema computacional capaz de realizar a análise de forma automática. As técnicas mais adequadas seriam técnicas probabilísticas ou que façam uso de lógica nebulosa. A justificativa é que o enquadramento de um enunciado em uma cena não é exato e está sujeito a uma gradação, o que é tratado por essas técnicas. Na próxima seção são exemplificados este mapeamento e a técnica selecionada: a técnica probabilística das *redes bayesianas*.

## 6.7 Distinção ontológica por meio de Redes Bayesianas

O Teorema de Bayes e as redes bayesianas já foram apresentados na seção que trata da metodologia empregada nesta pesquisa. Aqui 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, seguindo o que preconizam as definições de 7 a 11, a respeito dos tipos ontológicos.

As redes bayesianas foram aplicadas sobre a saída produzida pelo programa de anotação, cujos passos de anotação também foram descritos na seção que trata da metodologia empregada nesta pesquisa. Aqui nos limitaremos a apresentar os resultados da anotação, as entradas e as saídas dos programas. O *subcorpus* utilizado nos testes, com 57 sentenças, está apresentado na FIG. 42.

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 Montevidéu 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 Cadillac 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 42 – *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** e, propositadamente, não se preocupou em incluir apenas sentenças bem comportadas no *subcorpus*, do tipo “*alguém viajou para algum lugar*”. 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 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.

Foi criada uma ontologia de domínio com base no *subcorpus* para a anotação das sentenças. A ontologia de domínio criada serve apenas o propósito de se testar a anotação do *subcorpus* e apoiar a análise das sentenças, não devendo ser considerada como uma ontologia 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 FIG. 43 mostra os 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;agentivo=;telico=papel\_social  
 Atlantis/formal=veiculo;constitutivo=estrutura;agentivo=artefato;telico=transporte  
 Brasil/formal=local\_geopolitico  
 Cadillac/formal=veiculo;constitutivo=estrutura;agentivo=artefato;telico=transporte  
 Calçada/formal=local  
 Cassini/formal=entidade\_concreto;constitutivo=estrutura;agentivo=artefato;  
 Chateaubriand/formal=humano;constitutivo=estrutura  
 Columbia/formal=veiculo;constitutivo=estrutura;agentivo=artefato;telico=transporte  
 Comprei/formal=processo  
 Coronel/formal=humano;constitutivo=estrutura;agentivo=;telico=papel\_social  
 Embarcou/formal=evento  
 Eu/formal=humano;constitutivo=estrutura  
 Europa/formal=local\_geopolitico  
 Fez/formal=evento  
 Foi/formal=evento  
 Foz/formal=local  
 França/formal=local\_geopolitico  
 Itajaí/formal=local\_geopolitico  
 Japão/formal=local\_geopolitico  
 Jerusalém/formal=local\_geopolitico  
 Lisboa/formal=local\_geopolitico  
 Manaus/formal=local\_geopolitico  
 Maputo/formal=local\_geopolitico  
 noivos/formal=humano;constitutivo=agrupamento;agentivo=;telico=papel\_social  
 Noruega/formal=local\_geopolitico  
 Pernambuco/formal=local\_geopolitico  
 Pássaros/formal=animal;constitutivo=agrupamento  
 Recife/formal=local\_geopolitico  
 Rodoviária/formal=local;constitutivo=estrutura;agentivo=artefato  
 Sol/formal=local  
 Soldados/formal=humano;constitutivo=agrupamento;agentivo=;telico=papel\_social  
 Temuco/formal=local\_geopolitico  
 Terra/formal=local  
 Tijuca/formal=local\_geopolitico  
 Túnis/formal=local\_geopolitico  
 Universidade/formal=entidade\_concreto;constitutivo=agrupamento;agentivo=;telico=ensino  
 Ural/formal=local\_geopolitico  
 Velho/formal=propriedade;constitutivo=;agentivo=;telico=  
 Viagem/formal=processo  
 Viajavam/formal=processo  
 Viajei/formal=processo  
 Viajo/formal=processo  
 acidente/formal=evento  
 activistas/formal=humano;constitutivo=agrupamento;telico=papel\_social  
 agosto/formal=entidade\_abstrata  
 albatrozes/formal=animal;constitutivo=agrupamento  
 ano/formal=entidade\_abstrata  
 anos/formal=entidade\_abstrata  
 asas/formal=entidade\_concreto;constitutivo=parte-de;telico=movimento  
 atacante/telico=papel  
 automóvel/formal=veiculo;constitutivo=estrutura;agentivo=artefato;telico=transporte  
 averbar/formal=evento  
 ave/formal=animal  
 aves/formal=animal;constitutivo=agrupamento  
 avião/formal=veiculo;constitutivo=estrutura;agentivo=artefato;telico=transporte  
 axilas/formal=entidade\_concreto;constitutivo=parte-de  
 bagagem/formal=entidade\_concreta;agentivo=artefato  
 bailarina/formal=humano;constitutivo=agrupamento;agentivo=;telico=papel\_social  
 baixa/formal=propriedade  
 bilhões/constitutivo=quantidade  
 boate/formal=local;constitutivo=estrutura;agentivo=artefato  
 bonde/formal=veiculo;constitutivo=estrutura;agentivo=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  
 cá/formal=local  
 calado/formal=propriedade  
 camelô/formal=humano;telico=papel\_social  
 caminhonete/formal=veiculo;constitutivo=estrutura;agentivo=artefato;telico=transporte  
 caravela/formal=veiculo;constitutivo=estrutura;agentivo=artefato;telico=transporte  
 carro/formal=veiculo;constitutivo=estrutura;agentivo=artefato;telico=transporte  
 carroças/formal=veiculo;constitutivo=estrutura;agentivo=artefato;telico=transporte  
 cesto/formal=entidade\_concreta;;constitutivo=estrutura;agentivo=artefato;telico=recipiente  
 cidade/formal=local\_geopolitico  
 cometas/formal=entidade\_concreto  
 constelações/formal=local  
 crianças/formal=humano;constitutivo=agrupamento;telico=papel  
 espaçonave/formal=veiculo;constitutivo=estrutura;agentivo=artefato;telico=transporte  
 estudante/formal=humano;constitutivo=estrutura;telico=papel\_social  
 fazenda/formal=local  
 feijão/formal=entidade\_concreto  
 filhos/formal=humano;constitutivo=agrupamento;agentivo=;telico=papel\_social  
 foguete/formal=veiculo;constitutivo=estrutura;agentivo=artefato;telico=transporte  
 formigas/formal=animal;constitutivo=agrupamento  
 fótons/formal=entidade\_física  
 general/formal=humano;constitutivo=estrutura;telico=papel\_social  
 gente/formal=humano  
 historiador/formal=humano;telico=papel\_social  
 homem/formal=humano  
 imaginação/formal=abstrato  
 informações/formal=abstrato  
 jaca/formal=entidade\_concreto  
 lábios/formal=entidade\_concreta;constitutivo=parte\_corpo  
 locomotiva/formal=veiculo;constitutivo=estrutura;agentivo=artefato;telico=transporte  
 longa/formal=propriedade\_temporal  
 Lua/formal=local  
 militar/formal=propriedade  
 moto/formal=veiculo;constitutivo=estrutura;agentivo=artefato;telico=transporte  
 motoristas/formal=humano;constitutivo=agrupamento;telico=papel\_social  
 naus/formal=veiculo;constitutivo=estrutura;agentivo=artefato;telico=transporte  
 nave/formal=veiculo;constitutivo=estrutura;agentivo=artefato;telico=transporte  
 navio/formal=veiculo;constitutivo=estrutura;agentivo=artefato;telico=transporte  
 naus/formal=veiculo;constitutivo=estrutura;agentivo=artefato;telico=transporte  
 negro/formal=propriedade  
 negros/formal=propriedade  
 notícias/formal=abstrato  
 ônibus/formal=veiculo;constitutivo=estrutura;agentivo=artefato;telico=transporte  
 pai/formal=humano;agentivo=;telico=papel\_social  
 partiram/formal=evento  
 partículas/formal=entidade\_física  
 passado/formal=abstrato  
 passageiro/telico=transportado  
 passaporte/formal=entidade\_concreta;constitutivo=;agentivo=artefato;telico=identidade  
 países/formal=local\_geopolitico;constitutivo=agrupamento  
 pele/formal=entidade\_concreta;constitutivo=parte\_corpo  
 pensamento/formal=abstrato  
 planetas/formal=local;constitutivo=agrupamento  
 planeta/formal=local  
 policiais/formal=humano;constitutivo=agrupamento;agentivo=;telico=papel\_social  
 porcos/formal=animal;constitutivo=agrupamento  
 porcos-voadores/formal=abstrato  
 poucas/formal=quantidade  
 poucos/formal=quantidade  
 prefeito/formal=humano;telico=papel\_social  
 presidente/formal=humano;telico=papel\_social  
 primeira/formal=propriedade  
 primeiro/formal=propriedade  
 primeiros/formal=propriedade  
 problema/formal=propriedade  
 provocou/formal=evento  
 radar/formal=entidade\_concreto;constitutivo=estrutura;agentivo=artefato;telico=perceber  
 secretário/formal=humano;telico=papel\_social  
 século/formal=entidade\_abstrata

selva/formal=local
semana/formal=entidade_abstrata
semente/formal=entidade_concreto
setembro/formal=entidade_abstrata
sexta/formal=entidade_abstrata
sondas/formal=entidade_concreto;constitutivo=estrutura;agentivo=artefato;telico=perceber
temperatura/formal=propriedade
tempo/formal=abstrato
transferência/formal=evento
trem/formal=veiculo;constitutivo=estrutura;agentivo=artefato;telico=transporte
três/formal=propriedade_numeral
turma/formal=propriedade;constitutivo=agrupamento;
universo/formal=local
vagão/formal=veiculo;constitutivo=estrutura;agentivo=artefato;telico=transporte
velha/formal=propriedade
velho/formal=propriedade
velocidade/formal=propriedade
velocidades/formal=propriedade
vermiformes/formal=propriedade
veiculos/formal=entidade_concreto;constitutivo=estrutura;agentivo=artefato;telico=transporte
viagem/formal=processo
viajam/formal=processo
viajar/formal=processo
viajaram/formal=processo
viajava/formal=processo
viajavam/formal=processo
viajo/formal=processo
viajou/formal=processo
vinham/formal=evento
viva/formal=propriedade
vítimas/formal=humano;constitutivo=agrupamento;telico=papel_social
viúva/formal=humano;telico=papel_social
voadores/formal=propriedade
vou/formal=evento
vôo/formal=evento
época/formal=propriedade
órbitas/formal=local
única/formal=propriedade

FIGURA 43 – Ontologia de domínio.

A anotação deste *subcorpus* com as camadas de POS e ontológica gerou uma tabela parcialmente mostrada na FIG. 44. A tabela completa consta no Anexo.

Viajou	hoje	no	bonde	um	homem	embriagado	meio	dormindo		
V	ADV	PREPDA	CN	UM	CN		PNM	ADJ	GER	
A	imaginação	de	Verne	viajou	por	todos	os	quadrantes	do	planeta
DA	CN	PREP	PNM	V	PREP	QNT	DA	CN	PNT	CN
ela	fez	uma	viagem	de	volta	ao	fim	do	século	passado
PRS	V	UM	CN	PREP	CN	LPREP1	LPREP2	LPREP3	CN	PPA
Um	arrepio	viajou	lhe	pelo	corpo					
UM	CN			PREPDA	CN					
Elesbão	viajava	de	ônibus	para	São Paulo					
PNM	V	PREP	PNM	PREP	PNM					

FIGURA 44 – Parte da anotação das sentenças do *subcorpus*.

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

A criação das redes semânticas seguiu dois caminhos distintos para que fossem possíveis análises usando dois softwares diferentes de análise bayesiana. O uso de dois softwares foi empregado para permitir os pontos fortes de cada ambiente: facilidade de visualização em um caso, e facilidade de análise de grandes volumes de dados no outro caso.



Para o primeiro software as probabilidades de cada nó foram calculadas manualmente, a partir da planilha mostrada na FIG. 44. Já o segundo software calculou automaticamente as probabilidades a partir de um arquivo, em um formato específico, gerado pelo programa anotador. As análises geradas por essas duas ferramentas computacionais serão descritas a seguir.

### 6.7.1 Análise com a Ferramenta MSBNx

O MSBNX (*Microsoft Bayesian Network Editor*) (KADIE; HOVEL; HORVITZ, 2001) é uma ferramenta para análise probabilística com uso de redes bayesianas disponibilizada gratuitamente pela Microsoft® no site<sup>28</sup>. Apesar de ser gratuito MSBNx não é uma ferramenta de código aberto. Como ferramenta de análise probabilística, ela possui a vantagem de ser fácil de ser usada e de produzir gráficos, visualmente, interessantes. Como desvantagem, a ferramenta é pouco flexível para a análise de arquivos padronizados e ajustes de testes. Por essa razão foi decidido a realização de um teste com um caráter mais didático com a MSBNx e, posteriormente, uma análise como uma ferramenta mais robusta, que será apresentada na próxima subseção. A versão utilizada da MSBNx foi a 1.4.2.

Em função da ferramenta não realizar cálculo das probabilidades e de não gerar a rede de dependência automaticamente, essas etapas foram feitas manualmente. Veremos na próxima subseção, que a segunda ferramenta utilizada realiza essas etapas automaticamente.

A ideia é estabelecer probabilidade de um enunciado  $S$  pertencer a um Frame  $F$  (hipótese  $H$ ) dadas as evidências de ocorrência de tipos ontológicos  $E$ , expressa por  $P(H/E)$ . A probabilidade da ocorrência de um tipo ou tipos ontológicos  $E$  implicar no Frame  $F$  são obtidas do *subcorpus* de treinamento. Ela é calculada pela divisão do número de ocorrências de  $E$  quando ocorre  $F$ , pelo número de ocorrências de  $E$ , expressa pela razão:  $(ocorrências(E \wedge F) / ocorrências(E))$ . Dessa forma, se em 100

---

<sup>28</sup> <http://research.microsoft.com/en-us/um/redmond/groups/adapt/msbnx/>.

ocorrências de um determinado tipo ontológico, em 70 a sentença estava associada a uma determinada cena, a ocorrência do tipo ontológico indica uma chance 70% da sentença estar associada à cena.

A análise do *subcorpus* gerou uma rede de dependência probabilística, que está expressa na Fig. 45. A rede foi construída a partir das evidências de *corpus*, portanto, um *corpus* maior e com outros enunciados gerará outra rede e outras probabilidades condicionais. A dependência entre os nós foi estabelecida pela ocorrência dos tipos ontológicos concomitantemente com os frames. Os tipos ontológicos são as variáveis independentes e os nós **Viagem\_proto**, **Viagem\_metafora** e **Viagem\_quasiProto**, representam as cenas de viagem prototípica, viagem como metáfora e viagem com um certo grau de distanciamento da cena prototípica.

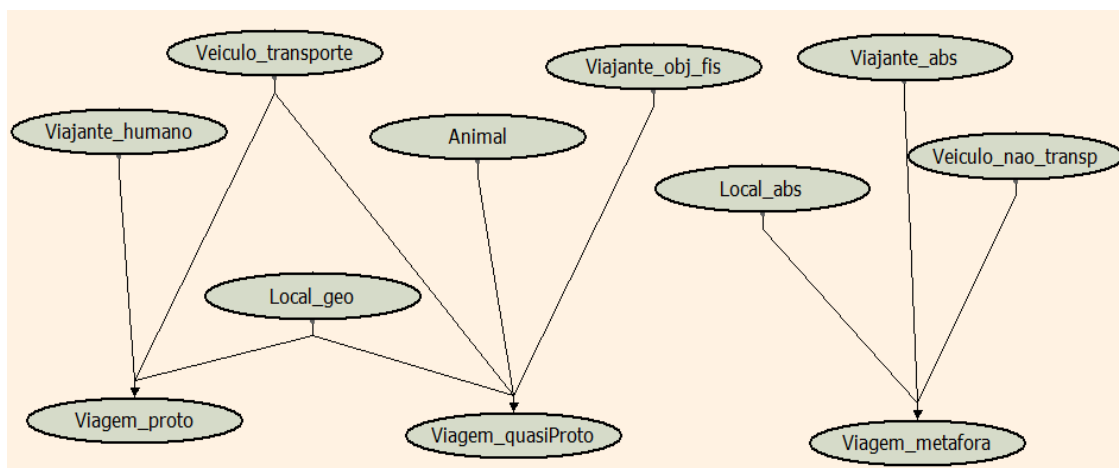


FIGURA 45 - Rede bayesiana para o Frame de VIAGEM.

A FIG. 46 mostra as probabilidades para a cena na qual os lexemas são usados metaforicamente.

Parent Node(s)			Viagem_metafora		
Viajante_abs	Local_abs	Veiculo_nao_transp	Yes	No	bar charts
Yes	Yes	Yes	1,0	0,0	
		No	1,0	0,0	
	No	Yes	1,0	0,0	
		No	0,0	1,0	
No	Yes	Yes	0,906	0,094	
		No	0,885	0,115	
	No	Yes	0,167	0,833	
		No	0,0	1,0	

FIGURA 46 - Probabilidades de uma cena de viagem ser metáfora.

A execução do sistema sobre a anotação de um enunciado gera a probabilidade do enunciado evocar uma determinada cena. Então, para o enunciado da FIG. 47, obteve-se um grau de pertinência de 78% para a Viagem\_quasiProto e zero para as demais cenas.

Enunciado:

“com semente de jaca e do feijão camaratu, que **viajaram** nas naves espaciais Columbia e Atlantis.” (19N:Br:SCat)

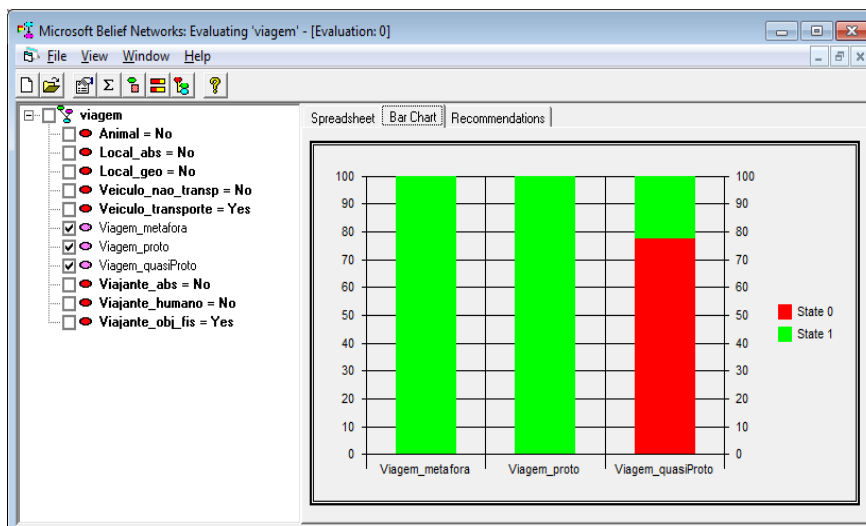


FIGURA 47 - Avaliação de um enunciado na rede bayesiana para o Frame de VIAGEM.

A apresentação da análise usando o software MSBNx foi realizada para mostrar de forma gráfica como a análise é realizada. Para a realização de uma análise mais completa foi utilizado o software Weka, cujos resultados serão apresentados a seguir.

## 6.7.2 Análise com a Ferramenta Weka

O Weka<sup>29</sup> é um conjunto de programas escritos na linguagem de programação Java voltados para tarefas de mineração de dados. O Weka 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 ferramentas para pré-processamento, classificação, regressão, agrupamento e visualização de dados. A FIG. 48 mostra a tela inicial da ferramenta.

<sup>29</sup> <http://www.cs.waikato.ac.nz/ml/weka>.



FIGURA 48 - Tela inicial da Weka.

Particularmente, para o problema em questão, foi decidido 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*. Neste 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. O anotador ontológico foi adaptado para gerar este arquivo e a FIG. 49 mostra o arquivo para as 57 sentenças do arquivo de teste.

```
@relation FrameViagem

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

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

```

V,F,F,F,V,prototipico
F,F,F,F,V,quasi
V,F,F,F,F,quasi
V,F,F,F,V,prototipico
V,V,F,F,F,quasi
V,F,F,F,V,prototipico
V,F,F,F,V,prototipico
F,F,F,F,F,quasi
F,V,V,F,F,quasi
V,F,F,F,V,prototipico
V,F,F,F,F,prototipico
F,F,F,F,V,prototipico
F,V,F,F,F,quasi
F,F,F,F,F,quasi
V,F,F,F,F,prototipico
F,V,F,F,F,quasi
F,V,F,F,V,prototipico
F,V,F,F,F,quasi
F,F,F,F,V,quasi
F,F,F,F,F,metaforico
F,F,V,F,F,quasi
F,F,F,V,F,metaforico
F,F,F,V,F,quasi
V,F,F,F,F,quasi
V,F,F,F,F,quasi
F,F,F,V,F,metaforico
F,F,F,V,F,metaforico
V,F,F,F,F,quasi
V,F,F,F,V,prototipico
F,F,V,F,F,quasi
F,F,F,F,V,quasi
F,F,F,F,F,quasi
V,F,F,F,F,quasi
V,F,F,F,V,prototipico
F,F,F,F,F,quasi
F,F,F,F,F,quasi
F,V,F,F,F,quasi
F,F,V,F,F,quasi
F,F,F,F,F,quasi
V,V,F,F,V,prototipico
V,V,F,F,F,quasi
F,F,F,V,V,metaforico
F,F,F,V,F,metaforico
F,F,F,F,F,metaforico
V,V,F,F,V,prototipico

```

FIGURA 49 – 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 foram definidos manualmente. Isto é necessário para treinar a rede e para verificar seu desempenho no conjunto de teste. A definição se uma sentença pertence à cena prototípica e o seu enquadramento na definição padrão do frame.

O treinamento da rede gerou a rede bayesiana mostrada pela FIG. 50. 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.

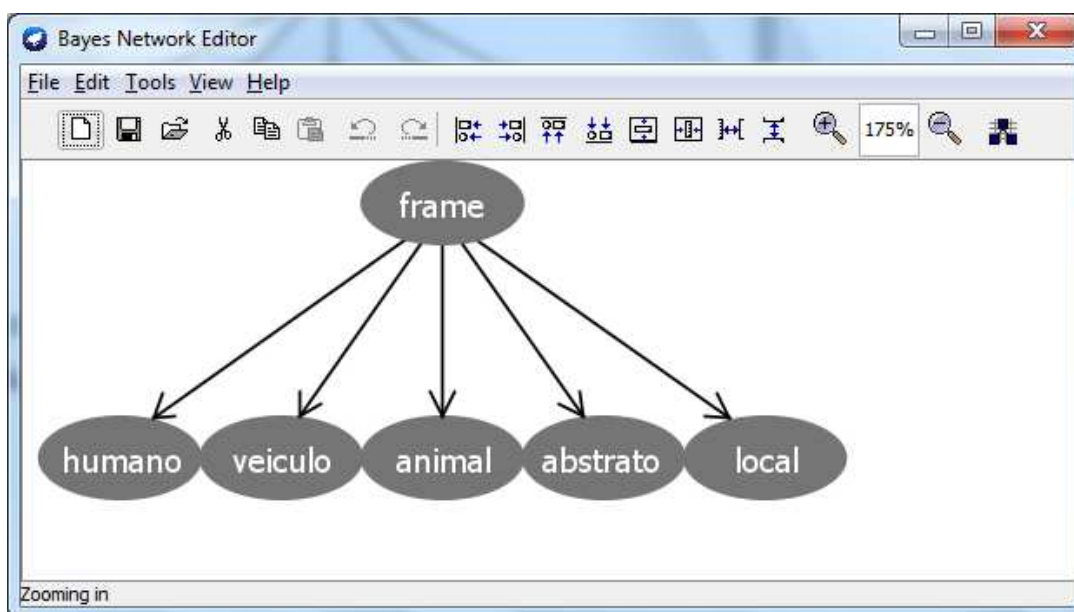
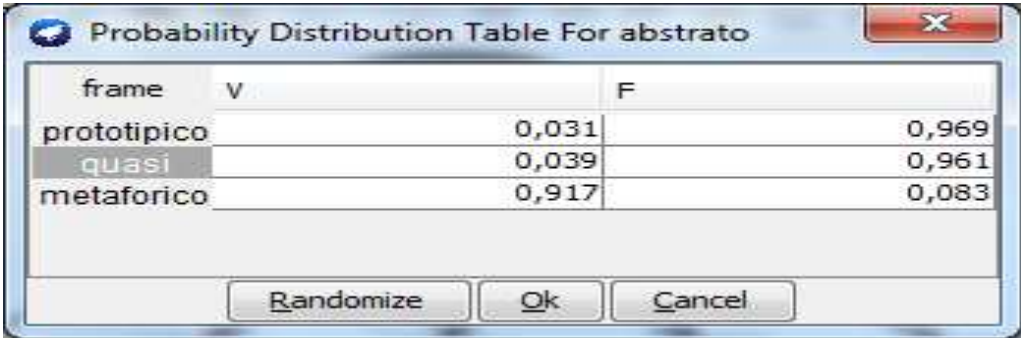


FIGURA 50 – Rede bayesiana gerada pelo Weka.

As probabilidades condicionais também são geradas automaticamente. A FIG. 51 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 51 – 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 FIG. 52 mostra o resultado da predição da rede.

```

=== Run information ===

Scheme:weka.classifiers.bayes.BayesNet          -D          -Q
weka.classifiers.bayes.net.search.local.K2     --  -P  1  -S  BAYES  -E
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
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

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:quasi    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

=== 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 52 – Saída da rede para o conjunto de teste.

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 do 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.

## 6.8 Discussão dos resultados

Neste ponto é importante avaliar os resultados obtidos pela proposta. A aplicação da metodologia, utilizando as regras formalizadas à base permitiu uma abordagem mais segura para a avaliação das sentenças.

Fica mais fácil visualizar inconsistências entre a cena e as instâncias de enunciados. Isso não significa que a formalização estabelece uma verdade indiscutível, mas, sim, uma base segura para a argumentação. Os passos inseridos na metodologia de elaboração da FrameNet adicionam a análise ontológica dentro do trabalho lexicográfico, o que é muitas vezes negligenciado ou feito de forma implícita.

A metodologia proposta não é mandatória. Ela pode ser adaptada às necessidades especiais da equipe de trabalho e funciona como um lembrete que o aspecto ontológico não deve ser negligenciado. A etapa 6 da metodologia (*montar a rede de sentidos dos frames*) não foi explicitada no exemplo do frame VIAGEM mas, com base na aplicação das regras da ontologia a rede resultante seria diferente da apresentada pela FrameNet de Berkeley.

A principal diferença seria a impossibilidade do Frame VIAGEM herdar do frame MOVIMENTO-PRÓPRIO em função da violação da regra 9. Uma vez que tudo que é verdade para o frame pai é verdadeiro para o frame filho, toda viagem teria que ser realizada por movimento próprio, sem exceções. Além disso, seria importante ter uma ligação de uso com o frame de TRANSPORTE e, poderíamos arriscar uma ligação metafórica (ainda não existente na FrameNet) com um frame abstrato, tal como DEVANEIO. A FIG. 53 mostra como seria essa sugestão de rede obtida pela metodologia proposta.

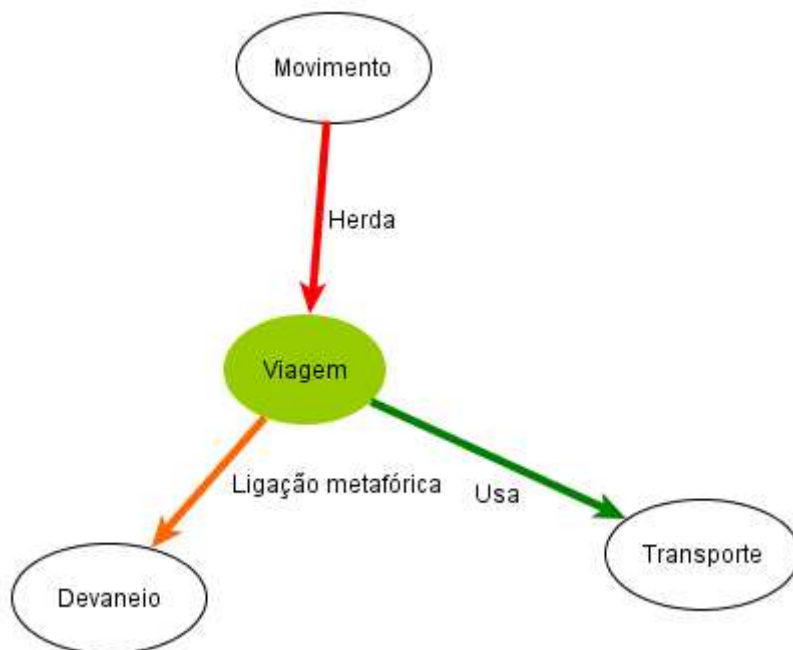


FIGURA 53 – Sugestão de rede para o frame VIAGEM.

A aplicação automática da formalização foi realizada apenas como prova de conceito. Ou seja, não está sendo proposto um sistema completo, até porque isso foge ao escopo do trabalho. O que está sendo proposto é um teste de viabilidade da automação do processo de enquadramento de enunciados a cenas usando um sistema probabilístico que se baseia nos aspectos ontológicos. Para uma maior precisão um número maior de elementos e sentenças deve ser adicionado ao sistema. Porém, como prova de conceito, os resultados obtidos mostram a viabilidade da proposta, mesmo avaliando apenas alguns poucos aspectos ontológicos, sem usar sua relação com os elementos de frames.

## 7 CONCLUSÕES

A interpretação de um item lexical é uma tarefa complexa e que envolve vários aspectos. O significado do item lexical não é inerente ao item e sim evocado por ele, como verificado pelos pesquisadores pioneiros da linguística cognitiva: Fillmore, Fauconnier, Turner e outros.

Além disso, suas interpretações são múltiplas e dinâmicas. O seu significado só pode ser estabelecido a partir do posicionamento em um frame conceitual. Essa ligação na mente humana é indissociável, formando um todo complexo, que não pode ser compreendido somente pela análise de suas partes. Todos os aspectos do significado, sua natureza, sua função, seus aspectos emocionais, fazem parte desse complexo conceitual.

No entanto, quando tentamos capturar essa estrutura conceitual em uma linguagem de representação, por mais expressiva que seja essa linguagem, parte do significado é perdido. O que se pode esperar é enriquecer a representação o suficiente para cumprir os propósitos almejados. Muitas vezes é necessário lançar mão de mais de um formalismo, para expressar os diversos aspectos de um significado que se fazem necessários para uma tarefa. Pode-se enriquecer uma representação simbólica, para tentar aproximar o máximo possível da representação conceitual interna, porém a equiparação é impraticável. Mesmo que essa tarefa fosse factível, a representação resultante seria difícil de ser compreendida e manipulada, em virtude de suas complexidades. Brachman e Levesque (2004, p.328) levantam e resumem essa questão declarando que “um fato fundamental da vida é que existe um balanço entre a expressividade da linguagem de representação e a tratabilidade computacional que está associada com a tarefa de raciocínio.”

A Semântica de Frames, proposta por Charles Fillmore, é uma teoria abrangente, capaz de abarcar e unir diversas teorias sobre as estruturas conceituais que se formam na mente humana, tais como esquemas imagéticos (Mark Johnson), modelos cognitivos idealizados (George Lakoff) e espaços mentais (Gilles Fauconnier). Ajuda também a explicar fenômenos linguísticos, tais como metáforas, metonímias, com base na

operação dessas estruturas conceituais. A Semântica de Frames, como teoria cognitiva tem sido verificada e validada pelos experimentos na área da Neurociência realizados pelo grupo de Teoria Neural da Linguagem, liderado por Jerome Feldman.

A FrameNet (Charles Fillmore e Collin Baker) é uma base lexical semântica baseada 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. Fortes evidências disso podem ser observadas nas pesquisas realizadas pelo o Instituto que abriga a FrameNet, o *International Computer Science Institute*, tais como Scheffczyk e Lönneker-Rodman. A associação de tipos semânticos aos elementos da FrameNet é, de certa forma, uma análise ontológica, mas, como ressaltam Scheffczyk *et al.* (2006), os tipos semânticos são limitados, não formalizados e não foram atribuídos a todos os elementos de frames.

A Teoria do Léxico Gerativo proposta por Pustejovsky procura se basear na hipótese de que a natureza, uso, estruturação e forma de geração de um elemento associado a um item lexical permite predizer a participação do item em um enunciado. Esses aspectos analisados pela TLG possibilitam a construção de dicionários léxicos de multicamadas e multifacetados, como é o caso do projeto SIMPLE-CLIPS, desenvolvido pelo grupo liderado por Antonio Zampolli. Mas os projetos que adotam essa abordagem deixam de lado um componente importante do significado lexical, que é a cena onde o item ocorre, o que é capturado pela Semântica de Frames de Fillmore. A junção de mais esse aspecto à semântica lexical permitirá a construção de léxicos mais completos.

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. A proposta inclui uma alteração na metodologia de desenvolvimento de frames. Todo o trabalho se baseia nas seguintes hipóteses:

- A Semântica de Frames é, atualmente, a teoria que melhor descreve as estruturas conceituais presentes no aparato cognitivo humano.
- A FrameNet é uma representação parcial dessas estruturas.
- A análise ontológica dos itens lexicais permite uma melhor análise e enquadramento dos enunciados.

A metodologia empregada na pesquisa foi a de analisar a metodologia correntemente empregada para a criação da FrameNet; propor uma alteração na metodologia, inserindo passos relacionados com a análise ontológica. Formalizar os elementos da FrameNet para atender às necessidades da análise ontológica. Definir um domínio (deslocamento espacial) e um frame (VIAGEM) para a realização de um estudo de caso. Definir uma técnica para automatização da análise. O frame selecionado para a análise foi o frame VIAGEM. A técnica selecionada para realizar a automatização foi a de redes bayesianas.

Com os resultados, a análise ontológica mostrou que o lexema *viagem* é usado em português no sentido de uma cena de Viagem prototípica (deslocamento de pessoa), no sentido metafórico (deslocamento numa dimensão abstrata) e no sentido de transporte (sem frame correspondente na FrameNet) e por isso não evocaria a cena de Viagem. Este sentido pode ser evidenciado pela distinção do tipo ontológico do agente, que, neste caso, exerce o papel de transportador. A ligação entre os dois frames (VIAGEM e TRANSPORTE) por meio de um *link* de distinção ontológica, do agente (consciente *versus* artefato) pode ajudar na compreensão dos fenômenos relacionados à polissemia. E também quando o lexema é usado no sentido metafórico, evocando uma cena distinta, novamente, as distinções ontológicas se fazem presentes.

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.

As contribuições e impactos do trabalho podem ser observados nas áreas da linguística cognitiva e da linguística computacional. Para a linguística cognitiva, uma contribuição importante deste trabalho é formalização da FrameNet. Ela permite uma compreensão mais precisa de seus conceitos pela comunidade de pesquisadores e, conseqüentemente, uma melhor comunicação para quem atua na área. Além disso, assegura um procedimento de geração de uma base lexical mais completa e bem formada.

A revisão bibliográfica também fornece uma contribuição importante, ao relatar como as descobertas da linguística cognitiva tem influenciado o desenvolvimento de sistemas computacionais e robóticos mais flexíveis e com melhor interação com os seres humanos.

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.

Como trabalhos futuros propomos a realização de uma análise automática usando os papéis semânticos dos frames e uso de um raciocinador para utilizar a cadeia hierárquica da ontologia nas inferências sobre os tipos ontológicos. A importância de uma análise automática que combina os papéis semânticos com o tipos ontológicos se baseia na hipótese de que a conceptualização é multifacetada e que, em decorrência disso, o entendimento de um enunciado envolve a combinação dessas facetas. Como destacado no presente trabalho, o conhecimento dos tipos ontológicos contribui fortemente para compreensão de sentenças em linguagem natural. No entanto, este é apenas um aspecto do conhecimento associado à sentença. Os micro papéis temáticos são de fundamental importância para composição do significado total. Portanto, a combinação destes dois aspectos tem o potencial de gerar análises automáticas mais precisas sobre o enquadramento de um enunciado em uma cena.

Dessa forma, na sentença “seu pensamento viajava por caminhos sem fim”, a informação de que o “pensamento” é uma entidade abstrata, associada à informação de está exercendo o papel de viajante, remete o enunciado a uma cena metafórica de forma mais categórica.

Já o uso de um raciocinador para utilizar a cadeia hierárquica da ontologia nas inferências sobre os tipos ontológicos se faz necessário para a correta automação da aplicação das regras 10 e 11 da formalização. Particularmente, no caso dessas regras é necessário percorrer a estrutura hierárquica da ontologia para a verificação se um tipo subjuga outro e se estão em linhas hierárquicas distintas. Este tipo de navegação estrutural é facilmente implementado por racionadores de lógica de descrições.

Ambas as propostas mencionadas têm o potencial de auxiliar o trabalho lexicográfico e poderão ser objetos de investigação em uma pesquisa de pós-doutoramento.



## REFERÊNCIAS

- ABRAÇADO, J. Entrevista com Maria Margarida Martins Salomão sobre a Linguística Cognitiva e suas relações com outras ciências. **Cadernos de Letras da UFF – Dossiê: Letras e cognição**, n.4, p. 15-25, 2010. (Entrevista)
- AMARO R.; MENDES S.; MARRAFA P. Lexical-conceptual relations as qualia role encoders. In: INTERNATIONAL CONFERENCE ON TEXT, SPEECH AND DIALOGUE (TSD'10), 13, 2010, Heidelberg, Berlin. **Proceedings...Heidelberg: Springer-Verlag, Berlin, 2010**, p. 29-36.
- ARMSTRONG, S. L. *et al.* What some concepts might not be. **Cognition**, n.13, p. 263-308, 1983.
- AUGUSTO, M. R. A. Dados de percepção/compreensão e de produção na aquisição: representações gramaticais distintas? **Letras de Hoje**. Porto Alegre, v. 42, n. 1, p. 113-130, março, 2007.
- BAKER, Collin F., FILLMORE, Charles J., LOWE, John B. The Berkeley FrameNet project. In: COLING-ACL, 1998, Montreal, Canada. **Proceedings of the COLING-ACL**, Montreal, Canada : [s.n.], 1998.
- BARSALOU, L.W. Ad hoc categories. **Memory & Cognition**, v.11, p. 211-227, 1983.
- BĀRZDIŅŠ, G. *et al.* Multidimensional Ontologies: Integration of Frame Semantics and Ontological Semantics. In: EURALEX INTERNACIONAL CONGRESS, 13, Barcelona, 2008. **Proceedings...** BERNAL, E.; DECESARIS, J. (eds.). Universitat Pompeu Fabra, Barcelona, July, v. 20, p. 277-284, 2008. (Serie Activitats).
- BIRD, Steven; KLEIN, Ewan; LOPER, Edward. **Natural Language Processing with Python**. O'Reilly Media, 2009.
- BOAS, Hans C. **Multilingual FrameNets in Computational Lexicography: Methods and Applications**. Berlin: Mouton de Gruyter, 2009. p. X+352.
- BORGES NETO, J. Ontology, Language, and Linguistic Theory. In: SIMPÓSIO LANGUAGE AND ONTOLOGY, julho de 2007, IEL/UNICAMP. **Proceedings...** IEL/UNICAMP, julho de 2007.
- BORIN, L. *et al.* Thinking Green: Toward Swedish FrameNet++. In: FRAMENET MASTERCLASS AND WORKSHOP, 2009, Milano. **Proceedings..** Milano, 2009.
- BRACHMAN, R. J. What IS-A is and isn't: an analysis of taxonomic links in semantic networks. **IEEE Computer**, 16, 10, p. 30-36, October. 1983.

\_\_\_\_\_. On the epistemological status of semantic networks. In: N.V. Findler (Ed.) **Associative Networks: representation and use of knowledge by computers**. Academic Press : [s.l.], 1979. p. 3-50.

\_\_\_\_\_; LEVESQUE, Hector J. **Knowledge Representation and Reasoning**, May 2004. 381 p.

\_\_\_\_\_; SCHMOLZE J. An overview of the KL-ONE Knowledge Representation System. **Cognitive Sci**, v.9, n.2, p. 171-216, 1985.

BRATMAN, M. **Intention, Plans, and Practical Reason**. Cambridge, Mass.: Harvard University Press, 1987.

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.

CHISHMAN, R. L. O sentido polissêmico dos verbos eventivos segundo a Teoria do Léxico Gerativo. **Fórum Linguístico**, v. 3, n. 2, 2003.

\_\_\_\_\_. Integrando léxicos semânticos e ontologias: uma aproximação a favor da web semântica. **Inf.Inf.**, Londrina, v. 14, n. esp., p.104-124, 2009.

\_\_\_\_\_; ALVES, Isa Mara da Rosa; BERTOLDI, Anderson. O conhecimento semântico representado em ontologias aplicadas à busca. In: SIMPÓSIO NACIONAL & I SIMPÓSIO INTERNACIONAL DE LETRAS E LINGÜÍSTICA, 11, Uberlândia, 2006. **Anais do XI Simpósio Nacional & I Simpósio Internacional de Letras e Lingüística**, 2006. p. 334-345.

CHOMSKY, N. **Aspects of a Theory of Syntax**. Cambridge, MA: MIT Press, 1965.

CHOMSKY, N. (1995). **The Minimalist Program**. Cambridge, MA: MIT Press, 1995.

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.

CIENKI, A. Frames, idealized cognitive models, domains. In: GEERAERTS, Dirk; CUYCKENS, Hubert (Eds.) **The Oxford Handbook of Cognitive Linguistics**. Oxford: Oxford University Press, 2007. p.170-187.

CIMIANO, P. **Ontology learning and population from text: algorithms, evaluation and applications**. Springer, 2006.

\_\_\_\_\_.; WENDEROTH, Johanna. Automatically Learning Qualia Structures from the Web. In: ACL-SIGLEX WORKSHOP ON DEEP LEXICAL ACQUISITION, 2005. **Proceedings...**, 2005. p. 28-37.

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

CROFT. Metonymy. In: GEERAERTS, D. *et al.* (Eds). **Cognitive Linguistics: Basic Readings**. Berlin : Mouton de Gruyter, 2006.

\_\_\_\_\_. Connecting frames and constructions: a case study of 'eat' and 'feed'. **Constructions and Frames**, v. 1, n.1, p.7-28, 2009.

\_\_\_\_\_; CRUSE, D. Alan. **Cognitive Linguistics**. Cambridge : Cambridge University Press, 2004. 374p.

CRUSE, D. Alan. Aspects of the micro-structure of word meanings. In: RAVIN, Y. LEACOCK, C. (Eds). **Polysemy: Theoretical and Computational Approaches**. Oxford: Oxford University Press, 2002. p. 30–51.

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

DAVIS, Wayne. Implicature. In: ZALTA, Edward N. (Ed.). **The stanford encyclopedia of philosophy**. Winter 2010 Edition.  
Available from Internet:  
<<http://plato.stanford.edu/archives/win2010/entries/implicature/>>.

DHARMENDRA S. M. *et al.* Cognitive Computing: unite neuroscience, supercomputing, and nanotechnology to discover, demonstrate, and deliver the brain's core algorithms. **Communications of the ACM**, v. 54, n. 8, p. 62-71, 2011.

DI PELLEGRINO, G. *et al.* Understanding motor events: a neurophysiological study. **Experimental Brain Research**, v. 91, n. 1, p. 176-180, 1992.

EVANS, V.; GREEN, M. **Cognitive Linguistics**: an introduction. Edinburgh : Edinburgh University Press, 2006. 830 p.

FAUCONNIER, Gilles; TURNER, Mark. Conceptual Integration Networks. **Cognitive Science**, v. 22, n. 2, p. 133-187, April-June 1998.  
Available at SSRN: <<http://ssrn.com/abstract=1292966>>.

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

FELDMAN, J. A. **Categories and concepts**- introduction. Spring 2008.102 p. (Slide).

FELLBAUM, C. **WordNet. an electronic lexical database**, MA: The MIT Press. 1998.

FIKES, R.; NILSSON, N. STRIPS: a new approach to the application of theorem proving to problem solving. **Artificial Intelligence**, 2, p. 189-208, 1971.

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. Volume 280: 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.

\_\_\_\_\_. **FrameNet: the lexicon**. ppt, 2008.

Available from Internet:<[http://www.hf.uib.no/forskingskole/FillmoreJune2\(a\)](http://www.hf.uib.no/forskingskole/FillmoreJune2(a))>

FILLMORE, C. J. Frame Semantics. In: GEERAERTS, Dirk (Ed.). **Cognitive Linguistics: basic readings**. 2006.

FODOR, J. **The language of thought**. Thomas Y. Crowell Co. Paperback, Harvard University Press. 1975.

\_\_\_\_\_; LEPORE, Ernie. The red herring and the pet fish: why concepts still can't be prototypes. **Cognition**, v.58, n.2, p. 253–70, 1996.

GALLESE, V.; LAKOFF, G. The brain's concepts: the role of the sensorimotor system in conceptual knowledge. **Cognitive Neuropsychology**, v.21, p. 455-479, 2005.

GAWRON, Jean Mark. **Frame Semantics**. San Diego State University, 2008. (Manuscript)

GEDIGIAN, Matt; BRYANT, John; CIRIC, Branimir. Catching metaphors. In **WORKSHOP ON SCALABLE NATURAL LANGUAGE UNDERSTANDING**. 2006. **Proceedings...**, 2006.

Available from Internet:<[www.icsi.berkeley.edu/~snarayan/metaphor.pdf](http://www.icsi.berkeley.edu/~snarayan/metaphor.pdf)>

GEERAERTS, D. **Cognitive Linguistics: Basic Readings**. Berlin: Mouton de Gruyter. 2006.

GEERAERTS, Dirk; CUYCKENS, Hubert (Eds.) **The Oxford Handbook of Cognitive Linguistics**. Oxford: Oxford University Press, 2007.

GIBBS, R. W. **The poetics of mind: figurative thought, language, and understanding**. Cambridge : Cambridge University Press., 1994.

GILDEA, Daniel; JURAFSKY, Daniel. Automatic Labeling of Semantic Roles. **Computational Linguistics**, v. 28, n.3, p. 245-288, 2002.

GLUCKSBERG, S. How metaphors create categories – quickly. In: GIBBS, R. (ed.) **Cambridge Handbook of Metaphor and Thought**. New York: Cambridge University Press. 2008. p.67-83.

GIVÓN, T. The visual information-processing system as an evolutionary precursor of human language. In: GIVÓN, Talmy; MALLE, Bertram F. (Eds.). **The Evolution of**

**Language out of Pre-language.** Amsterdam/Philadelphie : John Benjamins. 2002. p. 3-50.

GOLDBERG, A. Verbs, constructions and semantic frames. In: HOVAV, M. Rappaport; DORON, E.; SICHEL, I. (Eds.). **Syntax, Lexical Semantics and Event Structure.** Oxford: Oxford University Press. 2009.

GONZALEZ, M.; LIMA, V. L. S. de. Redefining Traditional Lexical Semantic Relations with Qualia Information. **Revista Palavra**, [S.l.], n. 12, p.25-36, 2004.

GRADY, Joseph. **Foundations of meaning: primary metaphors and primary scenes.** Berkeley : Linguistics Dept, University of California, 1997. (Doctoral thesis). Available from Internet:<: [www.il.proquest.com/umi/dissertations/](http://www.il.proquest.com/umi/dissertations/)>.

\_\_\_\_\_. Image schemas and perception: refining a definition. In: HAMPE, Beate (Ed.) **From perception to meaning: image schemas in Cognitive Linguistics.** Berlin: Mouton de Gruyter. 2005. p.35-56.

\_\_\_\_\_. Metaphor. In: GEERAERTS, Dirk; CUYCKENS, Hubert (Eds.) **The Oxford Handbook of Cognitive Linguistics.** Oxford: Oxford University Press, 2007. p. 188-213.

GRUBER, T. R. **Ontolingua: a mechanism to support portable ontologies.** Stanford: Knowledge Systems Laboratory Stanford University, 1992. (Relatório KSL91-66).

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

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

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

\_\_\_\_\_; WELTY, C. A Formal Ontology of Properties. In: INTERNATIONAL CONFERENCE KNOWLEDGE ENGINEERING AND KNOWLEDGE MANAGEMENT: METHODS, MODELS AND TOOLS (EKAW2000), 12, 2000. In: DIENG, R.; CORBY, O. (Eds.). **Conference...**[s.l.] : Springer Verlag, 2000. p. 97-112.

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

HAMM, F. Frame Semantics. In: **The Cambridge Encyclopedia of the Language Sciences.** Cambridge University Press. 2009.

HAMPE, Beate. **From perception to meaning: image schemas in Cognitive Linguistics.** Berlin: Mouton de Gruyter. 2005.

HARI, R; KUJALA, M. Brain basis of human social interaction: from concepts to brain imaging. **Physiol Reviews**, v. 89, n.2, 2009. p. 453–479. [Cited 12 fev. 2012]. . Available from Internet: <<http://physrev.physiology.org/content/89/2/453.full>>

HAUSER, M. D.; CHOMSKY, N.; FITCH, W. T. The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? **Science**, v. 298, p.1569--1579, 2002.

HODGES, WILFRID. Model Theory. In: ZALTA, Edward N. (Ed.). **The Stanford Encyclopedia of Philosophy**. Fall 2009 Edition. [Cited 15 Dez. 2011]. Available from Internet: <<http://plato.stanford.edu/archives/fall2009/entries/model-theory/>>.

JOHNSON, M. **The Body in the Mind**: the bodily basis of meaning, imagination and reason. Chicago University Press. Chicago, 1987.

JOHNSON, M. **The Meaning of the Body**. University of Chicago Press, 2008.

KADIE C.M.; HOVEL, D.; HORVITZ, E. MSBNx: a component-centric toolkit for modeling and inference with Bayesian Networks. Microsoft MSR-TR-2001-67, July 2001. (Research Technical Report)

KASAMA, D. Y.; ZAVAGLIA, C.; ALMEIDA, G. M. de B. Do termo à estruturação semântica: representação ontológica do domínio da Nanociência e Nanotecnologia utilizando a Estrutura Qualia. **LinguaMÁTICA**. v. 2, n. 3, p. 43-58, 2010.

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

KIPPER, Karin *et al.* Extending VerbNet with novel verb classes. In: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION LREC'06, 5, 2006, Genoa, Italy. **Proceedings...** Genoa, Italy, May, 2006.

LAKATOS, E. M.; MARCONI, M. de A. **Fundamentos de metodologia científica**. 4. ed. rev.ampl. São Paulo : Atlas, 2001. 288 p.

LAKOFF, G. **Women, fire, and dangerous things: what categories reveal about the mind**. University of Chicago, 1987.

\_\_\_\_\_. Some empirical results about the nature of concepts. **Mind and Language**, Oxford, v. 4, n. 1/2, p. 103-129, 1989.

\_\_\_\_\_. The invariance hypothesis: is abstract reason based on imageschemas?, **Cognitive Linguistics**, v.1, n. 1, p. 39–74, 1990.

\_\_\_\_\_. The Neural Theory of Metaphor. In: GIBBS, Raymond W. **Handbook of Metaphor and Thought**, Cambridge University Press, 2008a. p. 17-38.

\_\_\_\_\_; JOHNSON, Mark. **Metaphors We Live By**. Chicago: University of Chicago Press, 1980.

LANGACKER, R. W. **Foundations of cognitive grammar**. Stanford, California : Stanford University Press, 1987. Vol. I.

\_\_\_\_\_. Cognitive Grammar. In: GEERAERTS, Hubert; CUYCKENS, Hubert (Eds.). **The Oxford Handbook of Cognitive Linguistics**. Oxford University Press, 2007. p. 421-462.

LAURENCE, S.; MARGOLIS, E. Concepts and cognitive science. In: MARGOLIS, E.; LAURENCE, S. (Eds). **Concepts: core readings**. Cambridge, MA: MIT Press, 1999. p. 3-81.

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.* From text to content: computational lexicons and the semantic web. In: AAAI - ASSOCIATION FOR THE ADVANCEMENT OF ARTIFICIAL INTELLIGENCE (AAAI) (WS 2-16), 2002. **Proceedings...**, 2002. p.1-10. (Technical Report).

\_\_\_\_\_. **SIMPLE** work package 2 - Linguistic specifications, Deliverable D2.1. , Pisa :ILC-CNR, March 2000.

LEVIN, J. Functionalism. In: ZALTA, Edward N. (Ed.). **The Stanford Encyclopedia of Philosophy**. Summer, 2010. [Cited 10 Dez. 2011]. Available from Internet:<<http://plato.stanford.edu/archives/sum2010/entries/functionalist/>>.

LI, Y.H.; JAIN, A.K. Classification of text documents. **Computer Journal**, v.41, n. 8, p. 537-46, 1998.

LITKOWSKI, K. Senseval-3 task: automatic labeling of semantic roles. In INTERNATIONAL WORKSHOP ON THE EVALUATION OF SYSTEMS FOR THE SEMANTIC ANALYSIS OF TEXT, 3, July 2004, Barcelona, Spain. **Proceedings of Senseval-3: The Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text**. Barcelona, Spain, July 2004.

LÓPEZ, J. M. *et al.* Towards an ontology for describing emotions. In: WORLD SUMMIT ON THE KNOWLEDGE SOCIETY: Emerging technologies and information systems for the knowledge society, Athens, Greece, September 24-26, 2008. **Proceeding Lecture Notes In Artificial Intelligence**. LYTRAS, M. D. *et al.* (Eds.). Springer-Verlag, Berlin, Heidelberg, 2008. Vol. 5288, p. 96-104.

LYONS, J. **Semantics**. Cambridge : Cambridge University Press. 1977.

MCCARTHY, J. *et al.* **A Proposal for the dartmouth summer research project on Artificial Intelligence**. 1955.

Available from Internet: <<http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>>. Último acesso em 24/03/2012.

MALONEY, J. C. Functionalism. In: WILSON, Robert A.; KEIL, Frank C. (Eds). **The MIT Encyclopedia of the Cognitive Sciences**. Cambridge: MIT Press, 1999.

MENDES, Sara; CHAVES, Rui Pedro. Enriching WordNet with Qualia Information. In: WORKSHOP ON WORDNET AND OTHER LEXICAL RESOURCES: APPLICATIONS, EXTENSIONS, E CUSTOMIZATIONS at the II NAACL Meeting 2001, Carnegie Mellon University, Pittsburgh, PA, EUA. **Proceedings...** Carnegie Mellon University, Pittsburgh, PA, EUA, 2001. p. 108-112.

MEURS, Marie-Jean; LEFEVRE, Fabrice; MORI, Renato de. 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.

MINATO, T. *et al.* CB2: A child robot with biomimetic body for cognitive developmental robotics. In: IEEEERAS INTERNATIONAL CONFERENCE ON HUMANOID ROBOTS, 7, 2007. **Proceedings...** 2007. p. 557-562. Available from Internet:<<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4813926>>

MINSKY, M. A Framework for Representing Knowledge. In: WINSTON, Patrick (Ed.) **The Psychology of Computer Vision**. New York : McGraw-Hill. 1975. p. 211-277.

MORA, J. F. On the early history of Ontology. **Philosophy and Phenomenological Research**, [s.l.], v. 24, p. 36-47, 1963.

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.

MORRIS, J.; HIRST, G. Lexical cohesion computed by thesaural relations as an indicator of the structure of text. **Computational Linguistics**, v.17, n.1, p. 21-48, 1991.

NARAYANAN, S. **KARMA**: Knowledge-based action representations for metaphor and aspect. UC Berkeley, 1997. (PhD Thesis).

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

NARIOKA, Kenichi; HOSODA, Koh. Motor development of an pneumatic musculoskeletal infant robot. In: IEEE INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION, 2011. **Proceeding...**, 2011. p.963-968.



NEWELL, A.; SIMON, H. A. GPS: a program that simulates human thought. In: FEIGENBAUM, E.A.; FELDMAN, J. (Eds.). **Computers and thought**, McGraw-Hill. 1963.

NILES, I.; PEASE, A. Linking lexicons and ontologies: mapping WordNet to the suggested upper merged ontology. In: IEEE INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE ENGINEERING. **Proceedings of the IEEE International Conference on Information and Knowledge Engineering**, 2003. p. 412-416.

OFOGHI, B. **Enhancing Factoid Question Answering Using Frame Semantic-based approaches**. University of Ballarat, 2009.

\_\_\_\_\_; YEARWOOD, J.; GHOSH, R. A Within-Frame Ontological Extension on FrameNet: Application in Predicate Chain Analysis and Question Answering. In: Conference on Artificial Intelligence, Australian, 2007. **Conference...** Australian, 2007. p. 404-414.

OHARA, K.; NIKIFORIDOU, K. Background: the constructional view of language. **Constructions and Frames**, v.1, n.1, p.7-28. 2009.

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. (Forthcoming in)

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

PALMER M.; KINGSBURY, P.; GILDEA, D. The proposition bank: an annotated *corpus* of semantic roles. **Computational Linguistics**, 31, 1, p.71–106. 2005.

PANTHER, K. U.; THORNBURG, L. L. Metonymy. In: GEERAERTS, D.; CUYCKENS, H. (Eds.). **The Oxford Handbook of Cognitive Linguistics**. Oxford: Oxford University Press. 2007. p. 236-263.

PARADIS, C. Ontologies and construals in lexical semantics. **Axiomathes**, 15, 2005. p. 541-573.

PEDERSEN, B.S. *et al.* DanNet - A WordNet for Danish. In: INTERNATIONAL CONFERENCE ON GLOBAL WORNETS, 3, 2006, Jeju, South Korea. **Proceedings...** Jeju, South Korea, 2006. p.329-330.

PETRUCK, Miriam R. L. Frame Semantics. In: VERSCHUEREN, Jef, *et al* (eds.). **Handbook of Pragmatics**. Philadelphia : John Benjamins, 1996.

Available from Internet:

<[http://framenet.icsi.berkeley.edu/index.php?option=com\\_content&task=view&id=39&Itemid=42](http://framenet.icsi.berkeley.edu/index.php?option=com_content&task=view&id=39&Itemid=42)>.

POLLARD, C.; SAG, Ivan S. **Information-based syntax and semantics**. Centre for the Study of Language and Information, Stanford University, CA. 1987. x + 227 p. (Vol. 1: Fundamentals).

PULLUM, Geoffrey K.; SCHOLZ, Barbara C. Empirical assessment of stimulus poverty arguments. **The Linguistic Review**, v. 19, p. 9–50. 2002.

PUSTEJOVSKY, James . The generative lexicon. **Comput. Linguist.** 17, 4, p. 409-441, December 1991.

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

\_\_\_\_\_; ANICK, P. On the Semantic Interpretation of Nominals. In: COLING, 1988. **Proceedings...**, v.2, 1988, p. 518-523.

QUILLIAN, R. **Semantic Memory**. Carnegie Institute of Technology, 1966. (PhD Thesis).

RATNAPARKHI, Adwait. A maximum entropy part-of-speech tagger. In: EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING CONFERENCE, May 17-18, 1996. **Proceedings...** May 17-18, 1996. REITER, N. **Towards a Linking of FrameNet and SUMO**. Saarland University, 2007. (Master's thesis).

RICH, E.; KNIGHT K. **Inteligência Artificial**. 2 ed. São Paulo : Makron Books editor, 1993.

ROBINSON, H. Dualism. In: ZALTA, Edward N. (Ed.). **The Stanford Encyclopedia of Philosophy**. Summer 2009 Edition.  
Available from Internet: <<<http://plato.stanford.edu/archives/sum2009/entries/dualism/>>

ROSCH, Eleanor. Cognitive representations of semantic categories. **Journal of Experimental Psychology**: general, v. 104, p. 192–233, 1975.

\_\_\_\_\_ *et al.* Basic objects in natural categories. **Cognitive Psychology**, v. 8, p. 382-439, 1976.

\_\_\_\_\_. Human categorization. In: WARREN, N. (Ed.). **Studies in Cross-linguistic Psychology**. London: Academic Press, 1977. p. 1–49.

\_\_\_\_\_. Principles of categorization. In: LOYD, B.; ROSCH, E. (Eds). **Cognition and Categorization**. Hillsdale, NJ: Erlbaum, p. 27–48. In: MARGOLIS, E.; LAURENCE, S. (Eds). **Concepts: Core Readings**. Cambridge, MA: MIT Press, [1978], 1999. p. 189–206. (Reprinted)

\_\_\_\_\_; OLIVIER, D. C. The structure of the color space in naming and memory for two languages, **Cognitive Psychology**, v. 3, n. 2, p. 337-354, April 1972.

ROSSA, A. A; ROSSA. C. R. P. A perspectiva enatista e relações sociointeracionistas da aquisição da linguagem. **Letras de Hoje**, Porto Alegre, v. 46, n. 3, p. 18-21, jul./set. 2011.

RUI-LING, Zhang; HONG-SHENG, Xu. Using bayesian network and neural network constructing domain ontology. In: WRI WORLD CONGRESS ON COMPUTER SCIENCE AND INFORMATION ENGINEERING (CSIE), 2009, IEEE Computer Society, Washington, DC, USA. **Proceedings...** IEEE Computer Society, Washington, DC, USA, 2009. V. 6, p. 116-120.  
Available from Internet:<<http://dx.doi.org/10.1109/CSIE.2009.328>>

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

RUIMY, N. *et al.* Clips, a multi-level italian computational lexicon: a glimpse to data. In: LREC - INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION, 3, 29<sup>th</sup> 30<sup>th</sup> e 31 May 2002, Las Palmas de Gran Canaria, Spain. **Proceedings...** Las Palmas de Gran Canaria, Spain, 29th, 30th & 31 May 2002.

RUPPENHOFER, Josef *et al.* **FrameNet II**: extended theory and practice. ICSI Technical Report. ICSI. 2006. (Technical Report).

\_\_\_\_\_ *et al.* **FrameNet II**: extended theory and practice. ICSI Technical Report. ICSI. 2010. (Technical Report).

RUSSELL, Stuart; NORVIG, Peter. **Artificial intelligence**: a modern approach. 3. ed. New Jersey : Prentice Hall, 2010. 1.152 p.

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

\_\_\_\_\_. **Linguagem e cognição linguística**: notícias do debate contemporâneo. Juiz de Fora: UFJF, 2009b. 12 p. (Notas de aula).

\_\_\_\_\_. **As bases neurobiológicas da cognição linguística II**: o papel do sistema sensório-motor na cognição e na linguagem. Juiz de Fora: UFJF, 2009c. 29 p. (Notas de aula).

\_\_\_\_\_. **Teorias da linguagem**: a perspectiva sociocognitiva. Rio de Janeiro : UFRJ, 2006. (Palestra no II FÓRUM DE LINGUAGEM: LINGUAGEM NATUREZA E CULTURA). [Citado 29 maio 2010]  
Disponível na Internet: <<http://www.forumde linguagem.com.br/textos/>>.

SAMET, J. Nativism, History of. In: WILSON, Robert A.; KEIL, Frank C. (Eds). **The MIT Encyclopedia of the Cognitive Sciences**. Cambridge: MIT Press, 1999.

SCHANK, R.C. **Conceptual Information Processing**. New York: Elsevier. 1975.

\_\_\_\_\_; ABELSON, R. **Scripts, plans, goals, and understanding**: an inquiry into human knowledge structures. Hillsdale, NJ: Lawrence Erlbaum. 1977.

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.

SCHMID, Hans-Jörg. **English abstract nouns as conceptual shells**: from *corpus* to cognition. Berlin, New York : Mouton de Gruyter. 2000.

SCHULZ, R. *et al.* Lingodroids: studies in spatial cognition and language In: ICRA - THE INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION, Shanghai, China, May 2011. **Proceedings...** Shanghai, China, May 2011.

SHANAHAN, M. The Frame Problem. In: ZALTA, Edward N. (Ed.). **The Stanford Encyclopedia of Philosophy**. Winter 2009.  
Available from Internet: <<http://plato.stanford.edu/archives/win2009/entries/frame-problem/>>.

SMITH, N. **Chomsky ideas and ideals**. Cambridge, United Kingdom: Cambridge University Press. 1999. 278 p.

SOWA, J. F. **Knowledge Representation: Logical, Philosophical and Computational Foundations**. Pacific Grove, USA: Brooks/Cole ed. 2000. 594 p.

STEUP, M. Epistemology. In: ZALTA, Edward N. (Ed.). **The Stanford Encyclopedia of Philosophy**. Spring 2010 Edition.  
Available from Internet:  
<<http://plato.stanford.edu/archives/spr2010/entries/epistemology/>>.

TALMY, L. How language structures space. In: PICK, H. L.; ACREDOLO L P. (Eds). **Spatial Orientation**: theory, research, and application. New York: Plenum Press, 1983. p. 225-282.

TAYLOR, John R. How many meanings does a word have? **Stellenbosch Papers in Linguistics**, v.25, p 133–168, 1992.

TOMASELLO M. **Origins of Human Communication**. Cambridge, MA/London: The MIT Press, 2008. XIII, 393 S.

VALENTE, A. **Legal knowledge engineering**: a modelling approach. Amsterdam : IOS Press, 1995.

VOSSSEN, P. **EuroWordNet**: a multilingual database with lexical semantic networks. Springer, 1998. 179 p.

\_\_\_\_\_; FELLBAUM C. Universals and idiosyncrasies in multilingual WordNets. In: BOAS, H. C. **Multilingual FrameNets in Computational Lexicography**: methods and applications. Berlin: Mouton de Gruyter. 2009.

W3C. **OWL 2 Web Ontology Language: Document Overview**. 27 October 2009. (W3C Recommendation); [citado 17/01/2011]. Available from Internet: <<http://www.w3.org/2009/pdf/REC-owl2-overview-20091027.pdf>>

WELTY, C.; GUARINO, N. Supporting Ontological Analysis of Taxonomic Relationships. **Data and Knowledge Engineering**, v.39, n.1, p. 51-74. 2001.

WERMTER, S.; PALM, G.; ELSHAW, M. Biomimetic neural learning for intelligent robots. Springer, Heidelberg, Germany. 2005.

WILSON, Robert A.; FOGLIA, Lucia. Embodied Cognition. In: ZALTA, Edward N. (Ed.). **The Stanford Encyclopedia of Philosophy**. Fall 2011 Edition. [Cited 14 Dez. 2011]. Available from Internet:<[plato.stanford.edu/archives/fall2011/entries/embodied-cognition/](http://plato.stanford.edu/archives/fall2011/entries/embodied-cognition/)>.

WOOD, Mary McGee. **Categorial Grammars**. New York: Routledge, 1993.

WOODS, W. A. What's in a link: foundations for semantic networks. In: Bobrow, D. G.; Collins, A. M. (Ed.). **Representation and Understanding: Studies in Cognitive Science**. Academic Press, 1975.

YU N. **The Contemporary Theory of Metaphor: a perspective from Chinese**. Amsterdam, Philadelphia: John Benjamins, 1998. (Human Cognitive Processing series, 1).

ZADEH, L. A. Fuzzy logic and approximate reasoning. **Synthese**, v. 30, p. 407–428, 1975.

ZAVAGLIA, C.; GREGHI, J. G. Homonymy in natural language processes: a representation using Pustejovsky's Qualia structure and ontological information. In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL PROCESSING OF THE PORTUGUESE LANGUAGE (PROPOR'03), 6, 2003, Berlin, Heidelberg. **Proceedings...** . NUNO, J. Mamede; TRANCOSO, Isabel; BAPTISTA, Jorge; NUNES, Maria das Graças Volpe (Eds.). Berlin, Heidelberg : Springer-Verlag, 2003. p. 86-93.

## ÍNDICE

### C

Conhecimento, 7, 8, 13, 14, 31, 33, 35, 37, 46, 57, 67,  
68, 69, 70, 74, 77, 78, 79, 81, 82, 90, 91, 92, 93,  
96, 114, 115

### D

DOLCE (ontologia), 80, 84, 85, 89, 94, 97, 125, 126

### E

Elemento de frame, 64, 93, 94, 117, 118, 121, 122,  
123, 137, 138

### F

FrameNet, 53, 54, 55, 57, 59, 60, 61, 62, 63, 64, 65,  
66, 80, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92,  
93, 94, 96, 114, 115, 116, 117, 118  
Frames, 28, 29, 30, 31, 34, 35, 36, 37, 38, 39, 40, 41,  
42, 46, 47, 54, 55, 56, 57, 58, 59, 60, 61, 63, 65,  
68, 69, 70, 71, 77, 83, 85, 86, 87, 88, 89, 90, 91,  
92, 93, 94, 96, 115, 116  
*Framework*, 1, 2, 4, 5, 6, 96, 98, 106, 115

### I

Informação, 7, 21, 50, 53, 55, 61, 63, 67, 70, 73, 85,  
90  
Inteligência artificial, 11, 49, 53, 67, 73, 74

### M

Metáfora, 23, 42, 43, 44, 45, 47, 65, 66

Metonímia, 46, 47

### O

Ontologia, 65, 66, 73, 75, 76, 77, 80, 81, 82, 83, 84,  
85, 86, 87, 94, 97, 103, 114, 115, 116, 117, 144  
Ontologias, 81, 82, 90

### R

Relação, 8, 11, 14, 15, 22, 28, 29, 30, 34, 35, 39, 44,  
46, 52, 53, 57, 58, 59, 65, 73, 77, 81, 83, 84, 85,  
89, 94, 115  
Representação de conhecimento, 3, 75, 96, 119

### S

Semântica, 8, 9, 10, 13, 19, 25, 37, 38, 39, 40, 41, 54,  
58, 68, 77, 78, 83, 85, 90, 116  
Semântica de Frames, 1, 2, 3, 4, 5, 17, 37, 54, 95, 96,  
103, 107, 118, 119, 120, 157, 158, 159, 160  
SIMPLE-CLIPS (ontologia), 97, 98, 100, 103, 106,  
108, 109, 110, 115, 117, 118, 131, 141, 145, 158,  
160, 172

### U

Unidades léxicas, 55, 56, 57, 59, 61, 83, 86, 97

### W

WordNet, 53, 83, 84, 89

**ANEXO – PLANILHAS DA ANOTAÇÃO DAS SENTENCAS**