

Ely Edison da Silva Matos

LUDI:
**Um *framework* para desambiguação lexical
com base no enriquecimento da
Semântica de Frames**

Juiz de Fora

2014

Ely Edison da Silva Matos

LUDI:
Um *framework* para desambiguação lexical
com base no enriquecimento da
Semântica de Frames

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

Orientadora: Profa. Dra. Maria Margarida Martins Salomão

Coorientador: Prof. Dr. Tiago Timponi Torrent

Juiz de Fora

2014

Ficha catalográfica elaborada através do Programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Matos, Ely Edison da Silva.

LUDI: um framework para desambiguação lexical com base no enriquecimento da Semântica de Frames / Ely Edison da Silva
Matos. -- 2014.

200 p. : il.

Orientadora: Maria Margarida Martins Salomão

Coorientador: Tiago Timponi Torrent

Tese (doutorado) - Universidade Federal de Juiz de Fora, Faculdade de Letras. Programa de Pós-Graduação em Linguística, 2014.

1. Desambiguação do sentido de lexemas. 2. Semântica Computacional. 3. Semântica de Frames. 4. Léxico Gerativo. I. Salomão, Maria Margarida Martins, orient. II. Torrent, Tiago Timponi, coorient. III. Título.

Ely Edison da Silva Matos

LUDI:
**Um *framework* para desambiguação lexical
com base no enriquecimento da
Semântica de Frames**

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

Trabalho aprovado. Juiz de Fora, 27 de junho de 2014:

Profa. Dra. Maria Margarida Martins Salomão - UFJF
Orientadora

Prof. Dr. Tiago Timponi Torrent - UFJF
Coorientador

Prof. Dr. Carlos Subirats-Rüggeberg - UAB/Espanha

Prof. Dr. Oto Araújo Vale - UFSCAR

Profa. Dra. Regina Maria Maciel Braga - UFJF

Juiz de Fora
2014

*Para Edison e Rosemary, que me deram a oportunidade desta vida;
para Ligia, que me ensina diariamente a viver;
para os meninos Felipe, Paulo e Daniel, que dão sentido a tudo isso.*

Agradecimentos

Quando estacionamos os olhos diante da tela de um computador, procurando a solução de um problema persistente ou o melhor termo que transmita nossas ideias, o trabalho de pesquisa parece ser extremamente solitário. Mas não é. Trazemos nossa bagagem individual, mas ela é enriquecida pelo auxílio de milhares de seres que nos ajudaram a chegar onde chegamos. Por isso, é preciso agradecer; e eu agradeço

a Deus, Causa Primária e Inteligência Suprema;

aos meus pais, Edison e Rosemary, por terem me dado tudo que um filho precisa e muito mais;

à Ligia, por compartilhar mais uma existência, em aprendizado mútuo;

aos meninos Felipe, Paulo e Daniel, que me mostram, a cada dia, o que os livros não ensinam;

à Prof^a Margarida, por ter aceitado o desafio de orientar, com paciência e segurança, um "estrangeiro" na área da Linguística;

ao Prof. Tiago, pelo amplo conhecimento, transmitido sempre com tolerância, humildade e espírito científico;

à Prof^a Fernanda e à Prof^a Regina, por me apresentarem o projeto FrameNet Brasil e incentivarem mais um vôo intelectual em um céu desconhecido;

ao Prof. Oto Vale e ao Prof. Carlos Subirats, pela avaliação das ideias apresentadas neste trabalho;

à Coordenação do Programa de Pós-graduação em Linguística e aos professores do curso, pelo apoio constante;

às colegas de curso, Júlia, Simone, Maucha e Ludmila, pelo intercâmbio de ideias, certezas, dúvidas e conhecimentos;

aos amigos do CGCO e da FEAK, pelas oportunidades de trabalho;

às centenas de pesquisadores espalhados pelo mundo, cujas ideias, aplicações, artigos e livros possibilitaram que este fosse um trabalho de muitas mentes.

*Uma peônia branca, dizem,
mas me parece
rosa claro.*

Haikai de Kyoshi(poeta japonês, séc. XIX)

Resumo

Enquanto no âmbito da Sintaxe, as técnicas, os algoritmos e as aplicações em Processamento da Língua Natural são bem estudados e já estão relativamente bem estabelecidos, no âmbito da Semântica não é possível observar ainda a mesma maturidade. Visando, então, contribuir para os estudos em Semântica Computacional, este trabalho busca maneiras de implementar algumas das ideias e dos *insights* propostos pela Linguística Cognitiva, que é, por si, uma alternativa à Linguística Gerativa. A tentativa é reunir algumas das ferramentas disponíveis, seja no viés computacional (Bancos de Dados, Teoria dos Grafos, Ontologias, Mecanismos de inferências, Modelos Conexionistas), seja no viés linguístico (Semântica de Frames e Teoria do Léxico Gerativo), seja no viés de aplicações (FrameNet e ontologia SIMPLE), a fim de abordar as questões semânticas de forma mais flexível. O objeto de estudo é o processo de desambiguação de Unidades Lexicais. O resultado da pesquisa realizada é corporificado na forma de uma aplicação computacional, chamada Framework LUDI (*Lexical Unit Discovery through Inference*), composta por algoritmos e estruturas de dados usados na desambiguação. O *framework* é uma aplicação de Compreensão da Língua Natural, que pode ser integrada em ferramentas para recuperação de informação e sumarização, bem como em processos de Etiquetagem de Papéis Semânticos (SRL - *Semantic Role Labeling*).

Palavras-chaves: Desambiguação do Sentido de Lexemas, Semântica Computacional, Semântica de Frames , Léxico Gerativo.

Abstract

While in the field of Syntax techniques, algorithms and applications in Natural Language Processing are well known and relatively well established, the same situation does not hold for the field of Semantics. Aiming at contributing to the studies in Computational Semantics, this work implements ideas and insights offered by Cognitive Linguistics, which is itself an alternative to Generative Linguistics. We attempt to bring together contributions from the computational domain (Databases, Graph Theory, Ontologies, inference mechanisms, Connectionists Models), the linguistic domain (Frame Semantics and the Generative Lexicon), and the application domain (FrameNet and SIMPLE Ontology) in order to address the semantic issues more flexibly. The object of study is the process of disambiguation of Lexical Units. The results of the research are embodied in the form of a computer application, called Framework LUDI (Lexical Unit Discovery through Inference), and composed of algorithms and data structures used for Lexical Unit disambiguation. The framework is an application of Natural Language Understanding, which can be integrated into information retrieval and summarization tools, as well as into processes of Semantic Role Labeling (SRL).

Keywords: Lexeme Sense Disambiguation, Computational Semantics, Frame Semantics, Generative Lexicon.

Lista de ilustrações

| | |
|--|-----|
| Figura 1 – Estrutura do item lexical <i>build</i> (contruir) | 37 |
| Figura 2 – <i>Frames</i> representados como MVA | 46 |
| Figura 3 – Estrutura do <i>Frame</i> CENÁRIO_DO_TURISMO_PARTIDA. | 49 |
| Figura 4 – Relações entre <i>Frame</i> e <i>Elementos de Frame</i> | 51 |
| Figura 5 – Abordagem baseada em gramática. | 58 |
| Figura 6 – Modelo para ontologia. | 61 |
| Figura 7 – Arquitetura de um sistema DL. | 68 |
| Figura 8 – Top-level da ontologia SIMPLE. | 71 |
| Figura 9 – Estrutura <i>qualia</i> para <nadador>. | 72 |
| Figura 10 – Estrutura <i>qualia</i> para <peixe>. | 73 |
| Figura 11 – Definição parcial de <Artifact_food>. | 74 |
| Figura 12 – Tipos Unificados na SIMPLE-OWL. | 74 |
| Figura 13 – Exemplo de Ativação Propagada. | 77 |
| Figura 14 – LUDI - Integração de ontologias. | 84 |
| Figura 15 – Modelo conexionista: and-node. | 88 |
| Figura 16 – Modelo conexionista: Restrições. | 88 |
| Figura 17 – Elementos da FrameNet. | 92 |
| Figura 18 – Elementos da ontologia SIMPLE. | 92 |
| Figura 19 – Classe TYPE. | 93 |
| Figura 20 – Type.Aspect. | 93 |
| Figura 21 – Exemplo de uso da Classe TYPE. | 94 |
| Figura 22 – Ontologia de eventos. | 95 |
| Figura 23 – Espaços Conceptuais. | 99 |
| Figura 24 – Estrutura para Qualidades. | 101 |
| Figura 25 – Integração entre <i>Frames</i> e Relações <i>Qualia</i> | 109 |
| Figura 26 – Relações <i>qualia</i> e classes genéricas. | 109 |
| Figura 27 – Papéis temáticos e classes genéricas. | 110 |
| Figura 28 – Integração entre relações <i>qualia</i> e papéis temáticos. | 110 |
| Figura 29 – <livro pesado> após ativação da rede. | 112 |
| Figura 30 – <livro triste> após a ativação da rede. | 113 |
| Figura 31 – <comida rapida> e <carro rapido> após a ativação | 114 |
| Figura 32 – <i>SemSpec</i> de <homem comer banana>. | 115 |
| Figura 33 – <i>SemSpec</i> de <homem ler livro>. | 116 |
| Figura 34 – <i>SemSpec</i> de <homem ler livro> - possivelmente com erro. | 116 |
| Figura 35 – <i>SemSpec</i> de <homem ler livro mulher> - com ambiguidade. | 117 |
| Figura 36 – <i>SemSpec</i> de <homem ler livro mulher> - após desambiguação. | 118 |

| | |
|--|-----|
| Figura 37 – Esquema para <abrir.v> | 123 |
| Figura 38 – Cluster de <i>Frames</i> herdeiros de ABRIR_ESQUEMA. | 125 |
| Figura 39 – Rede de <i>Frames</i> herdeiros de ABRIR_ESQUEMA. | 126 |
| Figura 40 – <i>Frames</i> por Clusters. | 126 |
| Figura 41 – Abrir_boca | 144 |
| Figura 42 – Abrir_olho | 145 |
| Figura 43 – Abrir_mão | 146 |
| Figura 44 – Abrir_mão (colocação) | 147 |
| Figura 45 – Abrir_carteira | 148 |
| Figura 46 – Abrir_porta | 149 |
| Figura 47 – Porta_aberta | 150 |
| Figura 48 – Abrir_jornal | 151 |
| Figura 49 – Abrir_museu | 152 |
| Figura 50 – Abrir_museu_domingo | 153 |
| Figura 51 – Abrir_exposicao | 154 |
| Figura 52 – Abrir_shopping | 155 |
| Figura 53 – Abrir_caminho (físico) | 156 |
| Figura 54 – Abrir_caminho (possibilidade) | 157 |
| Figura 55 – Abrir_cerveja | 158 |
| Figura 56 – Abrir_cerveja (2) | 159 |
| Figura 57 – Abrir_marcador | 160 |
| Figura 58 – Notações em DF. | 190 |
| Figura 59 – Padrões steady-state. | 191 |
| Figura 60 – Padrões de deslocamento. | 192 |
| Figura 61 – O "self dividido". | 193 |
| Figura 62 – Padrões de fases no domínio inter-psicológico. | 195 |
| Figura 63 – Ontologia SIMPLE - Topo | 198 |
| Figura 64 – Ontologia SIMPLE - Entidades | 199 |
| Figura 65 – Ontologia SIMPLE - Relações | 200 |

Lista de tabelas

| | |
|---|-----|
| Tabela 1 – <i>Unidades</i> da rede conexionista. | 86 |
| Tabela 2 – <i>Links</i> da rede conexionista. | 86 |
| Tabela 3 – Descrição das classes da Ontologia de eventos. | 96 |
| Tabela 4 – Qualidades: Tipos semânticos. | 101 |
| Tabela 5 – Qualidades: Frames. | 101 |
| Tabela 6 – Operações de composição da TLG. | 104 |
| Tabela 7 – Classificação de participantes. | 105 |
| Tabela 8 – Terminologia Sowa vs. TLG. | 106 |
| Tabela 9 – Papéis temáticos. | 107 |
| Tabela 10 – Relações SIMPLE e relações inversas. | 108 |
| Tabela 11 – Corpora utilizados. | 121 |
| Tabela 12 – Elementos do esquema para <abrir.v>. | 123 |
| Tabela 13 – Abordagens escolhidas no <i>framework</i> LUDI. | 163 |
| Tabela 14 – Peso do <i>link</i> por Tipo Semântico SIMPLE. | 174 |

Lista de abreviaturas e siglas

| | |
|---------|--|
| AGO | Agonista |
| ANT | Antagonista |
| BD | Banco de Dados |
| CT | Corpus CETENFolha/CETEMPUBLICO |
| DF | Dinâmica de Forças |
| DOLCE | <i>Descriptive Ontology for Linguistic and Cognitive Engineering</i> |
| DP | Corpus Domínio Público |
| DUL | DOLCE <i>Ultra Lite</i> |
| EF | Elemento de Frame |
| FEE | <i>Frame Evoking Element</i> |
| FG | Função Gramatical |
| MVA | Matriz de Valores de Atributos |
| NURC | Corpus do Projeto da Norma Urbana Oral Culta do Rio de Janeiro |
| OWL2 | Web Ontology Language, versão 2 |
| PLN | Processamento da Língua Natural |
| SRL | <i>Semantic Role Labeling</i> |
| SemSpec | <i>Semantic Specification</i> (Especificação Semântica) |
| TLG | Teoria do Léxico Gerativo |
| TS | Tipo Sintagmático |
| UL | Unidade Lexical |
| USem | Unidade Semântica |

Tipografia

| | |
|----------------|--|
| <i>itálico</i> | Usado para palavras e expressões em língua estrangeira. |
| negrito | Usado para destacar nomes de conceitos e teorias. |
| sanserif | Usado para destacar formas de palavras, <i>Unidades Lexicais</i> e <i>Unidades Semânticas</i> . |
| <i>slanted</i> | Usado para destacar expressões que possuem interpretações específicas dentro das teorias utilizadas (p.ex. <i>Frame</i> , <i>Elemento de Frame</i> , <i>Unidade Semântica</i>). |
| SMALLCAPS | Usado para destacar nomes de <i>Frames</i> , <i>Elementos de Frames</i> e <i>Tipos Ontológicos</i> . |

Sumário

| | | |
|------------|--|-----------|
| | Introdução | 18 |
| I | PREPARAÇÃO DA PESQUISA | 20 |
| 1 | O PROBLEMA | 21 |
| 1.1 | A polissemia de abrir | 21 |
| 1.2 | Terminologia | 23 |
| 1.3 | Pressupostos teóricos e hipóteses | 24 |
| 1.4 | O trabalho realizado | 25 |
| 1.5 | Aplicação dos resultados | 26 |
| 2 | METODOLOGIA | 27 |
| 2.1 | Objeto de estudo | 27 |
| 2.2 | Métodos | 27 |
| 2.3 | Metodologia de pesquisa em PLN | 29 |
| 2.4 | Etapas da Pesquisa | 29 |
| 2.4.1 | Levantamento bibliográfico | 29 |
| 2.4.2 | Proposta de Implementação | 30 |
| 2.4.3 | Implementação Inicial | 30 |
| 2.4.4 | Implementação corrigida | 30 |
| II | REFERENCIAIS TEÓRICOS | 31 |
| 3 | ASPECTOS LINGUÍSTICOS | 32 |
| 3.1 | O Significado Lexical | 32 |
| 3.2 | A Teoria do Léxico Gerativo | 33 |
| 3.3 | A Linguística Cognitiva | 39 |
| 3.3.1 | Categorização e Polissemia | 40 |
| 3.3.2 | Semântica de Frames | 43 |
| 3.4 | FrameNet | 47 |
| 3.4.1 | <i>Frames</i> | 48 |
| 3.4.2 | Relações entre <i>Frames</i> | 50 |
| 3.4.3 | Banco de dados | 51 |
| 3.4.4 | Versão atual | 52 |
| 3.5 | Causação e Dinâmica de Forças | 52 |

| | | |
|------------|---|------------|
| 3.6 | Composicionalidade | 53 |
| 4 | ASPECTOS COMPUTACIONAIS | 56 |
| 4.1 | Processamento de Línguas Naturais (PLN) | 57 |
| 4.2 | Ontologias | 60 |
| 4.3 | Ontologias e Léxico | 62 |
| 4.4 | Lógica Descritiva e OWL | 65 |
| 4.5 | Ontologia SIMPLE | 70 |
| 4.6 | Modelo Conexcionista | 74 |
| 4.6.1 | Modelos Conexcionistas Estruturados | 75 |
| 4.6.2 | Ativação Propagada | 77 |
| III | O FRAMEWORK LUDI | 79 |
| 5 | PROPOSTA DE UMA ONTOLOGIA LUDI | 80 |
| 5.1 | Introdução | 80 |
| 5.2 | Interface Ontolex | 80 |
| 5.3 | Entidades | 81 |
| 5.4 | <i>Elementos de Frame</i> | 82 |
| 5.5 | Integração de Ontologias | 83 |
| 6 | LUDI COMO MODELO CONEXIONISTA PARA PROCESSAMENTO DA LÍNGUA NATURAL | 85 |
| 6.1 | Unidades e Conexões da Rede | 85 |
| 6.2 | Construção da Rede | 87 |
| 6.3 | Ativação condicional | 88 |
| 6.4 | Valências e Funções Gramaticais | 89 |
| 7 | TRATAMENTO DOS RECURSOS LEXICAIS | 91 |
| 7.1 | Estrutura de Traços | 91 |
| 7.2 | A classe Type | 93 |
| 7.3 | Eventos e Qualidades | 94 |
| 7.3.1 | Eventos | 95 |
| 7.3.2 | Qualidades | 97 |
| 7.4 | Mecanismos gerativos | 102 |
| 7.5 | Papéis temáticos e relações qualia | 104 |
| 8 | EXEMPLOS DE INFERÊNCIAS | 111 |
| 8.1 | Qualidades | 111 |
| 8.2 | Eventos | 114 |

| | | |
|-------------|---|------------|
| IV | VERIFICAÇÃO EMPÍRICA DO <i>FRAMEWORK</i> LUDI | 119 |
| 9 | PONTO DE PARTIDA: OS SENTIDOS DE <ABRIR.V> | 120 |
| 9.1 | A análise das sentenças | 120 |
| 9.1.1 | Alternância causativo-incoativo | 120 |
| 9.1.2 | Metonímia | 122 |
| 9.2 | O esquema básico de <abrir.v> | 122 |
| 9.2.1 | Dinâmica de forças | 123 |
| 9.2.2 | Perfilamento | 124 |
| 9.3 | Os <i>Frames</i> de <abrir.v> | 124 |
| 9.3.1 | O esquema básico | 124 |
| 9.3.2 | Corpo físico | 127 |
| 9.3.3 | Entidades físicas | 128 |
| 9.3.4 | Aspecto | 129 |
| 9.3.5 | Modalidade | 132 |
| 9.3.6 | Estado Mental, Emocional ou Psicológico | 133 |
| 9.4 | Funções gramaticais | 136 |
| 9.4.1 | Argumento Externo | 137 |
| 9.4.2 | Objeto Direto | 137 |
| 9.4.3 | Objeto Indireto | 138 |
| 9.4.4 | Dependente | 138 |
| 9.5 | Valências | 139 |
| 9.5.1 | Corpo físico | 139 |
| 9.5.2 | Entidades Físicas | 140 |
| 9.5.3 | Aspecto | 141 |
| 9.5.4 | Modalidade | 141 |
| 9.5.5 | Estado Mental, Emocional ou Psicológico | 142 |
| 10 | A DESAMBIGUAÇÃO DE <ABRIR.V> USANDO LUDI | 143 |
| 10.1 | Corpo físico x Estado mental | 143 |
| 10.1.1 | Abrir_boca | 143 |
| 10.1.2 | Abrir_olho | 144 |
| 10.1.3 | Abrir_mão | 145 |
| 10.2 | Entidades físicas | 147 |
| 10.2.1 | Abrir_carteira | 147 |
| 10.2.2 | Abrir_porta | 148 |
| 10.2.3 | Abrir_jornal | 150 |
| 10.3 | Entidades físicas x Aspecto | 152 |
| 10.3.1 | Abrir_museu | 152 |
| 10.3.2 | Abrir_exposição | 153 |

| | | |
|-------------|---|------------|
| 10.3.3 | Abrir_shopping | 154 |
| 10.4 | Entidades físicas x Modalidade | 156 |
| 10.4.1 | Abrir_caminho | 156 |
| 10.5 | Metonímias | 158 |
| 10.5.1 | Abrir_cerveja | 158 |
| 10.5.2 | Abrir_marcaador | 159 |
| | Conclusão | 161 |
| | Referências | 165 |

APÊNDICES 173

| | | |
|------------|--|------------|
| | APÊNDICE A – MODELO CONEXIONISTA | 174 |
| A.1 | Ponderação dos links | 174 |
| A.2 | Redes de <abrir.v> | 175 |
| A.2.1 | Definição do <i>Frame</i> Esquemático | 175 |
| A.2.2 | Definição do Esquema Básico | 176 |
| A.2.3 | Definição do Cluster II.Movimento_corporal | 177 |
| A.2.4 | Exemplos de Unidades da Ontologia | 180 |
| | APÊNDICE B – CAUSAÇÃO | 184 |
| B.1 | Teorias da dependência | 184 |
| B.2 | Teorias da produção | 185 |
| B.3 | Semântica da Causação | 186 |
| B.3.1 | Eventos autônomos | 186 |
| B.3.2 | Situação causativa básica | 186 |
| B.3.3 | Situações causativas complexas | 187 |
| B.3.4 | Agentividade | 188 |
| | APÊNDICE C – DINÂMICA DE FORÇAS | 189 |
| C.1 | Notações | 189 |
| C.2 | Padrões básicos de DF | 190 |
| C.2.1 | Padrões <i>Steady-State</i> | 190 |
| C.2.2 | Padrões <i>Shifiting</i> | 191 |
| C.3 | Domínios da DF | 193 |
| C.3.1 | Domínio intra-psicológico | 193 |
| C.3.2 | Domínio inter-psicológico | 194 |
| C.4 | DF e causação | 195 |

ANEXOS **197**

ANEXO A – ONTOLOGIA SIMPLE **198**

A.1 **Ontologia SIMPLE** **198**

Introdução

Creio ser possível classificar em três, as reações dos interlocutores quando descobrem que minha área de pesquisa é a Linguística Computacional: "Os computadores *nunca* vão compreender a linguagem humana!"; são, geralmente, linguistas. Outros dizem: "Os computadores *já* compreendem a linguagem humana!"; são, geralmente, cientistas da computação. O terceiro grupo, mais numeroso, simplesmente ignora sobre o que estou falando.

O desafio de construir programas que possam compreender, ainda que de maneira limitada e rústica, "alguma coisa" da linguagem usada pelos seres humanos, acompanha os pesquisadores desde que os primeiros computadores foram desenvolvidos. Os objetivos são os mais variados e diversos possíveis: da indexação de documentos à sumarização de textos; da análise sintática e semântica de sentenças à representação do discurso; da melhoria da interface humano-computador à comunicação com robôs via inteligência artificial; de sistemas de pergunta-resposta à interpretação de metáforas conceptuais.

Porém, uma conclusão que pode ser obtida a partir de todos estes esforços (não enunciada claramente pelos pesquisadores da área) é que o Processamento da Língua Natural (PLN) é difícil. E é difícil porque a linguagem humana é difícil de ser processada nos modelos computacionais atualmente utilizados: ela é vaga, ambígua, dependente do contexto pragmático, criativa, repleta de polissemias, metonímias, metáforas e idiomatismos, amiga de exceções e não de regras. Enfim, o oposto de nossos modelos computacionais, justamente baseados em algoritmos, regras, limites precisos e lógicas formais.

Talvez por isso, os maiores sucessos do PLN tenham sido (e continuam sendo) na área da Sintaxe. As raízes da abordagem sintática preponderante atualmente, apoiada na Linguística Gerativa, se confundem com as raízes da Ciência da Computação. Não é exagero dizer que, pelo menos em seus primeiros anos, uma alimentou a outra e vice-versa. No entanto, dado o caráter "rebelde" da linguagem humana, nota-se claramente o crescimento, nos últimos anos, de modelos estatísticos e probabilísticos em PLN. O uso intensivo da probabilidade nas técnicas de PLN expõe claramente que não podemos esperar que a linguagem esteja presa apenas a princípios e parâmetros.

Algo semelhante, mas em escala bem menor, ocorre com a Semântica. Em um estudo abrangente sobre Semântica Computacional, [Bos \(2011\)](#) mostra que o objetivo desta área de pesquisa é a captura do significado de expressões em língua natural. Este significado deve ser representado de uma forma adequada para a realização de inferências, na tentativa de compreensão da linguagem humana. O instrumento por excelência, para isso, é a lógica formal, em especial a Lógica de Primeira Ordem. Mas neste campo a

dificuldade não é menor, pois fenômenos como a estrutura de eventos, a aspectualidade, os contextos modais, a quantificação e a resolução de anáforas, também resistem a uma formalização.

O objetivo deste trabalho é estudar a desambiguação de itens lexicais. De alguma forma ele pode ser relacionado às técnicas de WSD (*Word Sense Desambiguation*) e, portanto, enquadrado na área do PLN, em geral, e na área da Semântica, em particular. Porém, dado o quadro exposto anteriormente, a abordagem se afasta do *mainstream* em três aspectos principais. Em primeiro lugar, busco adotar a filosofia, os conceitos e a terminologia proposta pela Linguística Cognitiva, que é uma alternativa à Linguística Gerativa e na qual o processo de significação tem um papel de destaque. Em segundo lugar, ao invés de trabalhar com uma Semântica Formal, baseada em lógica, me apoio na Semântica de Frames, uma abordagem que privilegia a compreensão da linguagem em lugar de sua formalização. Por fim, recorro ao paradigma de pesquisa do grupo NTL (*Neural Theory of Language*), da Universidade da Califórnia em Berkeley, para a criação de um Modelo Conexionista Estruturado, como forma de implementação computacional das ideias desenvolvidas durante a pesquisa. Este modelo possibilita o tratamento das questões *fuzzy* da linguagem, sem apelar (diretamente) para uma abordagem baseada em probabilidades. Cada um destes aspectos é apresentado em detalhes, no decorrer do texto.

O resultado da pesquisa realizada é corporificado na forma de uma aplicação computacional, chamada Framework LUDI (*Lexical Unit Discovery through Inference*), composto por algoritmos e estruturas de dados usados na desambiguação. O *framework* é uma aplicação de Compreensão da Língua Natural, que pode ser integrada em ferramentas para recuperação de informação e sumarização, bem como em processos de Etiquetagem de Papéis Semânticos (SRL - *Semantic Role Labeling*).

O texto foi dividido em quatro partes. A primeira parte aborda o problema que motivou a pesquisa, os pressupostos considerados e as hipóteses adotadas. Além disso, é discutida também a metodologia. A segunda parte apresenta os referenciais teóricos, que servem de fundação para as ideias desenvolvidas na concepção do *framework* e na sua implementação. Optei por fazer uma divisão clara entre os aspectos linguísticos e computacionais, seguindo a metodologia adotada. A terceira parte descreve o *framework* LUDI, abordando o modelo conexionista, a construção do recurso lexical e alguns exemplos de aplicação. Finalmente, a quarta parte descreve os experimentos realizados, através da aplicação dos algoritmos em diversos contextos.

Parte I

Preparação da pesquisa

1 O Problema

1.1 A polissemia de abrir

Para um leitor ou ouvinte humano, as seguintes sentenças, baseadas em exemplos discutidos em [Searle \(1983\)](#) não devem apresentar dificuldades de compreensão:

- i João abriu a janela.
- ii João abriu sua boca.
- iii João abriu o livro.
- iv João abriu as cortinas.
- v O cirurgião abriu o peito de João.

Como Searle observa, em exemplos como estes o significado de <abrir> é função do que ele chama de *background*, ou seja, o conhecimento que temos de como as entidades e objetos de diferentes tipos são "abertos". Além deste conhecimento enciclopédico, é necessário também considerar o contexto da enunciação. A contribuição semântica de <abrir>, portanto, é bastante variável.

Se, além do sentido literal, consideramos usos mais figurativos de <abrir>, a dependência do contexto se torna ainda mais marcante:

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

O significado de <abrir> agora se relaciona com uma diversidade de ações, eventos, situações e agentes. Apesar desta diversidade, que pode incluir usos idiomáticos, metonímicos e metafóricos, e é analisada sob diversas perspectivas no âmbito da Linguística, um leitor humano ainda teria pouca ou nenhuma dificuldade de compreensão das sentenças.

A questão é diferente se considerarmos a legibilidade por máquinas, no âmbito computacional do PLN (Processamento de Línguas Naturais). Desprovidas de um "conhecimento do mundo" e de posse de um contexto limitado (quando este é fornecido), as

aplicações computacionais têm, sim, muita dificuldade para compreender um texto, se comparadas com os seres humanos.

Naturalmente um grande esforço tem sido realizado no campo da Linguística Computacional para minimizar este problema. Podemos, no entanto, observar que dois fatos se destacam e se sucedem quando buscamos a bibliografia nesta área: (i) o foco em questões relacionadas à Sintaxe e (ii) quando não é este o caso e a Semântica é considerada, o uso quase exclusivo de teorias associadas a Semântica Formal. A explicação para as duas questões é simples: em termos de Sintaxe, prevalece hoje ainda a abordagem Gerativista, que tem suas raízes associadas à Computação, desde a década de 50. Em termos de Semântica, o formalismo oferecido pelos diversos tipos de Lógica tem sido usado extensivamente no próprio fundamento da Computação.

Há exceções. O projeto FrameNet ([FILLMORE; JOHNSON; PETRUCK, 2003](#)), uma iniciativa do International Computer Science Institute (ICSI) de Berkeley, Califórnia (EUA), vem, desde 1997, implementando computacionalmente uma base de dados lexical fundada na Semântica de Frames. A Semântica de Frames ([FILLMORE, 1982](#); [FILLMORE, 2006](#)) propõe que um item lexical só pode ter seu significado esclarecido quando relativizado a um *background*, ou *frame*. Na busca de uma "semântica da compreensão" ([FILLMORE, 1985](#)), ela adota algumas ideias associadas a Semântica Cognitiva e destaca algumas das limitações das chamadas "condições de verdade" e da categorização baseada em condições "necessárias e suficientes", pontos chave das semânticas de orientação formalista.

No nosso caso, trabalhando em uma aplicação real ([SALOMÃO et al., 2011](#)), que visa à implementação de um dicionário multilingual baseado em frames, nos deparamos com uma questão eminentemente prática: uma vez que a Semântica de Frames trata a questão da polissemia relacionando o lexema polissêmico com mais de um *Frame*, que *Frame* deve ser exibido ao usuário quando ele consultar o dicionário informando um lexema polissêmico, como no caso de <abrir>?

Em verdade, a resposta a esta questão é óbvia: devem ser exibidos todos os *Frames* associados com aquele lexema. Neste caso, o dicionário baseado em *Frames* não se distinguiria dos dicionários comuns (online ou não), que ou exibem todos os sentidos de um lexema na mesma entrada ou criam entradas repetidas com o lexema para cada sentido. Em ambas as situações, cabe ao usuário o processo da desambiguação (o que não deixa de ser um tanto paradoxal, pois se ele já conhece o sentido do que procura, o dicionário pouco pode lhe ajudar).

A questão motivadora deste trabalho na realidade é outra: se o usuário fornecer "pistas" a respeito do que ele procura, por exemplo informando duas ou três palavras, não é possível fornecer a ele uma resposta mais precisa? Afinal cada palavra pode evocar um ou mais *Frames* e os *Frames* estão relacionados entre si. Estas relações poderiam ajudar a estabelecer o *background* proposto pela Semântica de Frames, contra o qual o sentido mais

correto (ou mais próximo) da palavra poderia ser descoberto.

O problema proposto para esta pesquisa pode então ser formulado da seguinte forma:

Como utilizar uma base de dados de Frames, como a FrameNet, para simular, em uma aplicação computacional, o processamento cognitivo realizado para compreender uma expressão linguística (palavras, enunciados, sentenças) com base no Frame evocado por esta expressão?

1.2 Terminologia

Este trabalho se enquadra na área da Semântica Lexical. Segundo Baker, Fillmore e Cronin (2003), diferentes pesquisadores nesta área usam diferentes terminologias. Assim, é importante definir como diversos termos serão usados ao longo deste texto.

As **palavras** são expressões linguísticas que podem representar entidades, fatos ou conceitos referentes às situações em que são usadas. O conjunto de palavras de uma língua forma um **léxico**, que está, desta forma, associado à capacidade cognitiva do uso das palavras. Já um **recurso lexical** possui um viés tecnológico, geralmente referenciando bases de dados lexicais, também chamadas **léxicos computacionais**.

Uma **forma de palavra** (*word form*) representa uma das formas que uma palavra pode assumir, variando com a flexão. Um **lexema** representa a forma não flexionada da palavra (substantivo singular, verbo no infinitivo, a forma básica de um adjetivo). Um lexema pode apresentar ambiguidade, ou seja, sentidos diferentes dependendo do contexto em que esteja sendo usado. Expressões multi-palavras (*MWE - MultiWord Expressions*) são organizadas em um nível mais alto de abstração, chamado **lema**. Um lema é composto por um ou mais lexemas (sendo que uma MWE sempre tem um núcleo). Cada lema possui uma **parte do discurso** (POS - *Part of Speech*), que categoriza o lema, assim como cada lexema. Cada lexema é associado com uma ou mais formas de palavra, mas cada forma de palavra é associada apenas com um lexema. Uma **Unidade Lexical** é o pareamento de um lexema com um significado (no contexto da FrameNet, o pareamento de um lema com um *Frame*). As sentenças em (1.1) ilustram estes conceitos.

- (1.1) (a) Eles cantaram o Hino Nacional de costas para a bandeira.
(b) Ele cantou a passageira a seu lado.

Nas sentenças (1.1), <cantaram> e <cantou> são **palavras** (ou formas de palavras), associadas ao **lexema** <cantar>. Como a **POS** de <cantar> é "verbo" (v), neste texto o lexema será descrito como <cantar.v>. As sentenças apresentam duas UNIDADES LEXICAIS distintas: <cantar.v> no sentido de "usar a voz musicalmente" e <cantar.v> no sentido de "tentar seduzir".

Neste texto, usamos a expressão **tipo semântico** no seu sentido mais abrangente, envolvendo um processo de categorização qualquer, que classifique o item lexical sob análise. Já a expressão **tipo ontológico** é usada em sentido mais restrito, para referenciar um tipo semântico que esteja formalmente registrado em uma ontologia.

1.3 Pressupostos teóricos e hipóteses

Situada no contexto do projeto FrameNet, esta pesquisa naturalmente adota os pressupostos teóricos inerentes à Semântica de Frames. No entanto, a fim de que o trabalho apresentado possa ser avaliado adequadamente, ressaltam-se aqui alguns pressupostos mais específicos:

- a) Os lexemas só tem significação quando associadas a um *Frame*. Mais claramente, é a associação do lexema a um *Frame* que dá significação a ele. Esta é uma questão importante, já que, de maneira quase unânime, as teorias sobre léxico postulam uma significação para os lexemas independentemente do uso, como pode ser observado em Geeraerts (2009). Dito de outra forma, o significante é insuficiente para fornecer o significado;
- b) Muitas teorias lexicais assumem uma certa independência do léxico em relação aos processos cognitivos, considerando, quando muito, a existência de um "léxico mental" correspondente ao léxico linguístico. Aqui estamos assumindo que tal independência não existe. O léxico é visto como uma extensão dos modelos cognitivos (no caso, dos *Frames*). Mais claramente, as unidades lexicais são tratadas igualmente como dispositivos cognitivos que têm uma estrutura interna. Ainda que esta estrutura seja insuficiente para atribuir um significado por si só, é ela que permite a composição da unidade lexical em um cenário mais amplo;
- c) O significado de uma expressão é alcançado através do processo de composicionalidade. Isto implica que não apenas o significado da sentença está relacionado com a composição dos significados das palavras que a formam, mas que o próprio significado da palavra (via *Unidade Lexical*) dentro da sentença é resultado da composicionalidade. Ou seja, enquanto o significado da sentença é função dos *Frames* evocados pelos lexemas presentes na sentença, o *Frame* escolhido para cada lexema é aquele que melhor se encaixa na composição.

Assim, o significado, ou melhor, a construção do significado em uso, decorre tanto do potencial semântico das *Unidades Lexicais*, quanto do seu potencial combinatório (associado às valências das *Unidades Lexicais* nas construções em que ela ocorre), quanto da determinação (dinâmica) feita pelo contexto/situação em que os enunciados acontecem.

Considerando estes pressupostos, as hipóteses básicas deste trabalho, visando a investigação do problema proposto na seção anterior, são:

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

1.4 O trabalho realizado

Considerando o valor da abordagem proposta pela Semântica de Frames para o tratamento de várias questões em Linguística, é natural que pesquisadores de todo o mundo tenham buscado utilizá-la de alguma forma. Em especial, no estudo da polissemia são diversos os trabalhos realizados (BOAS, 2001), (DURAN-MUNOZ, 2011), (LIEN, 2000), (LIU; CHIANG; CHOU, 2005). Estes trabalhos, no entanto, se desenvolvem apenas no aspecto linguístico. Quando se trata do aspecto computacional, com o uso da base de dados da FrameNet, diversos desafios são levantados.

Naturalmente, para uma pesquisa realizada no âmbito da língua portuguesa, como esta, o primeiro obstáculo foi o fato de a base de dados estar em inglês. Nos últimos anos, grupos localizados em diversos países vem trabalhando na criação de suas próprias "framenets". No Brasil está em andamento o projeto FrameNetBrasil (SALOMÃO, 2009), em cujo contexto este trabalho foi desenvolvido. Uma vez superada a questão da língua, os desafios estiveram relacionados à consistência dos dados da FrameNet, à formalização da estrutura da rede, bem como à metodologia adotada para criação e manutenção dos *Frames*. Estes tópicos são bem discutidos em (MOREIRA; SALOMAO, 2012; OSSWALD; Van Valin Jr, 2014; OVCHINNIKOVA, 2012) e não foram tratados neste trabalho.

Uma das premissas adotadas para a pesquisa foi buscar não alterar a estrutura básica da FrameNet como ela se encontra hoje. Os acréscimos realizados devem ser vistos como uma "camada", em que se registraram novas associações entre os elementos já existentes. Assim, o trabalho foi desenvolvido em dois aspectos complementares: a implementação computacional de um processo de desambiguação de *Frames* e a integração da rede FrameNet com outros recursos lexicais a fim de promover a melhoria deste processo.

A implementação do processo de desambiguação compreendeu o desenvolvimento de uma aplicação em PLN sobre a base de dados da FrameNet e a realização de experimentos

com o lexema <abrir>. Esta aplicação envolve a criação de um algoritmo que considere os *Frames* evocados por um conjunto de lexemas (lexemas coocorrentes) e retorne o frame "mais adequado" à compreensão do conjunto, ou seja, aquele que, probabilisticamente, represente melhor o cenário descrito pela composição dos lexemas.

Um segundo aspecto foi a criação de mecanismos que ajudam no estudo do processo de composicionalidade. Realizou-se uma integração da Semântica de *Frames* com a Teoria do Léxico Gerativo (TLG) (PUSTEJOVSKY, 1995), usada para explorar a estrutura interna dos itens lexicais.

1.5 Aplicação dos resultados

Pretende-se que os resultados desta pesquisa possam ser usados em diversas tarefas em PLN:

- a) Desambiguação de lexemas: em aplicações onde a busca de *Frames* é feita a partir de lexemas polissêmicos (por exemplo, dicionários baseados em *frames*) a desambiguação é um processo importante para fornecer respostas mais adequadas às consultas feitas pelo usuário;
- b) SRL (*Semantic Role Labeling*): em aplicações que realizam a etiquetagem semântica de sentenças, através de processos automáticos ou semi-automáticos, a correta identificação dos *Frames*, *Elementos de Frames* e tipos semânticos associados às palavras na sentença é fundamental;
- c) Processo de anotação de texto corrido: este processo de anotação, usado no âmbito do projeto FrameNet, pode ter uma fase inicial semi-automática, propiciando a anotação de um maior número de textos;
- d) Sumarização de textos: o processo de descoberta dos *Frames* evocados em um texto, associado a medidas estatísticas, pode auxiliar na construção (semi)automática de sumários;
- e) Processos de tradução semi-automáticos: com o suporte de um dicionário multilíngue baseado em *Frames*, é possível construir equivalentes de tradução (ou pelo menos sugestões de equivalentes de tradução), uma vez que os *Frames* utilizados sejam definidos;
- f) Análise de ocorrências de metonímias: situações em que palavras associadas a um *Elemento de Frame* ou a um subframe sejam usadas em uma sentença, no lugar da palavra original, podem ser estudadas através do processo de composição dos lexemas.

2 Metodologia

Esta pesquisa tem como meta investigar como uma base de dados de *Frames*, no caso a FrameNet, pode ser utilizada em uma aplicação computacional que simule o processo cognitivo de compreensão de expressões linguísticas, com base na teoria da Semântica de Frames. Uma aplicação deste tipo pode ser útil em diversas tarefas de PLN, além de auxiliar na avaliação das teorias linguísticas que lhe dão suporte. Atenção especial é dada ao problema da polissemia, que sempre se apresenta como um desafio tanto para os estudos linguísticos quanto para as aplicações em PLN.

2.1 Objeto de estudo

O objeto de estudo é a própria FrameNet, sua estrutura e as relações estabelecidas entre os diversos elementos desta estrutura. Sendo um projeto com características tanto linguísticas quanto computacionais, para a realização da pesquisa foi realizado um levantamento bibliográfico na área da Semântica Lexical Cognitiva (onde se enquadra a Semântica de Frames, teoria base do projeto FrameNet) e em algumas áreas da Ciência da Computação, uma vez que para o experimento proposto (a implementação de uma aplicação computacional) são usados conceitos referentes a Redes Semânticas, Ontologias, Modelos Conexionistas e Processos de Inferência. Naturalmente, a bibliografia disponível sobre a FrameNet também foi revista. Um sumário dos aspectos mais relevantes de cada área, para esta pesquisa, será apresentado na Parte II.

2.2 Métodos

Como discutido no capítulo 1, o problema sob investigação é a simulação computacional do processo cognitivo de compreensão da língua natural usando uma base de dados de *Frames*. A hipótese levantada é que a inclusão de novas estruturas e relações na base de dados, interligando *Elementos de Frames* a outros *Frames*, a *Unidades Lexicais* e a *Tipos Ontológicos*, pode melhorar a resposta computacional oferecida pela simulação.

Como o trabalho aborda a organização conceptual da linguagem, ele se enquadra no campo da Semântica Cognitiva, no sentido apresentado por (TALMY, 2000, p. 4). Para a Semântica Cognitiva, o principal objeto de estudo é o fenômeno qualitativo mental, como ele existe no inconsciente cognitivo e como é trazido à consciência. Mais especificamente, é o estudo do conteúdo conceptual e sua estruturação na linguagem. A proposta de Talmy é que a única instrumentalidade disponível para acessar tal conteúdo fenomenológico é a **introspecção**.

Esta proposta é desenvolvida em [Talmy \(2006\)](#), argumentando que a metodologia da introspecção tem ocupado um lugar central no desenvolvimento da Linguística Cognitiva e continua sendo a principal metodologia nesta área. A introspecção linguística, segundo ele, é a atenção consciente dirigida por um usuário da língua para aspectos específicos da linguagem, como manifesta em sua cognição. Em ([TALMY, 2000](#), p. 5), ele exemplifica este ponto, mostrando que alguém pode estar plenamente consciente de um significado específico de uma palavra que ouviu, ao mesmo tempo que tem apenas uma leve consciência (ou nenhuma) dos aspectos homonímicos ou polissêmicos daquela palavra. Assim, estes dois aspectos semânticos diferentes de uma palavra (o significado específico e a faixa de significados) são diferentes em seu acesso à consciência. Em geral, os aspectos mais acessíveis à consciência são mais sujeitos ao método da introspecção, enquanto os aspectos menos acessíveis requerem outros métodos indiretos de análise. Ou seja, o método de introspecção pode (e deve) ser complementado por uma metodologia que envolva o pensamento analítico, incluindo a manipulação sistemática de ideias, a abstração, a comparação e o raciocínio.

Além disso, Talmy alerta que a introspecção deve ser empregada com rigor. Isto implica no controle do material linguístico cujo significado será estudado e na correlação dos resultados encontrados com os resultados de outras metodologias (tais como a introspecção usada por outros pesquisadores, a análise do discurso e de corpora, a análise diacrônica, técnicas de observação e experimentais da psicolinguística, entre outras).

Neste trabalho a metodologia da introspecção está sendo usada, no sentido de se buscar a compreensão da língua natural a partir da experiência do autor como usuário da língua. No entanto este passo inicial é complementado pela análise dos sentidos já oferecidos pelos léxicos utilizados na pesquisa. Além disso, a proposta de métodos alternativos para a desambiguação dos lexemas envolve o uso do pensamento analítico para realizar abstrações e comparações com as técnicas já existentes.

A pesquisa foi concretizada através de um sistema computacional que implementou as ideias e técnicas estudadas. Embora este trabalho não envolva diretamente a análise de corpus, os experimentos são realizados com exemplos de sentenças retiradas de corpus, ao invés de simples exemplos construídos (estes são usados esporadicamente, apenas para explicação da metodologia). A avaliação tem um caráter qualitativo, considerando-se que a hipótese apresentada no capítulo 1 pode ter um grau maior de veracidade se houver um número significativo de acertos, ou seja, se a resposta computacional se aproxima significativamente da compreensão do texto realizada por um humano.

Uma limitação que afetou diretamente a pesquisa é a ausência de corpus em Português Brasileiro anotados semanticamente usando a Semântica de Frames. Como esta pesquisa é realizada no âmbito do Projeto FrameNet Brasil ([SALOMÃO et al., 2011](#)), acreditamos que o uso de sentenças anotadas permitirá, futuramente, realizar mais uma validação dos resultados obtidos na implementação, uma vez que eles poderiam ser

comparados com as anotações realizadas.

Como o trabalho envolve o desenvolvimento de um sistema computacional, e se enquadra como uma pesquisa na área de PLN, faz-se necessário, também, a adoção de uma metodologia específica para esta área.

2.3 Metodologia de pesquisa em PLN

DIAS-DA-SILVA (2006) propõe uma estratégia de pesquisa em PLN que envolve equacionar os problemas em três domínios de investigação: Linguístico, Linguístico-Computacional e Computacional. O **Domínio Linguístico** envolve questões relacionadas à explicitação do conhecimento e do uso linguísticos. É nele que os fatos de língua e de uso são especificados. Conceitos, termos, regras, princípios e formalismos linguísticos são trabalhados neste nível. O **Domínio Linguístico-Computacional** abrange questões de sistemas de representação, que incluem, por exemplo, cálculo proposicional e de predicados, redes semânticas, *frames*, além de tratar das estratégias de codificação dos elementos do domínio anterior. Finalmente, no **Domínio Computacional** estão as questões relativas à implementação das representações através de programas de computador e dos sistemas computacionais usados por estes programas.

A investigação realizada neste trabalho enquadra-se bem nesta proposta. No Domínio Linguístico estão os estudos realizados na área da Linguística Cognitiva, incluindo a Semântica de Frames, e sobre a Teoria do Léxico Gerativo. No Domínio Linguístico-Computacional está o uso de ontologias e o desenvolvimento de um modelo conexionista estruturado para redes semânticas, que é a base do *framework* LUDI. Por fim, no Domínio Computacional está a implementação propriamente dita.

A adoção desta estratégia se reflete na estrutura de apresentação do trabalho. Na Parte II, os referenciais teóricos são subdivididos em aspectos linguísticos e computacionais. A Parte III discute o Domínio Linguístico-Computacional.

2.4 Etapas da Pesquisa

A pesquisa foi realizada em quatro grandes etapas: (i) Levantamento bibliográfico; (ii) Proposta de implementação; (iii) Implementação inicial e (iv) Implementação corrigida. Este texto apresenta o resultado das etapas (i), (ii) e (iv).

2.4.1 Levantamento bibliográfico

A etapa de levantamento bibliográfico teve por objetivos aprender novos conceitos e técnicas, revisar conceitos estudados anteriormente e analisar trabalhos relacionados ao

problema da pesquisa. Um sumário do levantamento é apresentado na Parte II e envolveu basicamente o estudo de bibliografia relacionada a:

- a) Semântica Cognitiva, Semântica lexical, Semântica de Frames e FrameNet;
- b) Teoria do Léxico Gerativo e Ontologia SIMPLE;
- c) PLN (Processamento de Línguas Naturais), em especial tópicos relacionados ao processo de desambiguação;
- d) Ontologias e Modelos Conexionistas.

2.4.2 Proposta de Implementação

A proposta de implementação foi apresentada na forma de um projeto, denominado Projeto LUDI (*Lexical Unit Discovery through Inference*). A ideia do projeto foi agregar as diversas soluções computacionais sob um único *framework*, facilitando a referência e o uso.

2.4.3 Implementação Inicial

Na implementação inicial foram criados os programas componentes do *framework* LUDI. As aplicações referentes a cada etapa da implementação foram construídas isoladamente, com uso de tecnologias diferentes. Esta etapa teve por objetivo basicamente testar a viabilidade das propostas, através de experimentos. Nesta etapa foram criadas novas entradas na ontologia SIMPLE e na base da FrameNet (para os lexemas em Português) e testados diferentes valores para os parâmetros usados nos algoritmos. Para análises relativas a composicionalidade, foi usado o programa **Constructional Analyser** (BRYANT, 2008).

Para os experimentos desta etapa foram usados apenas sintagmas da forma V-SN. Nos experimentos dos algoritmos foram usados exemplos de expressões linguísticas que envolviam a ocorrência de polissemias, metonímias e metáforas, a fim de se analisar a necessidade ou possibilidade de introdução de novas estruturas que pudessem melhorar os resultados.

2.4.4 Implementação corrigida

Três alterações principais foram realizadas para a implementação corrigida: (i) possibilidade de informar mais de um argumento do verbo usado como *Unidade Lexical*; (ii) o uso das valências no processo de desambiguação e (iii) a implementação de um modelo conexionista estruturado para a realização das inferências. Além disso, um número maior de sentenças e de *Frames* foi usado nos experimentos, cobrindo tanto interpretações literais quanto metafóricas. A integração da FrameNet com a ontologia SIMPLE também foi revista e ampliada.

Parte II

Referenciais teóricos

3 Aspectos Linguísticos

O presente trabalho está relacionado com a questão de como as palavras "significam" alguma coisa. O significado é tradicionalmente objeto de estudo da Semântica. Mais especificamente, uma vez que o conjunto de palavras de uma língua forma um léxico, este trabalho se enquadra na área da Semântica Lexical.

Por outro lado, com o desenvolvimento e popularização dos recursos da computação, é crescente o número de **léxicos computacionais** e grandes bases de dados disponibilizadas para os pesquisadores. A criação, implementação e uso de léxicos computacionais trazem uma série de questões desafiadoras para a Semântica Lexical. Teorias satisfatórias quando aplicadas a dados construídos artificialmente mostram-se frágeis quando aplicadas em larga escala ou com dados reais (*usage data*). As aplicações em PLN tornam-se mais complexas, por um lado implementando as teorias propostas, por outro apontando a necessidade de revisão ou complementação destas teorias.

As questões tratadas por esta pesquisa foram levantadas no contexto do uso computacional da base de dados do projeto FrameNet ([FILLMORE; JOHNSON; PETRUCK, 2003](#)) em uma aplicação real ([SALOMÃO et al., 2011](#)), possibilitando uma visão crítica não apenas do projeto mas também da própria Semântica de Frames ([FILLMORE, 1982](#)). A fim de desenvolver uma proposta que buscasse tratar aquelas questões, foi realizado um levantamento bibliográfico que possibilitasse a fundamentação teórica da proposta. Um sumário deste levantamento é apresentado nas seções deste capítulo e do seguinte.

Este capítulo revisa a questão do significado lexical, introduz a Teoria do Léxico Gerativo, apresenta as principais ideias da Linguística Cognitiva e da Semântica de Frames, destaca alguns aspectos do processo de Causação e da Dinâmica de Forças e, finalmente, aborda o processo da composicionalidade semântica. Os aspectos computacionais são apresentados no capítulo 4.

3.1 O Significado Lexical

Como as palavras significam alguma coisa? Ou, dito de outra forma, como compreendemos o significado das palavras? Para a maioria das pessoas (ou pelo menos para a maioria que não é linguista) esta parece ser uma questão trivial, e falar em uma Teoria do Significado Lexical parece sem propósito. Afinal, um <gato> é um gato. Um <livro> é um livro. O que mais há para dizer sobre isso?

Na realidade, há muita coisa. Principalmente depois das pesquisas de linguistas como Charles Fillmore, George Lakoff, James Pustejovsky, entre muitos outros, e de

implementações de léxicos computacionais como a Wordnet (FELLBAUM, 1998) e a FrameNet, e ontologias como a DOLCE (MASOLO et al., 2003). Todos estes trabalhos nos dão uma descrição muito sugestiva, como diz Asher (2011), do que as palavras realmente significam. Longe de terem um significado relativamente bem comportado e estável (embora as vezes se apresentem assim), as palavras são "proteanas" (EVANS, 2009, p. xi). Ou seja, elas podem mudar seu significado nos diferentes contextos de uso. Este é, claro, um grande desafio quando se pretende construir alguma teoria de compreensão da linguagem.

Não apenas isto, mas (um fato reconhecido facilmente, mas pouco compreendido ainda) quando os significados de duas palavras diferentes são combinados, o significado resultante pode ser diferente da simples "soma" dos significados das duas palavras. A ideia que os significados das palavras, ou mais propriamente o resultado da interação semântica entre as palavras, vai mudar, dependendo de outros elementos na predicação ou dependendo do contexto do discurso é um tanto óbvia. No entanto, este é um campo de pesquisa aberto na Linguística, e é também um dos temas desta pesquisa.

A abordagem adotada aqui, dado o viés computacional, se enquadra no campo daquelas que usam uma representação intermediária entre a língua natural e o seu uso, para especificar o significado. A palavra, ou mais corretamente uma entrada lexical em dado léxico, deve então especificar (ou estar submetida a) algum tipo de estrutura que permita, quando combinada com outras palavras em uma sentença bem formada, gerar uma representação que possa ser "interpretada". Embora Asher (2011, p. 7) chame esta representação de **forma lógica**, vamos evitar esta expressão uma vez que ela está associada, no campo da semântica formal, à representação da sentença usando algum tipo de fórmula baseada em Lógica. A geração desta representação, a partir de uma sentença fornecida em língua natural, é um dos processos básicos do PLN. Neste trabalho, a estruturação das palavras estará baseada na Teoria do Léxico Gerativo.

3.2 A Teoria do Léxico Gerativo

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

Segundo Pustejovsky (1998b), a TLG é largamente motivada pelo desejo de ver qual a contribuição que a informação lexical fornece para a interpretação semântica completa da sentença. A TLG busca estudar a habilidade que temos de tomar uma

categoria e refinar ou redefinir seu uso de uma forma nova ou em um novo contexto. O refinamento ou a redefinição contínua dos papéis que um objeto exerce em nosso ambiente e como conceptualizamos as diferentes propriedades daquele objeto em diferentes contextos é um dos mecanismos analisados pela TLG. A tese é que o potencial semântico da linguagem não deve ser explicado em termos de mecanismos *ad hoc*, mas sim através de um **sistema gerativo restrito**, criado especialmente com esta finalidade. Assim, os mecanismos subjacentes à semântica permanecem inalterados, enquanto dão origem a novos significados. De fato, afirma Pustejovsky, o que faz a linguagem tão unicamente expressiva é o modo como ela parece abraçar as mudanças de significados, tais como a polissemia. A polissemia (mais especificamente a chamada polissemia regular, sistemática ou lógica) ocupa, realmente, uma posição de destaque nos estudos da TLG, nos quais ela é considerada central à linguagem.

Assim, ao invés de tratar a polissemia através de uma abordagem enumerativa, na TGL a metodologia tem sido a construção de estruturas diferenciais, que têm comportamentos funcionais. Mais especificamente, a semântica dos itens lexicais é considerada subespecificada; cada elemento em um sintagma pode funcionar ativamente na composição do sintagma e a noção de "sentido da palavra" somente emerge no nível da sentença, ou seja, é o "sentido no contexto" (PUSTEJOVSKY, 1991).

Crucialmente a TLG, ao invés de assumir um conjunto fixo de **primitivas** para descrever o significado das palavras, assume um número fixo de **mecanismos gerativos** que são usados para construir as expressões semânticas. Desta forma, um **léxico gerativo** pode ser caracterizado como um sistema envolvendo pelo menos quatro níveis básicos de representação linguística:

- a) Estrutura argumental (A), que define o número e o tipo de argumentos de um item lexical;
- b) Estrutura de eventos (E), que se refere às propriedades de um evento associado ao item lexical;
- c) Estrutura *qualia* (Q), que provê os objetos semânticos que definem o significado de um item lexical;
- d) Estrutura de herança lexical (I), que lista as relações existentes entre uma dada estrutura lexical e outras estruturas no léxico.

Um dado item lexical (IL) é então representado pela informação nestes níveis distintos pela expressão (3.1).

$$(3.1) \quad IL = \langle A, E, Q, I \rangle$$

Por outro lado, um conjunto de mecanismos gerativos conecta estes quatro níveis, provendo a interpretação composicional das palavras em contexto (PUSTEJOVSKY,

1995, p. 105). A natureza destes mecanismos determina a expressividade polimórfica da semântica de maneira bastante restritiva. Estes mecanismos incluem a coerção de tipo, a subseleção e a co-composição.

A **coerção de tipo** consiste em uma operação semântica que converte um argumento de um dado tipo no tipo esperado pelo predicado. Apesar da coerção de tipo poder ser refletida na expressão sintática do predicado, esta operação semântica não envolve mudanças para o tipo sintático do item, mas sim para seu tipo semântico, que pode ser associado a formas sintáticas canônicas diferentes. O tipo mais simples de coerção é a coerção de subtipo, quando um argumento é de um subtipo do tipo esperado pelo predicado. A sentença (3.2) é compreensível porque <Honda> é um subtipo de <veículo> (o complemento esperado para <dirigir>). Na mesma sentença pode-se ver que o tipo de <serviço> (uma ação) sofre uma coerção para <local>, que é o tipo esperado para um dos argumentos de <chegar>.

(3.2) Pedro dirigiu seu Honda até chegar ao serviço.

A **subseleção** é um mecanismo relacionado à descrição da relação entre um modificador e o item modificado. A subseleção permite ao modificador selecionar seus argumentos de um conjunto de objetos no conteúdo semântico do item modificado. Este mecanismo possibilita, por exemplo, explicar a sensibilidade ao contexto de adjetivos como <rápido>, dependendo do substantivo, uma vez que a interpretação resultante emerge da modificação do valor do *qualia* selecionado presente no conteúdo semântico do substantivo. Em <um carro rápido> e <um digitador rápido>, o adjetivo seleciona, claramente, diferentes interpretações.

Já a **co-composição** está associada a ocorrência da chamada **polissemia lógica**. Pustejovsky (1995, p. 28)) apresenta a seguinte definição para a polissemia lógica:

*I will define logical polysemy as a complementary ambiguity where there is no change in lexical category, and the multiple senses of the word have overlapping, dependent, or shared meanings.*¹

Assim, são descritas como casos de polissemia lógica as várias alternâncias que sistematicamente ocorrem em nomes, tais como a relação entre contáveis e não contáveis; recipiente e conteúdo; processo e resultado; lugar e pessoas etc. Esta recorrência justifica o tratamento dado a esse tipo de polissemia como "lógica", pois parece haver algo intrínseco a natureza lexical desses termos que os torna polissêmicos, e esse é um fenômeno produtivo na língua .

¹ "Vou definir a polissemia lógica como uma ambiguidade complementar, onde não há mudança na categoria lexical e os múltiplos sentidos da palavra têm significados sobrepostos, dependentes ou compartilhados". (Tradução nossa)

A co-composição vai, então, congregiar as operações semânticas que permitem completar formas subespecificadas com o conteúdo semântico de seus argumentos. Dentre estas operações estão, por exemplo, a co-composição de maneira, a unificação de *features* e a especificação de verbos leves. Moura e Pereira (2004), estudando o verbo <preparar.v>, argumentam que este apresenta polissemia lógica, pois possui sentidos diversos que podem ser correlacionados de maneira regular. Esta polissemia seria produzida pela regra de co-composição, ou seja, os diferentes sentidos do verbo derivam de sua combinação com os sintagmas nominais em posição de complemento. Por exemplo, quando os argumentos, em posição de complemento, são eventos, a ação verbal designa atividades de planejamento. Já quando os argumentos são pessoas, a ação verbal se refere, única e exclusivamente, a uma mudança de estado. As sentenças (3.3) ilustram esta situações.

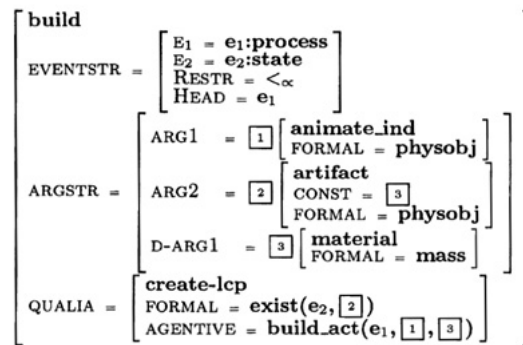
- (3.3) (a) A empresa prepara o lançamento de um novo produto.
 (b) O curso prepara o aluno para o processo seletivo.

Por outro lado, a TLG considera que o significado das palavras está estruturado na base de quatro fatores gerativos, chamados "papéis *qualia*". Cada **papel *qualia*** captura como os seres humanos entendem os objetos e as relações no mundo e provê uma explicação mínima para o comportamento linguístico dos itens lexicais. A determinação dos valores de cada papel *qualia* não é obrigatória; para um dado item lexical, somente os valores dos papéis pertinentes devem ser definidos. (PUSTEJOVSKY, 1995, p. 85) define quatro papéis *qualia*:

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

A Figura 1 ilustra três dos níveis descritos anteriormente, considerando a semântica do verbo <construir> 'build'. Os números representam a unificação dos traços de cada estrutura. O processo e_1 é identificado como um ato agentivo, envolvendo tanto um sujeito *ARG1* e um argumento *default D – ARG1*, que é relacionado com o objeto relação constitutiva de *ARG2*.

Dado o viés computacional da TLG, os valores dos papéis *qualia* são especificados através de expressões lógicas, com tipos bem definidos e estruturas relacionais, indicando a

Figura 1 – Estrutura do item lexical *build* (contruir)

Fonte: Pustejovsky (1995, p. 82)

associação apropriada com o termo predicador. O sistema de tipos, usado para categorizar os itens lexicais, é assim uma questão importante na TLG. Em especial, em muitos sistemas de tipos usados para estruturas conceituais (e as lexicalizações associadas) é necessário empregar um esquema de herança múltipla. Porém, este esquema sofre de uma noção muito limitada de estrutura lexical; em particular, uma consequência deste modelo é a ambiguidade em relação a participação de um item em uma classe.

Para tratar esta questão, a TLG propõe o uso de uma **herança ortogonal**. Neste modelo de herança, ao invés de uma única estrutura organizacional para os tipos, diferentes estruturas são usadas ortogonalmente, baseadas na estrutura *qualia*. Um item lexical herda informação de acordo com a estrutura *qualia* que ele carrega. Por exemplo, <livro> é um <objeto físico> (*qualia* formal), mas <livro> é uma <informação> (*qualia* télico). Os tipos para um dado item lexical podem então ser unificados. Esta estratégia permite a criação de conceitos cada vez mais específicos. Tipos unificados podem ser vistos como estruturados por dimensões ortogonais, ou perspectivas.

Um item lexical pode também apresentar sentidos que aparentemente são contraditórios ou distintos, mas que na realidade estão ligados mais profundamente. Nas sentenças em (3.4) a palavra <livro> parece ter sentidos distintos (informação e objeto comercial), mas que estão relacionados. Na TLG esta situação é representada pelo uso de **tipos complexos**. Um **tipo complexo** é uma relação que associa um par de sentidos de uma forma bem definida, permitindo que, em dado contexto, um ou outro sentido seja usado na interpretação.

- (3.4) (a) Maria não acreditou naquele livro.
 (b) Paulo vendeu seu livro para João.

Pustejovsky (1998a) afirma que a posição adotada pela TLG é de que a capacidade linguística humana é reflexo de nossa habilidade em categorizar e representar o mundo

de maneira específica. O que é unicamente humano não é uma linguagem extensional *per se*, mas a habilidade gerativa de construir o mundo conforme ele é revelado através da linguagem e das categorias que são empregadas. Esta posição claramente reflete a relação entre a capacidade gerativa e o processo cognitivo, teoricamente assentada por Chomsky desde seus primeiros trabalhos e apresentada mais detalhadamente em (HAUSER; CHOMSKY; FITCH, 2002), com a ressalva de que, neste caso, a geratividade é estritamente sintática.

Em especial, na TLG pode-se perceber a ligação entre a ideia da geratividade sintática e da geratividade semântica. A chamada **infinitude discreta** pode ser explicada, na sintaxe, pela manipulação simbólica da forma, através de fórmulas, algoritmos, regras de transformação. Dispondo de um conjunto de unidades combinatórias e de um conjunto de princípios que combinem estas unidades, sentenças bem formadas e interpretáveis podem ser produzidas. Projetada na âmbito da semântica, a TLG propõe que o significado também pode ser assim representado, deduzido, computado, uma vez que se disponha de certas unidades que podem ser combinadas (o léxico) e de mecanismos que realizem a combinação (os mecanismos gerativos).

Como destaca Salomão (2002), é interessante notar que, a par de avanços inegáveis em relação às teorias que privilegiam apenas os aspectos sintáticos (pois ocorrem refinamentos nos processos de composição semântica e a polissemia passa a ser considerada), permanece ainda a ideia que o significado do todo pode ser obtido pela composição dos significados das partes; apenas as partes, agora, possuem uma estrutura um pouco mais complexa (não "atômica", mas "molecular"). Nesta etapa, segundo Salomão (2002), a significação continua ser tratada como representação simbólica, derivada algorítmicamente.

Paralelamente, um grupo de pesquisadores com experiência em semântica gerativa (como Fillmore, Lakoff, Langacker, entre outros) passaram a buscar um viés teórico mais consistente para as relações entre sintaxe e semântica, investigando principalmente as relações entre forma e significado (FERRARI, 2011, p. 13). A ideia fundamental nesta abordagem alternativa é que a forma serve para evocar o significado, mas não explica os mecanismos de sua construção. As operações associadas ao significado, tais como a identidade, a integração dos conceitos e o uso da imaginação, apresentam extremas dificuldades de formalização e possuem um caráter essencialmente inconsciente, usando metáforas, simbolismos e analogias. Os estudos sob esta perspectiva, em que a linguagem não é vista como uma capacidade cognitiva à parte das demais, mas visa estruturar linguisticamente o conteúdo conceptual, deram origem a chamada Linguística Cognitiva.

3.3 A Linguística Cognitiva

Geeraerts e Cuyckens (2007) definem a Linguística Cognitiva como o estudo da linguagem em sua função cognitiva de intermediar estruturas informacionais em nossas relações com o mundo. 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. Sob esta perspectiva, a análise das bases conceitual e experiencial das categorias linguísticas é de fundamental importância na Linguística Cognitiva: as estruturas formais da linguagem não são estudadas como se fossem autônomas, mas como reflexos de uma organização conceitual geral, com princípios de categorização, mecanismos de processamento e influências experienciais e ambientais.

A partir desta caracterização geral, Geeraerts e Cuyckens (2007) derivam três características fundamentais da Linguística Cognitiva:

- 1 A primazia da semântica nas análises linguísticas: a função básica da linguagem é a categorização e o significado é o fenômeno linguístico primário.
- 2 A natureza enciclopédica do significado linguístico: se a linguagem é um sistema para a categorização do mundo, é desnecessário postular um outro nível de significado linguístico diferente do nível em que o conhecimento do mundo está associado a formas linguísticas.
- 3 A natureza perspectivizada do significado linguístico: o mundo não é objetivamente refletido na linguagem; pelo contrário, é a função de categorização da linguagem que impõe uma estrutura ao mundo.

Ainda que estas características não sejam exploradas aqui, não é difícil perceber que a abordagem cognitivista da linguagem, em geral, e da semântica, em particular, se contrapõe às abordagens de caráter mais "realistas", "objetivas" ou "formais", que propõem a existência de uma realidade externa objetiva, com significados pré-existentes, e estruturas formais (inclusive linguísticas) para descrever estes significados. É óbvio que o mundo existe objetivamente, mas ele é, de certa forma, constantemente construído e reconstruído subjetivamente através dos processos cognitivos. A linguagem, no seu aspecto cognitivo, é uma ferramenta para esta (re)construção. Assim as palavras, enquanto expressões linguísticas objetivas, dão nome às "coisas da mente" e não às "coisas do mundo". O processo de significação, ao invés de ser um processo em que uma "realidade" é "representada", é um processo em que são evocadas estruturas cognitivas complexas, que se articulam em redes e que categorizam ou analisam a experiência do mundo. Estas estruturas não são apenas linguísticas, mas transitam do linguístico ao neurobiológico, ao psíquico (inconsciente cognitivo, consciência), ao social (modelos culturais, ideologia,

ética). Dado este caráter tão fluido do significado, é natural que sua representação não possa estar restrita aos limites das estruturas formais.

Considerando que uma das questões subjacentes ao problema apresentado para investigação nesta pesquisa é a desambiguação de lexemas polissêmicos no contexto do projeto FrameNet, nas seções seguintes restringimos a discussão ao processo de categorização via categorias radiais e redes esquemáticas e à representação de conceitos através de *frames*, via Semântica de Frames.

3.3.1 Categorização e Polissemia

Se consideramos que os conceitos não existem de forma isolada na mente, mas estão organizados de alguma forma, uma consequência é a de que o léxico também deve possuir uma estrutura, uma organização. O léxico não é apenas uma longa lista de lexemas isolados e desconectados. Ao contrário, é possível reconhecer vários tipos de relações e conexões entre os elementos do léxico, o que nos ajuda a estabelecer regularidades e reconhecer padrões. Estas relações são definidas com base no significado do lexema e definem o que é conhecido como **estrutura semântica**.

Uma observação deve ser feita aqui: se, conforme as hipóteses apresentadas anteriormente, um lexema não tem um significado *per si* (pois este é construído a partir da evocação de um *Frame*), podemos considerar que as relações lexicais representam relações conceptuais, ou relações entre conceitos. Inversamente, a classificação ou categorização de conceitos também deve se refletir na categorização dos lexemas que os evocam.

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

Esta caracterização tipo "tudo ou nada" (o conceito deve atender a todas as características para ser classificado em dada categoria) é útil na matemática e na computação, mas não se deve esperar que todas as categorias que aprendemos consistam de características necessárias e suficientes. As pessoas evoluíram para ver as coisas em graus (FELDMAN, 2006), ou seja, para experimentar o mundo com flexibilidade cognitiva. Os estudos feitos por Rosch e descritos em Lakoff (1987) mostram que as categorias são tanto graduadas em termos de tipicidade quanto estruturadas internamente; nem todos os membros de uma categoria são igualmente representativos daquela categoria. Por exemplo, se consideramos um pardal, uma galinha e um pelicano, testes psicológicos mostram que, embora sejam todos pássaros, o pardal é um exemplo melhor de pássaro que a galinha ou o pelicano.

Em uma categoria, alguns membros são mais típicos que outros. Estes funcionam como **pontos de referência cognitiva** ou **protótipos**.

Por outro lado, enquanto as categorias clássicas tem limites bem definidos (pela própria definição da categoria), as categorias cognitivas possuem limites imprecisos. Claramente um pardal e uma galinha podem ser classificados como pássaros (embora pardais sejam melhores exemplos da categoria que galinhas), mas uma categoria como <pe^{so}oa velha> apresenta uma incerteza quanto ao seu limite: quando exatamente uma pessoa pode ser considerada velha?

Esta questão é tratada pela **Teoria dos Conjuntos Nebulosos** (*Fuzzy Set Theory*). Segundo Lakoff (1987), esta teoria pode ser usada para modelar as categorias em que há gradações no pertencimento de um elemento a categoria. Por exemplo, uma categoria como <homens altos> é graduada, simplesmente porque existem gradações de altura para os seres humanos. Enquanto nos conjuntos clássicos a participação de um elemento no conjunto só pode ser definida como 0 (não participa) ou 1 (participa), em conjuntos nebulosos são permitidos valores intermediários entre 0 e 1, refletindo a ideia de que existem homens que não são totalmente altos nem totalmente baixos, mas ocupam alguma posição intermediária. Assim, os conjuntos nebulosos podem ser usados para modelar a incerteza nos limites de uma categoria.

Outra questão que aponta para a necessidade de categorias prototípicas diz respeito a completude das condições necessárias e suficientes. Um pássaro empalhado, por não voar, deixa de ser classificado como pássaro? É necessário definir uma nova categoria de pássaros, ou ampliar e detalhar as condições de classificação como <pá^{ss}aro>? Categorias prototípicas são definidas menos rigidamente que as categorias clássicas. Ao invés de uma única descrição consistindo de características individualmente necessárias e conjuntamente suficientes, a definição toma a forma de um *cluster* de descrições parciais. Os membros destas categorias "parciais" são relacionados entre si por uma espécie de "semelhança familiar" (LEWANDOWSKA-TOMASZCZYK, 2007). Além disso, os membros desta família possuem pesos diferentes: algumas características ou *clusters* de características podem ser mais importantes para a descrição da categoria do que outros.

Partindo desta perspectiva de categorização dos conceitos e seu reflexo na categorização dos lexemas, é preciso considerar também a ambiguidade lexical: um mesmo lexema pode apontar para mais de um conceito. Neste caso, pode se tratar de homonímia ou de polissemia. Embora não exista um consenso na diferenciação dos dois casos (LYONS, 1979, p. 431), pode-se considerar, simplificadaamente, que na homonímia não há relação semântica entre os conceitos evocados pela palavra. Nestas circunstâncias, é comum afirmar que na realidade são duas palavras diferentes que calharam de ter as mesmas propriedades formais (fonológicas e/ou gráficas) (CRUSE, 2010). Na polissemia, os sentidos evocados pela palavra estão de alguma forma relacionados. Mas qual a estrutura destas relações?

Os significados podem ser vistos na forma de um conjunto de categorias radiais. O modelo de categorias radiais é descrito em Brugman e Lakoff (1988) e Lakoff (1987). Os lexemas polissêmicos apresentam sua significação organizada radialmente, ainda que cada um dos sentidos polissêmicos possa, por si mesmo, apresentar uma complexa estrutura na qual se observam efeitos de prototipia. O membro central do conjunto provê um modelo cognitivo que motiva os sentidos mais periféricos. Os sentidos estendidos, em volta da categoria central, são relacionados por diversos tipos de links (transformações de esquemas imagéticos, metáforas, metonímias, etc). Por exemplo, alguns subtipos da categoria <mãe> são extensões previsíveis do membro central da categoria, tais como <mãe biológica> e <madrasta>, enquanto outros estão sujeitos a condições culturais e convenções e dão origem a categorias imprevisíveis, como <mãe de aluguel>.

Lakoff (1987) apresenta algumas características da estrutura radial:

- i não há uma representação única para este tipo de estrutura (as subcategorias central e periféricas tem suas próprias representações);
- ii as categorias periféricas são motivadas pela categoria central (elas não são nem arbitrárias, nem previsíveis);
- iii é necessária uma teoria sobre os tipos de links entre a categoria central e as periféricas.

Observa-se assim que a distinção semântica forma um *continuum*, com os casos claros de homonímia de um lado, os casos claros de vagueza no outro e os casos de polissemia localizados entre esses dois extremos (GRIES, 2006). As sentenças (3.5) com o lexema <banco> exemplificam esta situação, com os sentidos de "banco como artefato", "banco como construção" "banco como instituição" e "banco como estado". A sentença (3.5)(a) parece se referir ao artefato (homonímia); a sentença (3.5)(b) mostra a questão da polissemia (construção-instituição); a sentença (3.5)(c) exemplifica a vagueza, uma vez que é possível uma leitura em que o locutor se refere ao artefato em que ele está sentado e uma leitura em que ele se refere a instituição financeira; por fim, a sentença (3.5)(d) mostra o uso metonímico do lexema <banco> em um domínio específico (o domínio do futebol), em que "estar no banco" pode ser interpretado como "estar como reserva". Assim, a clareza na distinção dos sentidos de um lexema é considerada uma questão de grau, não sendo, como vemos, uma questão trivial.

- (3.5) (a) Me lembro do banco cinza da velha praça.
(b) Estou esperando abrirem a porta do banco.
(c) Agora posso esperar sentado no banco, até me atenderem.
(d) O técnico resolveu deixar o craque no banco.

Outras abordagens reagiram contra a proposta de Brugman e Lakoff (1988), defendendo que a contribuição semântica associada com as palavras é construída no contexto

(CROFT; CRUSE, 2004, p. 97). Ou seja, ao invés de terem sentidos pré-definidos, as palavras teriam um "significado potencial" ou um "uso potencial". Se os significados são conceituais por natureza (CRUSE, 2010), se os conceitos podem ser construídos dinamicamente, via *construals*, e se um item lexical pode ser associado a diferentes *construals*, os significados também podem ser construídos dinamicamente.

Uma forma mais rica de representação considera que cada significado lexical é um ponto de acesso para uma rede de categorias relacionadas. A estrutura radial é um tipo de rede, mas o Modelo de Redes Esquemáticas (LANGACKER, 1987) vai um passo além: ele apresenta níveis diferentes de abstração no modelo. Se existe um esquema mais abstrato que cobre um esquema mais concreto, eles podem coexistir na rede, em diferentes níveis hierárquicos.

Por outro lado, não devem ser ignorados os usos metonímicos e metafóricos dos lexemas que, de certa forma, estendem os seus sentidos centrais. Em especial, para a **Teoria (Cognitiva ou Conceptual) da Metáfora**, apresentada inicialmente em Lakoff e Johnson (1987) e desenvolvida em trabalhos posteriores, as metáforas são mapeamentos conceituais, que integram o sistema conceptual, e não apenas expressões linguísticas. As metáforas permitem entender um campo da experiência com as palavras de outro campo. No caso de metáforas convencionais (cotidianas), as definições metafóricas encontram-se em domínios básicos da experiência, como dinheiro e guerra. Isto implica dizer que qualquer estrutura usada para representar os sentidos das palavras (seja uma estrutura radial, redes esquemáticas, redes semânticas, etc.) deverá considerar a questão da metáfora. Por exemplo, a sentença (3.6) é uma expressão da conhecida metáfora O AMOR (OU A RELAÇÃO AMOROSA) É UMA VIAGEM. A partir da sentença, fica claro que na rede de sentidos para <encruzilhada> deve ser incluído um sentido que evoque o processo de "decisão" ou "escolha", além do sentido prototípico de cruzamento de duas vias.

(3.6) O casamento de Pedro chegou a uma encruzilhada.

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

3.3.2 Semântica de Frames

Da discussão anterior, observa-se que os conceitos não ocorrem isoladamente: eles estão presentes em uma rede de relações com outros conceitos, de tal modo que, para se

compreender um conceito e usá-lo, é preciso entender o seu posicionamento na rede. Além disso, cada conceito pode ser visto como possuindo uma estrutura interna, composta por outros conceitos e pelas relações entre estes. Lakoff (1990, p. 12) apresenta um exemplo, discutindo o significado de <terça-feira>. Para se entender o conceito de <terça-feira> é necessário se entender o conceito de <semana>. Mas o conceito de <semana> não existe na natureza; ele é produto da imaginação criativa da mente humana e da experiência cultural compartilhada (o que possibilita que diferentes culturas tenham diferentes conceitos para <semana>). Da mesma forma, podemos considerar, recursivamente, que o conceito de <semana> depende do conceito de <mês> e este depende do conceito de <ano>. O tipo de estrutura imaginativa necessária para definir um conceito como <terça-feira> é justamente o que Lakoff (1990) chama de **esquema** ou **frame**.

Fillmore e Baker (2010) vão utilizar o mesmo exemplo, para mostrar que o significado de <terça-feira> não pode ser compreendido sem conhecimento sobre como o tempo é dividido na cultura ocidental, incluindo o ciclo de sete dias e os nomes dados a cada membro deste ciclo. Os conceitos de <dia útil> ou <fim de semana> dependem mais ainda de outros conhecimentos, como os períodos da semana que são tipicamente dedicados ao trabalho e ao descanso. Entender uma expressão como "Até que enfim é sexta!" depende de conhecer os períodos da semana e uma generalizada preferência pelo descanso ao invés do trabalho. Ou seja, a compreensão de um conceito, expresso por um item lexical, demanda o conhecimento da situação em que aquele item está sendo usado. Este *background* é o que Fillmore chama de **Frame** e o estudo destes *Frames* é denominado **Semântica de Frames**.

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

Para a Semântica de Frames as palavras são categorizações das experiências e cada uma destas categorias é sustentada por uma situação motivadora, ocorrendo contra o cenário de conhecimento e experiência representado pelo *Frame*. O significado de uma expressão linguística é interpretado de acordo com o *Frame* de fundo que representa a cena no momento da enunciação. Por exemplo, o lexema <quebrar> pode significar <partir, romper> ("O copo quebrou") ou <falir> ("A empresa quebrou"), dependendo do *Frame*

aplicado (ou "evocado", na terminologia comumente usada na Semântica de Frames) e dos elementos linguísticos que ocorrem simultaneamente no mesmo enunciado.

A Semântica de Frames pode então ser definida como o estudo de como as formas linguísticas evocam ou ativam o conhecimento estruturado em *Frames*, e como os *Frames* assim ativados podem ser integrados para fornecer a compreensão das passagens que contém estas formas linguísticas (FILLMORE; BAKER, 2010). Por isso, a Semântica de Frames é também chamada de "Semântica da Compreensão" (FILLMORE, 1982). O estudo completo do processo de compreensão, destaca Fillmore, deve considerar as maneiras pelas quais a informação não-linguística é integrada no processo.

Além disso, como ressalta Salomão (2009), cada *Unidade Lexical*, o pareamento de uma palavra com um de seus significados (CRUSE, 2010), ao evocar um *Frame* impõe sobre ele uma certa perspectiva ou perfilamento. Tomando-se como exemplo o *Frame* associado a uma transação comercial, os lexemas <comprar> e <vender> perfilam a transação sob perspectivas diferentes (do comprador e do vendedor, respectivamente). A ideia central é que um lexema, ao evocar um *Frame*, destaca algum elemento deste *Frame* de forma particular. Isto significa que, quando compreendemos uma palavra, simultaneamente reconhecemos a informação relevante no *background* no qual aquela palavra desempenha um papel interpretativo.

O tratamento da polissemia em um estudo como a Semântica de Frames envolve a escolha entre manter significados distintos agrupados em uma só acepção ("*lumping*") ou separar os significados em diferentes acepções ("*splitting*"). Embora a Semântica de Frames não direcione esta escolha, a abordagem que considera cada *Unidade Lexical* (UL) como o pareamento de um lexema com um sentido (o que implica o pareamento de uma UL com o *Frame* no qual o significado é definido) possibilita o processo de "*splitting*". Isto acontece porque cada UL está localizada em um sistema de relações com outras palavras do *Frame* e com suas propriedades combinatoriais (valência). Por exemplo, <pequeno> em contraste com <grande> é uma UL (evoca um *Frame*), enquanto <pequeno> em contraste com <alto> é outra UL (evoca outro *Frame*).

O estudo dos vários sentidos que podem ser atribuídos a um único lexema deve, então, ser feito procurando lexemas que tenham a mesma forma linguística em mais de um *Frame*. Fillmore e Baker (2010) exemplificam a questão com o *Frame* COMPLIANCE. O lexema <adhere> aparece tanto no *Frame* COMPLIANCE quanto no *Frame* ATTACHMENT, estando relacionado com coisas que conectam-se umas as outras ("o curativo adere à ferida"). Embora possa ser argumentado que <adhere> em COMPLIANCE é um sentido 'motivado' pelo sentido de ATTACHMENT, a evidência que as duas UL <adhere> pertencem a *Frames* diferentes é encontrada na relação morfológica com as nominalizações correspondentes. Através da separação das nominalizações, podemos dizer que a nominalização para o sentido em COMPLIANCE é <adherence>, enquanto o

substantivo correspondente no *Frame* ATTACHMENT é <adhesion>. Fillmore e Baker (2010) explicam que uma das vantagens em separar as UL que têm a mesma forma é que fatos sobre o significado, valência e outras propriedades distribucionais, bem como as correspondências entre as formas das palavras que compartilham uma base morfológica, pode ser expostos no nível da UL e não do lexema.

Este é um ponto fundamental para o presente trabalho, uma vez que a melhoria da interpretação por máquinas pressupõe a possibilidade de desambiguação dos lexemas polissêmicos. O fato de cada UL possuir seu próprio sentido, ou seja, evocar um *Frame* específico já se constitui em parte do processamento que deve ser realizado.

É importante entender também que os *Frames*, enquanto mecanismos cognitivos, não estão restritos à dimensão cultural. Qualquer conceptualização, linguística ou não-linguística, processa-se como evocação de um *Frame* (SALOMÃO, 2009), uma posição que também é assumida pela **Teoria Neural da Linguagem** (FELDMAN, 2006). Todos os conceitos podem ser esquematizados como *Frames* na forma de uma MVA (Matriz de Valores de Atributos), em que os atributos representam os elementos (entidades, propriedades, participantes, etc) que podem ou devem estar presentes para a caracterização do *Frame*. Esquemas imagéticos, cinestésicos e executivos podem ser considerados *Frames* elementares, em contraste com os *Frames* "culturais". A figura Figura 2 mostra três exemplos de *Frames*, em níveis de abstração diferentes: na Figura 2(a) COMÉRCIO_COMPRAR pode ser caracterizado como um *Frame* cultural; na Figura 2(b), ATIVIDADE é um *Frame* abstrato; na Figura 2(c), CONTAINER é uma esquema imagético.

Figura 2 – *Frames* representados como MVA

| COMÉRCIO_COMPRAR | ATIVIDADE | CONTAINER |
|--|---|---|
| <div style="display: flex; align-items: center;"> <div style="border-left: 1px dashed black; border-right: 1px dashed black; padding: 0 5px;"> <i>Comprador</i> </div> <div style="margin-left: 10px;">□</div> </div> | <div style="display: flex; align-items: center;"> <div style="border-left: 1px dashed black; border-right: 1px dashed black; padding: 0 5px;"> <i>Agente</i> </div> <div style="margin-left: 10px;">□</div> </div> | <div style="display: flex; align-items: center;"> <div style="border-left: 1px dashed black; border-right: 1px dashed black; padding: 0 5px;"> <i>Interior</i> </div> <div style="margin-left: 10px;">□</div> </div> |
| <div style="display: flex; align-items: center;"> <div style="border-left: 1px dashed black; border-right: 1px dashed black; padding: 0 5px;"> <i>Mercadoria</i> </div> <div style="margin-left: 10px;">□</div> </div> | <div style="display: flex; align-items: center;"> <div style="border-left: 1px dashed black; border-right: 1px dashed black; padding: 0 5px;"> <i>Atividade</i> </div> <div style="margin-left: 10px;">□</div> </div> | <div style="display: flex; align-items: center;"> <div style="border-left: 1px dashed black; border-right: 1px dashed black; padding: 0 5px;"> <i>Exterior</i> </div> <div style="margin-left: 10px;">□</div> </div> |
| <div style="display: flex; align-items: center;"> <div style="border-left: 1px dashed black; border-right: 1px dashed black; padding: 0 5px;"> <i>Dinheiro</i> </div> <div style="margin-left: 10px;">□</div> </div> | <div style="display: flex; align-items: center;"> <div style="border-left: 1px dashed black; border-right: 1px dashed black; padding: 0 5px;"> <i>Duração</i> </div> <div style="margin-left: 10px;">□</div> </div> | <div style="display: flex; align-items: center;"> <div style="border-left: 1px dashed black; border-right: 1px dashed black; padding: 0 5px;"> <i>Fronteira</i> </div> <div style="margin-left: 10px;">□</div> </div> |
| <div style="display: flex; align-items: center;"> <div style="border-left: 1px dashed black; border-right: 1px dashed black; padding: 0 5px;"> <i>Finalidade</i> </div> <div style="margin-left: 10px;">□</div> </div> | | |
| (a) | (b) | (c) |

Em termos linguísticos, os elementos de um *Frame* emergem na valência lexical e construcional das expressões que evocam este *Frame* e podem ser considerados, segundo Salomão (2009), funções microtemáticas, postuladas em relação ao *Frame* caracterizados por elas. O *Frame* de transação comercial, citado anteriormente, pode, por exemplo, ter como elementos o VENDEDOR e o COMPRADOR, ao invés de uma função temática mais geral, como AGENTE. Além de responder as inadequações das nomenclaturas anteriores, os *Elementos de Frame* são importantes no processo de desambiguação, ajudando a diferenciar situações como exemplificadas pelas sentenças (3.7) (adaptadas de Salomão (2009)).

- (3.7) (a) Dunga *técnico* substituiu o atacante Robinho *jogador* no último jogo.
 (b) Neymar *jogador* substituiu o atacante Robinho *jogador* no último jogo.

A relação entre a Semântica de Frames e a TLG, discutida anteriormente, é explicitada em [Fillmore e Baker \(2010\)](#). Com substantivos que designam um objeto que tem uma função, o contexto linguístico pode em alguns casos ajudar a distinguir entre os casos que se referem meramente ao objeto físico e os casos que apresentam o objeto como servindo a sua função. Por exemplo, nas sentenças (3.8) <sobre a> e <na> ajudam a distinguir a <televisão> como um eletrodoméstico e como um meio de comunicação, o que efetivamente evoca *Frames* diferentes.

- (3.8) (a) Ele pôs o copo sobre a televisão
(b) Ele viu o jogo na televisão.

Na TLG esta é uma distinção entre os papéis *qualia*. Um objeto pode ter um *qualia* formal e um *qualia* télico, o que apontaria para sua representação em Semântica de Frames como duas *Unidades Lexicais* diferentes. [Moreira e Salomao \(2012\)](#) apresenta o exemplo do lexema <livro>, que pode estar associado ao *Frame* TEXTO, através do seu *qualia* formal, e ao *Frame* INFORMAÇÃO, através do seu *qualia* télico. Já [Salomão \(2009\)](#) mostra que <pegar um ovo> e <pegar uma pedra> são diferentes não pelo *Frame* MANIPULAÇÃO evocado por <pegar>, mas pelos diferentes *Frames* que poderiam ser evocados por <ovo> e <pedra>, devido a distinção do seu *qualia* constitutivo.

Um ponto comum às duas teorias (Semântica de Frames e TLG) é o tratamento da significação lexical considerando o seu contexto sintático, ou seja, o registro da Estrutura Argumental (no caso da TLG) e a descrição das valências (no caso da Semântica de Frames). Assim, de alguma forma as duas teorias se complementam na tarefa de possibilitar a compreensão mais completa de um enunciado. Esta é também a perspectiva adotada nos experimentos deste trabalho.

As ideias de Fillmore em relação a Semântica de Frames deram origem a uma implementação computacional, o projeto FrameNet, apresentado na próxima seção.

3.4 FrameNet

O Projeto FrameNet (referido neste texto como "FrameNet" ou simplesmente "FN") é um projeto de lexicografia computacional, iniciado em 1997, que busca extrair informações sobre as propriedades sintáticas e semânticas das palavras, a partir de um corpus eletrônico suficientemente grande. Estas informações são armazenadas digitalmente em um banco de dados e apresentadas em vários relatórios web ([FILLMORE; JOHNSON; PETRUCK, 2003](#)).

A FN é baseada na teoria da Semântica de Frames. A semântica do item lexical é formalizada por uma estrutura que associa os sentidos da palavra a *frames* semânticos, ou simplesmente **Frames**. A FN identifica e descreve os *Frames*, e analisa os significados das

palavras tendo estes *Frames* como *background*. O principal foco do projeto é a descrição das valências em termos sintáticos e semânticos, ou seja, o estudo da correlação entre a semântica do item lexical e a sua realização sintática. Segundo (FILLMORE; BAKER, 2010):

*The general purposes of the project are both to provide reliable descriptions of the syntactic and semantic combinatorial properties of each word in the lexicon, and to assemble information about alternative ways of expressing concepts in the same conceptual domain.*²

Segundo Ruppenhofer et al. (2010), o trabalho de anotação na FN consiste em o anotador se concentrar em uma *Unidade Lexical* específica e extrair sentenças, de diferentes *corpora*, que contenham aquela *Unidade Lexical*. Para um conjunto de sentenças escolhidas, então, é feita a anotação, em diversas camadas, sobre as informações sintático-semânticas. Um outro tipo de anotação, realizado em menor escala, é a anotação de texto corrido, em que, para uma mesma sentença, são feitas anotações sobre todas as *Unidades Lexicais* presentes na sentença.

3.4.1 *Frames*

A estrutura de um *Frame* na FN consiste de: nome do *Frame*, uma definição informal do *Frame*, os elementos que compõe o *Frame* (ou seja, o conjunto de papéis semânticos associados com o *Frame*) - chamados *Elementos de Frame* (*Frame Elements*) - e as *Unidades Lexicais* (*Lexical Units*) que evocam aquele *Frame*.

Uma distinção feita pela FN classifica os *Elementos de Frame* em **nucleares** (*core*) e **não-nucleares** (*non-core*). A diferença básica é que *Elementos de Frame* nucleares são necessários para o significado central do *Frame* (FILLMORE, 2008, p. 133), embora outros critérios como a obrigatoriedade de sua realização sintática ou sua realização em funções gramaticais básicas também sejam usados (ATKINS et al., 2003). Já os elementos não-nucleares são divididos em **periféricos** (*peripheral*) e **extratemáticos** (*extra-thematic*). *Elementos de Frame* periféricos são aqueles que não introduzem eventos adicionais, independentes ou distintos do evento principal reportado na sentença; geralmente marcam noções como tempo, lugar, maneira ou grau (RUPPENHOFER et al., 2010). Eles não caracterizam exclusivamente um *Frame* e podem ser instanciados em qualquer *Frame*, onde sejam semanticamente apropriados. *Elementos de Frame* extra-temáticos, por sua vez, são usados para anotar elementos que introduzem uma informação que não é necessariamente parte da descrição do *Frame* central (FILLMORE; BAKER, 2010). Em muitos casos, os sintagmas que representam este tipo de EF se referem a um outro *Frame*.

² "Os propósitos gerais do projeto são tanto fornecer descrições confiáveis das propriedades combinatoriais sintáticas e semânticas de cada palavra no léxico quanto agregar informações sobre maneiras alternativas de expressar os conceitos no mesmo domínio conceitual." (Tradução nossa)

Frames e *Elementos de Frame* podem ser associados a um *Tipo Semântico*. Atualmente esta é uma marcação genérica usada para indicar o tipo básico de preenchimento (*filler*) dos *Elementos de Frame*, ou para atuar como uma meta-descrição do *Frame* (por exemplo, para marcar o *Frame* como **não-lexical**) ou ainda para marcar dimensões da variação semântica entre *Unidades Lexicais* (por exemplo, se uma UL representa um julgamento positivo ou negativo, no contexto do *Frame*). Além dos tipos semânticos, é também possível registrar algum tipo de restrição entre os *Elementos de Frame* de um *Frame*, por exemplo, indicar que a ocorrência de um dado *Elemento de Frame* requer a ocorrência de outro Elemento ou que exclui a ocorrência de outro Elemento. A [Figura 3](#) apresenta um exemplo da estrutura de um *Frame* (neste caso definido para o Projeto Copa 2014 (SALOMÃO et al., 2011)).

Figura 3 – Estrutura do *Frame* CENÁRIO_DO_TURISMO_PARTIDA.

| Cenário_do_turismo_partida | |
|---|---|
| Definição [Definition]: | |
| O Turista sai da Origem com o objetivo de chegar ao seu Destino , que, geralmente, é o local onde reside. Para que isso aconteça, é necessário que ele finalize a estadia no alojamento escolhido. | |
| Muitos turistas SAEM do Rio de Janeiro pela Linha Vermelha. | |
| Tipo Semântico [Semantic Type]: Non-perspectivalized_frame | |
| Elementos de Frame [Frame Elements]: | |
| Nuclear [Core]: | |
| Meio de transporte [Mode_of_transportation] | O meio de transporte utilizado pelo Turista para deixar a Origem . Os ônibus PARTEM todos os dias pela manhã . |
| Origem [Source] | Localização do Turista antes de ele se deslocar. Semantic Type: Location |
| Turista [Tourist] | Indivíduo ou grupo que parte do local turístico. Semantic Type: Sentient |
| Periféricos [Non-Core]: | |
| Acompanhante [Co-participant] | Aquele que se retira da Origem juntamente com o Turista . Semantic Type: Sentient |
| Destino [Goal] | Localidade aonde o Turista pretende chegar. O ponto final do deslocamento. Semantic Type: Goal |
| Distância [Distance] | A distância do deslocamento até o Destino . |
| Finalidade [Purpose] | O propósito do Turista ao se deslocar da Origem . Semantic Type: State_of_affairs |
| Frequência [Frequency] | Quantas vezes um Turista ou um Meio de transporte se retiram da Origem . |
| Lugar [Place] | Localização a partir da qual se dá a partida do Turista quando esta não for especificada pela Origem . Semantic Type: Location |
| Maneira [Manner] | Detalhes da realização da partida no sentido de especificar o estado do Turista e/ou o modo como se deu a partida em si. Semantic Type: Manner |
| Meio [Means] | Ações ou meios que possibilitam a partida do Turista . |
| Tempo [Time] | O tempo em que ocorre a partida do Turista . Semantic Type: Time |
| Trajectoria [Path] | A rota na qual o deslocamento para acontece. Semantic Type: Path |
| Velocidade [Speed] | A velocidade com a qual o Turista se desloca. Semantic Type: Speed |
| Conjunto(s) de EF Nucleares [FE Core set(s)]: {Meio_de_transporte, Turista} | |
| Unidades Lexicais [Lexical Units]: deixar.v, partir.v | |

3.4.2 Relações entre *Frames*

Os *Frames* na FN não são um conjunto de entidades isoladas. Eles estão associados com outros *Frames* por um sistema que envolve centenas de relações *Frame-a-Frame*. Existem oito tipos de relações (incluindo uma chamada *Ver_também* (*See_also* - não discutida aqui). As demais são classificadas em três grupos: generalização, estrutura de eventos e sistemáticas (FILLMORE; BAKER, 2010):

a) Generalização

- **Herança (*inheritance*)**: é considerada a relação mais forte entre *Frames*. Segundo Ruppenhofer et al. (2010), através desta relação tudo que é estritamente verdade sobre a semântica do *Frame-pai* deve corresponder a um fato tão ou mais específico sobre o *Frame-filho*. Todos os *Elementos de Frame* do *Frame-pai* são relacionados a *Elementos de Frame* do *Frame-filho*, mas os EF do *Frame-filho* não precisam ter necessariamente o mesmo nome. É possível também que o *Frame-filho* tenha mais elementos que o *Frame-pai*. A FN também considera a possibilidade de herança múltipla.
- **Perspectiva (*Perspective_on*)**: Esta relação implementa a ideia de perfilamento e da relação figura/fundo. Itens lexicais diferentes (por exemplo, <comprar>, <vender>) evocam o mesmo *Frame* de um evento (transação comercial) mas com diferentes perspectivas.
- **Uso (*Using*)**: Esta é uma relação "genérica", que teve seu uso diminuído a partir da definição de relações mais específicas. Segundo Ruppenhofer et al. (2010), ela é usada quase que exclusivamente para os casos em que uma parte da cena evocada pelo *Frame-filho* se refere ao *Frame-pai*. Fillmore e Baker (2010) explicam que o *Frame-filho* depende do conhecimento em *background* provido pelo *Frame-pai*; pelo menos alguns dos *Elementos de Frame* nucleares do *Frame-pai* devem ser relacionados a *Elementos de Frame* do *Frame-filho*, mas não necessariamente todos eles.

b) Estrutura de Eventos

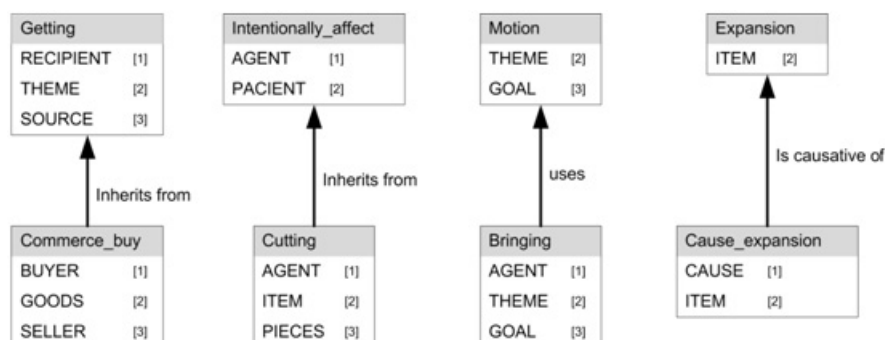
- **Subframe**: Esta relação representa o *Frame-filho* como um subevento de um evento mais complexo, o *Frame-pai*. Geralmente há uma ordenação temporal dos eventos (por exemplo, no *Frame* GIVING_SCENARIO da FN), mas não necessariamente (por exemplo, o *Frame* COMMERCIAL_TRANSACTION da FN).
- **Precedência (*Precedes*)**: basicamente, esta relação indica que há uma ordenação temporal entre os *Frames* (a ocorrência do *Frame-pai* precede a ocorrência do *Frame-filho*).

c) **Relações sistemáticas**

- **Causativo_de** (*Causative_of*): o *Frame-pai* representa o causativo referente ao *Frame-filho*.
- **Incoativo_de** (*Inchoative_of*): o *Frame-pai* representa o incoativo e o *Frame-filho* representa o estativo.

É importante ressaltar que todas as relações *Frame-a-Frame* são acompanhadas das respectivas relações entre *Elementos de Frame*, bem como são refletidas no sistema de tipos semânticos. Ainda que com exceções, isto significa que se um *Elemento de Frame* não tem um tipo semântico definido, ele pode herdar o tipo referente ao *Elemento de Frame* do *Frame-pai*, com o qual se relaciona. Próximo ao topo da hierarquia de *Frames*, os *Elementos de Frame* são similares aos papéis temáticos tradicionais (agente, tema, instrumento, etc) (BAKER, 2012). A Figura 4, adaptada de Osswald e Van Valin Jr (2014), mostra exemplos simples das relações *Frame-a-Frame* e EF-a-EF (os números entre colchetes mostram a correspondência entre os *Elementos de Frame*).

Figura 4 – Relações entre *Frame* e *Elementos de Frame*.



Fonte:Osswald e Van Valin Jr (2014)

3.4.3 Banco de dados

Como a discussão neste trabalho envolve a implementação computacional da FN para fins de PLN, é importante frisar que a informação exposta pela FN não é expressa no mesmo nível de profundidade existente em dicionários impressos ou com características presentes em outros léxicos computacionais (FILLMORE; JOHNSON; PETRUCK, 2003). Por exemplo, a Wordnet (FELLBAUM, 1998) é projetada para fornecer informação sobre relações paradigmáticas. A FN, por outro lado, é mais ou menos limitada a registrar informações sobre os requisitos combinatoriais das palavras estudadas (ATKINS et al., 2003). O banco de dados foi organizado para associar as palavras com os *Frames* que evocam e para registrar as relações entre os *Frames*, mas o trabalho da Wordnet não é duplicado.

Os dados da FN são armazenados em um banco de dados (BD) relacional MySQL database (BAKER; FILLMORE; CRONIN, 2003). A estrutura do BD tenta refletir as bases teóricas do projeto. Uma vez que diferentes tipos de informações são representadas no BD, ele pode ser visto como dividido em duas partes relacionadas: o BD lexical e o BD de anotações.

O **BD lexical** contém informações sobre o que é considerada a "estrutura" da FN: *Frames*, elementos de *Frame*, lemas, lexemas, unidades lexicais, forma de palavras (*word forms*) e partes do discurso. Assim, o BD lexical basicamente possui tudo que é necessário para caracterizar as unidades lexicais.

O **BD de anotações** contém as sentenças anotadas, juntamente com os subcorpora dos quais elas foram selecionadas para anotação. Ele apresenta o que são considerados os "dados" da FN. Para cada Unidade Lexical existem sentenças que exemplificam seu uso. Para cada sentença existe um conjunto de camadas de anotação para os *Elementos de Frame*, os Tipos Sintagmáticos e as Funções Gramaticais. O processo de anotação da FN não é discutido neste trabalho, mas é apresentado detalhadamente em Ruppenhofer et al. (2010).

3.4.4 Versão atual

A versão atual da FN, distribuída publicamente, é a Release 1.5. O banco de dados é dividido em um conjunto de três arquivos XML (para os *Frames*, unidades lexicais e as sentenças anotadas), juntamente com as definições dos XML Schemas e os scripts XSL para renderização dos relatórios. Segundo Baker (2012), em abril de 2012 o banco de dados continha 1.043 frames (116 não-lexicais) e 12.601 unidades lexicais. Os *Frames* possuem, no total, 10.077 elementos de *Frame*. Estavam anotadas 193.846 sentenças, com uma média de 20 sentenças por UL.

3.5 Causação e Dinâmica de Forças

Um aspecto importante para a definição dos *Frames* usados neste trabalho foi a consideração da chamada **alternância causativo-incoativo**. Considerando que a causação tem um papel importante no estudo do significado e que muitos *Frames* estão organizados em torno de eventos e suas relações causais, um estudo mais detalhado da causação é apresentado no Apêndice B. O estudo da alternância causativo-incoativo é feito na seção 9.1.1.

Por outro lado, a ideia de causação está, muitas vezes, associada à ideia do exercício de alguma força (ou forças). Para considerarmos estas questões na definição da estrutura da rede de *Frames*, nos baseamos na **Dinâmica de Forças**. A *Dinâmica de Forças* (apresentada em Talmy (2000) e detalhada no Apêndice C) está relacionada a como as entidades interagem com respeito à força, o que inclui o exercício da força, a resistência

a tal força, a superação de tal resistência, o bloqueio da expressão da força e a remoção de tal bloqueio. Assim, a *Dinâmica de Forças* compreende a expressão da força como a oposição entre entidades e os resultados de tais interações com o tempo, generalizando a noção tradicional de "causativo" (TALMY, 2000, p. 409).

3.6 Composicionalidade

Esta pesquisa investiga maneiras de melhorar a compreensão, por aplicações computacionais, de sentenças ou textos expressos em língua natural. É importante ressaltar que estamos usando termo "compreensão" no sentido aplicado às tarefas em PLN, ou seja, a identificação da situação, contexto ou domínio abordado pela sentença, com o objetivo de construir uma "interpretação", uma representação computacional daquela situação, contexto ou domínio. Através de uma análise semântica (e sintática, quando aplicável), a ideia é reconhecer "sobre o que" a sentença trata e não "o que" a sentença efetivamente significa, já que o significado, no seu sentido mais amplo, depende de um conhecimento de mundo, do contexto do discurso e de fatores pragmáticos que não são abordados neste trabalho. Porém, mesmo limitada, a "compreensão textual" por máquinas é útil em diversas aplicações de PLN, tais como sumarização de textos e dicionários eletrônicos, classificação de documentos, refinamento de pesquisas em corpus e realização de inferências lexicais.

A compreensão por parte da máquina envolve a criação de uma representação do significado da sentença. Um princípio básico que governa a interpretação de expressões linguísticas é o **Princípio da Composicionalidade** (CRUSE, 2010, p. 83). O que este princípio estabelece é que o significado de uma expressão semanticamente complexa é uma função composicional dos significados de seus constituintes semânticos. Segundo Cruse, esta definição se apoia em duas fortes pressuposições:

1. Uma língua tem um número infinito de sentenças gramaticais;
2. A linguagem tem um poder de expressão ilimitado, ou seja, qualquer coisa que possa ser concebida pode ser expressa via linguagem.

Não é difícil observar aqui alguns pressupostos da **Gramática Gerativa e Transformacional**, uma abordagem linguística que vem sofrendo várias elaborações e adaptações, desde a sua apresentação no final da década 1950, por Noam Chomsky. Ela é denominada *Gerativa* porque pretende fornecer uma descrição estrutural (finita) para todos os objetos linguísticos (infinitos) que são gramaticais dentro da língua natural sob exame; e é denominada *Transformacional*, porque concebe e descreve estruturas de superfície como resultado de transformações operadas em estruturas profundas. Ou seja, esta abordagem estuda as gerações e as transformações que ocorrem nas sentenças.

Em relação às estruturas das sentenças, a *Estrutura de Superfície* é aquela diretamente relacionada com a forma fonética do enunciado; sua representação é linear, possui um arranjo hierárquico dos elementos e pode levar à ambiguidade. Já a *Estrutura Profunda* não está representada diretamente na forma do signo, é abstrata e não possui ambiguidade. Estas estruturas (sintáticas) são representadas através de *Diagramas Arbóreos* (ou simplesmente, *árvores*).

Como o gerativismo privilegia a sintaxe, é natural a ideia de que o significado de uma expressão complexa pode ser predito pelo uso de regras gerais, aplicadas aos seus constituintes, porque ele é uma função do significado daqueles constituintes e da forma sintática associada a eles. Ou seja, a linguagem humana seria produtiva e sistemática. Produtiva porque o sistema de representação inclui um número infinito de unidades simbólicas diferentes, sintáticas e semânticas. Sistemática porque cada unidade simbólica do sistema de representação possui o mesmo valor em cada estrutura em que ele é usado.

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

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

Não obstante estas dificuldades, adotamos como pressuposto que o processo de composicionalidade é fundamental na busca do significado da sentença. Para obter este significado, um passo essencial é olhar além das palavras individualmente, reconhecendo que o significado de uma dada palavra tipicamente inclui conceitos que não são únicos para nenhum item linguístico (DODGE, 2010). Outro passo é o reconhecimento que o significado não está associado apenas com o lexema, mas também com os padrões construcionais onde estes lexemas ocorrem (como *Unidades Lexicais*). Embora não tratado diretamente nesta pesquisa, o estudo das construções gramaticais é importante, pois elas desempenham um papel preponderante na compreensão do enunciado.

A estratégia desenvolvida ao longo deste trabalho é o uso de *Frames*. Os *Frames* permitem simultaneamente localizar a palavra num contexto mais amplo de significado (expresso pela definição do *Frame*) e relacionar a palavra com outras partes da sentença,

através dos *Elementos de Frame*. Assumindo que o significado principal da sentença como um todo está associado com um *Frame* (dependendo de qual palavra é tomada como foco da análise - uma operação de *construal*), podemos identificar quais elementos do *Frame* estão sendo expressos pelas diferentes partes da sentença.

O uso da Semântica de Frames permite que nos afastemos da Hipótese Forte da Composicionalidade (HFC) que, como visto, considera obter o significado de uma expressão linguística através da soma do significado literal das suas partes constituintes, tentando excluir da constituição do significado todo e qualquer conhecimento de mundo ou efeito pragmático. Um modelo assim não admite ambiguidade, sinonímia, homonímia e vagueza. Já o uso de *Frames* nos possibilita o acesso a um parcial conhecimento de mundo e o entendimento de que a soma do significado das partes pode não ser igual ao significado do todo. A Semântica de Frames possibilita inferências explicadas (ou motivadas) pelo *Frame*, diante das quais a HFC nada tem a dizer. Associada a estas inferências, a integração de conceitos através de processos cognitivos, usando as relações entre *Frames*, nos possibilita uma abordagem mais ampla para a construção do significado.

Segundo Dodge (2010), a análise usando *Elementos de Frame* possui duas vantagens sobre o uso de papéis temáticos isolados. Uma vantagem se refere à profundidade do significado. Através do uso de diferentes *Frames* e *Elementos de Frame* para representar os significados das sentenças, podemos capturar significados mais específicos do que aqueles associados tipicamente aos papéis temáticos (como Agente e Paciente). Outra vantagem se refere à integração do significado. Tipicamente os papéis temáticos são definidos como entidades separadas, enquanto Comprador e Mercadorias (elementos do *Frame* COMÉRCIO_COMPRAR) estão claramente relacionados um com o outro.

A FrameNet, portanto, oferecendo um grande conjunto de *Frames* (e uma metodologia para criação de novos *Frames* quando necessário), se constitui um recurso valioso para o tipo de análise aqui proposta. No entanto, apesar de existirem *Frames* para verbos, nomes, adjetivos e preposições, historicamente uma atenção maior vem sendo dada ao verbos e ao estudo das valências verbais. Apesar do reconhecimento de que substantivos associados a artefatos e entidades naturais possuem uma estrutura de *Frame* mínima (RUPPENHOFER et al., 2010), estes *Frames* raramente dominam as orações em que eles ocorrem, e raramente são selecionados como objetos de anotação.

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

4 Aspectos Computacionais

Como visto na seção 3.4.3, os dados da FrameNet (versão 1.5) estão armazenados em um banco de dados relacional. Este armazenamento é útil para tarefas computacionais simples (como cadastro de estruturas da FrameNet, registro das anotações e emissão de relatórios). No entanto o modelo relacional é limitado para realização de tarefas computacionais mais complexas, como inferências lógicas e classificação dinâmica de dados, um processo geralmente conhecido por *reasoning* e executado por programas específicos, chamados *reasoners*.

Para este tipo de tarefa estamos usando ontologias. É importante ressaltar que o uso de lógica formal, via ontologias, não significa que estamos advogando o uso da Semântica Formal para representação do conhecimento de mundo. Nosso modelo conceitual, como discutido no capítulo anterior, é baseado na Semântica de Frames . As ontologias estão sendo usadas aqui para o que, em Ciência da Computação, é chamado **modelo lógico**, ou seja, a representação implementável computacionalmente do **modelo conceitual** usado. Dito de outra forma, as ontologias são usadas, neste caso, no seu viés meramente computacional e não como representação de classes do Universo de Discurso.

Isto implica dizer que, diferentemente de trabalhos anteriores associando FrameNet e Ontologias ((DOLBEY; ELLSWORTH; SCHEFFCZYK, 2006),(SCHEFFCZYK; PEASE; ELLSWORTH, 2006),(SCHEFFCZYK; BAKER; NARAYANAN, 2010)) , não se está buscando aqui transformar a FrameNet em uma ontologia (ou associá-la a uma Ontologia de Topo existente) a fim de representar fatos do mundo, mas sim criar uma ontologia para representar a estrutura da FrameNet. Por exemplo, as classes nesta ontologia são "*Frames*" e "*Elementos de Frame*" e não eventos (como Comprar ou Vender) ou entidades (como Pessoa ou Veículo). Também não é um objetivo específico formalizar as relações existentes na FrameNet, como proposto em Ovchinnikova (2012) e Moreira e Salomao (2012), embora isto possa ser um subproduto deste trabalho.

O que se busca é a criação de uma camada, representada formalmente, que permita inferências lógicas, que atue sobre as informações já presentes na FrameNet e que acrescente outras, que sejam úteis para o processo de desambiguação, que é o objetivo deste trabalho. Estes acréscimos incluem, por exemplo, a representação de *Frames* como estruturas de *features* e a criação de referências que permitam uma integração com as estruturas propostas pela TLG.

A realização destas tarefas implica o uso de diversas teorias e técnicas. As seções seguintes apresentam o conceito e o uso de ontologias em computação, discutem a integração entre Ontologias e Léxico (a fim de mostrar onde este trabalho se situa neste campo),

fazem uma breve revisão sobre a Lógica Descritiva (DL – *Description Logics*), a linguagem OWL e a Ontologia SIMPLE, e finalmente apresentam o Modelo Conexcionista Estruturado, que é o paradigma usado na implementação deste trabalho.

4.1 Processamento de Línguas Naturais (PLN)

Por usar um léxico computacional (a FrameNet), este trabalho se enquadra no campo do PLN (Processamento de Línguas Naturais). O PLN é um domínio de estudo que visa a modelagem computacional da linguagem humana, ou seja, a criação de um modelo computacionalmente tratável do uso do léxico e da gramática de uma língua natural, nas diversas situações de comunicação (DIAS-DA-SILVA, 2006). O PLN é geralmente dividido em seis grandes áreas (ROSA, 2011, p. 187):

- a) Interfaces em língua natural para bases de dados.
- b) Tradução de uma língua natural para outra usando máquinas.
- c) Programas de indexação para sumarização de grandes volumes de texto.
- d) Geração de texto para produção de documentos padronizados.
- e) Sistemas de interação via voz com os computadores.
- f) Aplicações específicas.

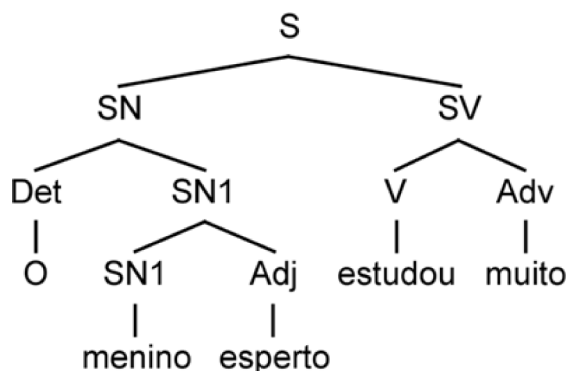
Em todas estas áreas, um problema predominante é a transformação de um texto de entrada, potencialmente ambíguo, em uma forma não ambígua que possa ser manipulada internamente em um sistema computacional para atender os objetivos de uma dada aplicação. Esta transformação é conhecida como *parsing* (análise). Em PLN, *parsing* é geralmente um processo de combinar os símbolos de um texto em um grupo que pode ser substituído por outro símbolo mais geral. Este novo símbolo pode, por sua vez, ser combinado em um outro grupo, e assim por diante, até que surja uma estrutura tratável computacionalmente.

Rosa (2011, p. 145) apresenta cinco tipos diferentes de *parsers* (analisadores): casamento de padrões, baseado em gramática, semântico, baseado em conhecimento e analisadores por redes neurais. Cada tipo possui uma abordagem única no contexto do PLN, discutida a seguir.

O **casamento de padrões** (*pattern matching*) se apoia na ideia de que os *parsers* poderiam procurar padrões linguísticos nas sentenças sem usar nenhum formalismo gramatical explícito. O sistema apenas procura por um possível casamento com um número fixo, pré-determinado, de padrões. Se um padrão é encontrado, o sistema executa determinada ação. Um dos programas de casamento de padrões mais conhecido é ELIZA, projetado para simular um diálogo com um psicólogo.

A **abordagem baseada em gramática** pressupõe um conjunto de regras sintáticas que descrevem uma sentença. O objetivo destas regras é construir uma estrutura em árvore que mostra (pretensamente sem ambiguidade) como as palavras interagem em uma sentença. A [Figura 5](#) mostra um exemplo de uma estrutura em árvore.

Figura 5 – Abordagem baseada em gramática.



Fonte: Adaptado de [Rosa \(2011\)](#)

Já a **abordagem semântica** não desconsidera a análise sintática, mas busca considerar também o significado de cada estrutura. Uma opção é usar a Gramática de Casos ([FILLMORE, 1968](#)), com a definição dos papéis de agente, instrumento, etc. Outra possibilidade é a definição de classes semânticas específicas; por exemplo, ao invés de usar classes como verbos ou adjetivos, são usadas classes como pessoas, artefato, veículos, etc. Uma alternativa mais recente é o uso da FrameNet e outros léxicos computacionais, bem como da Gramática das Construções ([HOFFMANN; TROUSDALE, 2013](#)).

Os ***parsers* baseados em conhecimento** consideram que a linguagem não pode ser reduzida somente a várias regras de reescrita. Assim, eles tem acesso a uma base de conhecimento sobre o domínio específico que está sendo tratado pela aplicação. Este conhecimento considera que a interpretação de uma palavra é dependente do contexto em que ela é usada. Em outros casos, são usados os chamados **primitivos semânticos**, que buscam uma representação canônica de uma sentença, possibilitando que sentenças diferentes, que tenham um mesmo significado, tenham também uma mesma representação. Por exemplo, as sentenças em (4.1) teriam a mesma representação canônica mostrada em (4.2).

- (4.1) (a) Alfredo come doce.
 (b) O doce foi comido por Alfredo.

- (4.2) Alfredo ↔ COMER ↔ doce.

As **redes neurais**, baseadas em um modelo conexionista, buscam se aproximar do modelo humano de processamento de informação linguística. A ideia básica é que unidades individuais não transmitem grande quantidade de dados, mas o processamento é possível porque um grande número de unidades similares são conectadas entre si. Um modelo de análise por rede neural pode, por exemplo, conter três níveis de neurônios (ou três camadas). O primeiro nível é lexical e serve como o nível de entrada da rede. Os neurônios são mapeados em determinadas palavras. No segundo nível, as entradas são combinadas para ativar neurônios que representam o significado das palavras. No terceiro nível, os significados são combinados para formar predicados e objetos.

Na maioria dos sistemas de *parsing* atuais, a análise do texto é realizada de forma modular (OVCHINNIKOVA, 2012, p. 74). O texto é processado sequencialmente por um *tokenizer* (que divide o texto em *tokens*), por um *tagger* morfológico (que cria uma etiqueta para cada *token*), por um componente sintático e, finalmente, por um componente semântico. Cada componente usa a saída do componente prévio como entrada e gera sua própria saída, enriquecida com mais informação linguística. O *parser* semântico é geralmente o último estágio do processamento do texto. Como é possível perceber, o que se busca com todo este processamento, até chegar ao nível semântico, é uma **interpretação** (representada formalmente) do texto sob análise. Este é um ramo específico dentro da PLN, chamado NLU (*Natural Language Understanding*), ou Compreensão de Língua Natural (CLN).

Dentro do campo interdisciplinar da Linguística Computacional, a Compreensão de Língua Natural tem por objetivo construir sistemas capazes de compreender e produzir expressões em língua natural, para tornar os computadores mais facilmente usáveis. Mais especificamente, os aspectos tecnológicos são definidos em áreas como o desenvolvimento de representações gramaticais apropriadas, técnicas de *parsing* e mecanismos para representação do conhecimento. Incidentalmente, esta é uma área de interesse também para os linguistas teóricos e para psicolinguistas. Aplicações para compreensão de língua natural incluem a recuperação de informação (*information retrieval*), sistemas de pergunta-resposta, sistemas de recomendação e sistemas de interfaces com bancos de dados.

A interpretação (ou a representação da interpretação) obtida através do *parser* semântico é essencialmente baseada na teoria semântica usada pelo *parser* e, portanto, totalmente dependente desta teoria. A maneira como o significado de cada item lexical na sentença é considerado, a influência ou contribuição deste significado para o significado da sentença e a desambiguação de itens que possuem mais de um sentido constituem aspectos relevantes na criação da interpretação.

Além disto, também deve ser considerado o processo de inferência lexical, ou seja, o mecanismo através do qual o conhecimento do significado de um ou mais itens na sentença pode levar a compreensão do contexto no qual a sentença está inserida ou revelar um

conhecimento que não está explícito. Como exemplo, é possível inferir que uma sentença que apresente as palavras <criança>, <bolo> e <presente> muito provavelmente se refere a uma festa de aniversário, mesmo que as palavras <festa> ou <aniversário> não estejam presentes. De forma semelhante, na sentença (4.3) conhecendo-se que a função de escritor é tipicamente escrever livros e que livros podem ser vendidos, pode-se inferir que o escritor escreveu um livro que foi muito vendido, mesmo que a palavra <livro> não esteja presente na sentença.

(4.3) Aquele escritor vendeu muito.

4.2 Ontologias

Historicamente o termo ontologia tem origem no grego *ontos*, ser e *logos*, estudo. É um termo introduzido na filosofia com o objetivo de distinguir o estudo do ser humano como tal, do estudo de outros seres das ciências naturais. A origem é a palavra aristotélica "categoria", que pode ser usada para classificar e caracterizar alguma coisa.

Na área de Ciência da Informação, duas definições para ontologia são as mais famosas: "uma ontologia é uma especificação explícita de uma conceptualização" (GRUBER, 1993, p. 1); e uma elaboração desta, a partir das idéias em Gruber (1994) e Borst (1997): "uma ontologia define uma especificação formal e explícita dos termos de um domínio e das relações entre eles, de uma conceptualização compartilhada". Nesta definição **formal** significa legível para computadores; **especificação explícita** diz respeito a conceitos, propriedades, relações, funções, restrições, axiomas que são explicitamente definidos; **compartilhado** quer dizer conhecimento consensual; e, **conceptualização** diz respeito a um modelo abstrato de algum fenômeno do mundo real.

Outra definição, apresentada por Guarino (1998) desenvolve o aspecto formal intensional da noção de ontologia:

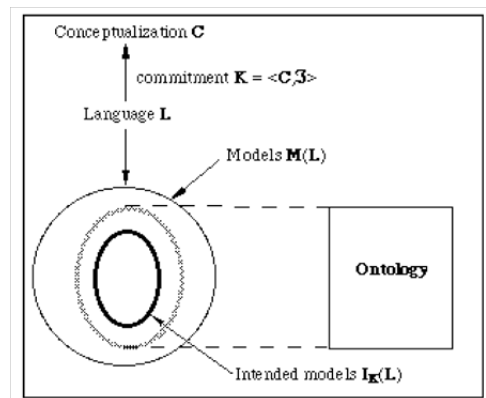
An ontology is a logical theory accounting for the intended meaning of a formal vocabulary, i.e. its ontological commitment to a particular conceptualization of the world. The intended models of a logical language using such a vocabulary are constrained by its ontological commitment. An ontology indirectly reflects this commitment (and the underlying conceptualization) by approximating these intended models.¹

Segundo Ovchinnikova (2012), para Guarino uma **conceptualização** é uma visão de mundo independente de linguagem (um conjunto de relações conceituais definidas em

¹ "Uma ontologia é uma teoria lógica considerada para o significado pretendido de um vocabulário formal, ou seja, seu compromisso ontológico comum a conceptualização específica do mundo. Os modelos pretendidos de uma linguagem lógica usando tal vocabulário são restringidos por seu compromisso ontológico. Uma ontologia reflete, indiretamente, este compromisso (e a conceptualização subjacente) através da aproximação destes modelos pretendidos." (Tradução nossa)

um espaço de domínio), enquanto uma **ontologia** é um artefato cognitivo dependente de linguagem, que é associado a uma certa conceptualização do mundo através de uma dada linguagem (entendida esta como uma linguagem de representação, não uma língua natural). As relações entre a linguagem (ou vocabulário), a conceptualização, o compromisso ontológico e a ontologia são ilustrados na [Figura 6](#)².

Figura 6 – Modelo para ontologia.



Fonte: [Guarino \(1998\)](#)

Uma ontologia provê um mecanismo para capturar a compreensão comum sobre objetos e seus relacionamentos, em um certo domínio de interesse, oferecendo um modelo formal e manipulável desse domínio. Através de axiomas, é feita a especificação formal do significado dos termos utilizados, o que possibilita a criação de novos termos através da combinação dos já existentes, bem como permite a integração com outras ontologias. Por exemplo, a expressão <Ilha no Pacífico> pode ser definida como em (4.5).

$$(4.5) \quad \forall x(ILHANOPACIFICO(x)) \\ \rightarrow ILHA(x) \wedge \exists y(LOCALIZADA_EM(y) \wedge OCEANO_PACIFICO(y))$$

O estudo e o uso de ontologias na área de software foram popularizados com a ideia da **web semântica** introduzida em Berners-Lee et al. (2001). Nesse artigo os autores apresentam a visão da **web semântica** como uma extensão da web atual, na qual a informação recebe um significado bem definido, permitindo o trabalho cooperativo de computadores e pessoas. Neste contexto, o tipo mais comum de ontologia consiste de uma taxonomia e um conjunto de regras de inferência, que permitem capturar o conhecimento que não está explícito na taxonomia. Na verdade, as ontologias podem (e geralmente vão) estender os relacionamentos hierárquicos das taxonomias, permitindo relacionamentos horizontais entre os termos, em uma estrutura tipo grafo. Desta forma, as ontologias facilitam a modelagem de requisitos de informação típicos do mundo real.

² "O modelo pretendido de uma linguagem lógica reflete o seu compromisso com uma conceptualização" ([GUARINO, 1998](#))

Uma importante propriedade das ontologias é que sua representação permite seu processamento computacional. Sendo representadas em linguagens lógicas, as ontologias permitem a definição formal da semântica dos conceitos. O quanto pode ser expresso na ontologia vai depender do poder de expressividade da linguagem lógica usada na representação daquela ontologia, ou seja, o conjunto de construtos que a linguagem oferece. Esta expressividade pode ser ampliada com o uso de **máquinas de inferência** (*reasoners*). A capacidade de inferência está relacionada à decidibilidade da lógica subjacente à ontologia, o que vai impactar na capacidade de fazer inferências corretas. Por exemplo, a Lógica de Primeira Ordem (FOL - *First Order Logic*), e suas extensões, não são decidíveis.

Apesar de existirem diferentes formalismos para representação do conhecimento para implementação de ontologias, cada um com componentes diferentes, um conjunto mínimo é comum a todos:

- a) **Classes**: representam conceitos do domínio; são definidas por termos geralmente organizados em uma taxonomia;
- b) **Relações**: representam um tipo de associação entre as classes. Geralmente as associações são binárias, com o primeiro argumento sendo chamado domínio (*domain*) da relação e o segundo argumento sendo o escopo (*range*). As relações são instanciadas segundo o conhecimento disponível sobre o domínio. As relações binárias são também usadas para expressar atributos ou propriedades das classes. Atributos são diferentes de relações, porque eles estão limitados a um tipo de dados (e não a uma classe).
- c) **Instâncias**: representam elementos ou indivíduos em uma ontologia.

Em Ciência da Informação também é útil a distinção entre **Ontologias Fundacionais** (ou de Topo) e **Ontologias de Domínio**. As primeiras modelam conceitos gerais, tais como espaço, tempo, objeto, evento, etc. que são aplicáveis na maioria dos domínios. As Ontologias de Domínio, como o nome indica, modelam um domínio específico, por exemplo, Geografia ou Genética, e representam os significados dos termos próprios daquele domínio (tais como <bacia> ou <cromossomo>).

4.3 Ontologias e Léxico

Uma vez que este trabalho propõe a integração de um recurso lexical (a FrameNet) com uma ontologia (a Ontologia SIMPLE), a discussão sobre a interface entre léxico e ontologias - denominada *Interface OntoLex*, em Huang et al. (2010) - torna-se relevante. O resumo de Hirst (2009) continua sendo um bom ponto de partida para a discussão sobre OntoLex como um todo :

"A lexicon is a linguistic object and hence is not the same thing as an ontology, which is non-linguistic. Nonetheless, word senses are in many ways similar to ontological concepts and the relationships found between word senses resemble the relationships found between concepts."³

Assim, embora haja um certo consenso de que os léxicos não são realmente ontologias, tanto as diferenças quanto as semelhanças entre os dois tipos de recursos mantêm em andamento a investigação sobre uma possível interface entre eles.

No campo das diferenças, pode-se citar, por exemplo, que enquanto a sinonímia e a quase-sinonímia são relações muito importantes para os léxicos semânticos, não há espaço para elas nas ontologias formais, onde os conceitos não podem ser ambíguos e onde termos sinônimos são agrupados sob o mesmo conceito. Outro exemplo é a informação sobre como as palavras são usadas (seu registro) que é oferecida pelos léxicos e que não é relevante para as ontologias tradicionais. Afinal, os recursos linguísticos, como os léxicos, são constituídos de expressões linguísticas e não dos conceitos subjacentes, enquanto ontologias linguísticas contêm estes conceitos. É fácil perceber que o tipo de conhecimento que estes recursos tentam capturar são de naturezas muito diferentes.

No campo das semelhanças, embora um léxico seja sobre palavras e ontologias sejam sobre conceitos, ambos representam uma conceptualização compartilhada, da perspectiva da convencionalização (HUANG et al., 2010). Para aplicações referentes à tecnologia da linguagem humana, um léxico estabelece uma interface entre agentes humanos e o conhecimento. Para aplicações na Web Semântica, uma ontologia possibilita à máquina o processamento do conhecimento diretamente.

Assim, ontologias e recursos lexicais estão associados, historicamente, a tipos diferentes de aplicações e só recentemente começaram a ser considerados simultaneamente. Do ponto de vista ontológico, os blocos de construção básicos de ontologias são **conceitos** e **relações**. Identificar estes objetos e decidir sobre sua natureza é uma tarefa fundamental na análise ontológica. Algo similar ocorre nos recursos lexicais, referente a **termos** e **relações**, uma vez que supõe-se que as relações contribuem de maneira significativa para o significado das entradas lexicais. Conceitos (ou palavras) e relações são, assim, os dois primeiros objetos a serem considerados, quando se pensa em um trabalho que reúna ontologias e recursos lexicais.

A interface OntoLex possui também uma grande importância quando se pensa no projeto de recursos multilinguais. Estes recursos, via de regra, são constituídos de vários recursos monolíngues mapeados para uma interlíngua. A possibilidade de que esta interlíngua seja estruturada, correspondendo a um nível conceitual e, portanto,

³ "Um léxico é um objeto linguístico e, assim, não é a mesma coisa que uma ontologia, que é não-linguística. Apesar disto, os sentidos das palavras são, de várias maneiras, semelhantes a conceitos ontológicos e os relacionamentos encontrados entre os sentidos das palavras lembram os relacionamentos encontrados entre os conceitos." (Tradução nossa)

representável via ontologias, mostra o quanto uma interface entre léxico e ontologias pode ser útil.

Porém, esta interface entre recursos não é trivial. Considere-se, por exemplo, os dois objetos básicos (conceitos e termos, de um lado e relações, do outro). A diferença entre um léxico baseado em termos e um léxico baseado em conceitos é bastante clara. Mas um léxico baseado em sentidos embaça a distinção. Neste tipo de léxico (como a FrameNet ou a Wordnet) os nós do recurso lexical (por exemplo, as **unidades lexicais** ou os *synsets*) nem são simples termos, nem são entidades puramente conceituais: estão associados a sentidos de um lexema que correspondem ao seu uso convencionalizado, possivelmente vindo de exemplos atestados em corpus. Estes recursos podem ser considerados léxicos, uma vez que suas entradas são expressões linguísticas, mas tem sido muitas vezes usados como ontologias. Claramente um nível intermediário parece ser necessário, sendo discutido com detalhes em [Huang et al. \(2010\)](#).

Com as relações, a questão também não é simples. Nas ontologias os conceitos são integrados em um todo coerente com relações. Estas relações são conceitualmente dirigidas e tomam os conceitos como argumentos. Por outro lado, os recursos lexicais estão preocupados com a organização dos itens lexicalizados. As relações usadas por eles não estão, geralmente, focadas nos conceitos. Além disso, relações com o mesmo nome em ontologias formais e linguísticas parecem ser muito diferentes, quando analisadas com mais cuidado. No entanto, deve-se ressaltar que a classificação destas relações em paradigmáticas ou sintagmáticas é comum tanto nas abordagens conceituais quanto lexicais.

Relações paradigmáticas existem entre elementos da mesma natureza que pertencem a um paradigma comum. No domínio lexical são bem conhecidas as relações de sinonímia, antonímia, meronímia, hiperonímia e hiponímia. Em ontologias, as relações de oposição conceitual, parte-de, é-um-tipo-de são formalmente definidas. Como estas relações estão associadas entre si ainda é tema de investigação, uma vez que o uso de recursos lexicais como ontologias (ou vice-versa) pode gerar uma interpretação errônea das relações.

Relações sintagmáticas existem entre entidades de naturezas diferentes. Os itens relacionados por estas relações co-ocorrem frequentemente, mas não podem ser substituídos por um outro. Geralmente estas relações são lexicalizadas por palavras que pertencem a categorias sintáticas diferentes. Na Semântica Lexical, relações sintagmáticas são associadas a estudos da interface sintaxe-semântica, com foco na predicação e papéis temáticos, e constituem a maioria das relações na FrameNet.

Outro ponto importante na discussão da interface OntoLex está no próprio processo de categorização (ou classificação). Uma abordagem de cunho mais cognitivo, como a adotada neste trabalho, vê a associação entre palavras e conceitos como uma condição de gradiência (*graded membership*). Assim, a determinação dos limites de uma categoria (ou

a própria existência destes limites) é um problema relevante, já que são diversos fatores que podem influenciar esta limitação. A definição de uma lista de características (*feature list*), que é a solução geralmente adotada nas implementações computacionais, é tida como uma solução simplista demais e levanta a questão sobre como as próprias características devem ser definidas.

Finalmente, [Huang et al. \(2010\)](#) mostram também que, apesar da realização de estudos com base filosófica sobre a linguagem e os sistemas de língua natural, em busca de uma ontologia que possa representar a linguagem, estas pesquisas tem tido pouco impacto direto no desenvolvimento de recursos práticos. A exceção apontada é a TLG (discutida na [seção 3.2](#) deste texto). Segundo aqueles autores ([HUANG et al., 2010](#), p. 20), a contribuição de Pustejovsky é combinar uma base filosófica com uma teoria semântica-lexical rica, apresentando uma abordagem sofisticada para a predicação. Ao invés de trabalhar com a enumeração de sentidos e tentar uma longa lista de palavras, a TLG argumenta que os sentidos das palavras são infinitos, uma vez que os falantes podem facilmente criar novos sentidos. Por outro lado, a visão multidimensional do significado, que é as vezes capturada por uma herança múltipla massiva é resolvida pela TLG através de uma herança ortogonal cuidadosa.

4.4 Lógica Descritiva e OWL

A razão preponderante para propormos o uso de ontologias no contexto deste trabalho é a possibilidade da especificação formal das relações semânticas e a obtenção de informações adicionais através do processo de inferência. Embora esta seção não tenha por objetivo uma descrição de caráter técnico a respeito da notação e terminologia das Lógicas Descritivas, alguns conceitos importantes são apresentados a seguir, adaptados de [Kokar, Matheus e Baclawski \(2009\)](#).

Em Matemática, uma relação é um subconjunto do produto cartesiano de um certo número de conjuntos. Por exemplo, o produto cartesiano de dois conjuntos A e B é o conjunto de todos os pares ordenados (a,b) , formalizados em (4.8).

$$(4.8) \quad A \times B = \{(a, b) \mid a \in A, b \in B\}$$

Uma relação R é então um subconjunto do produto cartesiano, $R = A \times B$. Uma relação pode ser especificada tanto extensionalmente quanto intensionalmente. Uma especificação extensional de uma relação é dada listando explicitamente todas as tuplas da relação. Uma especificação intensional de uma relação R é dada através de um predicado P , onde R contém todas as tuplas r para as quais o predicado P é verdadeiro. A definição formal é apresentada em (4.9).

$$(4.9) \quad R = \{(a, b) \mid P(a, b)\}$$

A questão é: como saber que o predicado P é verdadeiro para um dado par (a, b) ? É quando entra em cena o poder da inferência. Predicados são os componentes principais de **sentenças**, que por sua vez são partes de **teorias lógicas**. Relações, por outro lado, são partes de **modelos**, ou seja, interpretações de sentenças. O processo de inferência (ou *reasoning*) é executado dentro de uma teoria específica. A inferência baseada em computador é puramente sintática: o mecanismo de inferência trata "fatos" que estão armazenados como strings de texto. Ele tenta combinar **regras de inferência** com padrões na base de "fatos" e "deriva" novos fatos de acordo com a regra de inferência cujo padrão combinou. Para que isto seja possível, é preciso:

1. Uma linguagem formal em que todos os fatos usados no processo de inferência estejam expressos.
2. Uma especificação formal do processo de inferência.

Uma linguagem formal é dada por uma **gramática** e uma noção de **interpretação**. Uma gramática é dada por uma quantidade de **regras** para a construção de sentenças compostas a partir de sentenças elementares. Uma **interpretação** é uma função que mapeia todos os elementos da linguagem formal para uma estrutura relacional, chamada **modelo**. Especificamente, uma interpretação mapeia cada predicado para uma relação. Uma sentença é dita verdadeira para uma interpretação se sua relação correspondente existe no modelo. Para um conjunto de sentenças A , um modelo de A é qualquer estrutura relacional para qual existe cada sentença de A .

A especificação do processo de inferência é dada pela noção de acarretamento (*entailment*). Um conjunto de sentenças A acarreta uma sentença s se e somente se, para cada interpretação de A , sempre que todas as sentenças de A forem verdadeiras, a sentença s também for verdadeira.

Neste contexto, uma linguagem formal (na realidade, uma família de linguagens formais) usada para representar conhecimento formalmente é a Lógica Descritiva (ou de Descrição). As **Lógicas Descritivas** (DL - *Description Logics*), constituem uma família de formalismos que representam um subconjunto decidível da Lógica de Primeira Ordem (NARDI; BRACHMAN, 2003). Segundo LUTZ (2006), elas são consideradas o formalismo mais importante para representação de conhecimento, unificando e fornecendo uma base lógica para os sistemas tradicionais nesta área (*frames*, redes semânticas, representações orientadas a objeto, modelos semânticos de dados e sistemas de tipos). As Lógicas Descritivas são utilizadas em projetos de sistemas que possuem uma linguagem para definir uma base de conhecimento (KB - *Knowledge Base*) e ferramentas para realizar inferências sobre a base definida.

Com o uso de Lógicas Descritivas a representação do conhecimento é feita através de uma abordagem funcional, ou seja, são fornecidas especificações precisas das funcionalidades a serem providas pela base de conhecimento e das inferências a serem realizadas. Dentro da base de conhecimento pode ser feita uma clara distinção entre **conhecimento intensional** (o conhecimento geral sobre o domínio do problema ou Universo do Discurso) e **conhecimento extensional** (conhecimento específico em um problema em particular).

As Lógicas Descritivas descrevem o domínio em termos de conceitos (classes), papéis (propriedades, relacionamentos) e indivíduos. Em termos de lógica, os nomes dos conceitos são equivalentes a predicados unários - em geral, fórmulas com uma variável livre. Os nomes dos papéis são equivalentes a predicados binários - em geral, fórmulas com duas variáveis livres. Os nomes dos indivíduos são equivalentes a constantes. O conjunto de operadores é restrito, para que a linguagem seja decidível e com baixa complexidade.

Uma ontologia expressa em DL, então, consiste de um *Terminological Box* (TBox) e um *Assertion Box* (ABox). O **TBox** contém o conhecimento intensional, na forma de uma terminologia ou taxonomia, sendo construído através de declarações que descrevem os conceitos gerais e suas propriedades. Um Tbox T é um conjunto de axiomas na forma $C \subseteq D$ ou $C \equiv D$, onde C e D são descrições de conceitos. Exemplos da definição de conceitos são apresentados em (4.10).

$$(4.10) \quad \begin{aligned} \textit{Estudante} &\equiv \textit{Pessoa} \cap \exists \textit{EstudaEm.Universidade} \\ \textit{Doutorando} &\subseteq \textit{Estudante} \cap \textit{Pesquisador} \end{aligned}$$

O ABox contém o conhecimento extensional que é específico para os indivíduos do domínio. Um Abox A é um conjunto finito de asserções na forma $C(a)$ (asserções de conceitos) ou $R(a, b)$ (asserções de papéis), onde C é uma descrição de conceito, R é um nome de papel e a, b são indivíduos. Exemplos de asserções são apresentados em (4.11).

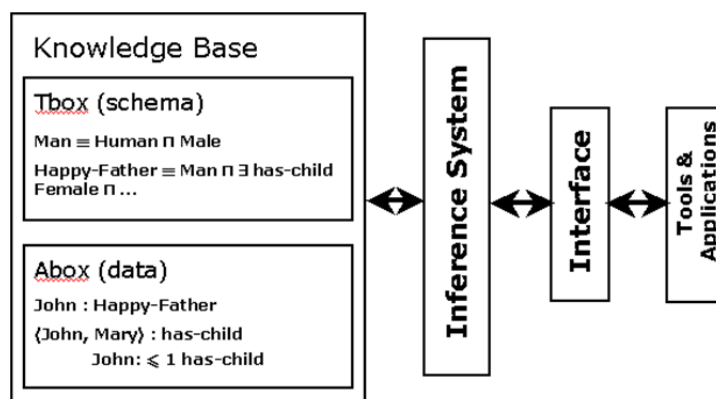
$$(4.11) \quad \begin{aligned} \textit{Doutorando}(\textit{Carlos}) \\ \textit{EstudaEm}(\textit{Carlos}, \textit{UFJF}) \end{aligned}$$

A [Figura 7](#) mostra a arquitetura típica de um sistema baseado em Lógicas Descritivas.

Um aspecto importante no uso de Lógicas Descritivas são os mecanismos de inferência (*DL Reasoners*). Através destes mecanismos, não apenas a base de conhecimento pode ser validada em relação a sua consistência, mas também o conhecimento implícito pode ser explicitado a partir do conhecimento expresso na base. Estes dois aspectos podem ser reduzidos às seguintes tarefas de inferência ([RUSSELL; NORVIG, 2010](#), p. 456):

1. **satisfatibilidade** (*satisfiability*): um conceito C pode ser satisfeito com respeito a um Tbox T se existe pelo menos um modelo de T onde a interpretação de C , C' , é

Figura 7 – Arquitetura de um sistema DL.



Fonte: [HORROCKS \(2002\)](#)

não-vazia. Esta tarefa é também denominada classificação, ou seja, a verificação se um objeto pertence a uma classe.

2. **subsunção** (*subsumption*): um conceito C é subordinado a um conceito D , com respeito a T , se e somente se $C' \subseteq D'$ para cada modelo de T . Ou seja, a verificação se uma classe é subconjunto de outra é feita pela comparação de suas definições.

A inferência na maioria das Lógicas Descritivas é decidível e existem algoritmos otimizados para algumas lógicas. Várias implementações destes algoritmos têm sido usadas com as linguagens de ontologia, tais como FACT++, Racer e Pellet. A linguagem de ontologia usada neste trabalho é a OWL2.

OWL2 Web Ontology Language, informalmente, **OWL2** ([W3C, 2012](#)), é a linguagem de ontologia padronizada pelo W3C⁴, uma versão ampliada da linguagem OWL. As ontologias OWL2 proveem classes, propriedades, indivíduos e valores de dados, e são armazenadas como documentos da Web Semântica.

As noções básicas para se compreender como o conhecimento é representado em OWL2 são:

- a) **Axiomas**: as declarações básicas que uma ontologia OWL2 expressa.
- b) **Entidades**: elementos usados para referenciar objetos do mundo real.
- c) **Expressões**: combinações de entidades para formar descrições mais complexas a partir das mais básicas.

A OWL2 pode ser considerada uma linguagem poderosa para modelagem de propósito geral de certas partes do conhecimento, mas o tipo de "conhecimento" que pode

⁴ W3C (*World Wide Web Consortium*) é a principal organização de padronização da World Wide Web, congregando atualmente quase 400 membros.

ser representado pela OWL2 obviamente não reflete todos os aspectos do conhecimento humano.

Para se formular explicitamente o conhecimento, é útil assumir que ele consiste de partes elementares, geralmente chamadas de **declarações** (*statements*) ou **proposições**. Declarações como <está chovendo> ou <todos os homens são mortais> são exemplos típicos de proposições básicas. As declarações feitas em OWL2 são chamadas **axiomas**, e uma ontologia pode ser vista como uma coleção de axiomas. A ontologia declara que todos os seus axiomas são verdadeiros.

Uma declaração X pode ser consequência de outra declaração Y. Essencialmente isso significa que sempre que a declaração Y for verdadeira, a declaração X também o será. Em termos de OWL2, diz-se que um conjunto de declarações A acarreta (*entails*) uma declaração a, se em qualquer situação onde todas as declarações de A são verdadeiras, a também for verdadeira. Além disso, um conjunto de declarações pode ser consistente (ou seja, há um estado de coisas possível em que todas as declarações no conjunto são verdadeiras) ou inconsistente (não há tal estado de coisas). A semântica formal da OWL2 especifica quais são os possíveis "estados de coisa" onde um conjunto específico de declarações é verdadeiro.

Raramente as declarações em OWL2 são monolíticas; frequentemente elas possuem uma estrutura interna que pode ser explicitamente representada. Normalmente estas declarações se referem a objetos do mundo e os descrevem, por exemplo, colocando-os em categorias (como <Maria é mulher>) ou dizendo algo sobre suas relações (<Carlos e Maria são casados>). Os constituintes atômicos das declarações, sejam objetos (Carlos, Maria), sejam categorias (Mulher) ou relações (Casado) são chamados **entidades**. Em OWL2, os objetos são **indivíduos**, as categorias são **classes** e as relações são **propriedades**. Propriedades em OWL2 são subdivididas em propriedades de objeto (relacionando objetos com objetos), propriedades de tipos de dados (associando valores a objetos) e propriedades de anotação (codificando informação sobre partes da própria ontologia - e não do domínio modelado).

Como uma característica fundamental da OWL2, nomes de entidades podem ser combinados em expressões, usando os chamados **construtores**. Como um exemplo básico (reproduzindo o exemplo apresentado na discussão sobre DL), as classes <Estudante> e <Pesquisador> podem ser combinadas conjuntivamente para descrever os estudantes de doutorado. Esta classe <Doutorando> pode ser usada em declarações ou em outras expressões. Neste sentido, as expressões podem ser vistas como novas entidades, definidas com base na sua estrutura. A linguagem OWL2 apresenta uma grande variedade de construtores.

4.5 Ontologia SIMPLE

Dentro da proposta de melhorar o processo de desambiguação utilizando a Teoria do Léxico Gerativo, neste trabalho estamos usando a ontologia SIMPLE como um recurso auxiliar complementar. SIMPLE (LENCI et al., 2000) é um projeto de larga escala, patrocinado pela União Europeia, com o objetivo de desenvolver um léxico semântico computacional, de ampla cobertura e múltiplos propósitos, para doze línguas da Europa (catalão, dinamarquês, holandês, inglês, finlandês, francês, alemão, grego, italiano, português, espanhol e sueco). A ontologia SIMPLE foi projetada neste contexto e desenvolvida para garantir a uniformidade e consistência entre os dicionários monolíngues, provendo um modelo comum harmonizado, que codifica tipos semânticos estruturados e um conjunto de templates (análogos a *Frames* semânticos).

Segundo Lenci (2001), mesmo que SIMPLE seja um projeto de construção de um recurso lexical, ele trata de questões desafiadoras para pesquisa e oferece um *framework* para testar e avaliar a maturidade no campo da semântica lexical apoiada em, ou conectada com, o projeto de uma ontologia de topo para tipos. SIMPLE pode ser considerado uma continuidade de um projeto anterior da União Europeia, o projeto LE-PAROLE (RUIMY; MONACHINI; DISTANTE, 2002), acrescentando uma camada semântica ao subconjunto das camadas de morfologia e sintaxe desenvolvidas no PAROLE.

Além de recomendações feitas em outros projetos, o modelo SIMPLE é baseado em extensões à Teoria do Léxico Gerativo. Uma característica essencial do Léxico Gerativo, discutida anteriormente, é a habilidade em capturar as várias dimensões do significado de uma palavra. O vocabulário básico se apoia em uma extensão da Estrutura *Qualia*, para organizar os tipos semântico/conceptuais como um mecanismo representacional que expresse a multidimensionalidade do significado. Isto fornece ao modelo um alto grau de generalidade, uma vez que os mesmos mecanismos são usados para conceitos de diferentes áreas semânticas (entidades, eventos, nomes abstratos, etc).

O modelo SIMPLE contém três tipos de entidades formais:

- a) **Unidades Semânticas:** as palavras são codificadas como *Unidades Semânticas* (ou SemU - *Semantic Units*). Cada SemU está associada a um Tipo Semântico da ontologia, mais algumas informações adicionais especificadas no *Template*, que contribui para a caracterização do sentido da palavra.
- b) **Tipo Semântico:** corresponde aos tipos semânticos associados as SemU. Cada tipo envolve uma informação estruturada, organizada segundo os quatro papéis *qualia* do Léxico Gerativo. A informação *qualia* é classificada em **informação definidora de tipo** e **informação adicional**. A primeira é a informação que intrinsecamente define o tipo semântico como ele é. Ou seja, a SemU não pode ser associada a certo tipo, a não ser que seu conteúdo semântico inclua a

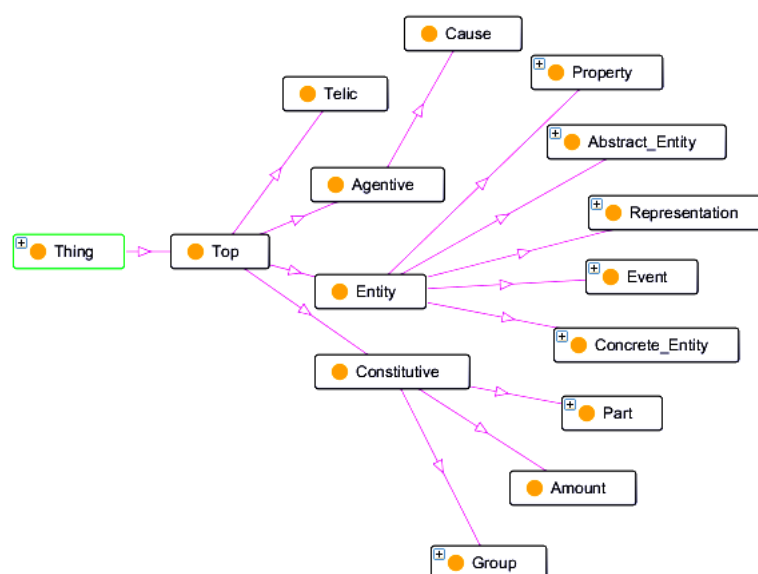
informação que define aquele tipo. Por outro lado, as informações adicionais especificam componentes da SemU que não participam de sua caracterização.

- c) **Template**: é uma estrutura esquemática usada pelo lexicógrafo para codificar um dado item lexical. O template expressa o tipo semântico mais algumas informações adicionais. O objetivo de seu uso é harmonizar, guiar e facilitar o trabalho lexicográfico.

A quantidade de informações codificadas no modelo SIMPLE é grande e inclui, entre outras: o tipo semântico; informações sobre o domínio; glossas lexicográficas; estrutura argumental para SemU predicativas; restrições seletivas para os argumentos; tipo de evento, que caracterize as propriedades aspectuais dos predicados verbais; links para os *Frames* de subcategorização sintática; a estrutura *qualia*; informações sobre alternativas polissêmicas regulares; sinônimos e colocações. No escopo deste trabalho, iremos utilizar apenas as informações sobre tipos semânticos e estrutura *qualia*.

Os tipos semânticos estão estruturados de forma a considerar os princípios da organização ortogonal, formalizados pelo Léxico Gerativo. Um tipo semântico pode ser visto como sendo um *cluster* de informações semânticas estruturadas. A hierarquia de tipos foi subdividida em duas camadas: a ontologia nuclear (*core*), que contém os tipos identificados como comuns aos léxicos das várias línguas, e a ontologia recomendada, formada por tipos mais específicos, fornecendo uma organização mais granular dos sentidos das palavras.

Figura 8 – Top-level da ontologia SIMPLE.



A Figura 8 mostra como os princípios da Estrutura Qualia foram adotados para organizar o nível mais alto da ontologia. O tipo CONSTITUTIVE, por exemplo, domina os tipos semânticos que descrevem sentidos de uma SemU associados via relações meronímicas a outras SemU. A seção A.1 apresenta a estrutura completa da ontologia.

A fim de fornecer uma linguagem mais expressiva para representação da informação semântica, o modelo SIMPLE apresenta um conjunto amplo de traços (*features*) e relações, que também são organizados segundo a estrutura *qualia* (Formal, Agentivo, Constitutivo e Télico). Os traços são usados para caracterizar os atributos para os quais uma faixa de valores restritos pode ser especificada. Por outro lado, as relações entre SemU são definidas para aqueles aspectos do significado lexical que não podem ser facilmente reduzidos a um conjunto fechado de pares atributo-valor, implementando e estendendo a estrutura *qualia*. As relações são também organizadas segundo uma hierarquia taxonômica, possibilitando a subespecificação na atribuição de tipos, bem como a introdução de subtipos mais refinados para uma dada relação.

O uso de relações *qualia* permite a captura de semelhanças entre aspectos semânticos de palavras pertencentes a domínios de áreas conceituais muito diferentes, bem como uma diferenciação mais fina (*fine-grained*) entre palavras pertencentes a tipos semânticos próximos. Como uma ilustração do uso da estrutura *qualia* estendida pelas relações, uma comparação entre os substantivos <nadador> e <peixe> é apresentada conforme Lenci et al. (2000). Um nadador é claramente um indivíduo cuja função típica é "nadar" (Figura 9).

Figura 9 – Estrutura *qualia* para <nadador>.

| | |
|----------------|---------------------------------------|
| Nadador | |
| FORMAL: | isa(<nadador>,<peessoa>) |
| TÉLICO: | is the activity of(<nadador>,<nadar>) |

Entretanto, no processo de codificar a semântica do item lexical <peixe>, pode-se querer codificar a informação que uma das atividades típicas dos peixes é nadar, apesar de ser necessário desmembrar como o mesmo tipo de evento desempenha um papel diferente caracterizando o tipo deste item lexical. É necessário, assim, observar o seu comportamento linguístico diferente. Por exemplo, modificado por um adjetivo, os dois substantivos comportam-se de forma bastante diferente, conforme as expressões em (4.13).

- (4.13) (a) Um velhador nadador
 Uma pessoa que é velha e que nada
 Uma pessoa que já nada há muito tempo
- (b) Um velho peixe
 Um peixe que é velho
 #Um peixe que já nada há muito tempo

A informação que um peixe tipicamente nada não é propriamente parte de sua dimensão télica, ou seja, não é concernente ao seu propósito inerente. Ao invés disso, esta informação é relacionada, de alguma forma, ao modo pelo qual o peixe é "construído constitucionalmente". A propriedade de nadar não adiciona informação télica ao tipo, mas especifica o peixe de acordo com uma dimensão constitutiva/física. Assim, a informação que um peixe tipicamente nada é expressa no qualia constitutivo (Figura 10).

Figura 10 – Estrutura *qualia* para <peixe>.

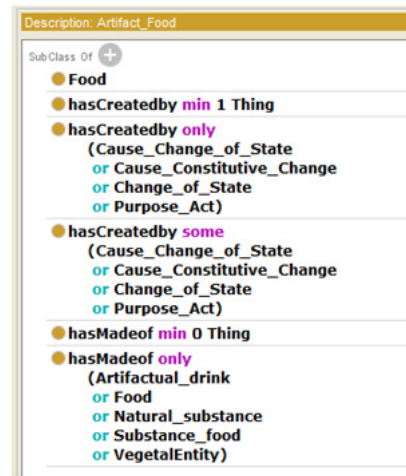
| | |
|----------------------|---|
| Peixe | |
| FORMAL: | <code>isa(<peixe>,<animal>)</code> |
| CONSTITUTIVO: | <code>constitutive_activity(<peixe>,<nadar>)</code> |

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

- a) Relações de um tipo *qualia* permitem identificar nós multidimensionais quando aplicadas como restrições a eles. Se uma classe da ontologia tem uma restrição mandatória sobre uma propriedade de objeto *qualia*, então podemos identificá-la como tendo este papel *qualia* como uma dimensão adicional. Por exemplo, a classe ARTIFACT_FOOD tem a dimensão agentiva adicional já que uma relação *qualia* agentiva (CreatedBy) é definida como uma restrição mandatória para esta classe (Figura 11).
- b) Foi criada uma classe OWL adicional para cada um dos quatro tipos *qualia*, com uma restrição de cardinalidade mínima necessária e suficiente com o valor 1, sobre a propriedade *qualia* correspondente. Por exemplo, a classe télica TELICTYPE teria como restrição a condição necessária e suficiente <hasTelic minimum 1> e assim, todas as classes que têm como uma restrição necessária qualquer relação télica seriam classificadas e automaticamente adicionadas como subclasses, através do mecanismo de inferência.

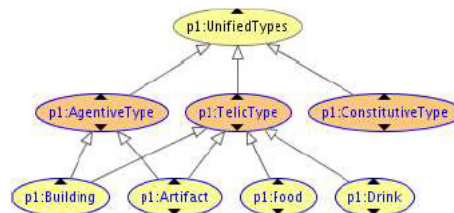
A Figura 12 mostra um exemplo da representação de tipos unificados através da inferência. AGENTIVETYPE, CONSTITUTIVETYPE e TELICTYPE são as classes adicionais. Como pode ser visto na figura, é inferido que as classes BUILDING e ARTIFACT tem

Figura 11 – Definição parcial de <Artifact_food>.



dimensões agentiva e tética adicionais, ao passo que FOOD e DRINK contém uma dimensão tética adicional.

Figura 12 – Tipos Unificados na SIMPLE-OWL.



Fonte: Toral e Monachini (2007)

4.6 Modelo Conexionista

O *framework* desenvolvido neste trabalho implementa um processo de desambiguação de item lexicais através de um processo *best-fit* de análise semântica. A ideia básica é tomar como entrada um conjunto de item lexicais (que formam um sintagma ou são constituintes de uma sentença) e oferecer como saída o *Frame* (ou o *Tipo Ontológico*) mais adequado para a interpretação dos itens lexicais no contexto fornecido. Além disso, a análise semântica busca oferecer também uma explicação para o *Frame* ou tipo escolhido, através da geração de uma **especificação semântica** (SemSpec - *Semantic Specification*).

Best-fit é um termo usado para descrever qualquer processo de tomada de decisão que combine informações de diversas fontes de maneira quantitativa. Classificamos o *framework* desenvolvido como um processo *best-fit* porque ele recebe como entrada, além dos itens lexicais, informações semânticas (através de *Frames* e *Tipos Ontológicos*) e informações sintáticas (através da descrição das funções gramaticais dos itens lexicais). A questão quantitativa está associada às ponderações usadas na rede semântica.

Modelos de processamento *best-fit* são também conhecidos como modelos interacionistas ou modelos baseados em restrições (*constraint-based models*) (BRYANT, 2008). **Modelos Baseados em Restrições** assumem que as múltiplas alternativas são avaliadas com base em algumas restrições impostas pelo modelo. No nosso caso, estas restrições são de caráter semântico: basicamente estamos associando tipos semânticos aos *Elementos de Frame* e aos itens lexicais.

Relacionados fortemente ao paradigma baseado em restrições estão os modelos conexionistas e, em particular, os modelos conexionistas estruturados. **Modelos Conexionistas Estruturados** estão associados à PLN desde a década de 80 (FELDMAN; BALLARD, 1982). Feldman (2006), por exemplo, apresenta um *framework* para reduzir a interpretação da linguagem a modelos conexionistas. Embora existam diversos modelos que podem ser classificados como conexionistas, estamos adotando um modelo *best-fit* que usa a técnica de *Ativação Propagada* (*Spread Activation*) para combinar as informações de múltiplos domínios e a competição entre as unidades da rede conexionista para modelar as hipóteses concorrentes.

Esta seção apresenta noções gerais sobre o uso de Modelos Conexionistas Estruturados de Redes Semânticas e sobre a técnica de *Ativação Propagada*.

4.6.1 Modelos Conexionistas Estruturados

Ao longo dos anos, vários esquemas de representação têm sido inventados para capturar o significado de enunciados linguísticos, nas aplicações de PLN. Segundo Jurafsky e Martin (2008), além da Lógica de Primeira Ordem, os esquemas mais usados tem sido as redes semânticas e as representações baseadas em *frames* (também chamadas *slot-filler*). Nas redes semânticas, os objetos são representados como nós em um grafo e as relações entre os objetos como *links* nomeados. Em sistemas baseados em *frames*, os objetos são representados como estruturas de traços.

Como um dos objetivos deste trabalho é a integração de recursos lexicais que estão disponíveis em diferentes formatos, estamos usando uma solução híbrida: podemos dizer que está sendo construída uma rede semântica (uma vez que os nós estarão representando conceitos) cujos objetos são estruturas de traço. A estrutura de traço utilizada é discutida em detalhes na seção 7.1. Além disso, queremos realizar um processo de inferência sobre

esta rede. Dados estes requisitos, a implementação realizada se baseou amplamente nas propostas de [Shastri \(1992\)](#), que apresenta as ideias básicas para a realização de um modelo conexionista estruturado de redes semânticas.

Os Modelos Conexionistas, por outro lado, podem ser caracterizados como modelos computacionais com habilidades de adaptar, aprender, generalizar, agrupar ou organizar dados, nos quais a operação é baseada em processamento paralelo. Neste modelo, uma rede é formada por um conjunto de unidades de processamento simples, que se comunicam entre si enviando sinais de uma para outra, através de conexões ponderadas.

O componente elementar deste modelo são as unidades de processamento, também chamadas de nós, neurônios ou células (e por isto, estas redes são comumente denominadas **Redes Neurais Artificiais**). Cada unidade é um modelo matemático, inspirado no modelo biológico de um neurônio. Para [Feldman \(2006\)](#) uma das belezas da teoria da computação neural é a sua simplicidade. Cada neurônio transmite sinais como respostas individuais que dependem do seu estado interno e da combinação dos sinais recebidos como entrada. A sinalização neural é capturada como um único valor numérico, representando tanto a "força" do sinal de entrada quanto do sinal de saída. Na maioria dos modelos, cada unidade computa o sinal de saída com base na soma ponderada dos sinais de entrada. O sinal de saída é transmitido para todos os neurônios vizinhos, e o processo continua.

Porém, uma questão se levanta: como unidades tão simples podem representar algum conhecimento? Ou, colocado de outra forma, quantos neurônios seriam necessários para termos uma unidade conceptual básica? A análise de [Feldman e Ballard \(1982\)](#) indica a possibilidade de construção de redes com um nível mais alto de abstração, em que as unidades poderiam possuir algum tipo de estrutura, necessária para a organização do conhecimento na rede. Este modelo é chamado de **Modelo Conexionista Estruturado**, e é o modelo utilizado neste trabalho.

Como visto, a representação de redes semânticas utiliza uma notação simples: elas codificam o conhecimento através de nós interconectados via links nomeados. Segundo [Shastri \(1992\)](#), esta notação permite que elas ressaltem a "vivacidade" da interdependência entre as diferentes partes do conhecimento conceptual. Elas se tornam assim análogas a um domínio: há uma correspondência uma-a-um entre os nós e os componentes da conceptualização do domínio e, mais importante, entre os *links* do grafo e as dependências inferenciais entre os componentes conectados pelo *link*. Isto reduz o problema da inferência ao problema de atravessar fisicamente uma grafo instanciado.

[Shastri e Ajjanagadde \(1989\)](#) apresentam a proposta de adotar o Modelo Conexionista Estruturado como uma abordagem para a implementação de redes semânticas que representem conhecimento e inferências. Ele ressalta duas consequências importantes nesta abordagem. Em primeiro lugar, em um sistema conexionista não há um interpretador distinto, externo à rede, para mediar a recuperação dos dados e as inferências: o padrão de

conexões, os pesos dos *links* e as características computacionais dos nós representam não apenas o conhecimento do domínio, mas também codificam o processos de recuperação e inferência. Em segundo lugar, se codificamos cada pedaço de informação como um nó e as dependências entre os pedaços de informação como *links* explícitos entre os nós apropriados, o processo de inferência pode ser visto como o processo de *ativação propagada* em um modelo conexionista.

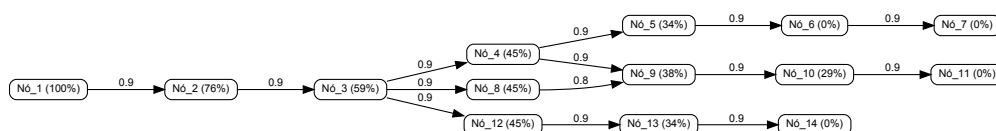
Este é exatamente o processo de inferência implementado neste trabalho: a construção do recurso lexical como uma rede semântica representada em um modelo conexionista estruturado e o uso da técnica de *ativação propagada* para a realização das inferências.

4.6.2 Ativação Propagada

O modelo conexionista permite usar uma técnica de travessia em redes chamado *Ativação Propagada* (AP) ou *Spreading Activation*. As redes em que esta técnica é aplicada são chamadas *Redes de Ativação Propagada* (RAP). Estas redes vêm sendo usadas há bastante tempo em processos de WSD (*Word Sense Desambiguation*), semelhantes ao que estamos tratando neste trabalho. Hirst (1988) apresenta uma proposta inicial, que foi posteriormente aplicada em grandes redes neurais (VERONIS; IDE, 1990). Diederich (1990) discute amplamente o uso da *Ativação Propagada* e modelos conexionistas no âmbito do PLN. Mais recentemente a técnica foi aplicada utilizando as relações semânticas da Wordnet para desambiguação de lexemas (TSATSARONIS; VAZIRGIANNIS; ANDROUTSOPOULOS, 2007).

De maneira simplificada, na RAP um ou mais nós iniciais recebem um determinado nível de ativação. Cada nó ativado então "espalha" sua ativação pelos nós vizinhos, que estejam ligados a ele. A cada passagem por um nó, o nível de ativação é diminuído. Quando o nível de ativação fica abaixo de um certo limiar previamente definido (*thresold*), ou quando uma condição de parada, previamente estabelecida, é alcançada, a ativação não é mais realizada. Outra condição comumente usada para terminar o processo é quando um mesmo nó é ativado por diferente vizinhos, através de caminhos diferentes. A Figura 13 ilustra o método. A propagação da ativação é iniciada no nó 1 (100%). Cada link tem o mesmo valor para o peso (0.9), exceto o *link* entre os nós 8 e 9 (0.8). O fator de decaimento é de 0.85 e o *thresold* é de 35%. A figura mostra a ocorrência de 5 ciclos de propagação da ativação.

Figura 13 – Exemplo de Ativação Propagada.



Três características são geralmente associadas à técnica de *Ativação Propagada*, embora sua implementação computacional possa ser realizada através de várias estratégias diferentes. A primeira questão se refere ao processamento massivo paralelo. O paralelismo, embora não seja obrigatório, é uma das características que ajuda a distinguir a AP de outras técnicas usadas no âmbito da Inteligência Artificial. Outra característica é que o tipo de comunicação entre as unidades de processamento é restrito a mensagens não-simbólicas e, portanto, o processo em si mesmo é independente de uma interpretação específica. Finalmente, o resultado do processo pode ser avaliado por um componente externo, um "avaliador de caminhos" (*path evaluator*), que compara e analisa os caminhos na rede que tenham unidades altamente ativadas.

A implementação da rede RAP neste trabalho é detalhada no capítulo 6.

Parte III

O framework LUDI

5 Proposta de uma Ontologia LUDI

5.1 Introdução

O trabalho de pesquisa, descrito neste texto, deu origem a uma aplicação computacional, à qual estamos denominando **Framework LUDI**. Este *framework* está baseado em um recurso lexical, apoiado em uma visão ontológica e implementado como um modelo conexionista estruturado. Neste capítulo (e nos demais da Parte III) são descritos os detalhes de cada componente do *framework* e apresentados exemplos de sua aplicação.

5.2 Interface Ontolex

A interface entre um recurso lexical e uma ontologia é denominada **Ontolex**, e foi apresentada na seção 4.3. No *framework* LUDI esta interface é criada entre a FrameNet e a ontologia SIMPLE.

Prévot, Borgo e Oltramari (2005) apresentam as seguintes opções metodológicas para a tarefa de interfacear recursos lexicais e ontologias:

- a) Reestruturar o léxico computacional com base em princípios dirigidos a ontologias;
- b) Enriquecer uma ontologia com informações lexicais;
- c) Alinhar uma ontologia com um recurso lexical.

A opção adotada neste trabalho é o **alinhamento**, por três razões. Primeiro, porque a SIMPLE já é uma ontologia com base lexical (as Unidades Semânticas representam sentidos de lexemas), ou seja, a ontologia SIMPLE já é, por si, um exemplo de alinhamento (PRÉVOT; BORGIO; OLTRAMARI, 2005). Segundo, porque a semântica representada na FrameNet requer uma lógica de mais alta ordem, para a qual existem poucas ferramentas computacionais disponíveis, o que dificulta a tarefa de "ontologizar" completamente a FrameNet. E, finalmente porque, neste projeto, os dados da FrameNet e da Ontologia SIMPLE serão armazenados usando a mesma representação (apresentada na seção seção 7.1); o uso da mesma representação facilita o alinhamento da conceptualização. Além disso, como ressaltam ainda Prévot, Borgo e Oltramari (2005), o alinhamento é o método de interfaceamento mais avançado, pois possui a vantagem de manter a ontologia e o recurso lexical como camadas distintas de informação semântica, embora melhorados pelas ligações mútuas.

Neste trabalho estamos assumindo que a seleção argumental é feita com base nas

Unidades Lexicais. Na FrameNet, cada *Unidade Lexical* está associada a uma ou mais valências. As preferências seletivas (SP - *Selectional Preferences*) podem ou não serem compartilhadas por todas as valências de uma *Unidade Lexical* que evoca dado *Frame*. De qualquer forma, é feita uma generalização dos argumentos, representados pelos *Elementos de Frame*. Segundo Tonelli et al. (2012) os modelos mais relevantes para a generalização dos argumentos podem ser agrupados em duas classes: **modelos distribucionais**, que se apoiam na co-ocorrência de palavras e **modelos baseados em hierarquias**, que usam uma ontologia ou uma hierarquia semântica, como a Wordnet. Este projeto se enquadra no segundo tipo de modelo, usando a ontologia SIMPLE como hierarquia.

O alinhamento é implementado através da associação dos *Tipos Ontológicos* (Classes) da ontologia SIMPLE com as os *Elementos de Frame* e *Frames* da FrameNet. A hipótese é que os tipos semânticos dos itens lexicais podem ser usados como *Tipos Ontológicos* no contexto da FrameNet. Para Ruppenhofer et al. (2010, s. 6.2.1), os tipos semânticos dos *Elementos de Frame* categorizam o tipo de *filler* (ou preenchimento) que é esperado para aquele *Elemento de Frame*. Cabe ressaltar também que as relações entre os *Frames* implicam em relações entre os *Elementos de Frame* e estas relações são exploradas pelo *framework*.

5.3 Entidades

A questão das Entidades (naturais ou artefatuais) no âmbito da FrameNet é discutida direta, ou indiretamente, em vários trechos de Ruppenhofer et al. (2010). Por serem revelantes para esta proposta, apresentamos duas destas discussões.

Já no início do texto (RUPPENHOFER et al., 2010, p. 5), se reconhece que nomes referentes a tipos artefatuais ou naturais possuem uma estrutura de *Frame* mínima, por si próprios. Artefatos geralmente ocorrem com outras expressões, indicando seu subtipo, o material de que é feito, a maneira de produção ou seu propósito e/ou uso¹. Esta estrutura é representada por *Elementos de Frame* nos *Frames* de vários tipos de artefatos. No entanto, os *Frames* evocados por nomes deste tipo raramente dominam as sentenças nas quais eles ocorrem e, portanto, dificilmente são selecionados para anotação.

Em Ruppenhofer et al. (2010, s. 3.8), na seção relativa à anotação dos chamados *slot fillers*, é explicado que nomes de entidades ocorrem, na maioria das vezes, preenchendo os *slots* em *Frames* evocados por verbos, adjetivos ou outros nomes. Há, no entanto, interesse do projeto FrameNet em registrar, para certas entidades, em quais *Frames* elas são usadas como *slot fillers*. Para isso é usada a anotação Gov-X, em que o anotador registra a regência sintática. Além disso é também proposto que algumas *Unidades Lexicais* de *Frames* de Entidades poderiam ser associadas a um tipo semântico, o que daria aos usuários do banco

¹ O leitor atento reconhece imediatamente a referência de Ruppenhofer et al. (2010) à estrutura *qualia*.

de dados da FrameNet informações sobre que classes de palavras atendem determinadas restrições seletivas.

Como se pode observar, estas colocações da equipe da FrameNet vêm ao encontro do alinhamento FrameNet/SIMPLE discutido anteriormente. A proposta é avançar mais no alinhamento, substituindo os *Frames* de Entidades atualmente registrados pela FrameNet, pelas classes da ontologia SIMPLE (seção A.1). Os *Frames* substituídos são apenas aqueles mais diretamente relacionados ao *Frame* ENTIDADE e constituem uma parte pequena da base de *Frames* (cerca de 15%)².

5.4 Elementos de Frame

A fim de aumentar a densidade da rede FrameNet, no intuito de ter um número maior de opções para a construção do modelo conexionista, foram feitas duas adições à estrutura da FrameNet:

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

Em relação aos *Tipos Ontológicos*, existem trabalhos que buscam generalizá-los a partir dos nomes dos *Elementos de Frame* e das relações entre eles (já que na versão atual da FrameNet – release 1.5 – apenas cerca de 50% dos *Elementos* têm tipos explicitamente anotados) (MATSUBAYASHI; OKAZAKI; TSUJII, 2009), ou que buscam uma melhor definição dos tipos semânticos (a ontologia da FrameNet conta apenas com 40 tipos), usando as sentenças anotadas e fazendo relações com os *synsets* da Wordnet (TONELLI et al., 2012). Outros trabalhos propõem algum tipo de alinhamento com ontologias existentes (SCHEFFCZYK; BAKER; NARAYANAN, 2010) ou a própria representação da FrameNet como uma ontologia (NUZZOLESE; GANGEMI; PRESUTTI, 2011). Na abordagem usada neste trabalho, atribuímos diretamente os *Tipos Ontológicos* aos *Elementos de Frame* usados nos experimentos.

A associação (b) necessita de uma justificativa, uma vez que ela não existe na estrutura da FrameNet atual e não foi ainda proposta em trabalhos anteriores. A ideia é registrar a referência que um *Elemento de Frame* pode fazer a um outro *Frame*. É importante ressaltar que esta referência (ou mapeamento) não implica necessariamente na definição do tipo semântico do *Elemento de Frame* (embora em casos específicos ela possa ser usada desta forma). O objetivo primário é estender a interpretação conceitual dos *Elementos de Frame*, mostrando que, além de atuarem em termos linguísticos como papéis microtemáticos, específicos para a situação descrita pelo *Frame*, eles podem desempenhar

² <http://www1.icsi.berkeley.edu/~warrenmc/FrameCategorization.html#EntityPool>

uma função cognitiva, como evocadores de outros *Frames*, a partir do *Frame* onde se situam. Pode-se argumentar que esta questão é atualmente tratada através da criação de novos *Frames* (associados através das relações de *Subframe* ou *Using* com o *Frame-pai*). Mas em muitos casos a “situação” evocada pelo *Elemento de Frame* ou já existe como um *Frame* definido na base da FrameNet, ou se relaciona apenas parcialmente a um *Frame* já definido, ou ainda possui um caráter muito específico, que pode ser generalizado por um *Frame* existente, ao invés de se criar um novo *Frame*.

De certa forma este mapeamento EF-F está implícito na FrameNet, quando se considera a anotação em texto corrido de uma sentença. Neste tipo de anotação, uma sentença pode possuir vários *targets*, elementos que evocam *Frames* (FEE - *Frame Evoking Elements*). Se considerarmos que fixando um *target* os demais sintagmas da sentença poderão ser anotados como *Elementos de Frame*, fica claro que é possível interpretar que um *Elemento de Frame* está referenciando outro *Frame* (geralmente pelo núcleo do sintagma). Deve ficar claro, no entanto, que nem sempre o núcleo do sintagma associado ao *Elemento de Frame* (em especial se Nuclear) e o *Frame* evocado compartilharão semântica.

Do ponto de vista do processo da conceptualização, um *Elemento de Frame* referenciar um *Frame* pode representar a focalização, uma operação de *construal*. É um processo cognitivo básico o fato de, a partir do momento que temos uma cena ou situação perfilada (um *Frame*), podermos focar nossa atenção em um aspecto específico desta cena ou situação (um elemento de *Frame*) e assim sermos conduzidos a uma outra cena ou situação (outro *Frame*).

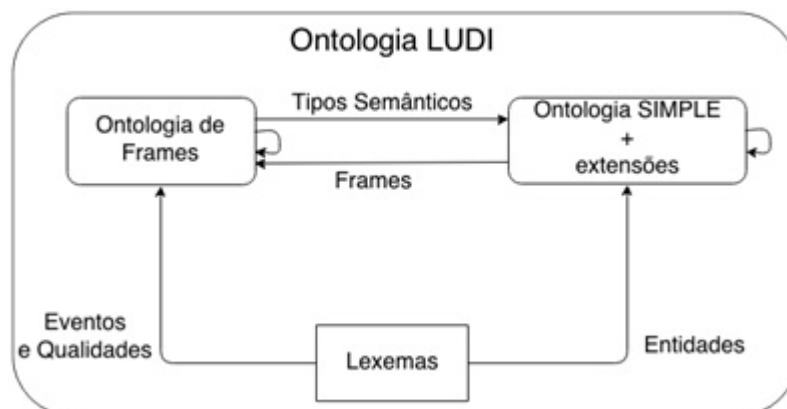
A fim de aproveitar a estrutura já existente na FrameNet, a proposta é a criação de uma relação adicional entre *Frames*, em que haveria uma única relação entre *Elementos de Frames*, associando o *Elemento de Frame* evocador com o seu correspondente no *Frame* evocado.

5.5 Integração de Ontologias

O recurso lexical usado no *framework* LUDI é implementado através de uma ontologia (Ontologia LUDI). Esta ontologia faz a integração de duas outras ontologias: uma Ontologia de *Frames* e a Ontologia SIMPLE. Cada uma atende a um determinado propósito. A [Figura 14](#) ilustra esta integração.

A Ontologia de *Frames* não é, na verdade, uma ontologia descrita formalmente, já que, como dito anteriormente, há dificuldades em formalizar a estrutura da FrameNet. O que estamos chamando de Ontologia de *Frames* consiste na representação de alguns elementos da FrameNet como classes ou propriedades. Esta ontologia não contém a totalidade das informações disponibilizadas pela FrameNet, mas apenas aquelas utilizadas nas operações do *framework* LUDI. Desta forma, são representados apenas alguns elementos da estrutura

Figura 14 – LUDI - Integração de ontologias.



semântica da rede: os *Frames*, os *Elementos de Frame*, as *Unidades Lexicais*, as *Relações* entre *Frames* e as *Relações* entre *Elementos de Frame*. Com base na discussão da seção anterior, na Ontologia LUDI, os *Elementos de Frame* podem fazer referências tanto a outros *Frames*, quanto a classes da Ontologia SIMPLE (indicando os tipos semânticos associados aqueles elementos). Como pode ser visto na [Figura 14](#), os *Frames* da Ontologia de *Frames* são evocados por Lexemas associados a *Eventos* (verbos e nomes de verbais) e a *Qualidades*.

A Ontologia SIMPLE (na verdade, uma versão adaptada da ontologia SIMPLE original, à qual foram acrescentadas algumas extensões) é usada para registrar os tipos semânticos estruturados. Estes tipos possuem relações entre si e são usados basicamente para categorizar os *Frames* e os *Elementos de Frames*. Os Lexemas referentes a entidades (tipos naturais e artefatos) são associados a tipos da ontologia SIMPLE. Os tipos da ontologia podem ser associados também a *Frames*, quando as relações indicam algum tipo de evento ou qualidade. Mecanismos de inferência são usados para a classificação dos lexemas em relação à estrutura *qualia*. A ontologia SIMPLE foi descrita na seção [4.5](#).

A Ontologia LUDI, portanto, é responsável pela integração dos recursos lexicais, registrando as relações entre os Lexemas e as Ontologias, bem como as relações entre as próprias ontologias.

6 LUDI como Modelo Conexionista para Processamento da Língua Natural

O fato de estarmos usando ontologias mostra que o *framework* LUDI trabalha em um nível bem alto de abstração para ser considerado simplesmente um modelo conexionista. No entanto, de acordo com a Teoria Neural da Linguagem (NTL - *Neural Theory of Language*) (FELDMAN, 2006), modelos em níveis mais altos de abstração ainda podem ser um tipo de modelo motivado neurologicamente. O grupo da NTL investiga como o cérebro computa a linguagem. Embora seja óbvio que o cérebro usa neurônios (que formam um modelo conexionista) para processar a linguagem, a NTL argumenta que os modelos conexionistas deste nível não são adequados para estudar este processamento. No outro extremo do espectro de complexidade, sistemas que processam a linguagem natural, como o *framework* LUDI, só podem ser considerados cognitivamente plausíveis se os algoritmos e estruturas de dados empregados puderem ser reduzidos a redes conexionistas. No entanto, Feldman (2006) mostra que *Estruturas de Traços (Feature Structures)* são estruturas computacionais que possuem uma implementação conexionista. Estes modelos são chamados **Modelos Conexionistas Estruturados** e foram apresentados na seção 4.6.

Nesta seção são apresentados alguns aspectos do modelo conexionista implementado. No capítulo seguinte é discutido como as *Estruturas de Traços* foram usadas para a construção do recurso lexical.

6.1 Unidades e Conexões da Rede

A Ontologia LUDI é implementada em uma rede conexionista, que é ativada quando uma análise deve ser realizada. As estruturas da ontologia constituem as *Unidades* da rede. Nas figuras a seguir, a rede é representada como um grafo dirigido, e os nós da rede recebem uma letra indicando o tipo de *Unidade* representada. A [Tabela 1](#) apresenta a descrição de cada *Unidade*.

A [Tabela 2](#) apresenta as conexões (*links*) entre as diversas unidades. O peso atribuído a cada *link* é sempre 1.0, exceto para os *links* em que o destino é um tipo semântico da ontologia SIMPLE. Neste caso, o peso do *link* é calculado em função da posição do tipo na hierarquia da ontologia. Os valores usados para cada conexão são apresentados no [Apêndice A](#).

Tabela 1 – Unidades da rede conexionista.

| Símbolo | Unidade | Descrição |
|---------|--------------------------------------|--|
| W | Word | Palavra (ou forma de palavra): uma das formas que uma palavra pode assumir, variando com a flexão. |
| X | Lexema | A forma não flexionada da palavra. |
| L | Lema | Expressões multi-palavras (MWE - <i>MultiWord Expressions</i>) são organizadas em um nível mais alto de abstração, chamado lema. Um lema é composto por um ou mais lexemas |
| U | Unidade Lexical Unidade Semântica | É o pareamento de um lexema com um significado. No contexto da FrameNet, o pareamento de um lema com um <i>Frame</i> é chamado <i>Unidade Lexical</i> (UL). No contexto da SIMPLE o pareamento é chamado de <i>Unidade Semântica</i> (USem). |
| F | Frame | Estrutura semântica usada para representar situações, eventos e qualidades. |
| E | Elemento de Frame | Usado para representar os participantes de um <i>Frame</i> . |
| V | Valência | Usado para representar a estrutura argumental relativa a uma <i>Unidade Lexical</i> . |
| T | Tipo Ontológico | Tipo semântico estruturado ontologicamente e baseado nas relações <i>qualia</i> . |

Tabela 2 – Links da rede conexionista.

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

6.2 Construção da Rede

O processo de construção e uso da RAP (Rede de Ativação Propagada) do *framework* LUDI é semelhante ao apresentado em Tsatsaronis, Vazirgiannis e Androutsopoulos (2007). A principal diferença é que estamos interessados em determinar o *Frame* evocado pelo lexema e não o *Frame* compartilhado por ambos os lexemas. Isto torna necessário (diferente da implementação comum de uma RAP) registrar o caminho entre o nó inicial e o nó sendo analisado.

O processo é iniciado com a construção da rede completa, chamada *Rede LUDI*. Para isso são usados todos nós e relações disponíveis na ontologia SIMPLE, na rede de *Frames* e nas redes construídas para os lexemas sob análise. A seguir, a partir da Rede LUDI, é construída uma rede parcial, em que constam apenas os nós e relações associados aos lexemas analisados. Esta rede é chamada *Rede Foco*.

O processo de **Ativação Propagada** é então aplicado sobre a *Rede Foco*. São ativados todos os lexemas, exceto o *target*. A estratégia de ativação consiste em iterações. A cada iteração é feita a análise dos nós ativados na interação anterior. O nó inicial tem um nível de ativação (A) igual a 1. Os demais nós têm o nível de ativação inicial igual a zero. A cada iteração, cada nó propaga sua ativação para os nós vizinhos, como uma função do valor atual do seu nível de ativação e dos pesos dos *links* que o conectam com cada vizinho. A cada iteração p , cada nó j tem um nível de ativação igual a $A_j(p)$ e uma saída (*output*) igual a $O_j(p)$, que é uma função do seu nível de ativação, conforme a equação (6.1).

$$(6.1) \quad O_j(p) = f(A_j(p))$$

A saída de cada nó afeta o nível de ativação da próxima iteração de qualquer nó k ligado ao j por um *link* direto. Assim, o nível de ativação de cada nó k da rede é função da saída (na iteração anterior $(p - 1)$) de cada nó vizinho j que tenha um *link* direto e_{jk} , bem como uma função do peso do *link* W_{jk} , como mostrado na equação (6.2).

$$(6.2) \quad A_k(p) = \sum_j O_j(p - 1)W_{jk}$$

A função de saída O (uma variação da função logística) foi escolhida cuidadosamente, para evitar um excesso de ativações nos nós da rede e é apresentada na equação (6.3). O valor τ é o *threshold*, ou seja, o valor limite, abaixo do qual a saída não é mais gerada, impedindo que nós com níveis de ativação muito baixos influenciem seus nós vizinhos.

$$(6.3) \quad O_j(p) = \begin{cases} 0, & \text{se } A_j(p) < \tau \\ \frac{1 - \exp(5 * (-A_j(p)))}{1 + \exp(-A_j(p))}, & \text{caso contrário} \end{cases}$$

Em termos de implementação computacional, no *framework* LUDI estamos trabalhando com um valor para o limite $\tau = 0.001$ e como condição de parada (além da

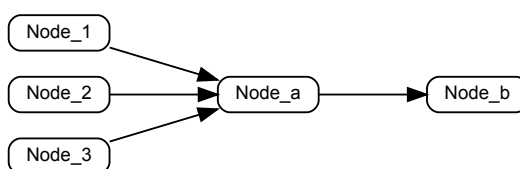
inatividade dos nós) um número máximo de iterações, que pode ser configurado pelo usuário.

6.3 Ativação condicional

A implementação do modelo conexionista neste trabalho está baseada, parcialmente, nas ideias apresentadas em Shastri (1992). Uma das características importantes para o processo de inferência é a ativação condicional de nós da rede. No *framework* LUDI são usados três tipos de condições: **and-nodes**, **restrições** e **links inibitórios**.

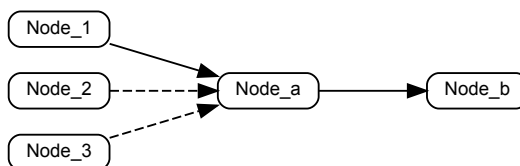
And-nodes são nós da rede que somente são ativados se todas as suas entradas forem ativadas. Por exemplo, na Figura 15, se o nó Node_a for um *and-node*, ele só será ativado se os nós Node_1, Node_2 e Node_3 forem ativados.

Figura 15 – Modelo conexionista: and-node.



Já as **restrições** indicam ativações condicionais mais específicas. Um nó, para ser ativado, precisa atender a uma condição específica, por exemplo, ter uma de suas entradas ativadas. Na Figura 16, o nó Node_a poderá ser ativado quando o nó Node_1 for ativado, sem a necessidade de ativação dos nós Node_2 e Node_3. Porém, sem a ativação do Node_1, o nó Node_a não é ativado, mesmo que Node_2 e Node_3 estejam ativados.

Figura 16 – Modelo conexionista: Restrições.



Links inibitórios são conexões entre nós em que a ativação do nó origem pode inibir (total ou parcialmente) a ativação do nó destino. De forma geral, a implementação de *links inibitórios* prevê que o valor de saída do nó origem é computado como um valor negativo no cálculo do nível de ativação do nó destino. No modelo conexionista implementado no *framework* LUDI, no entanto, é usada uma outra abordagem. O peso do *link* que faz a conexão entre o nó origem e o nó destino é considerado com um "percentual

de inibição", com valores reais na faixa $[0,1]$. Este percentual é aplicado ao valor do nível de ativação do nó destino. Assim, por exemplo, o peso 1 (100%) indica que o nó destino será completamente desativado, enquanto o peso 0.1 indica que o nó destino terá 90% do nível de ativação original. Desta forma, é possível na configuração da rede indicar o "quanto" a ativação de um nó vai inibir a ativação de outro. O "percentual de inibição" é propagado em cascata para os demais nós conectados.

And-nodes são usados para implementação de valências (discutidas a seguir) e restrições são usadas com *Elementos de Frame*. Por exemplo, um *Elemento de Frame* PARTE_CORPORAL pode ser conceptualizado como um container e ser associado aos tipos semânticos CONTAINER e BODY_PART. No entanto, ele só deve ser ativado se o nó BODY_PART for ativado e não apenas o nó CONTAINER. *Links inibitórios* são usados na competição entre *Frames* e na competição entre *Elementos de Frame* de um *CoreSet*.

6.4 Valências e Funções Gramaticais

Nesta implementação estamos usando o termo "valência" no mesmo sentido apresentado por Fillmore (2008): as valências registram as diversas possibilidades combinatórias da interface sintaxe-semântica. As valências indicam as várias configurações dos constituintes associados a uma determinada *Unidade Lexical*. É, portanto, uma estrutura complexa: ela está associada simultaneamente a uma *Unidade Lexical* e a vários *Elementos de Frame*, juntamente com as funções gramaticais atribuídas a cada um destes *Elementos de Frame*.

Como na implementação atual deste trabalho estamos considerando a análise de apenas uma única *Unidade Lexical*, optamos por representar cada valência como uma unidade da rede conexionista. Assim, cada *Frame* pode ser conectado a diversas valências (pois todas se referem a mesma *Unidade Lexical*) e cada valência pode ser conectada a diversos *Elementos de Frame*. As valências são implementadas como *and-nodes*, sendo ativadas somente quando todos os *Elementos de Frame* obrigatórios¹ forem ativados. O processo de análise consiste em determinar a valência mais provável para os dados fornecidos e, conseqüentemente, o *Frame* mais provável.

É possível configurar a rede para que um *Frame* prepondere sobre outro, independente do nível de ativação de cada um. No *framework* LUDI isto é usado para que a ativação dos *Frames* causativos (aqueles que possuem valências causativas/agentivas) se sobreponham a dos *Frames* incoativos, caso ambos os tipos sejam ativados no processo de inferência.

Links inibitórios também são usados para a escolha de um *Elemento de Frame* de

¹ "Obrigatório" aqui tem um sentido diferente de "nuclear". Como *Elementos de Frames* Nucleares podem ter instanciação nula, *Elementos de Frame* Obrigatórios são aqueles que devem estar expressos na sentença.

um *CoreSet*. **CoreSets** (RUPPENHOFER et al., 2010, p. 21) são grupos de *Elementos de Frame* s que atuam como um conjunto, onde a presença de um Elemento é suficiente para satisfazer a valência semântica do predicador. No *framework* LUDI, os Elementos de um *CoreSet* podem ser conectados via *links inibitórios* para que, caso de mais de um Elemento do conjunto seja ativado, apenas um seja considerado no processo de inferência.

As funções gramaticais (detalhadas na seção 9.4) são implementadas como traços dos *links* entre as *Valências* e os *Elementos de Frames* (*links VE*), e não como unidades da rede. Durante o processo de análise, a ativação ou não de um nó correspondente a uma *Valência* vai depender da compatibilidade entre a função gramatical atribuída aos nós origem (os lexemas) e a função gramatical atribuída ao *link VE*.

7 Tratamento dos Recursos Lexicais

7.1 Estrutura de Traços

Como os dados originais da FrameNet e da ontologia SIMPLE estão representados em formatos computacionais diferentes (a FrameNet está disponível como um banco de dados relacional e a ontologia SIMPLE no formato OWL), para a construção da Ontologia LUDI é realizada uma operação intermediária. Esta operação consiste em representar os dados da FrameNet e da SIMPLE formatados como MVAs (Matriz de Valores de Atributos), representadas via uma **Estrutura de Traços Tipados** (*Typed Features Structure*). A redução a um formato comum facilita a integração das duas fontes de dados. A implementação computacional, propriamente dita, é feita, por sua vez, usando um modelo conexionista estruturado.

Em relação à FrameNet, cada *Estrutura de Traços* representa um *Frame*. O tipo da estrutura é o nome do *Frame*. Cada traço representa um *Elemento de Frame*. O valor do traço é uma lista de apontadores (i) para outros *Elementos de Frame* (representando a relação entre *Elementos de Frames*), (ii) para as classes da Ontologia SIMPLE, que funcionam como *Tipos Ontológicos* para o *Elemento de Frame* correspondente ao *traço* (podendo ser uma lista vazia, caso o *Elemento de Frame* não tenha um tipo semântico associado) ou (iii) para outros *Frames*. Os Lexemas, os Lemas e as *Unidades Lexicais* também são representados por *Estruturas de Traços*, o que possibilita considerar as *Unidades Lexicais* como tipos estruturados com *traços* próprios (que podem ser unificados com os traços do *Frame*, por exemplo).

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

As figuras [Figura 17](#) e [Figura 18](#) mostram, respectivamente, exemplos de representações, em termos do modelo conexionista, para elementos da FrameNet e da ontologia SIMPLE.

Na [Figura 17](#) estão representados o lexema (X), o lema (L), os *Frames* (F), as valências (V), os *Elementos de Frame* e os *Tipos Ontológicos* (T) associados aos *Frames* e aos *Elementos de Frame*. Destacamos a representação dos *Elementos de Frame* como unidades "anônimas", com o objetivo de mostrar seu duplo papel: como relações dos constituintes da sentença com o *Frame* e como elementos que pode ter *Tipos Ontológicos* associados. Ressalte-se também que, na [Figura 17](#), não estão sendo exibidas as *Unidades*

Lexicais (que entram como unidades intermediárias entre o Lema e o *Frame*).

Figura 17 – Elementos da FrameNet.

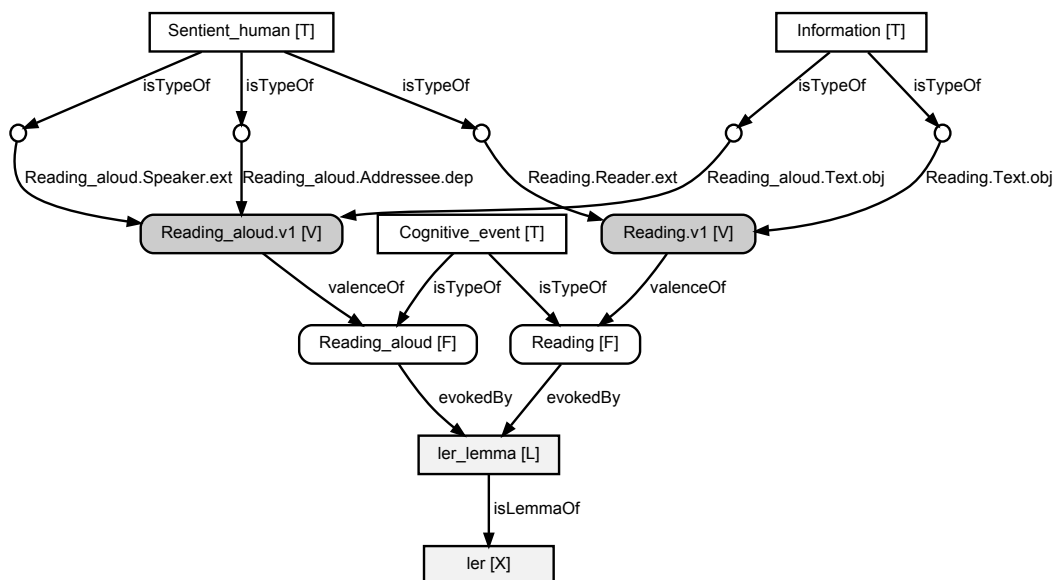
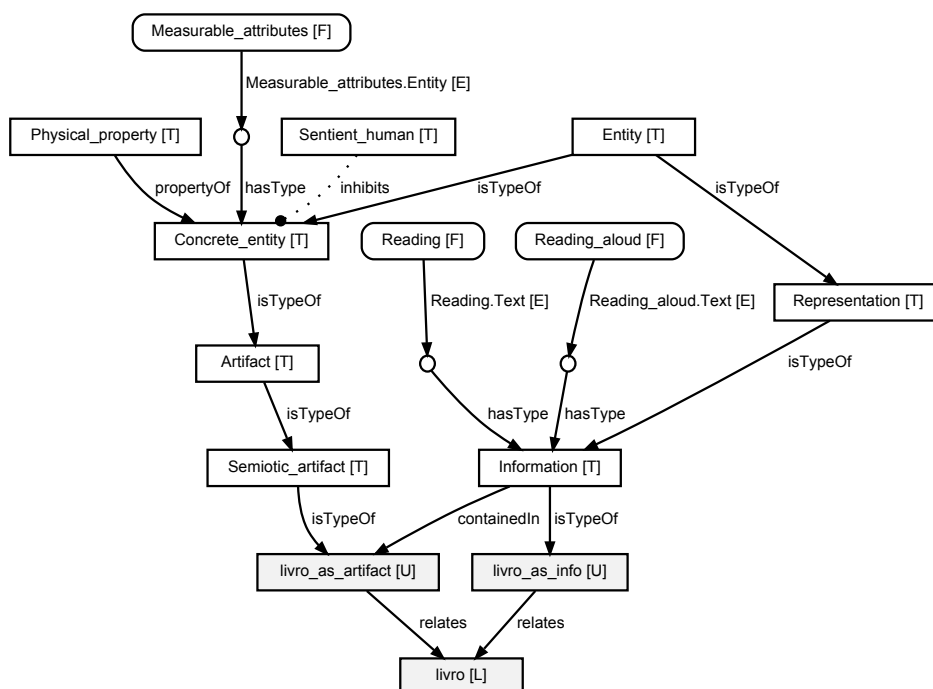


Figura 18 – Elementos da ontologia SIMPLE.



Na Figura 18, estão representados o lema (L), as *Unidades Semânticas* (U) e os *Tipos Ontológicos* (T). Note-se o fato que um mesmo lema pode estar associado a mais de uma *Unidade Semântica*: <livro> tem tanto um sentido de "livro como artefato" quanto

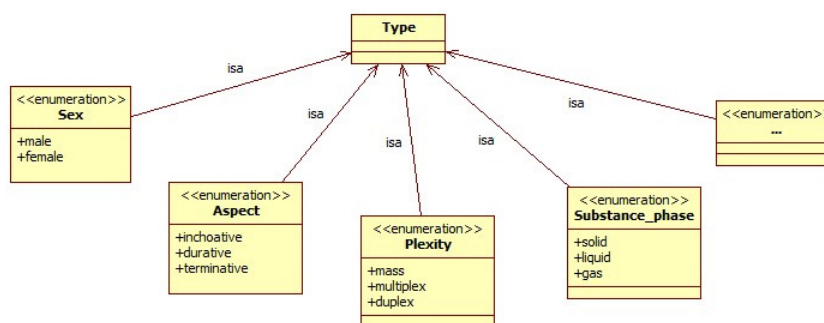
"livro como informação". Cada um destes sentidos está associado a *Tipos Ontológicos* distintos. A figura mostra também a associação de *Elementos de Frame* a *Tipos Ontológicos*, uma das maneiras de concretizar o alinhamento da FrameNet com a Ontologia SIMPLE.

7.2 A classe Type

A Ontologia LUDI possui uma classe especial chamada **Type**. O objetivo desta classe é fornecer, via suas subclasses, valores "constantes" que podem ser usados nas *Estruturas de Traços*, para caracterizar ou restringir um traço. Este valores basicamente implementam o conceito de "feature" conforme usado na Ontologia SIMPLE.

O diagrama da [Figura 19](#) mostra, como exemplo, algumas possíveis subclasses de TYPE: SEX, ASPECT, PLEXITY, SUBSTANCE_PHASE. Estas subclasses estão estereotipadas como *enumeration* porque a única função delas é enumerar os valores possíveis para um traço. Assim, o traço SEX pode assumir os valores *male* e *female*, enquanto o traço ASPECT pode assumir os valores *inchoative*, *durative* e *terminative*.

Figura 19 – Classe TYPE.



Em relação aos *Frames*, os traços associados a TYPE são implementados como *Elementos de Frame* "especiais", que auxiliam na caracterização do *Frame*. A [Figura 20](#) ilustra este uso. O tipo ASPECT é usado pelo *Frame* ACTIVITY_ONGOING. O tipo do *Elemento de Frame* ("especial") ASPECT em ACTIVITY_ONGOING assume o valor *durative*.

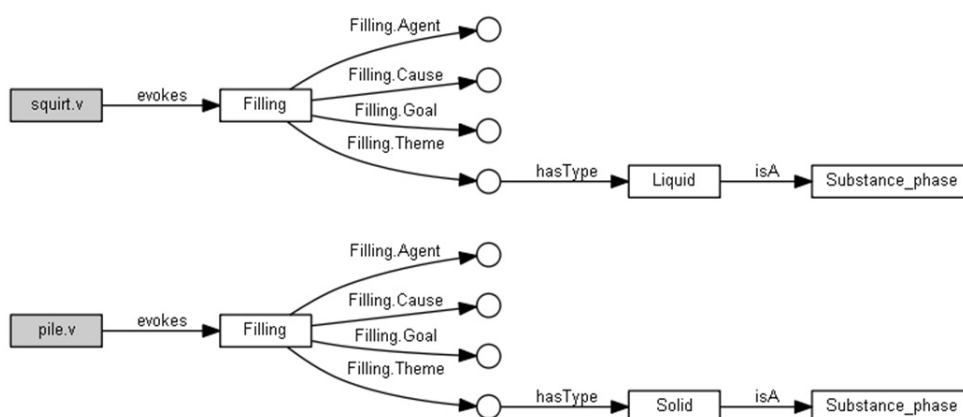
Figura 20 – Type.Aspect.



Os traços associados a TYPE também podem ser usados para caracterizar um *Elemento de Frame* em relação a uma *Unidade Lexical* específica (uma ideia descrita por

Michael Ellsworth no Wiki da FrameNet em 2007, mas ainda não implementada naquele projeto). Neste caso, o *Elemento de Frame* possui um *Tipo Ontológico* associado, e este tipo assume um valor mais específico dependendo da *Unidade Lexical*. Na Figura 21, o *Frame* FILLING possui o *Elemento de Frame* THEME, que tem como um de seus *tipos ontológicos* SUBSTANCE_PHASE, para indicar a fase da substância usada no preenchimento do container. Para a UL <pile.v> este tipo é restrito a SOLID, enquanto para a UL <squirt.v>, o tipo é restrito a LIQUID. Este pequeno acréscimo na estrutura da rede é usado para melhorar o desempenho das tarefas de inferência no *framework* LUDI.

Figura 21 – Exemplo de uso da Classe TYPE.



É importante ressaltar que, da mesma maneira que ocorre com as *features* da ontologia SIMPLE, a classe TYPE é usada somente quando o *tipo ontológico* do traço pode ser representado por um conjunto fechado de valores. Nos demais casos, é usada uma relação com outro *Tipo Ontológico* específico.

7.3 Eventos e Qualidades

A Ontologia LUDI faz a distinção entre três categorias básicas: Entidades, Eventos e Qualidades. Esta abordagem é comum em outras ontologias de topo. A DOLCE (MASOLO et al., 2003), parte da classe ENTITY e distingue entre entidades de informação (INFORMATIONOBJECT), entidades concretas (OBJECTS), entidades abstratas (ABSTRACT), eventos (EVENT) e qualidades (QUALITY). A Brandeis Ontology (PUS-TEJOVSKY et al., 2006), baseada na TLG, distingue entidades, eventos e propriedades. A SIMPLE, também baseada na TLG, parte da classe ENTITY e distingue entre entidades concretas (CONCRETE_ENTITY), entidades abstratas (ABSTRACT_ENTITY), propriedades (PROPERTY), representações (REPRESENTATION) e eventos (EVENT).

No *framework* LUDI, as entidades são categorizadas através da ontologia SIMPLE. Para categorização de eventos e qualidades, está sendo também usada uma abordagem ontológica, descrita nas seções seguintes.

7.3.1 Eventos

Buscando uma forma de organizar os *Frames* usados no *framework* LUDI, foi adotada uma perspectiva ontológica, ou seja, a organização dos *Frames* (e as relações entre eles) segundo uma ontologia. Como no *framework* LUDI a Ontologia de *Frames* está tratando de eventos, é necessário o uso de uma ontologia de eventos. Embora existam várias propostas neste sentido (por exemplo, (RAIMOND; ABDALLAH, 2007),(SCHERP et al., 2009),(KANEIWA; IWAZUME; FUKUDA, 2007)), neste trabalho optamos por adaptar a estrutura de eventos da própria ontologia SIMPLE, uma vez que a SIMPLE tem por motivação a organização do léxico e já está sendo usada neste trabalho para estruturar as entidades. É importante ressaltar que as classes de eventos não visam substituir os *Frames* da FrameNet, mas sim classificá-los.

É feita uma distinção entre **objetos** (chamados de **entidades**, na ontologia LUDI) e **eventos**. Embora a definição destas categorias varie na literatura, adotamos aqui a abordagem da ontologia DOLCE (MASOLO et al., 2003), em que eventos são perdurantes (*perdurants*), ou seja, **ocorrem** ou **acontecem** no tempo, e objetos são endurantes (*endurants*), ou seja, **existem** independente do tempo. A Figura 22 e a Tabela 3 apresentam a ontologia de eventos, definida pelo autor para este trabalho.

Figura 22 – Ontologia de eventos.

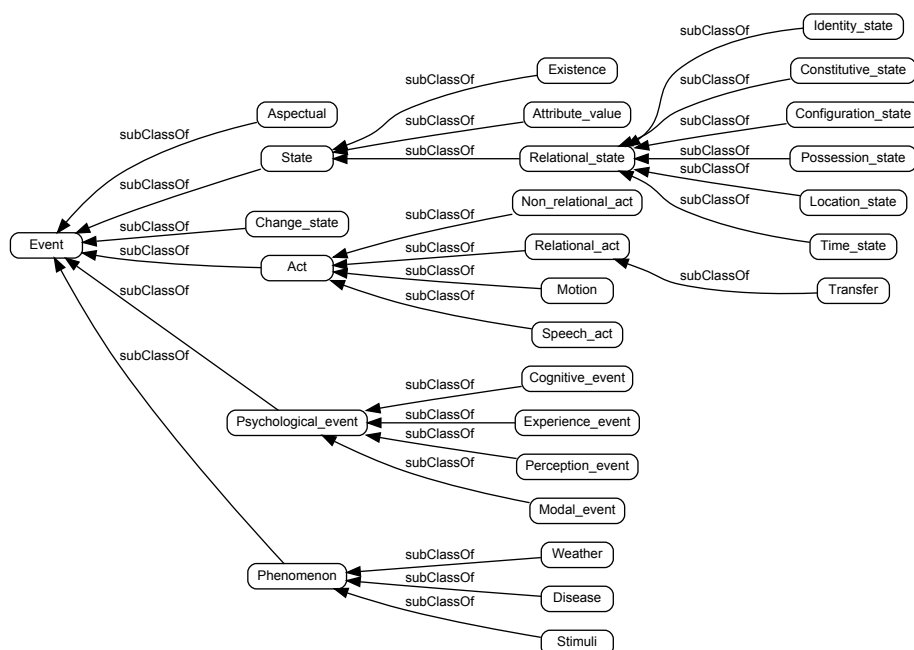


Tabela 3 – Descrição das classes da Ontologia de eventos.

| Classe | Descrição |
|---------------------|---|
| Event | Qualquer coisa que ocorre ou acontece no tempo: processos físicos, sociais ou mentais, atividades, estados e mudanças de estado. |
| Aspectual | Refere-se ao ponto de vista sobre o desenvolvimento do evento. Pode ser considerado sob sua duração, completamento ou repetição (iteração). |
| State | Situações referente ao estado das entidades em determinado instante do tempo. |
| Existence | Estado que representa a existência de uma entidade, independente do ambiente. |
| Attribute_value | O estado é apresentado através de atributos que podem ser valorados. |
| Relational_state | Situação em que o estado de uma entidade é definido em relação ao ambiente ou outra entidade. Associado também as relações não especificadas entre duas ou mais entidades. |
| Identity_state | Situação em que uma entidade é identificada através de uma propriedade específica ou através do modo ou maneira de se relacionar com outra entidade. |
| Constitutive_state | Situação em que é perfilada a constituição de uma entidade, por exemplo, suas partes componentes ou uma característica de sua constituição. |
| Configuration_state | Situação em que é perfilada a configuração da relação entre duas (ou mais) entidades. Geralmente associado a configuração de uma entidade atômica em relação a uma entidade composta. |
| Possession_state | Situação que descreve a relação entre duas entidades, em que uma pode ser conceptualizada como sendo posse ou propriedade de outra. Esta relação é estática e não envolve nenhuma transferência, podendo ser persistente ou temporária. |
| Location_state | Situação que descreve a posição de uma entidade em relação a uma referência (representada por uma ou mais entidades). Pode ser associada também a configuração espacial entre duas ou mais entidades. |
| Time_state | Situação que descreve a relação entre entidades (ou eventos conceptualizados como entidades), em que é perfilada a duração (um relação temporal) entre as ocorrências das mesmas. |
| Act | Situações relacionadas diretamente a uma entidade, ou que envolvem mais de uma entidade, podendo ou não provocar mudanças nos estados destas entidades. |
| Non_relational_act | Atividades que envolvem apenas uma entidade. |
| Relational_act | Atividades que envolvem mais de uma entidade. |
| Transfer | Situação em que ocorre transferência de algum elemento entre entidades. |
| Motion | Situações em que é perfilado o movimento de uma entidade, podendo ou não haver mudança de localização. |
| Speech_act | Situações que envolvem atos de fala. |

| | |
|---------------------|---|
| Psychological_event | Situações que são específicas a "mente humana", englobando tanto situações em que há alguma ação quanto as situações de percepção. |
| Cognitive_event | Situações que envolvem atividades mentais, relacionando um indivíduo a uma proposição. |
| Experience_event | Situações que surgem como resultado do indivíduo experimentar algum outro evento. |
| Perception_event | Situações que envolvem a percepção de um indivíduo. |
| Modal_event | Situações que envolve a relação de um indivíduo com um evento. Este evento é subordinado a uma modalidade gramatical (epistemica, deôntica, apreciativa). |
| Phenomenon | Situações referentes a fenômenos naturais. |
| Weather | Situações referentes a fenômenos ligados ao clima. |
| Disease | Situações referentes a enfermidades. |
| Stimuli | Situações referentes a entidades naturais que provocam algum estímulo. |
| Change_state | Situações que perfilam a mudança de estado das entidades. |

7.3.2 Qualidades

Qualidades podem ser consideradas como as entidades básicas que nós percebemos ou medimos: formas, cores, tamanhos, sons, cheiros, bem como pesos, comprimentos, cargas elétricas. Embora frequentemente "qualidade" seja usada como sinônimo de "propriedade", estas definições dependem de considerações ontológicas. A ontologia DOLCE, por exemplo, faz uma distinção entre "qualidades" e "propriedades": enquanto "qualidades" são particulares (entidades que não possuem instâncias), "propriedades" são universais (entidades que podem possuir instâncias). A cor de um carro específico é uma "qualidade" daquele carro, ao passo que "cor" é uma "propriedade" de carros.

Para a DOLCE, qualidades são inerentes às entidades: cada entidade possui certas qualidades, que existem enquanto a entidade existe. Qualidades físicas são inerentes a entidades físicas, qualidades temporais são inerentes a eventos, qualidades abstratas são inerentes a entidades abstratas. Por exemplo, uma qualidade "temperatura" é inerente a uma certa quantidade de matéria (como uma massa de ar) e uma qualidade "duração" é inerente a um evento, como uma tempestade. Uma qualidade pode ser abstraída de múltiplas instâncias para formar um "tipo de qualidade" (*quality type*) ou "propriedade", tais como cor, tamanho, cheiro. Uma "qualidade" pertence a um "tipo de qualidade". A ontologia DOLCE faz distinção entre uma "qualidade" (por exemplo, a cor de uma rosa específica) e o "valor" desta qualidade (por exemplo, um tom de vermelho específico). Este valor é chamado *quale* e descreve a posição de uma qualidade individual dentro de um espaço conceptual (chamado de "espaço de qualidade" ou "região de qualidade"). Assim, quando dizemos que duas rosas tem exatamente a mesma cor, queremos dizer que suas

"qualidade cor" (que são distintas) tem a mesma posição no "espaço de cor", ou seja, elas tem o mesmo *quale* de cor.

Embora a abordagem da DOLCE seja mais rica do ponto de vista ontológico, para organização do nível estrutural da ontologia LUDI, referente a qualidades, estamos adotando uma abordagem mais simples, baseada nos chamados *espaços conceptuais* (*conceptual spaces*), apresentados em Gärdenfors (2000).

Gärdenfors (2000) modela as "representações" usadas na ciências cognitivas através da introdução da noção de espaços conceptuais (*conceptual space*). **Espaços conceptuais** são coleções de domínios relacionados, onde cada domínio é uma coleção de dimensões (ou *dimensões de qualidade*). Como exemplo de *dimensões de qualidade* pode-se mencionar temperatura, peso, luminosidade, altura do som (grave/agudo) e as três dimensões espaciais (altura, largura e profundidade).

A teoria de espaços conceptuais é baseada na noção de similaridade [(GÄRDENFORS, 2004)]:

(...)judgments of similarity are central for a large number of cognitive processes. Judgments of similarity reveal the dimensions of our perceptions and their structures.¹

Dimensões correspondem às diferentes maneiras pelas quais os estímulos são julgados similares ou diferentes. Neste sentido, eles podem ser usados para representar as várias qualidades dos endurantes. Um ponto em uma *dimensão* "temperatura" pode representar, por exemplo, uma temperatura específica. Assim, a associação de dois endurantes ao mesmo ponto representa o fato experimental de que os dois endurantes são completamente similares em relação à temperatura. Pontos podem ser ordenados em uma escala (ex. Um tom pode ser "alto" ou "baixo") e geralmente assume-se que cada *dimensão* é dotada de uma estrutura matemática (geométrica ou topológica). Assumindo que as *dimensões* possuem alguma métrica, é possível falar de "distâncias" no *espaço conceptual*. Tais "distâncias" indicam o grau de similaridade entre os objetos representados no espaço.

A função primária das *dimensões* é representar as várias qualidades dos objetos em diferentes *domínios* (GÄRDENFORS, 2011). Um **Domínio** pode ser definido como um conjunto de dimensões integrais que são separáveis de todos os outros domínios. Um conjunto de *dimensões* é **integral** se um endurante que tem uma "posição" dentro de uma *dimensão* do conjunto necessariamente tem uma posição dentro de todas as outras *dimensões* do conjunto. Por exemplo, o conjunto {matiz, luminosidade} é integral porque se um endurante tem uma matiz específica, ele necessariamente tem também uma

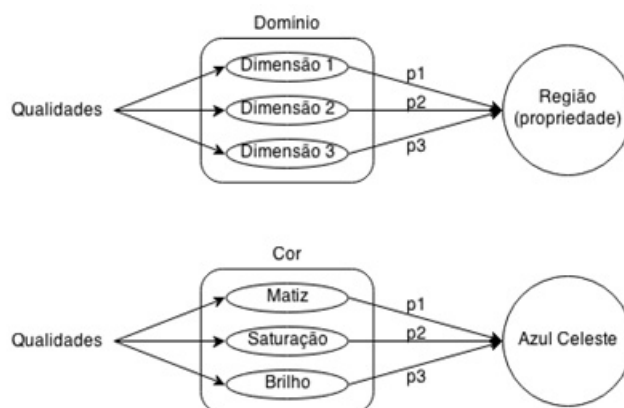
¹ "(...)julgamentos de similaridade são centrais para um grande número de processos cognitivos. Julgamentos de similaridade revelam as dimensões de nossa percepções e suas estruturas." (Tradução nossa)

luminosidade específica. Um conjunto de *dimensões* é **separável** se ele não é integral, por exemplo, {matiz, tamanho}.

Domínios são conjuntos maximais de dimensões integrais. Por exemplo, as três *dimensões* das cores - matiz, saturação e luminosidade - formam um *domínio* porque o conjunto {matiz, saturação, luminosidade} é integral. As *dimensões* podem ser usadas para atribuir "propriedades" para os endurantes, ou seja, classificar os endurantes: uma propriedade específica corresponde a uma *região* em um *domínio*. A restrição de separabilidade permite atribuir *propriedades* (*regiões* em um *domínio*) de maneira independente de outras *propriedades* (*regiões* em outros *domínios*). Isto captura o fato experimental de que o peso de um endurente é independente da cor do endurente.

A [Figura 23](#) busca ilustrar estes conceitos através de um exemplo, para melhor compreensão, sintetizando a descrição feita até aqui. As *qualidades* são representadas por *dimensões*. Um conjunto integral de *dimensões* forma um *domínio*. Um valor para a *dimensão* representa um *ponto* (p1, p2, p3, na figura) no *espaço conceptual*. Estes *pontos* formam uma *região*, representando uma *propriedade*. Assim, o *domínio* "cor" é formado pelas *dimensões* {matiz, saturação, brilho}. Um valor específico para cada *dimensão* constrói a "propriedade" azul celeste. Um enunciado como "o céu está azul" pode ser lido como "a entidade *céu* tem a propriedade *azul* para seu domínio *cor*", e o enunciado "a sopa está salgada" pode ser lido mais especificamente como "a entidade *sopa* tem a propriedade *alta* para a dimensão *sal* do seu domínio *gosto*".

Figura 23 – Espaços Conceptuais.



A FrameNet atualmente trata as qualidades/propriedades em três *clusters* de *Frames*: Atributos, Dimensões e Medições. O *Frame* (não-lexical) ATTRIBUTES define simplesmente que um objeto tem um determinado valor para um atributo. O *Frame* (não-lexical) MEASURES, que serve de pai na hierarquia de *Frames* relacionados a medições, define que uma entidade é medida em relação a um atributo, sendo feita a quantificação em uma unidade específica que serve para medir aquele atributo. Já o *Frame*

DIMENSION (lexical) define que ele trata de palavras que expressam a posição de um objeto em uma escala, ou seja, sua medida com respeito a algum atributo (sua dimensão).

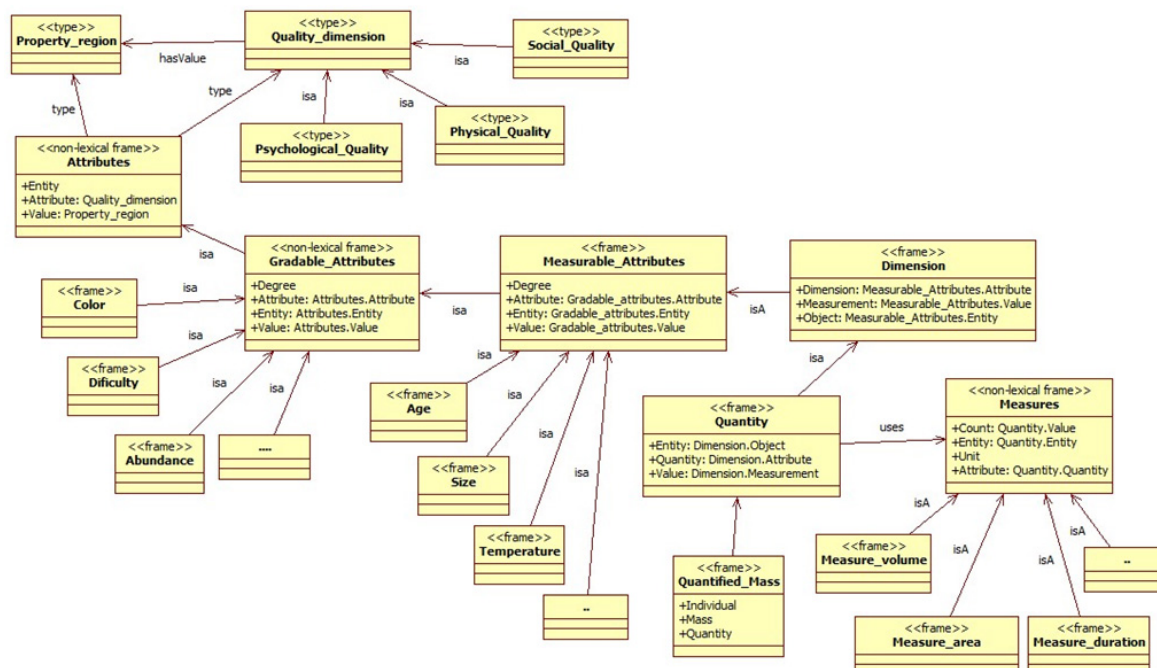
Como se pode observar, existe uma certa confusão relativa a estes três *clusters* (uma evidência disto é a quantidade de *unidades lexicais* marcadas com o *status* "Problem", no *Frame* DIMENSION). Além da questão terminológica (qual a diferença entre atributos e dimensões? e entre atributos e propriedades? e entre entidades e objetos?), estes *Frames*, embora estejam se referindo a propriedades, não estão conectados entre si (exceto por relações *See_also*). Registre-se também problemas na hierarquia: o *Frame* COLOR herda de GRADABLE_ATTRIBUTES, mas este não faz nenhuma referência ao *Frame* ATTRIBUTES.

É importante notar que, uma vez que a FrameNet, como um recurso lexical, não tem um compromisso ontológico ou um alinhamento com ontologias (FILLMORE; BAKER, 2010), algumas definições tornam mais difícil a tarefa de inferência. Por exemplo, ainda no âmbito das propriedades, não é feita uma distinção clara entre uma qualidade e o valor que ela pode assumir. No *Frame* COLOR encontramos tanto as *Unidades Lexicais* referentes as cores (red, blue, etc) quanto a própria *Unidade Lexical* color. Algo semelhante acontece com o *Frame* SHAPES onde, entre as *Unidades Lexicais* esperadas para este *Frame* (circular, oval, etc), encontramos a *Unidade Lexical* shape. Ainda que reconheçamos que color e shape possam, metonimicamente, serem usadas no lugar do valor da propriedade (ex. "Ele usou azul claro na parede da sala" e "Ele usou cor clara na parede da sala"), grande parte das anotações destas *Unidades Lexicais* na FrameNet se referem à qualidade e não ao valor da qualidade. Em termos de inferências, esta é uma distinção importante.

Para realizar as inferências no *framework* LUDI é fundamental que as relações semânticas estejam bem estabelecidas. Assim, as relações entre *Frames* estão sendo adaptadas, bem como os tipos semânticos da ontologia SIMPLE relativos à qualidades. Embora fosse necessária uma reestruturação dos *Frames* para se trabalhar mais completamente com a ideia de *espaços conceituais*, neste trabalho optou-se por uma simples adaptação das relações entre *Frames*, mantendo-se os *Frames* atuais, uma vez que a reconfiguração da rede de *Frames* da FrameNet está fora do escopo desta pesquisa. A Figura 24 apresenta a estrutura adaptada.

Basicamente a hierarquia de qualidades passa a estar organizada em torno do *Frame* (não-lexical) ATTRIBUTES. Em termos terminológicos, *atributos* e *dimensões* na FrameNet se referem a qualidades e os *valores de atributos* se referem a propriedades. Uma pequena ontologia de qualidades é criada, a fim de classificar os *Frames* (como foi feito em relação aos eventos). A Tabela 4 apresenta as definições dos tipos criados e a Tabela 5 traz a definição dos *Frames* (adaptando a definição presente na FrameNet, quando possível). Ressalte-se que, em uma ontologia mais completa, cada qualidade poderia ser representada como uma classe da ontologia. No LUDI estamos trabalhando com as qualidades de

Figura 24 – Estrutura para Qualidades.



maneira simplificada, usando a ideia de "tipos", apresentada na seção 7.2.

Tabela 4 – Qualidades: Tipos semânticos.

| Classe | Descrição |
|-----------------------|---|
| Quality_dimension | Se refere às "dimensões de qualidade", sendo subclassificadas em qualidades físicas, psicológicas ou sociais. |
| Psychological_quality | Se refere às qualidades psicológicas, tais como inteligência, coragem, atenção, cuidado, intuição, cinismo, etc. |
| Physical_quality | Se refere às qualidades físicas, tais como cor, gosto, peso, forma, altura, profundidade, comprimento, tamanho, carga, temperatura, velocidade, etc. |
| Social_quality | Se refere às qualidades sob uma perspectiva social, tais como autoridade, influência, representatividade, etc. |
| Property_region | Uma classe genérica para propriedades (valores de qualidades). Os <i>Frames</i> e <i>Elementos de Frame</i> referentes a propriedades devem ser classificados nesta classe. |

Tabela 5 – Qualidades: Frames.

| Classe | Tipo | Descrição | Exemplos de ULs |
|------------|-------------|--|-----------------|
| Attributes | Não-lexical | Uma Entidade tem um Atributo específico com algum Valor. Este <i>Frame</i> é usado primariamente para herança. | - |

| | | | |
|-----------------------|-------------|--|--------------------------------|
| Gradable_attributes | Não-lexical | Uma Entidade tem um Valor implícito para um Atributo. O EF Grau expressa uma comparação explícita ou um desvio do Valor quando comparado com outras entidades de um tipo similar. | - |
| Measurable_attributes | Lexical | Uma Entidade tem um Atributo escalar específico, com um valor que excede um padrão (especificado contextualmente ou geralmente compreendido). Adjetivos relacionados a quantidades mensuráveis são colocados neste <i>Frame</i> . Quando uma expressão de quantidade aparece junto com o adjetivo, ele é colocado no <i>Frame</i> Dimension. | alto, pesado, baixo, fundo |
| Dimension | Lexical | Este <i>Frame</i> se refere a palavras que expressam a posição de um Objeto numa escala (sua Medida) com respeito a algum atributo (sua Dimensão). | largura, comprimento, altura |
| Quantity | Lexical | Este <i>Frame</i> contém nome que denotam Quantidades de uma Entidade especificada e seus Valores numéricos. | quantidade, nível |
| Measures | Não-lexical | Uma Entidade é medida em relação a algum Atributo (implícito) através da quantificação (Contador) de uma Unidade específica para a qual o valor para o atributo é conhecido. | - |
| Quantified_mass | Lexical | Este <i>Frame</i> contém nomes transparentes (e alguns adjetivos) que denotam quantidades de Massa ou de Individuos. | um pouco, muito, pilha, vários |

7.4 Mecanismos gerativos

Uma questão fundamental, subjacente a este trabalho, é referente ao problema da seleção argumental (*argument selection*), ou seja, a relação existente entre o tipo de um argumento conforme requerido pelo predicado e o papel que este argumento tem na computação do significado da sentença. Pustejovsky (2006) discute a Decomposição Lexical e a Teoria de Tipos, mostrando que na teoria padrão de seleção considera-se que existe um inventário de tipos, associados com as entidades do domínio. As restrições seletivas impostas aos argumentos pelo verbo são herdadas do tipo associado com aquele argumento, de acordo com o tipo funcional atribuído ao verbo. Estas restrições são geralmente fracas e, se outras restrições devem ser impostas ao argumento, estas são construídas como pressuposições durante a interpretação.

Como apresentado na seção 3.2, a abordagem da TLG difere da teoria padrão em dois aspectos. Primeiro, as restrições seletivas impostas aos argumentos dos verbos tornam-se, de forma transparente, parte do tipo do próprio verbo. Isto implica em enriquecer o sistema de tipos manipulados pelas regras de composição (o que leva a TLG a trabalhar com um sistema de tipos em três níveis: **tipos simples** ou naturais, **tipos unificados** ou artefatuais e **tipos complexos** ou *dot-types*). Uma segunda diferença é que os mecanismos de seleção disponíveis não são apenas a aplicação de uma função aos seus argumentos, mas envolvem três operações sensíveis aos tipos: **co-composição** (*matching*), **seleção** e **coerção de tipo**, apresentadas brevemente na seção 3.2.

Jezeck e Lenci (2007), investigando a questão dos tipos semânticos e do fenômeno de coerção em relação a dados retirados de corpus, argumentam que os mecanismos gerativos no léxico, segundo a TLG, tem tanto uma perspectiva paradigmática quanto sintagmática. O aspecto paradigmático está relacionado a geratividade do sistema de tipos, uma vez que a estrutura *qualia* e as operações "dot" permitem criar tipos lexicais multidimensionais. Já a geratividade composicional, através das operações de composição que mudam e criam tipos no contexto, constitui o aspecto sintagmático.

Um dos objetivos no *framework* LUDI é fornecer subsídios que permitam a análise destes mecanismos gerativos. Como a definição destes mecanismos vem sendo adaptada desde a sua apresentação em Pustejovsky (1991), é importante ressaltar como eles estão sendo considerados aqui. Seguimos a classificação apresentada em Pustejovsky (2006) e usada por Jezeck e Lenci (2007), que consideram como questão chave a seleção argumental. Neste sentido, os mecanismos discutidos são a **seleção**, englobando as operações de **pura seleção** e **acomodação** e a **coerção de tipo**, subdividido nas operações de **exploração** e **introdução**.

A **pura seleção** ocorre quando o tipo que uma função (via verbo) requer dos seus argumentos é diretamente satisfeito pelo tipo do argumento. Na **acomodação** o tipo que a função requer é herdado através do tipo do argumento. Na **coerção** ocorre um choque sintagmático entre o tipo que uma função requer e o tipo do argumento. Quando isto acontece, ou a interpretação pode falhar completamente, ou então o tipo requerido é "imposto" ao argumento. Isto é obtido através da exploração ou da introdução. Na **exploração** o tipo requerido corresponde a parte da estrutura *qualia* do argumento. Um subcomponente do tipo do argumento é acessado para satisfazer os requisitos do predicado. Na **introdução**, o tipo requerido é mais rico que o tipo do argumento. O argumento é "empacotado" ou "envolvido" (*wrapped*) com o tipo requerido pelo predicado. Globalmente a teoria agora permite 9 possibilidades (Tabela 6). Além destas operações, que preservam o domínio, a TLG também inclui a co-composição, que licencia novas interpretações do predicado no contexto. Como ressaltado por Jezeck e Lenci (2007), uma vez que tanto a seleção quanto a co-composição podem operar simultaneamente no mesmo contexto

sintagmático, o quadro de "o que acontece onde" na combinação de palavras, conforme vai ocorrendo a construção do significado, não é algo fácil de reconstruir.

Tabela 6 – Operações de composição da TLG.

| Tipo do Argumento | Tipo requerido | | |
|------------------------|------------------|------------------------|----------------|
| | Simple (natural) | Unificado (artefatual) | Complexo (dot) |
| Simple (natural) | Seleção | Introdução | Introdução |
| Unificado (artefatual) | Exploração | Seleção | Introdução |
| Complexo (dot) | Exploração | Exploração | Seleção |

Fonte: Adpatado de (Jezek e Lenci, 2007).

Alguns exemplos, adaptados de Pustejovsky (2006), ajudam a compreender as interações entre estas operações. Nas sentenças em (7.1), os tipos dos argumentos são indicados entre parênteses, sendo que o símbolo \otimes representa um tipo unificado (com $x=T$ para tético e $x=A$ para agentivo) e o símbolo \cdot representa um tipo complexo.

- (7.1) (a) A pedra caiu no chão. (*phys*)
 (b) A comida estragou. (*phys* \otimes *comer*)
 (c) O livro caiu no chão. (*phys* \cdot *info* \otimes *ler* \otimes_A *escrever*)
 (d) A água estragou. (*phys*)

O mecanismo no exemplo (7.1)(a) é de **pura seleção**: o verbo <cair> requer um argumento do tipo físico e <pedra> é um tipo natural e físico. O mecanismo no exemplo (7.1)(b) também é **pura seleção**: o verbo <estragar> seleciona uma entidade artefactual como sujeito e <comida> atende este requisito. O mecanismo no exemplo (7.1)(c) é de **exploração**: <cair> seleciona um objeto físico e parte da estrutura *qualia* de <livro> (que é do tipo unificado) atende ao requisito. Finalmente, o exemplo (7.1)(d) apresenta uma **introdução**: <estragar> requer uma entidade artefactual, mas <água> é do tipo simples (natural). Para satisfazer os requisitos de tipo do predicado, é necessário a coerção, que vai "empacotar" o tipo natural com uma interpretação funcional, ou seja, para que a sentença faça sentido, a água deve estar sendo usada para alguma coisa (*qualia* tético). Este é o tipo de análise que o *framework* LUDI pretende possibilitar, através do alinhamento da FrameNet e da ontologia SIMPLE.

7.5 Papéis temáticos e relações qualia

Papéis temáticos representam relações semânticas que conectam indivíduos a um fenômeno. No contexto da Semântica de Frames, os elementos de um *Frame* podem ser considerados, na realidade, funções microtemáticas (SALOMÃO, 2009), já que representam relações bastante específicas entre a situação representada pelo *Frame* e os participantes desta

situação. Por outro lado, as relações *qualia* são também relações semânticas. A integração entre papéis temáticos e a estrutura *qualia* parece, então, ser bastante natural.

Para realizar esta integração no *framework* LUDI, estamos adotando, com algumas adaptações, a abordagem proposta por Sowa (1996, p. 12) e atualizada em Sowa (2000). A Tabela 7 mostra como os participantes de uma situação são classificados de acordo com a distinção binária entre determinante vs. imanente e fonte vs. produto.

Tabela 7 – Classificação de participantes.

| | Fonte | Produto |
|---------------------|--------------|----------------|
| Determinante | Iniciador | Objetivo |
| Imanente | Recurso | Essência |

Fonte: Sowa (2000) (traduzido pelo autor).

Cada participante é uma entidade que desempenha um papel em um processo. A divisão da Tabela 7 mostra estes participantes segundo o papel que desempenham. Na língua natural estas distinções são expressas através de marcadores gramaticais (ex. preposições), que ligam o verbo que expressa o processo aos nomes que expressam os participantes. Em lógica, estas distinções são expressas por relações ou predicados, que ligam os símbolos que identificam os processos aos símbolos que identificam os participantes. A distinção entre o tipo de participação é definida assim, por Sowa (2000):

- a) Um participante **Determinante** é aquele que determina a direção do processo, a partir do início, como um Iniciador, ou do fim, como um Objetivo;
- b) Um participante **Imanente** está presente através do processo, mas não controla ativamente o que acontece;
- c) Uma **Fonte** deve estar presente no início do processo, mas não precisa participar durante este processo;
- d) Um **Produto** deve estar presente no fim do processo, mas não precisa participar deste processo.

A partir das combinações destas distinções, Sowa estrutura os papéis temáticos em torno dos quatro *aitiai* (ou causas) de Aristóteles, de forma semelhante ao que Pustejovsky (1995) faz em relação a estrutura *qualia*:

1. **Iniciador** corresponde à causa eficiente, que inicia uma mudança ou um estado;
2. **Recurso** corresponde à causa material, que é a matéria ou substrato;
3. **Objetivo** corresponde à causa final, que é o propósito ou o benefício, o objetivo (*telos*) de qualquer geração ou movimento;

4. **Essência** corresponde à causa formal, que é a essência (*ousia*) ou "o que a coisa é".

Para Sowa (1996, p. 14) as relações *qualia* são, na verdade, relações inversas aos papéis temáticos. Ao invés de focar nos processos representados pelos verbos, as relações *qualia* focam nos objetos representados pelos nomes. Para um tipo específico de objeto, um romance, por exemplo, são analisados os tipos de processos cuja ação sobre o objeto pode ser esperada. Sowa sumariza esta diferença da seguinte forma:

- a) Papéis temáticos: da perspectiva de um processo, os papéis temáticos representam as quatro maneiras pelas quais os objetos podem participar no processo;
- b) Qualia: da perspectiva de um objeto, o *qualia* representa as quatro maneiras pelas quais os processos podem afetar (ou serem afetados) pelo objeto.

Ainda segundo Sowa (1996), uma vez que os processos são interações entre objetos, estas duas perspectivas levam a maneiras complementares de usar a *aitiai* de Aristóteles para analisar, classificar e interpretar as iterações.

No *framework* LUDI, estamos adaptando a terminologia de Sowa àquela usada na TLG, conforme a Tabela 8.

Tabela 8 – Terminologia Sowa vs. TLG.

| Sowa | Estrutura Qualia (TLG) |
|-----------|------------------------|
| Iniciador | Agentivo |
| Recurso | Constitutivo |
| Objetivo | Télico |
| Essencia | Formal |

Fonte: Adpatado de (Sowa, 1999).

A Tabela 9 apresenta as papéis temáticos definidos em Sowa (2000), já com a terminologia adaptada.

No caso do *framework* LUDI, os processos (em verdade, eventos em geral) são tratados via Semântica de Frames e os objetos (em verdade, entidades) são associadas a ontologia SIMPLE. A ontologia SIMPLE implementa tipos multidimensionais, chamados **Tipos Unificados**, que possibilitam a herança ortogonal. A herança ortogonal essencialmente enriquece os modelos de herança convencionais através da definição de relações semânticas em múltiplas dimensões, neste caso dadas minimamente pela estrutura *qualia*. Um conceito da ontologia é então apresentado através dos seus componentes de significado, que coexistem no conceito.

No *framework* LUDI os *Tipos Unificados* possuem um papel relevante, pois é principalmente através deles que os mecanismos gerativos são executados. *Tipos Unificados*

Tabela 9 – Papéis temáticos.

| Papel genérico | Papel temático | Descrição |
|------------------|----------------|--|
| AgentiveRole | Agent | Uma entidade ativa animada que voluntariamente inicia uma ação. |
| | Effector | Uma Fonte Determinante, animada ou inanimada, que inicia uma ação, mas sem intenção voluntária. |
| | Origin | Uma Fonte Determinante passiva associada a uma situação espacial ou ambiental. |
| | Start | Uma Fonte Determinante de uma situação temporal. |
| ConstitutiveRole | Instrument | Um Recurso que não é alterado por um evento. |
| | Matter | Um Recurso que é alterado pelo evento. |
| | Medium | Um Recurso físico para transmitir informações. |
| | Path | Um Recurso em uma situação espacial. |
| | Duration | Um Recurso em um processo temporal. |
| FormalRole | Patient | Um participante essencial que sofre alguma mudança estrutural como resultado de um evento. |
| | Theme | Uma Fonte Determinante que pode ser "movid", "dita", "experenciada", mas que não é alterada estruturalmente. |
| | Location | Um participante essencial de uma situação espacial. |
| | PointInTime | Um participante essencial de uma situação temporal. |
| TelicRole | Experiencer | Um Objetivo ativo animado de uma experiência. |
| | Recipient | Um Objetivo animado de um ato. |
| | Beneficiary | Um Recipiente que deriva uma benefício de um evento completado com sucesso. |
| | Result | Um Objetivo inanimado de um ato. |
| | Destination | Um Objetivo de um processo espacial. |
| | Completion | Um Objetivo de um processo temporal. |

Fonte: Adpatado de (Sowa, 1999).

precisam de múltiplas coordenadas para serem identificados e não apenas a sua posição na estrutura hierárquica. Além do supertipo ao qual eles são subordinados, as dimensões agentiva e télica entram também na constituição do tipo, ou seja, é explicitada a informação agentiva e/ou télica que inerentemente caracteriza a essência daquele tipo. Novos *Tipos Unificados* podem ser facilmente criados na ontologia, tomando um tipo simples como base e acrescentando alguma informação referente aos papéis agentivo ou télico, através das relações *qualia*. Um *Tipo Unificado*, por sua vez, pode ser tratado como o topo de uma nova hierarquia de tipos organizados de maneira taxonômica. Esta flexibilidade, derivada da herança ortogonal, permite estender facilmente a ontologia SIMPLE para um domínio mais específico.

Na Ontologia SIMPLE as dimensões agentiva e télica para um dado tipo semântico

são inferidas pela participação do tipo em relações de caráter agentivo ou télico. A ontologia apresenta várias relações associadas a cada dimensão. No *framework* LUDI estamos estendendo ainda mais esta ideia, adicionando relações inversas às relações SIMPLE. A [Tabela 10](#) apresenta estas relações que, de maneira geral podem ser lidas como:

- a) *Y is-agentive-of X*, expressando que o tipo/entidade Y representa a dimensão agentiva do tipo/entidade X;
- b) *X is-telic-of Y*, expressando que o tipo/entidade X representa a dimensão télica do tipo/entidade Y.

Essencialmente, no *framework* LUDI estamos considerando que as relações de caráter agentivo e télico são inversas entre si.

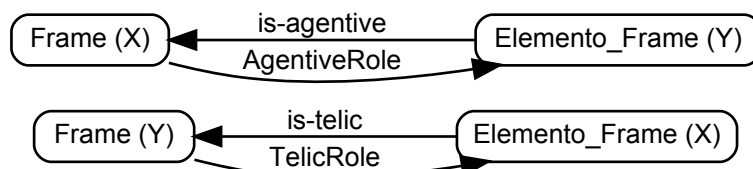
Tabela 10 – Relações SIMPLE e relações inversas.

| Relações SIMPLE (X is-telic-of Y) | Relações Inversas (Y is-agentive-of X) |
|-----------------------------------|--|
| Result_of | Results |
| Resulting_From | Results |
| Agentive_prog | Is_agentive_prog_of |
| Affected_by | Affects |
| Created_by | Creates |
| Derived_from | Derives |
| Agentive_experience | Is_agentive_experience_of |
| Caused_by | Causes |
| Source | Origins |
| Agentive_cause | Is_agentive_cause_of |
| Object_of_the_activity | Uses_object |
| Used_for | Uses |
| Used_against | Uses |
| Used_as | Uses |
| Used_by | Uses |
| Is_the_activity_of | Has_the_activity_of |
| Is_the_ability_of | Has_the_ability_of |
| Is_the_habit_of | Has_the_habit_of |
| Purpose | Has_as_purpose |

Finalmente, é possível fazer a integração com os papéis temáticos, conforme a [Figura 25](#). A figura mostra que, se um *Elemento de Frame* tem um papel temático agentivo (*AgentiveRole*) em relação a um *Frame*, então a relação *qualia* entre o *Elemento de Frame* e o *Frame* é uma relação agentiva; igualmente se um *Elemento de Frame* tem um papel

temático télico (*TelicRole*) em relação a um *Frame*, então a relação *qualia* entre o *Elemento de Frame* e o *Frame* é uma relação télica. Na prática, isto significa que o *Elemento de Frame* pode representar a dimensão agentiva ou télica de um tipo semântico da ontologia SIMPLE.

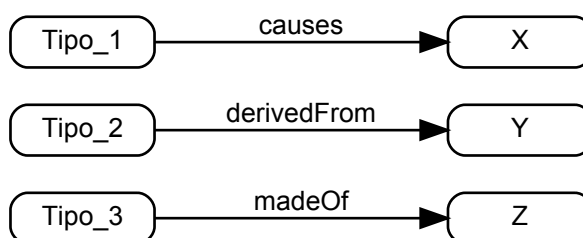
Figura 25 – Integração entre *Frames* e Relações *Qualia*.



Este fato é fundamental para o processamento das inferências no *framework* LUDI, pois ele permite que a associação de um *Elemento de Frame* com um *Tipo Ontológico* seja feita **dinamicamente**, com base no contexto. Crucialmente, ele permite implementar em um modelo conexionista, não-simbólico, as ideias que Pustejovsky (1998b) discute sob a subespecificação de itens lexicais do ponto de vista de uma *forma lógica* (*logical form*). A implementação do *framework* LUDI reforça a posição de Pustejovsky (1998b) de que as representações subespecificadas são parte do processo composicional em si mesmo, onde a especificação do significado emerge do contexto.

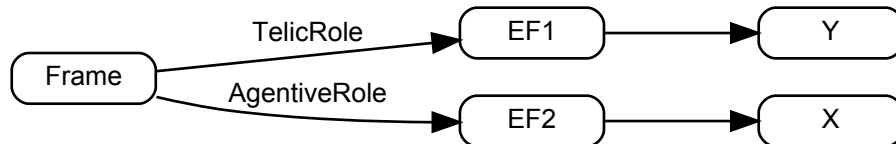
Para implementar estas ideias no modelo conexionista, a estratégia foi a criação de classes genéricas que representassem as dimensões agentiva, télica e constitutiva. Desta forma, são criadas as classes X (dimensão télica), Y (dimensão agentiva) e Z (dimensão constitutiva). As relações *qualia* podem assim serem representadas de maneira subespecificada, como na Figura 26. Os Tipos Unificados TIPO_1, TIPO_2 e TIPO_3 possuem, respectivamente, uma dimensão télica, agentiva e constitutiva, ainda que os argumentos das relações não sejam conhecidos. Estas relações podem ser parafraseadas como "TIPO_1 causa alguma coisa", "TIPO_2 deriva de alguma coisa" e "TIPO_3 é feito de alguma coisa".

Figura 26 – Relações *qualia* e classes genéricas.



Por outro lado, estas classes genéricas são usadas também para representar o papel temático dos *Elementos de Frame*. Um *Elemento de Frame* com papel temático que seja um *AgentiveRole* é associado com a classe X, enquanto um *Elemento de Frame* que seja um *TelicRole* é associado com a classe Y (Figura 27). Estes links representam a relação *qualia* do *Elemento de Frame* com o *Frame*.

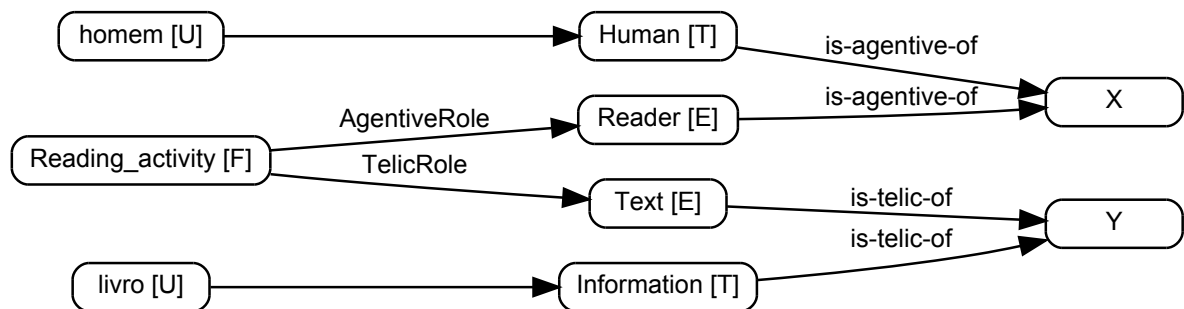
Figura 27 – Papéis temáticos e classes genéricas.



Durante o processamento da inferência, torna-se possível, então, associar dinamicamente um *Elemento de Frame* com um *Tipo Ontológico*, através destas classes genéricas. A Figura 28 mostra um exemplo da integração entre as relações *qualia* e os papéis temáticos para a sentença 7.2.

(7.2) O homem leu o livro.

Figura 28 – Integração entre relações *qualia* e papéis temáticos.



8 Exemplos de Inferências

A fim de ilustrar o funcionamento da rede conexionista do *framework* LUDI no processo de desambiguação de *Unidades Lexicais*, neste capítulo são apresentados exemplos de inferências envolvendo o uso de qualidades e eventos. É mostrado como o processo de composição de itens lexicais via relação de tipos semânticos pode ser usado na compreensão de fenômenos associados às restrições seletivas (como a coerção) e à atribuição de papéis temáticos (via *Elementos de Frames*).

8.1 Qualidades

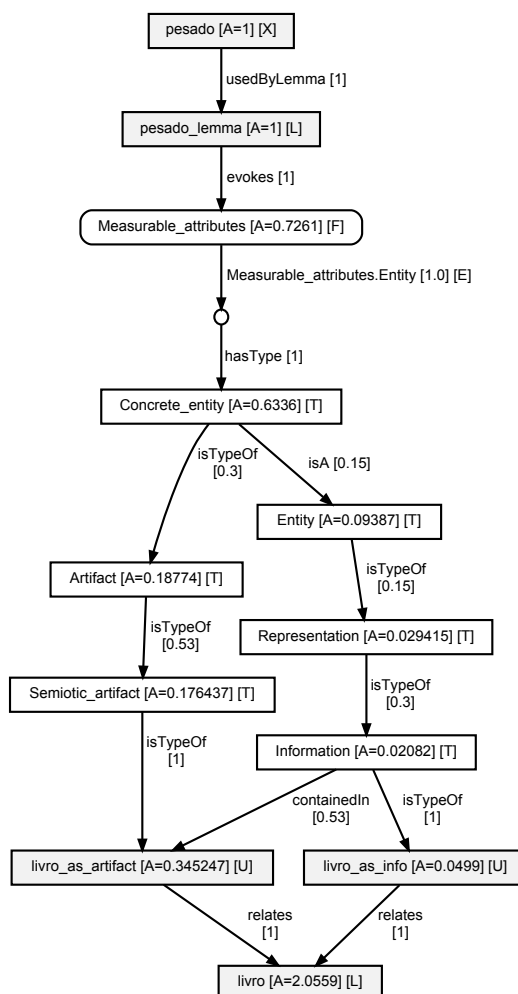
Em relação às qualidades, dois tipos de situação são exemplificados. Uma envolve o uso de diferentes qualidades associadas a um mesmo nome de entidade e a outra relaciona-se ao uso da mesma qualidade para diferentes entidades.

A primeira situação envolve a noção de *Tipos Complexos*, conforme apresentada por Pustejovsky (2006). Os *Tipos Complexos* são representados pelo operador *dot*. Por exemplo

livro : phys . info

indica que <livro> tem tanto um sentido de "objeto físico" quanto de "informação". Segundo Pustejovsky, é necessário um mecanismo gerativo de seleção, para a compreensão correta de uma expressão em que <livro> seja usado com um adjetivo. No *framework* LUDI esta seleção é feita de forma automática pela rede. Na Figura 29 é apresentada a rede para a expressão <livro pesado>, após o processo de ativação. Observe-se que o valor da ativação (indicado pela letra A) para *livro_as_artifact* é bastante superior a de *livro_as_info*. É importante ressaltar que o processo não apenas "realiza" a seleção, mas é possível observar "como" a seleção foi feita. A esta rede final gerada pelo processo é que estamos denominando *Especificação Semântica (SemSpec - Semantic Specification)*. Na Figura 29 observa-se que a seleção de *livro_as_artifact* deve-se não apenas a sua classificação como um artefato (*SEMIOTIC_ARTIFACT*) mas também a contribuição da relação de *qualia* constitutivo *containedIn* (faz parte da constituição de livros conterem informações).

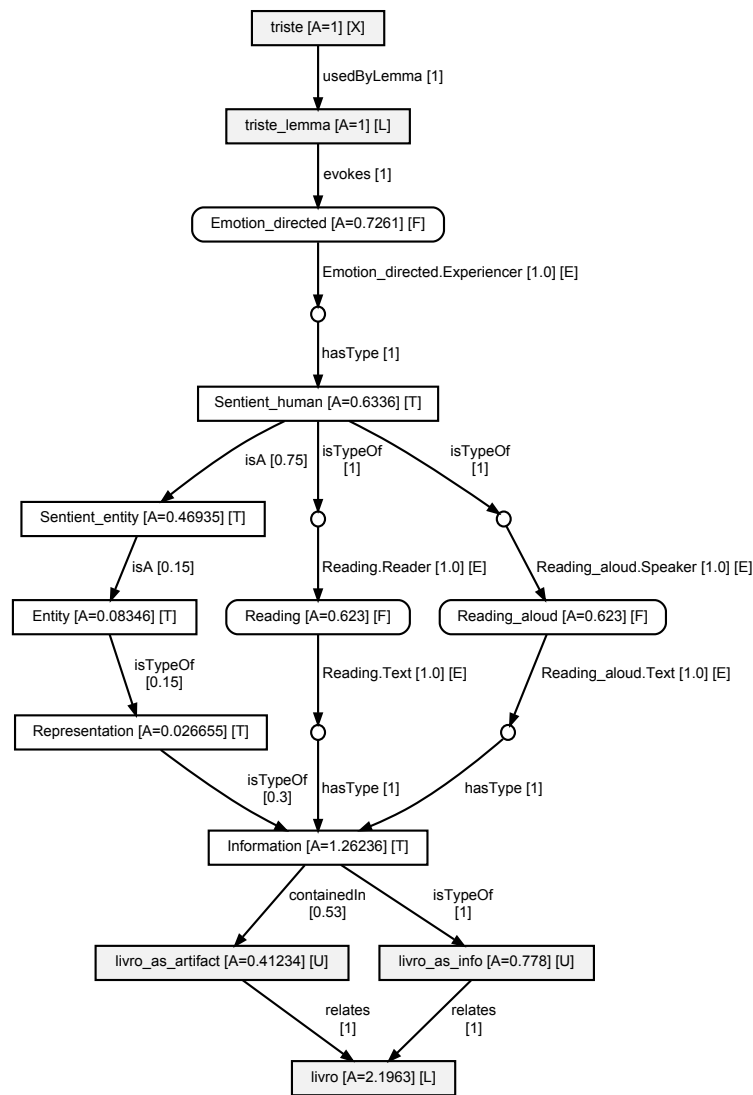
Figura 29 – <livro pesado> após ativação da rede.



Nas Figura 30 é apresentada a rede para a expressão <livro triste>, após o processo de ativação. Este exemplo mostra o uso de *links inibitórios* (*links* que quando ativados por um nó contribuem para reduzir a ativação de outros nós). Na rede (antes da ativação) há um *link inibitório* entre SENTIENT_HUMAN e CONCRETE_ENTITY, a fim de registrar que, neste modelo, entidades concretas não podem ser entidades sencientes. Este *link inibitório* é que diminui a ativação de livro_as_artifact. Esta figura mostra também a integração entre a FrameNet e a ontologia SIMPLE: pode-se observar a associação entre o *Frame* READING e o *tipo ontológico* INFORMATION, via o *Elemento de Frame* READING.TEXT. Através de relações como estas, pode ser feita a associação de Entidades com Eventos (e vice-versa). Além disso, na Figura 30 observa-se que os dois sentidos de livro (livro_as_artifact e livro_as_info) foram atingidos pelo processo de ativação da rede, no entanto com valores de ativação diferentes. Isto demonstra a característica *fuzzy* do processo, ou seja, a classificação é feita com base em "graus de verdade" e não em termos binários de verdadeiro/falso. O maior valor de ativação para livro_as_info demonstra que esta seria interpretação preferida, caso tivéssemos que escolher entre as duas. Esta parece ser também a interpretação mais

natural para um enunciado como "este livro é muito triste", que aponta para a informação e não para o objeto.

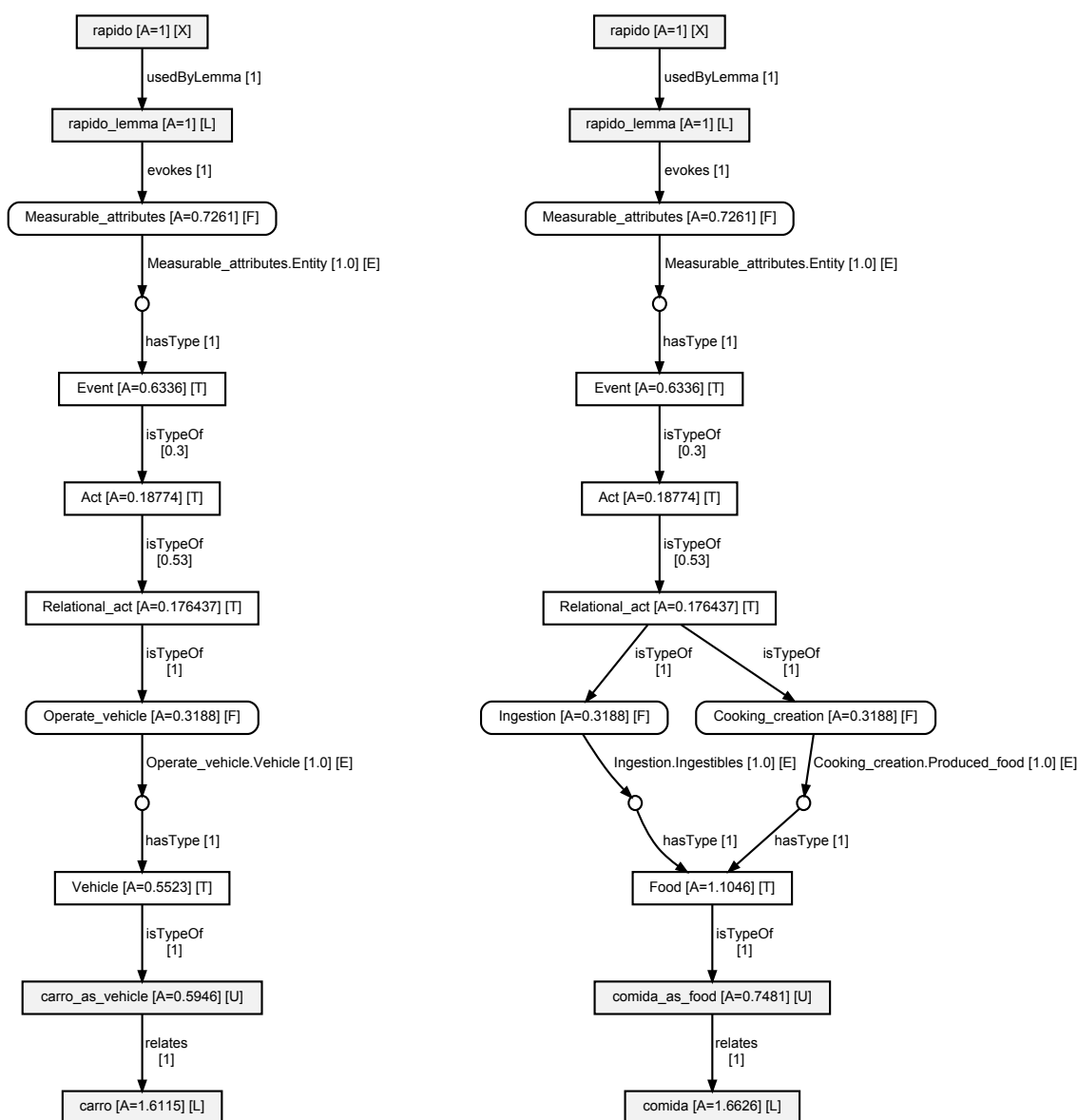
Figura 30 – <livro triste> após a ativação da rede.



A segunda situação, em relação a adjetivos, denominada **polissemia adjetival** em (PUSTEJOVSKY, 1995), se refere ao fato que a mesma qualidade pode ser compreendida de maneira diferente, dependendo da entidade que é qualificada. Ilustramos esta situação aqui, contrastando as expressões <carro rápido> e <comida rápida>. A TLG mostra que as interpretações destas duas expressões são muito diferentes, devido a estrutura *qualia* de <carro> e de <comida>. No *framework* LUDI isto fica claramente demonstrado. Na Figura 31, a rede para <carro rápido> mostra que <rápido> (que é uma qualidade para eventos) pode ser atribuído a <carro> (uma entidade) na interpretação de que este é o objeto da atividade <operar veículo> (representada pelo *Frame* OPERATE_VEHICLE). Não é o carro, em si, que é rápido: a sua operação é que está se processando rapidamente.

Algo semelhante ocorre em relação a <comida rapida>. A Figura 31 mostra que <comida> é objeto das ações representadas nos *Frames* cozinhar (COOKING_CREATION) ou ingerir (INGESTION). Mais uma vez, não é a comida, em si, que é rápida: ela pode ser "feita" rapidamente ou ela pode ser "ingerida" rapidamente. A especificação semântica representada pela rede expõe claramente estas interpretações. Uma interpretação mais específica - uma decisão entre cozinhar e ingerir - depende apenas de termos mais elementos do contexto.

Figura 31 – <comida rapida> e <carro rapido> após a ativação

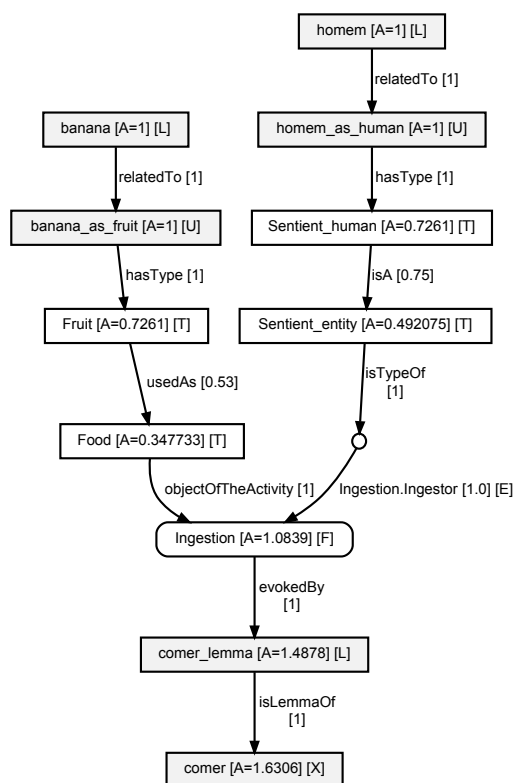


8.2 Eventos

Para iniciar os exemplos de aplicação em relação a eventos, a Figura 32 apresenta uma situação em que a interpretação é direta, ou seja, não é necessário um processo de

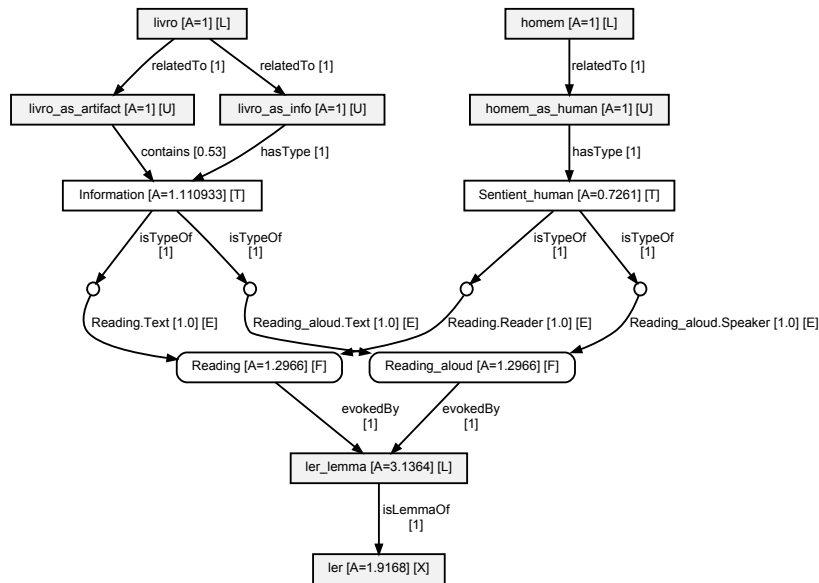
desambiguação. Na expressão <homem comer banana> o lexema <comer.v> evoca apenas um *Frame* (INGESTION) e, portanto, a expressão tem uma única interpretação. O *SemSpec* apenas mostra como esta interpretação é construída. Ressalte-se neste exemplo o uso da relação *qualia* entre FRUIT e FOOD (relação *usedAs*), e a atribuição automática dos *Elementos de Frame* (o processo de SRL - *Semantic Role Labeling*), com base nos *Tipos Ontológicos*.

Figura 32 – *SemSpec* de <homem comer banana>.



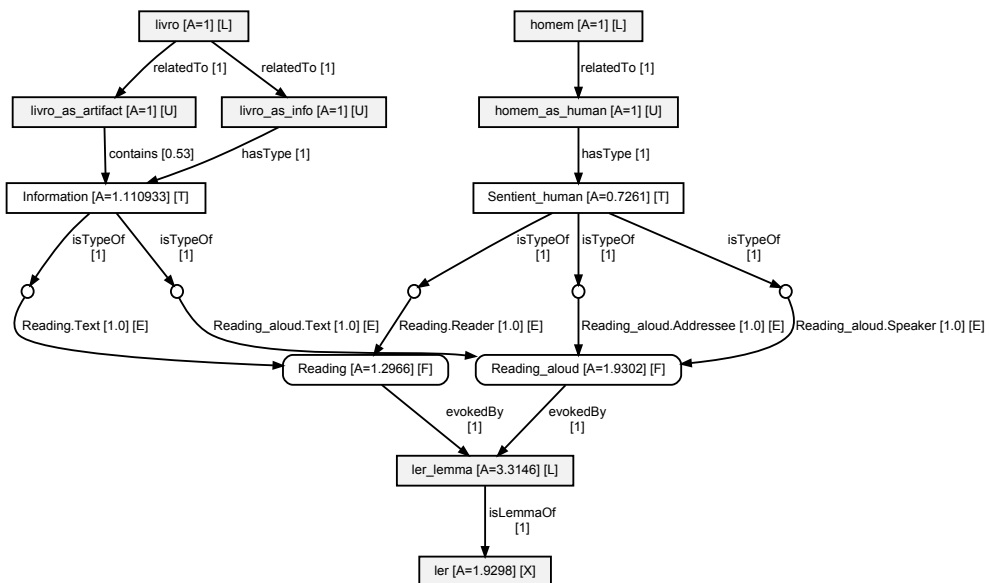
Uma situação mais interessante é aquela em que o verbo apresenta-se polissêmico. Na FrameNet isto é representado através da criação de uma *Unidade Lexical* referente a cada sentido, onde cada *Unidade Lexical* está associada a um *Frame* distinto. Como exemplo, a FrameNet apresenta dois sentidos para o verbo <ler.v> (*read.v*), sendo evocados os *Frames*: READING e READING_ALOUD. A Figura 33 apresenta o *SemSpec* para a expressão <homem ler livro>. Como foram fornecidos dois argumentos e todos os *Frames* possuem *Elementos de Frame* compatíveis com esses argumentos, o resultado é que os dois *Frames* são ativados igualmente (valor de ativação = 1.2966); ou seja, a ambiguidade permanece.

Figura 33 – *SemSpec* de <homem ler livro>.



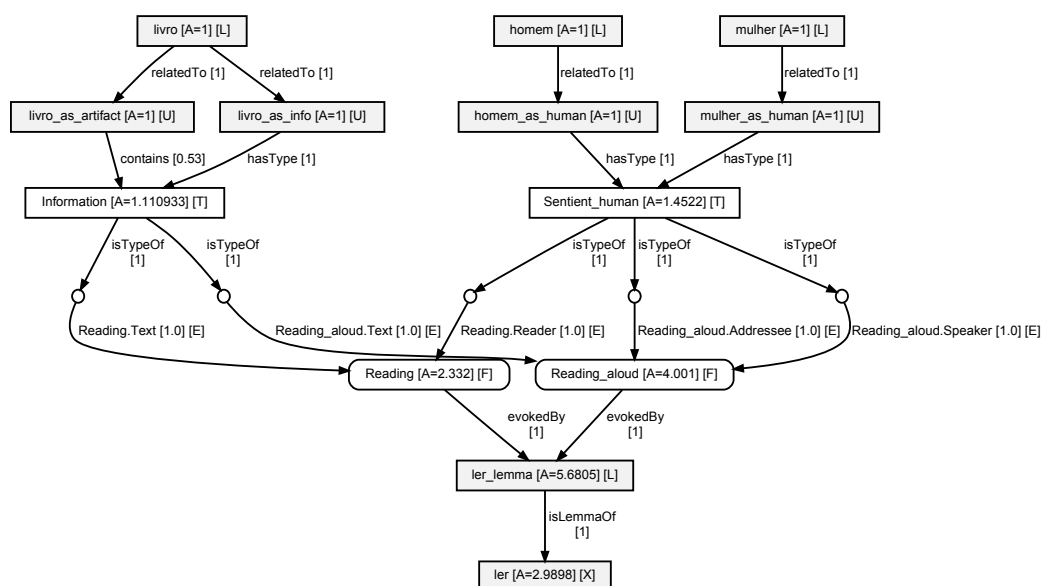
Como uma tentativa de desambiguação, podemos considerar outros *Elementos de Frame*. Por exemplo, o *Frame* READ_ALOUD possui o elemento ADDRESSEE (que indica o destinatário da mensagem que está sendo lida em voz alta). A *Figura 34* mostra o *SemSpec* com a inclusão do *Elemento de Frame* ADDRESSEE. Agora o *Frame* READ_ALOUD tem um valor de ativação maior. Mas, exceto se considerarmos que o homem está lendo em voz alta para si mesmo (o que é uma interpretação possível), temos apenas dois argumentos para três *Elementos de Frame*, o que pode gerar um resultado incorreto.

Figura 34 – *SemSpec* de <homem ler livro> - possivelmente com erro.



Consideremos então a inclusão de mais um argumento, por exemplo <homem ler livro mulher> (em uma sentença como "O homem lê o livro para a mulher"). A Figura 35 mostra o *SemSpec* com a inclusão do argumento. Novamente o *Frame* READ_ALOUD tem um valor de ativação maior, mas o problema agora é que não há uma distinção dos *Elementos de Frame*. Como <homem> e <mulher> estão associados ao *Tipo Ontológico* HUMAN, o *SemSpec* não deixa claro quem está associado ao *Elemento de Frame* SPEAKER e quem está associado a ADDRESSEE.

Figura 35 – *SemSpec* de <homem ler livro mulher> - com ambiguidade.



Para realizar a desambiguação podemos recorrer às funções gramaticais. Através das valências da *Unidade Lexical* em cada *Frame* (registradas na FrameNet), podemos refinar o processo. Consideremos, de forma simplificada (neste exemplo), apenas as funções Externo (Ext), Objeto (Obj) e Dependente (Dep). Se acrescentarmos, como dado de entrada, a função gramatical pretendida para cada argumento e incluirmos na rede conexionista as informações sobre as valências possíveis para a *Unidade Lexical* sob análise, o processo de desambiguação pode fornecer resultados mais confiáveis. Isto é ilustrado na Figura 36, onde o *SemSpec* mostra uma interpretação mais compatível com os dados de entrada.

As valências são representadas na rede através de nós ligados ao *Frame*. Isto possibilita representar várias valências associadas à mesma *Unidade Lexical*, quando necessário. Nestes casos, é possível também comparar os valores de ativação de cada valência. A associação dos argumentos com os *Elementos de Frame* é feita automaticamente, através de manutenção da "memória" da unidade ativadora. Desta forma, a cada passo da ativação é possível saber qual foi a unidade de origem. A relação *hasRole* mostra a associação do lema com o *Elemento de Frame* correspondente.

Parte IV

Verificação Empírica do *framework* LUDI

9 Ponto de partida: os sentidos de <abrir.v>

9.1 A análise das sentenças

De acordo com a metodologia adotada neste trabalho (capítulo 2), a busca de evidência em corpus é fundamental por dois grandes motivos. Em primeiro lugar, ainda que a intuição do analista como usuário falante da língua seja rica, o uso feito pelos outros milhões de falantes é muito mais rico. Isto implica dizer que uma categorização, como aquela que será apresentada a seguir, precisa ser testada contra um número razoável de sentenças em corpus, a fim de avaliar sua coerência. Em segundo lugar, ainda associado a riqueza do uso da língua, uma análise superficial de algumas centenas de sentenças é suficiente para concluir que qualquer categorização realizada é bastante limitada. Isto não é surpresa, se adotamos como hipótese de trabalho para representação das conceptualizações o uso de esquemas imagéticos (ao invés de uma solução meramente simbólica). A riqueza conceptual de cada esquema, acrescida das possibilidades de combinações e transformações, torna virtualmente impossível uma categorização completa. Assim, o uso de corpus para avaliar a categorização apresentada a seguir visa também testar a sua abrangência, sem a ilusão de que uma cobertura de 100% pode ser alcançada. No entanto, dado o viés computacional deste trabalho, "alguma" cobertura é necessária, e a análise de sentenças torna-se de importância capital.

Os *corpora* foram consultados através da aplicação SketchEngine, um sistema de consulta que incorpora esboços de texto, uma página, ou resumo. Os corpus usados são apresentados na Tabela 11 (as siglas foram definidas no escopo deste trabalho e são usadas para referência, nas sentenças apresentadas ao longo do texto).

Dois fenômenos de interesse para a presente análise se destacaram, entre vários outros, tornando-a mais complexa: a alternância causativo-incoativo e a numerosa ocorrência de metonímias.

9.1.1 Alternância causativo-incoativo

A alternância causativa é caracterizada por verbos que possuem um uso tanto intransitivo quanto transitivo, onde o uso intransitivo tipicamente denota um evento de mudança de estado sofrido por alguma entidade e o uso transitivo denota que este evento de mudança de estado foi acarretado ou causado por uma outra entidade (SCHÄFER, 2009). O uso transitivo é, por esta razão, frequentemente parafraseado como "causar V-intransitivo", como em (9.1).

Tabela 11 – Corpora utilizados.

| Sigla | Corpus | Descrição |
|-------|--|--|
| CT | CETENFolha/CETEMPúblico | O CETENFolha (Corpus de Extractos de Textos Electrónicos NILC/Folha de S. Paulo) é um corpus de cerca de 24 milhões de palavras em português brasileiro, criado pelo projecto Processamento computacional do português (projeto que deu origem a Linguateca) com base nos textos do jornal Folha de São Paulo que fazem parte do corpus NILC/São Carlos, compilado pelo Núcleo Interinstitucional de Linguística Computacional (NILC). O CETEMPúblico (Corpus de Extractos de Textos Electrónicos MCT/Público) é um corpus de aproximadamente 180 milhões de palavras em português europeu, criado pelo projeto Processamento Computacional do Português (projeto que deu origem à Linguateca), após a assinatura de um protocolo entre o Ministério da Ciência e da Tecnologia (MCT) português e o jornal PÚBLICO em Abril de 2000. |
| NURC | Projeto da Norma Urbana Oral Culta do Rio de Janeiro | Corpus constituído por entrevistas gravadas nas décadas de 1970 e 1990, num total de 350 horas, com informantes de nível superior completo, nascidos no Rio de Janeiro e filhos de pais, preferencialmente, cariocas. |
| DP | Domínio Público | Corpus composto de obras literárias do português brasileiro do século XIX em diante e obras traduzidas para o Português Brasileiro, presentes no site do Domínio Público, totalizando mais de 500 obras literárias. |

- (9.1) (a) A porta abriu.
 (b) O menino abriu a porta.

Enquanto a sentença intransitiva (9.1)(a) denota uma simples mudança de estado da entidade "porta", a sentença transitiva (9.1)(b) expressa que "o menino causou a porta estar aberta". Uma característica central desta alternância de transitividade é que o sujeito no uso intransitivo possui a mesma relação semântica com o verbo que o objeto no uso transitivo.

No uso intransitivo, os verbos que sofrem a alternância causativa são geralmente chamados de verbos anticausativos ou incoativos. No seu uso transitivo, eles são chamados de causativos (lexicais). A alternância causativa é também rotulada de alternância causativo-incoativo ou alternância anticausativa ou, ainda, de alternância ergativa.

A alternância causativo-incoativo é de interesse neste trabalho, pois é citada explicitamente na metodologia de desenvolvimento da FrameNet [Ruppenhofer et al. \(2010, c. 2\)](#). No caso da FrameNet a opção foi por registrar os relacionamentos entre os *Frames* estativos e os *Frames* causativo (onde a ocorrência de uma causa ou agente é explícita) e incoativo (onde o desencadeador dos eventos não é focalizado). As relações *Causative_of*

e *Inchoative_of* são usadas para este fim. Os *Frames* que participam destas relações como Causativos deveriam herdar do *Frame* TRANSITIVE_ACTION, os Incoativos deveriam herdar do *Frame* EVENT e os Estativos do *Frame* STATE ou GRADABLE_ATTRIBUTES. No entanto, muitas destas relações ainda não foram registradas diretamente na FrameNet.

9.1.2 Metonímia

Tradicionalmente, a metonímia é definida como o uso de uma palavra (ou um sintagma) para se referir ao significado de um conceito relacionado, mas que não é explicitamente mencionado. Ela está baseada na contiguidade e implica, de certa forma, um contato ou uma conexão física entre as entidades (SHUTOVA et al., 2013). A metonímia é geralmente explicada através de "padrões metonímicos" convencionalizados que operam sobre classes semânticas (LAKOFF; JOHNSON, 1987, p. 35), com o uso de uma entidade em lugar de outra. Estes padrões ocorrem sistematicamente na linguagem e são também denominados "polissemia regular" ou "extensão do sentido", por outros autores. De fato, as ocorrências de uso metonímico com o verbo <abrir.v>, encontradas durante a análise, são tão numerosas que quase configuram a regra, ao invés da exceção. Alguns padrões comuns de substituição com o verbo <abrir.v> são ilustrados a seguir.

- Parte-pelo-todo: A sala está aberta. [a porta da sala]
- Container-pelo-conteúdo: Abre o guaraná primeiro. [a garrafa com guaraná]
- Evento-por-entidade: O museu já está aberto. [a visita ao museu]
- Produto-por-produtor: No jantar ele abriu um Bordeaux. [a garrafa de vinho Bordeaux]
- Pessoas-por-instituição: A universidade está aberta a sugestões. [Os administradores da universidade]

O interesse pelas ocorrências metonímicas, neste trabalho, está relacionado não só a sua frequência, mas também como uma possibilidade de análise dos mecanismos gerativos apresentados na seção 7.4.

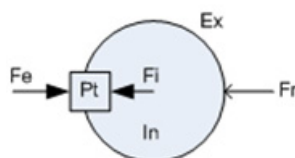
9.2 O esquema básico de <abrir.v>

Usando uma abordagem cognitiva, os muitos sentidos de <abrir.v> podem ser projetados a partir da junção do esquema imagético CONTAINER e da noção de dinâmica de forças (*force-dynamics*). Os sentidos são derivados através das operações de perfilamento, perspectivação e transformações no esquema imagético, bem como sua projeção em domínios mais abstratos, via metáforas.

O *Frame* esquemático para esta análise de <abrir.v> (Figura 37) é um CONTAI-

NER sobre cujo portal podem ser aplicadas duas forças (esquema FORCE), através de Agonista e de um Antagonista. Este é um esquema conceptual básico, no sentido que ele serve para a projeção dos vários sentidos de <abrir.v> apresentados neste trabalho. Assim, os elementos deste esquema somente são "realizados" nos *Frames*.

Figura 37 – Esquema para <abrir.v>.



Os elementos apresentados na [Figura 37](#) são detalhados na [Tabela 12](#).

Tabela 12 – Elementos do esquema para <abrir.v>.

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

O uso de um esquema básico, a partir do qual são conceptualizados os diversos sentidos de <abrir.v> , visa ressaltar o caráter polissêmico do lexema.

9.2.1 Dinâmica de forças

A noção de dinâmica de forças (apresentada no [Apêndice C](#)) é muito importante para a compreensão do esquema básico. Neste esquema a Força Externa (Fe) atua através da ação de um Agonista (AGO), enquanto a Força Interna (Fi) atua sobre um Antagonista (ANT), representado pelo Portal. A conceptualização de <abrir.v> está associada aos padrões de deslocamento do ANT. Por um lado, o exercício da Força Interna, no ato de abrir, conceptualiza a ideia de "permissão" ou "desbloqueio", através da mudança de estado dos

elementos da região interior. Por outro lado, o exercício da Força Externa, no ato de abrir, conceptualiza a ideia de "superação", obtida através de alguma ação dos elementos da região exterior. Em ambos os casos, é eliminada a "resistência" oferecida pelo Portal.

9.2.2 Perfilamento

A partir do esquema básico, três situações podem ser tomadas em perspectiva:

- a) O foco na ação: neste caso o foco está na força exercida sobre o portal, através de uma CAUSA, um AGENTE ou um INSTRUMENTO. Esta situação é perfilada nos *Frames* causativos.
- b) O foco no container/portal: neste caso o foco está no elemento (CONTAINER ou PORTAL) que sofre a ação da força. Esta situação é perfilada nos *Frames* incoativos.
- c) O foco no estado: neste caso o foco é no estado ("aberto") do container; este estado é alcançado através da ação que muda o estado de "fechado" para "aberto", tornando possível o acesso aos elementos interiores. Esta situação é perfilada nos *Frames* estativos.

9.3 Os *Frames* de <abrir.v>

9.3.1 O esquema básico

O *Frame* não-lexical ABRIR_ESQUEMA é usado para representar o esquema básico; sua definição é apresentada a seguir.

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

A discussão sobre causação ([Apêndice B](#)) é importante para uma definição mais precisa dos *Elementos de Frame* do *CoreSet* (Agente, Causa, Instrumento):

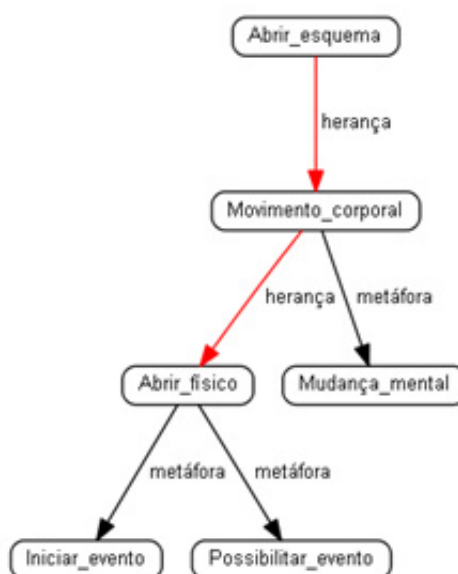
- a) Agente: representa uma entidade consciente - dotada de volição e que age com intenção - responsável pelo evento causador de <abrir.v>.

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

Em relação ao *CoreSet* (Container, Portal) é importante ressaltar a relação parte-todo existente. O Portal é "parte" do Container. Como discutido anteriormente o padrão metonímico "Parte-pelo-todo" é bastante comum, de maneira que as sentenças podem lexicalizar tanto o Portal quanto o Container.

A partir do esquema básico, uma rede de *clusters* de sentidos é apresentada na [Figura 38](#). Como pode ser observado, é definido um *cluster* para o sentido concreto e corporificado de <abrir.v> como Movimento_corporal, que é herdado pelo sentido, também concreto, de Abrir_físico. A partir deste sentido, os demais são obtidos através de interpretações que consideram o uso de metáforas.

Figura 38 – Cluster de *Frames* herdeiros de ABRIR_ESQUEMA.



A rede de *clusters* é expandida na [Figura 39](#). Basicamente é considerada a alternância causativo-incoativo e são adicionados alguns *Frames* já definidos pela FrameNet (o posicionamento destes *Frames* na rede foi definido pelo autor). A fim de facilitar a referência ao longo do texto, a [Figura 40](#) apresenta a numeração dos *clusters* e dos *Frames*. O detalhamento dos *Frames* é feito nas seções seguintes.

Figura 39 – Rede de Frames herdeiros de ABRIR_ESQUEMA.

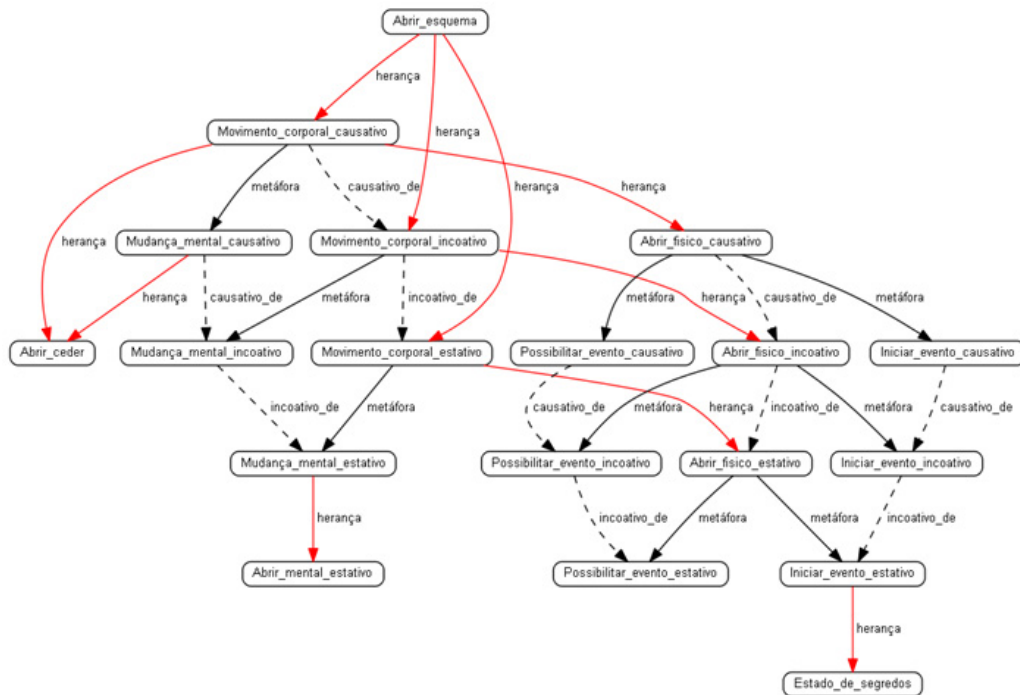


Figura 40 – Frames por Clusters.

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

9.3.2 Corpo físico

Considerando a hipótese da cognição corporificada (*embodied cognition*) (LAKOFF, 2012) e a ideia de esquemas imagéticos como padrões da experiência senso-motora (JOHNSON, 1989), o corpo físico de seres humanos ou de animais podem ser conceptualizados como containers: há uma "região interior" que é o interior do corpo, há uma "fronteira" estabelecida pela nossa extensão corporal, há "portais" como os órgãos de sentido (que permitem o acesso a elementos exteriores), a boca (que permite a introdução de elementos) e os órgãos excretórios (que possibilitam a saída de elementos). Como containers, os corpos são passíveis de serem "abertos"; partes do corpo podem ser conceptualizadas tanto como portais ou como containers. Neste sentido o ato de <abrir.v> também representa a possibilidade de entrada/saída de elementos externos/internos ao corpo.

| II.1 - MOVIMENTO_CORPORAL_CAUSATIVO | | |
|--|------------|----------------------|
| Um AGENTE, um INSTRUMENTO ou uma CAUSA exerce uma força sobre a PARTE_CORPORAL, modificando seu estado. O EXPERENCIADOR que possui a PARTE_CORPORAL pode ser informado. | | |
| Elemento de Frame | Tipo | EF do esquema básico |
| Agente | Core | Agente |
| Causa | Core | Causa |
| Instrumento | Core | Instrumento |
| Parte_corporal | Core | Portal |
| Experenciador | Periférico | Container |
| Unidades Lexicais: <abrir.v>,<abertura.n> | | |
| Exemplos construídos [João] AGENTE abriu [os olhos] PARTE_CORPORAL. [Maria] AGENTE abriu [os braços] PARTE_CORPORAL para recebê-lo. [O bisturi] INSTRUMENTO abriu [o abdomen]PARTE_CORPORAL em um único movimento. [A dor] CAUSA abriu [a boca] PARTE_CORPORAL [de João] EXPERENCIADOR. No passo seguinte, [ela] AGENTE fez uma abertura [de pernas] PARTE_CORPORAL perfeita. | | |
| Exemplos de uso em corpus e textos DP: Quando [eu] AGENTE abro [a boca] PARTE_CORPORAL e digo um soneto de Olavo Bilac (...) DP: No refeitório não se abria [a boca] PARTE_CORPORAL, não se falava nada (...) DP: Quando [eu] AGENTE abri [os olhos] PARTE_CORPORAL já era meia-noite (...) | | |

| II.2 - MOVIMENTO_CORPORAL_INCOATIVO | | |
|---|------------|----------------------|
| Uma PARTE_CORPORAL de um EXPERENCIADOR sofre uma ação (não especificada) que modifica seu estado. | | |
| Elemento de Frame | Tipo | EF do esquema básico |
| Parte_corporal | Core | Portal |
| Experenciador | Periférico | Container |
| Unidades Lexicais: <abrir.v> | | |
| Exemplos construídos [A boca] PARTE_CORPORAL [de João] EXPERENCIADOR abriu num grito. [O punho] PARTE_CORPORAL se abriu num gesto de paz. [Mãos] PARTE_CORPORAL se abriam para receber a comida. | | |
| Exemplos de uso em corpus e textos NURC: Quando [eu] EXPERENCIADOR durmo [minhas mãos] PARTE_CORPORAL se abrem. | | |

| II.3 - MOVIMENTO_CORPORAL_ESTATIVO | | |
|---|-------------|-----------------------------|
| Uma PARTE_CORPORAL de um EXPERENCIADOR tem seu estado descrito como “aberto”. | | |
| Elemento de Frame | Tipo | EF do esquema básico |
| Parte_corporal | Core | Portal |
| Experenciador | Periférico | Container |
| Unidades Lexicais: <aberto.a> | | |
| Exemplos construídos | | |
| [Ela] EXPERENCIADOR estava com [a boca] PARTE_CORPORAL aberta . | | |
| [João] EXPERENCIADOR mantinha [os olhos] PARTE_CORPORAL abertos . | | |
| [A mão] PARTE_CORPORAL aberta pedia ajuda. | | |
| Exemplos de uso em corpus e textos | | |
| NURC: [Dois palmos] PARTE_CORPORAL abertos embaixo e dois em cima | | |
| NURC: E [as pessoas] EXPERENCIADOR que têm [as pernas] PARTE_CORPORAL abertas (...) arcadas | | |

9.3.3 Entidades físicas

Nos *Frames* do domínio de entidades físicas, os *Elementos de Frame* (container, portal, etc) representam entidades do mundo físico. Este sentido de <abrir.v> envolve a noção de uma entidade concreta sendo perfilada como um container (um artefato, um local, uma entidade física), que está fechado ou vedado por alguma outra entidade que representa um portal ou a própria fronteira do container (podendo ser perfilado como uma barreira); <abrir.v> perfila a ação que pode ser aplicada sobre esta entidade (uma força externa), opcionalmente possibilitando o acesso de elementos que estão na região exterior a região interior (ou vice-versa). É possível também o registro de alguma resistência, representada por uma força interna, exercida direta ou indiretamente por elementos que estão na região interior.

| III.1 - ABRIR_FISICO_CAUSATIVO | | |
|---|------------|----------------------|
| Um AGENTE, um INSTRUMENTO ou uma CAUSA exerce uma força sobre o PORTAL de um CONTEINER, criando uma abertura, removendo uma barreira ou manipulando um fecho, o que possibilita acesso ao CONTEUDO ou à região INTERIOR do CONTEINER. | | |
| Elemento de Frame | Tipo | EF do esquema básico |
| Agente | Core | Agente |
| Causa | Core | Causa |
| Instrumento | Core | Instrumento |
| Conteiner | Core | Conteiner |
| Portal | Core | Portal |
| Conteúdo | Periférico | Conteúdo |
| Interior | Periférico | Interior |
| Exterior | Periférico | Exterior |
| Fronteira | Periférico | Fronteira |
| Unidades Lexicais: <abrir.v>, <abertura.n> | | |
| Exemplos construídos | | |
| [João] AGENTE abriu [a mochila] CONTEINER. | | |
| [O vento] CAUSA abriu [a janela] PORTAL [da sala] CONTEINER. | | |
| [Os tiros] CAUSA abriram [buracos] PORTAL [na parede] FRONTEIRA. | | |
| [Maria] AGENTE abriu [o pote] CONTEINER [de biscoitos] CONTEUDO. | | |
| [A chave] INSTRUMENTO abre [o baú] CONTEINER facilmente. | | |
| [Ele] AGENTE abriu [cada botão] PORTAL [da camisa] CONTEINER. | | |
| Vagarosamente [ele] AGENTE fez uma abertura [na parede] FRONTEIRA [da cela] CONTAINER. | | |
| Exemplos de uso em corpus e textos | | |
| NURC: [Você] AGENTE abria [a torneira] PORTAL [de água quente] CONTEUDO | | |
| NURC: [Você] AGENTE abre [a janela] PORTAL e dá de cada com um velhinho | | |
| NURC: [A gente] AGENTE foi abrir [caminho] PORTAL para o boi trazer a madeira | | |
| NURC: [O sujeito] AGENTE abre [um jornal] CONTEINER de domingo | | |

| III.2 - ABRIR_FISICO_INCOATIVO | | |
|--|------------|----------------------|
| Um PORTAL de um CONTEINER sofre uma ação (não especificada), criando uma abertura, removendo uma barreira ou manipulando um fecho, o que possibilita acesso ao CONTEUDO ou à região INTERIOR do CONTEINER. | | |
| Elemento de Frame | Tipo | EF do esquema básico |
| Conteiner | Core | Conteiner |
| Portal | Core | Portal |
| Conteúdo | Periférico | Conteúdo |
| Interior | Periférico | Interior |
| Exterior | Periférico | Exterior |
| Fronteira | Periférico | Fronteira |
| Unidades Lexicais: <abrir.v> | | |
| Exemplos construídos | | |
| [A porta] PORTAL abriu de repente. | | |
| Exemplos de uso em corpus e textos | | |
| CT: [O túnel Dois Irmãos] PORTAL que está para abrir todo dia | | |
| NURC: [A carteira] CONTEINER abria e [lá dentro] INTERIOR a gente botava as coisas da gente | | |
| NURC: [O carro] CONTEINER simplesmente não abria [a porta] CONTEINER, entendeu? | | |

9.3.4 Aspecto

Através da metáfora EVENTOS ESTÃO CONTIDOS EM CONTEINERS, é possível o uso de <abrir.v> como o aspecto inceptivo de uma atividade ou processo. Tomando como

| III.3 - ABRIR_FISICO_ESTATIVO | | |
|---|-------------|-----------------------------|
| Um CONTAINER tem seu PORTAL conceptualizado como uma abertura ou como uma passagem que está aberta, ou como uma barreira que foi removida, ou como um fecho que foi manipulado, o que possibilita acesso ao CONTEUDO ou à região INTERIOR do CONTAINER (ou o acesso do CONTEUDO à região EXTERIOR). | | |
| Elemento de Frame | Tipo | EF do esquema básico |
| Container | Core | Container |
| Portal | Core | Portal |
| Conteúdo | Periférico | Conteúdo |
| Interior | Periférico | Interior |
| Exterior | Periférico | Exterior |
| Fronteira | Periférico | Fronteira |
| Unidades Lexicais: <aberto.a> | | |
| Exemplos construídos | | |
| [A sala] CONTAINER está aberta . | | |
| Ele usou [o refrigerante] CONTEUDO [da garrafa] CONTAINER aberta . | | |
| [A fumaça] CONTEUDO saiu [pela janela] PORTAL aberta [na cozinha] CONTAINER. | | |
| Exemplos de uso em corpus e textos | | |
| NURC: Já [a igreja] CONTAINER é aberta [ao público em geral] EXTERIOR. | | |
| NURC: A gente tinha [uma picada] PORTAL aberta [no meio do mato] FRONTEIRA que nos levava a uma praia | | |
| NURC: Então, eu comia muita [esfirra] CONTAINER aberta ou fechada também. | | |

domínio-fonte o ESPAÇO e como domínio-alvo o TEMPO, podem ser traçadas algumas correspondências ontológicas: a fronteira de um container delimita certa porção do espaço (sua extensão), a fronteira de um evento (constituída pelos momentos de início e de fim) delimita certa porção de tempo (sua duração); um container pode ser aberto e fechado, um evento pode ser iniciado e terminado; para abrir um container é necessária a ocorrência de uma causa, para iniciar um evento também é necessária a ocorrência de uma causa; o acesso ao conteúdo de um container através portal é geralmente sequencial, a ocorrência dos eventos é geralmente sequencial.

Assim, neste sentido aspectual, <abrir.v> representa o evento inicial de uma possível série de outros eventos. O argumento de <abrir.v> será, via de regra, um evento (lexicalizado por um verbo ou um deverbal). Caso o argumento seja um nome de entidade, as ações tipicamente associadas com aquela entidade são evocadas (neste trabalho, as ações estariam associadas via relações *qualia*). Seguindo a prática proposta por Ruppenhofer et al. (2010), estão sendo criados dois *Elementos de Frame* (um para o evento, outro para a entidade).

| IV.1 - INICIAR_EVENTO_CAUSATIVO | | |
|---|------|----------------------|
| Um AGENTE inicia um EVENTO no qual ele poderá estar continuamente envolvido. O início do evento também pode ser representado por um EVENTO_INICIAL, um INSTRUMENTO ou uma CAUSA. O EVENTO pode ser representado por uma ENTIDADE. | | |
| Elemento de Frame | Tipo | EF do esquema básico |
| Agente | Core | Agente |
| Causa | Core | Causa |
| Evento_inicial | Core | Causa |
| Instrumento | Core | Instrumento |
| Evento | Core | - |
| Entidade | Core | - |
| Unidades Lexicais: <abrir.v>,<abertura.n> | | |
| Exemplos construídos | | |
| [João] AGENTE abriu [a sessão] EVENTO. | | |
| [A universidade] AGENTE abriu [as inscrições] EVENTO. | | |
| [O poema de Cecília Meireles] INSTRUMENTO abre [o livro] ENTIDADE. | | |
| [Um tiro] INSTRUMENTO abre [a temporada de caça] EVENTO. | | |
| [A redução da queda do dólar] EVENTO_INICIAL abre [a série de medidas do governo] EVENTO. | | |
| [Neymar] AGENTE abriu [o placar] ENTIDADE. | | |
| [Atentado] CAUSA abre [crise] EVENTO no Paraguai. | | |
| Na abertura [do concerto] EVENTO, [uma peça de Brahms] EVENTO_INICIAL. | | |
| Exemplos de uso em corpus e textos | | |
| NURC: [A pessoa] AGENTE sai dali e pode abrir [uma casa comercial] ENTIDADE | | |
| NURC: [Ele] AGENTE aí teria que abrir [uma conta] ENTIDADE num outro banco | | |
| NURC: E toca [as duas] AGENTE a abrir [o berreiro] EVENTO e chorar | | |
| CT: [O governo] AGENTE abriu [uma comissão de inquérito] ENTIDADE | | |
| CT: Em geral era [o número] EVENTO_INICIAL que abria [o programa] EVENTO | | |

| IV.2 - INICIAR_EVENTO_INCOATIVO | | |
|---|------|----------------------|
| Um EVENTO é iniciado. O EVENTO pode ser representado por uma ENTIDADE. | | |
| Elemento de Frame | Tipo | EF do esquema básico |
| Evento | Core | - |
| Entidade | Core | - |
| Tempo | Core | - |
| Unidades Lexicais: <abrir.v> | | |
| Exemplos construídos | | |
| [A sessão] EVENTO abriu [às 18 horas] TEMPO. | | |
| [As inscrições] EVENTO abriram [no dia 15 passado] TEMPO. | | |
| [O consultório] ENTIDADE abre [toda segunda-feira] TEMPO. | | |
| Exemplos de uso em corpus e textos | | |
| NURC: [Agora] TEMPO em Petrópolis abriu [uma churrascaria nova] ENTIDADE | | |
| NURC: Formava-se uma fila [logo cedo] TEMPO quando [o posto] ENTIDADE abria | | |
| NURC: [A loja] ENTIDADE abre [às nove] TEMPO então às sete já tem fila | | |
| NURC: [A exposição] EVENTO abre ao público [dia 24 de setembro] TEMPO | | |

Além do aspecto inceptivo, é possível também uma leitura de <abrir.v> no aspecto durativo (ou estativo) de um evento, indicando que elementos exteriores têm acesso a participação no evento.

| IV.3 - INICIAR_EVENTO_ESTATIVO | | |
|--|------------|----------------------|
| Um EVENTO está em um estado de acessibilidade. O EVENTO pode ser representado por uma ENTIDADE. | | |
| Elemento de Frame | Tipo | EF do esquema básico |
| Evento | Core | - |
| Entidade | Core | - |
| Participante | Periférico | - |
| Unidades Lexicais: <aberto.a> | | |
| Exemplos construídos [A sessão] EVENTO está aberta . Ele promoveu uma [discussão] EVENTO [aberta]. [A votação] EVENTO agora será aberta . [O torneio] EVENTO era aberto [a times pequenos] PARTICIPANTE. | | |
| Exemplos de uso em corpus e textos NURC: A que se conseguir que [o banco]ENTIDADE ficasse aberto o dia inteiro NURC: Porque aqui no Rio tem poucos [bares]ENTIDADE assim abertos CT: [Um mercado de trabalho]ENTIDADE aberto [pra todas as pessoas]PARTICIPANTE que quisessem trabalhar CT: [O debate]EVENTO será aberto [ao público]PARTICIPANTE NURC: Eles acreditavam em [casamento]ENTIDADE aberto | | |

A FrameNet apresenta uma especialização deste *Frame*, relacionado ao "estado de segredos" (SECRECY_STATUS).

| IV.4 - ESTADO_DE_SEGREDOS | | |
|---|------------|----------------------|
| Um FENÔMENO (que pode ser uma atividade, um estado ou um objeto) estava propositadamente escondido do conhecimento por um CONHECEDOR potencial e passa a ser revelado. | | |
| Elemento de Frame | Tipo | EF do esquema básico |
| Fenômeno | Core | Conteúdo |
| Conhecedor | Periférico | Exterior |
| Unidades Lexicais: <aberto.a> | | |
| Exemplos construídos [O mapa da instalação] FENÔMENO foi aberto [ao público] CONHECEDOR. [As contas da União] FENÔMENO estão abertas [aos contribuintes] CONHECEDOR. | | |
| Exemplos de uso em corpus e textos NURC: Quando estou num ambiente em que posso abrir [meu jogo]FENÔMENO facilmente... CT: Se o sucessor abrir [as contas]FENÔMENO e a Justiça não encontrar irregularidades | | |

9.3.5 Modalidade

Ainda usando a metáfora EVENTOS ESTÃO CONTIDOS EM CONTAINERS, vista na seção sobre Aspecto, é possível o uso de <abrir.v> para indicar a possibilidade de ocorrência de um ou mais eventos. Neste sentido, <abrir.v> representa um evento que remove o impedimento para que outro evento acontecesse, ou cria uma possibilidade para que outro evento aconteça. O "meio" para criação da possibilidade é comumente lexicalizado através de colocações como "abrir caminho", "abrir perspectiva", "abrir possibilidade", "abrir espaço", "abrir precedente", "abrir oportunidade", entre outras, e é representado no *Frame* a seguir através do *Elemento de Frame* MEIO, que conceptualiza um portal. Este *Elemento de Frame* é associado ao *Frame* POSSIBILITIES, da FrameNet.

| V.1 – POSSIBILITAR_EVENTO_CAUSATIVO | | |
|--|------|----------------------|
| Um AGENTE possibilita a ocorrência de um EVENTO. A possibilidade do evento também pode ser representada por um EVENTO_INICIAL, um INSTRUMENTO ou uma CAUSA. | | |
| Elemento de Frame | Tipo | EF do esquema básico |
| Agente | Core | Agente |
| Causa | Core | Causa |
| Evento_inicial | Core | Causa |
| Instrumento | Core | Instrumento |
| Evento | Core | - |
| Meio | Core | - |
| Unidades Lexicais: <abrir.v> | | |
| Exemplos construídos [Programas de Qualidade] CAUSA abrem [caminho] MEIO [para novos investimentos] EVENTO. [O governo] AGENTE abre [perspectiva] MEIO [de negociações] EVENTO. | | |
| Exemplos de uso em corpus e textos NURC: [Uma independência muito grande]CAUSA te abre assim [horizontes]MEIO incríveis NURC: Parece que [o curso dele] EVENTO_INICIAL abria [mais perspectiva]MEIO | | |

| V.2 - POSSIBILITAR_EVENTO_INCOATIVO | | |
|---|------|----------------------|
| Registra a possibilidade de ocorrência de um EVENTO. | | |
| Elemento de Frame | Tipo | EF do esquema básico |
| Evento | Core | - |
| Meio | Core | - |
| Unidades Lexicais: <abrir.v> | | |
| Exemplos construídos [Novas perspectivas] MEIO se abrem [para a Seleção] ENTIDADE. | | |
| Exemplos de uso em corpus e textos NURC: Em frente [às várias possibilidades]MEIO que se abrem, que tipo de coisa você vai escolher? CT: [Uma enorme oportunidade]MEIO se abriu para o Congresso | | |

| V.3 - POSSIBILITAR_EVENTO_ESTATIVO | | |
|---|------|----------------------|
| Registra a possibilidade de ocorrência de um EVENTO. | | |
| Elemento de Frame | Tipo | EF do esquema básico |
| Evento | Core | - |
| Meio | Core | - |
| Unidades Lexicais: <aberto.v> | | |
| Exemplos construídos Novas [perspectivas] MEIO estão abertas [para os jovens] ENTIDADE. Agora há um [precedente] MEIO aberto [para a condenação] EVENTO. | | |
| Exemplos de uso em corpus e textos CT: (...) mantem aberta a [possibilidade]MEIO de [reconciliação]EVENTO com os EUA (...) | | |

9.3.6 Estado Mental, Emocional ou Psicológico

Através da conjugação das metáforas A MENTE É UM CORPO e A MENTE É UM CONTAINER, o estado mental (emocional, psicológico) de uma entidade senciente (experenciador) é conceptualizado como a condição no interior de um container. A ação de <abrir.v> conceptualiza a mudança deste estado mental para uma condição que possibilita

maior acessibilidade do experienciador do fenômeno para as situações externas. O portal é conceptualizado como uma parte corporal.

| VI.1 - MUDANÇA_MENTAL_CAUSATIVO | | |
|--|------|----------------------|
| Um AGENTE, um INSTRUMENTO ou uma CAUSA exerce uma força sobre uma PARTE_CORPORAL de um EXPERENCIADOR, modificando seu estado mental. Um elemento mental (abstrato) do EXPERENCIADOR também pode ser conceptualizado como uma PARTE_CORPORAL. Esta situação é caracterizada como um “desbloqueio” ou uma “quebra de resistência” da parte do EXPERENCIADOR, que tem acesso a um EVENTO que ocorre no ambiente exterior. | | |
| Elemento de Frame | Tipo | EF do esquema básico |
| Agente | Core | Agente |
| Causa | Core | Causa |
| Instrumento | Core | Instrumento |
| Parte_corporal | Core | Portal |
| Experienciador | Core | Container |
| Evento | Core | Exterior |
| Unidades Lexicais: <abrir.v> | | |
| Exemplos construídos | | |
| [João] AGENTE abriu [os olhos] PARTE_CORPORAL para [o perigo]EVENTO. | | |
| [A carta] INSTRUMENTO abriu [os olhos] PARTE_CORPORAL [de João] EXPERENCIADOR para [a traição em curso] EVENTO. | | |
| [O discurso] INSTRUMENTO abriu [a consciência]PARTE_CORPORAL [de Carlos] EXPERENCIADOR para [aspectos ignorados] EVENTO. | | |
| [João] AGENTE abriu [seu coração] PARTE_CORPORAL para [o filho] EVENTO. | | |
| Exemplos de uso em corpus e textos | | |
| NURC: [Helena]AGENTE tenta abrir [os olhos]PARTE_CORPORAL de [Luiza]EXPERENCIADOR contra Shirley | | |
| CT: O site Hopeness fez uma lista de [dez documentários]AGENTE que podem abrir [a cabeça] PARTE_CORPORAL de [quem assiste sobre os mais diversos temas] EXPERENCIADOR | | |

| VI.2 - MUDANÇA_MENTAL_INCOATIVO | | |
|---|------|----------------------|
| Uma PARTE_CORPORAL de um EXPERENCIADOR sofre uma ação (não especificada) que modifica o estado mental do EXPERENCIADOR. Um elemento mental (abstrato) do EXPERENCIADOR também pode ser conceptualizado como uma PARTE_CORPORAL. Esta situação é caracterizada como um “desbloqueio” ou uma “quebra de resistência” da parte do EXPERENCIADOR, que tem acesso a um EVENTO que ocorre no ambiente exterior. | | |
| Elemento de Frame | Tipo | EF do esquema básico |
| Parte_corporal | Core | Portal |
| Experienciador | Core | Container |
| Evento | Core | Exterior |
| Unidades Lexicais: <abrir.v> | | |
| Exemplos construídos | | |
| [Seus ouvidos] PARTE_CORPORAL abriram para escutar [a revelação] EVENTO. | | |
| Finalmente [sua mente] CONTEUDO se abria para [verdades novas] EVENTO. | | |
| Exemplos de uso em corpus e textos | | |
| NURC: Quando [a cabeça]PARTE_CORPORAL se abre e adquire-se [mais informações sobre um tema] EVENTO é natural mudar de ideia | | |
| NURC: Na madrugada é onde [a mente]PARTE_CORPORAL se abre | | |

| VI.3 - MUDANÇA_MENTAL_ESTATIVO | | |
|--|------|----------------------|
| O estado "aberto" de uma PARTE_CORPORAL do EXPERENCIADOR conceptualiza o estado de receptividade a EVENTOS do exterior por parte do EXPERENCIADOR. Um elemento mental (abstrato) do EXPERENCIADOR também pode ser conceptualizado como uma PARTE_CORPORAL. Esta situação é caracterizada como um "desbloqueio" ou uma "quebra de resistência" da parte do EXPERENCIADOR. | | |
| Elemento de Frame | Tipo | EF do esquema básico |
| Parte_corporal | Core | Portal |
| Experenciador | Core | Container |
| Evento | Core | Exterior |
| Unidades Lexicais: <aberto.a> | | |
| Exemplos construídos | | |
| [O departamento] EXPERENCIADOR o recebeu de [braços] PARTE_CORPORAL abertos . [Esse cara] EXPERENCIADOR tem a [mente] CONTEUDO aberta . [Ele] EXPERENCIADOR está sempre de [ouvidos] PARTE_CORPORAL abertos para [as lamentações alheias] EVENTO. | | |
| Exemplos de uso em corpus e textos | | |
| NURC: De repente [o velho]EXPERENCIADOR tem [uma cabeça]PARTE_CORPORAL super aberta NURC: apesar da idade avançada, tinha [a mente]PARTE_CORPORAL aberta [para tudo que a modernidade oferecia] EVENTO | | |

A FrameNet possui um *Frame* que conceptualiza a ideia de "ceder", "renunciar", "desistir" (frame GIVING_IN). Este *Frame* pode ser evocado por colocações como "abrir mão" ou "abrir as pernas", entre outras. Em vista destas colocações usarem o verbo <abrir.v> e referenciarem partes do corpo, mas estarem relacionadas a um estado mental, estamos incluindo uma versão parcial deste *Frame* para análise.

| VI.4 - ABRIR_CEDER | | |
|---|------------|----------------------|
| Um CAPITULADOR, que estava se opondo a uma FORÇA, a qual é tipicamente exercida por um Oponente, cessa o exercício desta FORÇA, renunciando à sua posição sobre a QUESTÃO. Como AÇÃO_RESULTANTE a FORÇA segue seu curso e a QUESTÃO é resolvida contra os interesses do CAPITULADOR. Esta situação pode ser caracterizada como um "desbloqueio", ou uma "quebra de resistência" ou ainda uma "permissão" por parte do CAPITULADOR | | |
| Elemento de Frame | Tipo | EF do esquema básico |
| Capitulador | Core | Container |
| Força | Core | Portal |
| Questão | Core | Conteúdo |
| Oponente | Periférico | Agente ou Causa |
| Ação_resultante | Periférico | Exterior |
| Unidades Lexicais: <abrir mão.v>,<abrir as pernas.v> | | |
| Exemplos construídos | | |
| Para realizar [o acordo] AÇÃO_RESULTANTE com [a ex-esposa] Oponente, [ele] CAPITULADOR abriu mão de [vários bens] QUESTÃO. [A empresa] CAPITULADOR abriu mão [do processo] FORÇA contra [o funcionário] Oponente. | | |
| Exemplos de uso em corpus e textos | | |
| NURC: [Eu]CAPITULADOR preconizaria, abrindo mão [da minha condição de carioca]QUESTÃO CT: [Todos]CAPITULADOR abriram mão [do cache]QUESTÃO CT: [O ICN]CAPITULADOR resolveu abrir mão [das receitas das entrada]QUESTÃO para prevenir [eventuais reclamações] FORÇA [dos incoformados]OPONENTE | | |

A FrameNet também apresenta um *Frame* específico para a conceptualização de um estado de candura ou sinceridade no processo de comunicação (frame CANDIDNESS). Aqui estamos generalizando este *Frame*, focalizando o experienciador e o tópico a respeito do qual ele apresenta este estado de espírito.

| VI.5 - ABRIR_MENTAL_ESTATIVO | | |
|--|------|----------------------|
| O estado “aberto” do EXPERENCIADOR a respeito de um TOPICO conceptualiza a situação em que as barreiras para expressão das emoções ou para a sinceridade ou veracidade são removidas, permitindo tanto a exteriorização de sentimentos “interiores” quanto a interiorização de sentimentos expressos por agentes exteriores. | | |
| Elemento de Frame | Tipo | EF do esquema básico |
| Experenciador | Core | Container |
| Topico | Core | Exterior |
| Unidades Lexicais: <aberto.a> | | |
| Exemplos construídos | | |
| [Ela] EXPERENCIADOR sempre foi muito aberta em relação [a seus problemas] TOPICO. | | |
| [Eu] EXPERENCIADOR não consigo ser aberto [a mudanças] TOPIC. | | |
| Exemplos de uso em corpus e textos | | |
| NURC: Acho que [elas]EXPERENCIADOR têm um [relacionamento]TOPICO mais aberto com a mãe | | |
| NURC: [Eles]EXPERENCIADOR são muito abertos, têm confiança em mim e no pai | | |

9.4 Funções gramaticais

O trabalho de anotação realizado no projeto FrameNet é realizado através da divisão da sentença em segmentos. Como um dos propósitos da anotação é associar os valentes sintáticos de uma Unidade Lexical com seus valentes semânticos, mesmo os marcadores sintáticos e morfológicos dos sintagmas relevantes são incluídos nos segmentos (FILLMORE, 2008). Para cada segmento são registradas diversas informações, organizadas em camadas. As camadas mais usadas no processo de anotação são: *Target* (que indica a Unidade Lexical sob análise), Elemento de Frame (que indica qual o elemento de *Frame* associado com o segmento), Função Gramatical (FG) e Tipo Sintagmático (TS). Esta organização permite registrar as diversas possibilidades combinatórias da interface sintaxe-semântica (valências).

Como o foco deste trabalho é a desambiguação das Unidades Lexicais (no caso, as unidades lexicais associadas a <abrir.v>), e como para este processo estão sendo usados os núcleos dos sintagmas (sem os marcadores sintáticos e morfológicos), não julgamos relevante, a princípio, a anotação do Tipo Sintagmático dos segmentos. Desta forma, são considerados apenas o *Target* (a própria Unidade Lexical analisada), os *Elementos de Frame* (de cada *Frame*) e as Funções Gramaticais associadas a cada Elemento de *Frame*.

Para a FrameNet, as Funções Gramaticais assinaladas para cada constituinte (segmento) descrevem as maneiras pelas quais estes constituintes satisfazem os requisitos gramaticais abstratos de cada Unidade Lexical Ruppenhofer et al. (2010, p. 63). As etiquetas usadas para as funções gramaticais não descrevem as posições sintáticas de superfície de cada constituinte. Por exemplo, na sentença (9.2) [o doutor] é considerado o argumento externo de <tratar.v>, mesmo não sendo o sujeito da sentença.

(9.2) As circunstâncias forçaram [o doutor] a tratar seus inimigos.

Neste trabalho foram consideradas, parcialmente, as Funções Gramaticais usadas

no Projeto FrameNet Brasil (SALOMÃO, 2009), descritas em Torrent e Ellsworth (2013): Argumento Externo (Ext), Objeto Direto (DObj), Objeto Indireto (IObj) e Dependente (Dep).

9.4.1 Argumento Externo

Quando a Unidade Lexical alvo é um verbo, a etiqueta External (Ext) cobre as situações em que sintagmas que estão fora da projeção máxima da palavra alvo estão ligados funcionalmente aquela palavra alvo. Isto inclui tudo que satisfaça os requisitos de um Elemento de *Frame* nos seguintes contextos sintáticos:

a) como sujeito de um verbo alvo finito;

(9.3) [Maria] **comprou** frutas.

b) como sujeito alçado;

(9.4) [Maria] podia **comprar** frutas.

c) como expressões controlando sujeitos de verbos no infinitivo;

(9.5) Antônio viu [Maria] **comprar** frutas.

d) como argumento de um substantivo regente um verbo no infinitivo.

(9.6) Eu admiro a disposição [de Maria] para **ajudar** os amigos.

Quando a Unidade Lexical alvo é um adjetivo, a tag Ext é assinalada para o Elemento de *Frame* expresso como o sujeito de um verbo de ligação (9.7)a, ou como objeto em um predicado verbo-nominal (9.7)b.

(9.7) (a) [Maria] é **alta**.

(b) Antônio acha [Maria] **bonita**.

9.4.2 Objeto Direto

A etiqueta de objeto direto (DObj) é atribuída para os mesmos casos definidos pelo Projeto FrameNet:

a) qualquer objeto comum do verbo;

(9.8) Maria **comprou** [frutas].

b) objetos extraídos das construções QU-;

c) qualquer nome pós-verbal que controle o sujeito de um complemento verbal;

(9.9) (9) [O que] Maria **comprou**?

(9.10) Maria viu-[nos] **comprar** frutas.

(9.11) (a) Maria sabe [comprar frutas].
 (b) Maria sabe [que voce adora frutas].

d) Além destes casos, são incluídos também os objetos oracionais finitos (9.11)(a) e infinitivos (9.11)(b).

Em geral, no entanto, a etiqueta não é atribuída a dependentes do verbo que não possam ocorrer como sujeitos em uma construção passiva.

9.4.3 Objeto Indireto

Segundo [Castilho \(2010\)](#), o Português Brasileiro possui dois tipos de complementos preposicionais seguindo o verbo: objetos indiretos e oblíquos. Estas expressões são sempre preposicionadas e, portanto, é necessário distinguir seu uso. As principais diferenças entre elas seriam:

- a) Objetos indiretos são proporcionais aos pronomes dativos (me, te, lhe), enquanto oblíquos não;
- b) Objetos indiretos são preenchidos exclusivamente por sintagma preposicionado nucleado por "a" e "para", enquanto oblíquos podem ocorrer com outras preposições;
- c) Objetos indiretos podem coocorrer juntamente com o objeto direto, enquanto oblíquos tendem a ser o único argumento interno do verbo;

Além disso a construção em que figuram os objetos indiretos não é conversível à voz passiva.

A etiqueta IObj é aplicada exclusivamente para objetos indiretos, como nas sentenças em (9.12), enquanto os complementos oblíquos são associados a função Dependente.

(9.12) (a) Maria deu um livro [para mim].
 (b) Maria [me] deu um livro.
 (c) Maria deu um livro [para João] ler.

9.4.4 Dependente

A etiqueta genérica Dependente (Dep) é atribuída a todos os argumentos verbais não cobertos por Ext, Dobj ou IObj. Isto inclui advérbios, sintagmas preposicionados, sintagmas verbais e orações que ocorram depois de seus verbos, adjetivos ou nomes regentes, em sentenças declarativas normais, além dos complementos oblíquos. Esta etiqueta inclui

tanto o que é referenciado comumente como "argumentos" quanto o que é referenciado como "adjunto". No esquema de anotação da FrameNet, esta distinção é capturada via o status de Coreness do Elemento de Frame.

9.5 Valências

Usando a terminologia da FrameNet (FILLMORE, 2008), uma Unidade Lexical evoca um *Frame* e uma "descrição de valência" de uma dada Unidade Lexical apresenta o conjunto de maneiras pelas quais os acompanhantes sintáticos da Unidade Lexical introduzem informação sobre os elementos significantes do *Frame* evocado. Dito de outra forma, a "descrição de valência" mostra a associação dos *Elementos de Frame*s, com as Funções Gramaticais e com os Tipos Sintagmáticos.

As tabelas a seguir mostram as valências que foram utilizadas neste trabalho, em relação às Unidades Lexicais e *Frames* descritos na seção 9.3. Esta descrição é reconhecidamente incompleta, uma vez que ela é derivada da análise de algumas poucas dezenas de sentenças em corpus e de algumas sentenças contruídas. Além disso, de maneira geral foram considerados apenas os *Elementos de Frame* nucleares. Para simplificar a representação, as valências que se diferenciam apenas pela variação de *Elementos de Frame*s do mesmo CoreSet foram agrupadas em uma mesma linha. Por exemplo, o CoreSet (Agente, Causa, Instrumento) com a função gramatical "Argumento Externo" é representado como [Agente|Causa|Instrumento].ext. Para a implementação no *framework* LUDI, no entanto, cada valência é considerada individualmente.

9.5.1 Corpo físico

Frame II.1 - Movimento_corporal_causativo

UL: <abrir.v>

V1 [Agente|Causa|Instrumento].ext, Parte_corporal.dobj, Experenciador.dep

UL: <abertura.n>

V1 [Agente|Causa|Instrumento].ext, Parte_corporal.dep, Experenciador.dep

Frame II.2 - Movimento_corporal_incoativo

UL: <abrir.v>

V1 Parte_corporal.ext, Experenciador.dep

Frame II.3 - Movimento_corporal_estativo

UL: <aberto.a>

V1 Parte_corporal.ext, Experenciador.dep

9.5.2 Entidades Físicas

Frame III.1 - Abrir_físico_causativo

UL: <abrir.v>

V1 [Agente|Causa|Instrumento].ext, [Container|Portal].dobj,
[Conteudo|Interior|Exterior|Fronteira].depV2 [Agente|Causa|Instrumento].ext, Portal.dobj, Container.dep,
[Conteudo|Interior|Exterior|Fronteira].dep

UL: <abertura.n>

V1 [Agente|Causa|Instrumento].ext, [Container|Portal].dep,
[Conteudo|Interior|Exterior|Fronteira].dep**Frame III.2 - Abrir_físico_incoativo**

UL: <abrir.v>

V1 [Container|Portal].ext, [Conteudo|Interior|Exterior|Fronteira].dep

V2 Container.ext, Portal.dep, [Conteudo|Interior|Exterior|Fronteira].dep

V3 Container.dep, Portal.ext, [Conteudo|Interior|Exterior|Fronteira].dep

Frame III.3 - Abrir_físico_estativo

UL: <aberto.a>

V1 [Container|Portal].ext, [Conteudo|Interior|Exterior|Fronteira].dep

V2 Container.ext, Portal.dep, [Conteudo|Interior|Exterior|Fronteira].dep

V3 Container.dep, Portal.ext, [Conteudo|Interior|Exterior|Fronteira].dep

9.5.3 Aspecto

Frame IV.1 - Iniciar_evento_causativo

 UL: <abrir.v>

 V1 [Agente|Causa|Instrumento|Evento_inicial].ext, [Evento|Entidade].dobj

 UL: <abertura.v>

 V1 [Agente|Causa|Instrumento|Evento_inicial].ext, [Evento|Entidade].dobj

Frame IV.2 - Iniciar_evento_incoativo

 UL: <abrir.v>

 V1 [Evento|Entidade].ext, Tempo.dep

Frame IV.3 - Iniciar_evento_estativo

 UL: <aberto.v>

 V1 [Evento|Entidade].ext

Frame IV.4 - Estado_de_segredos

 UL: <aberto.a>

 V1 Fenômeno.ext, Conhecedor.dep

9.5.4 Modalidade

Frame V.1 - Possibilitar_evento_causativo

 UL: <abrir.v>

 V1 [Agente|Causa|Instrumento|Evento_inicial].ext, Meio.dobj, Evento.iobj

Frame V.2 - Possibilitar_evento_incoativo

 UL: <abrir.v>

 V1 Meio.ext, Evento.iobj

Frame V.3 - Possibilitar_evento_estativo

 UL: <aberto.a>

 V1 Meio.ext, Evento.iobj

9.5.5 Estado Mental, Emocional ou Psicológico

Frame VI.1 - Mudança_mental_causativo

 UL: <abrir.v>

 V1 [Agente|Causa|Instrumento].ext, Parte_corporal.dobj, Evento.iobj, Experenciador.dep

Frame VI.2 - Mudança_mental_incoativo

 UL: <abrir.v>

 V1 Parte_corporal.ext, Evento.iobj, Experenciador.dep

Frame VI.3 - Mudança_mental_estativo

 UL: <aberto.a>

 V1 Parte_corporal.ext, Evento.iobj, Experenciador.dep

Frame VI.4 - Abrir_ceder

 UL: <abrir_mão.v>

 V1 Capitulador.ext, [Força|Questao].iobj

VI.5 - Abrir_mental_estativo

 UL: <aberto.a>

 V1 Experenciador.ext, Topico.dep

10 A desambiguação de <abrir.v> usando LUDI

Este capítulo apresenta os resultados da aplicação do Framework LUDI para desambiguação de sentenças com o verbo <abrir.v>. A escolha das sentenças (coletadas em corpus ou construídas) se deu tanto pela sua potencial ambiguidade na interpretação (por máquina), quanto pela condição de prototypicalidade em relação aos *Frames* definidos no capítulo 9. O objetivo é avaliar a eficácia dos algoritmos utilizados, bem como levantar os pontos que ainda necessitam de correções ou merecem alguma melhoria.

Cada experimento é indicado pela expressão "abrir_x", onde "x" é uma entidade ou evento relacionado. Para cada experimento são apresentados:

- a) A motivação da escolha da expressão usada;
- b) A sentença e os segmentos usados como entrada para os algoritmos (juntamente com as funções gramaticais);
- c) O grafo que representa o resultado da aplicação do modelo conexionista estruturado. Embora, somente pelo grafo, não seja possível visualizar as diversas etapas da aplicação dos algoritmos, ele mostra a inferência final realizada.
- d) Comentários sobre o resultado obtido.

10.1 Corpo físico x Estado mental

10.1.1 Abrir_boca

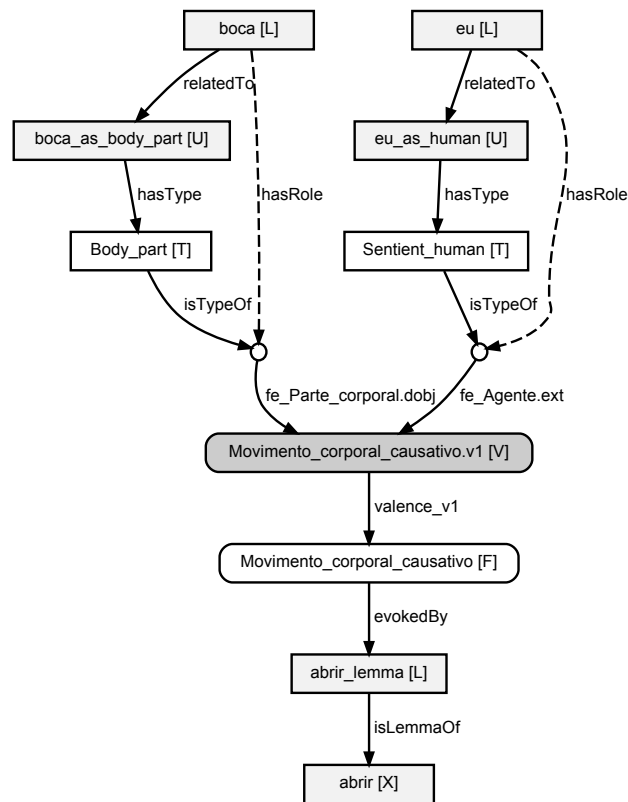
Motivação: Testar o *cluster* II, onde o corpo ou partes do corpo são conceptualizados como container ou portais.

Sentença: (...) então quando eu abro a boca e digo um soneto de Olavo Bilac (...)

Segmentos: abrir.target, boca.dobj, eu.ext

Grafo: [Figura 41](#)

Figura 41 – Abrir_boca



Comentários: Esta é uma inferência bastante direta no algoritmo, uma vez que os tipos dos segmentos são bastante compatíveis com os tipos dos *Elementos de Frame*. O destaque é a condição de obrigatoriedade do tipo BODY_PART para o *Elemento de Frame* PARTE_CORPORAL, implementada através de um *and-node*.

10.1.2 Abrir_olho

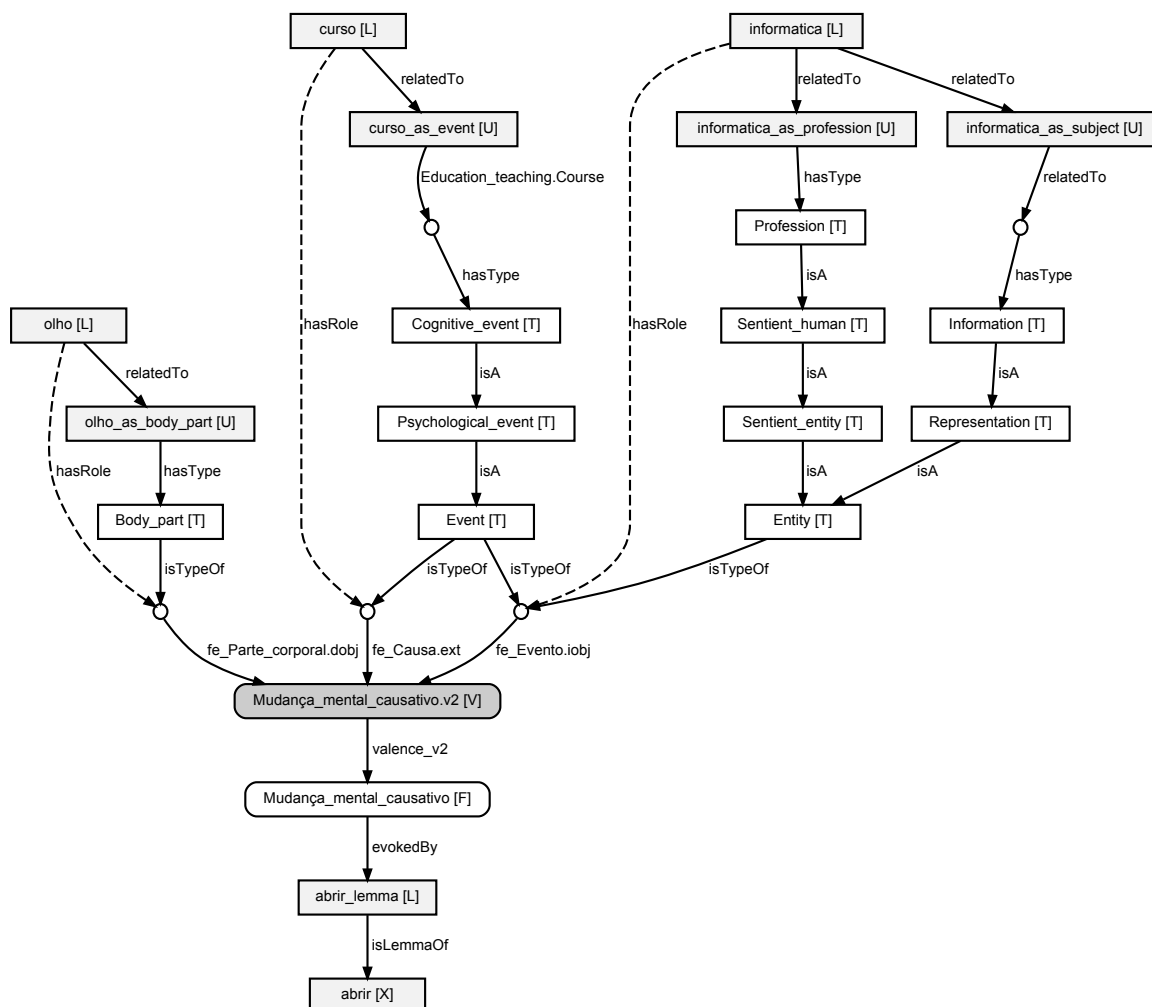
Motivação: Testar o cluster VI, onde partes corporais são usadas metaforicamente relacionadas a estados mentais.

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

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

Grafo: [Figura 42](#)

Figura 42 – Abrir_olho



Comentário: Dois fatos se destacam nesta inferência:

- A associação de <curso> (uma *Unidade Semântica*) com o *Elemento de Frame* EDUCATION_TEACHING.COURSE, mostrando a integração entre a FrameNet e a ontologia SIMPLE.
- A distinção entre os *Frames* VI.1.MUDANÇA_MENTAL_CAUSATIVO e II.1.MOVIMENTO_CORPORAL_CAUSATIVO só foi possível porque, na configuração da rede, as valências cadastradas para II.1.MOVIMENTO_CORPORAL_CAUSATIVO obrigam a expressão do *Elemento de Frame* EXPERENCIADOR, exceto quando o *Externo* é um AGENTE (que neste caso deve ser o próprio EXPERENCIADOR).

10.1.3 Abrir_mão

Motivação: Testar a diferença entre o uso da *Unidade Lexical multiword* <abrir_mao>, que evoca o *Frame* VI.4.ABRIR_CEDER e o uso de <abrir mao> como uma colocação,

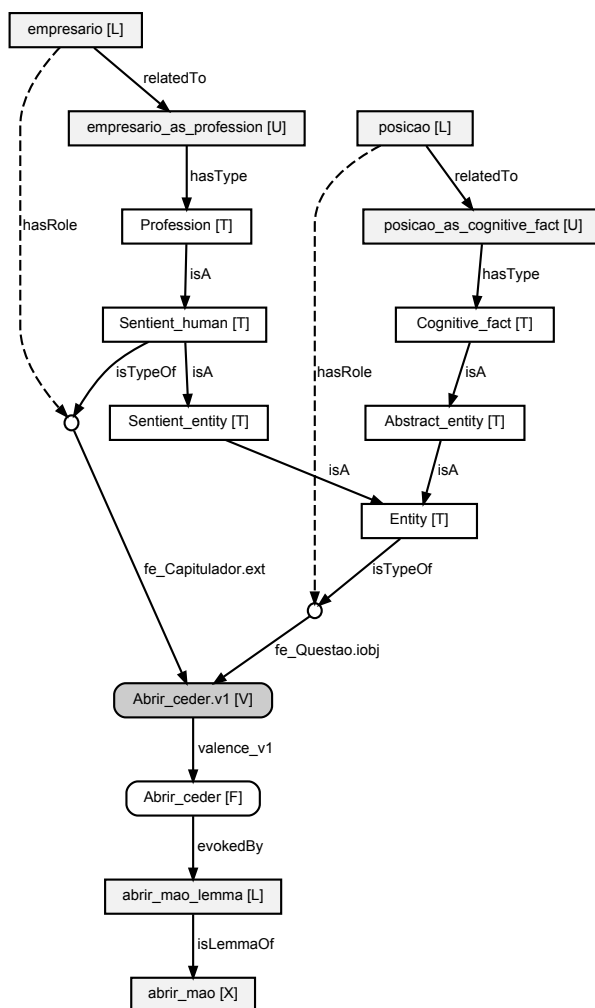
que pode evocar os *Frames* do *cluster* II.MOVIMENTO_CORPORAL.

Sentença: (...) Os empresários abriram mao de posições históricas (...)

Segmentos como MWE: abrir_mao.target, empresario.ext, posicao.iobj

Grafo: Figura 43

Figura 43 – Abrir_mão

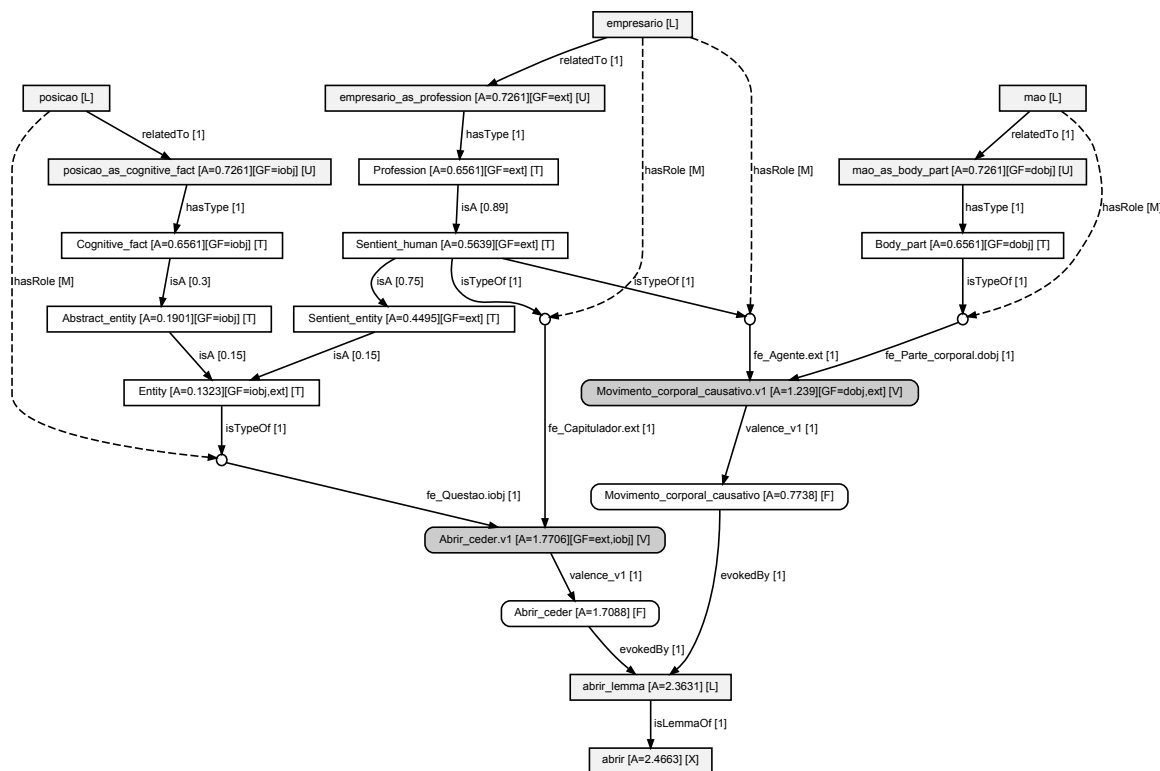


Segmentos como colocação: abrir.target, mao.dobj, empresario.ext, posicao.iobj

Grafo: Figura 44

Comentários: Como pode ser observado nas figuras, a inferência com a MWE é direta, uma vez que apenas um *Frame* é evocado pela expressão, e portanto não há necessidade de desambiguação. No entanto, quando <mao> é usado como um objeto direto de <abrir>, é possível a evocação dos *Frames* II.MOVIMENTO_CORPORAL_CAUSATIVO e VI.5.ABRIR_CEDER. Em termos da rede, a ambiguidade acontece pois a valência V1 de II.MOVIMENTO_CORPORAL_CAUSATIVO pede apenas um AGENTE e a PARTE_CORPORAL,

Figura 44 – Abrir_mão (colocação)



que estão presentes na sentença. A desambiguação é feita então com base no nível de ativação de cada *Frame*. O nível de ativação de VI.5.ABRIR_CEDER (1.7088) é mais que o dobro do nível de ativação de II.1.MOVIMENTO_CORPORAL_CAUSATIVO, apontando para uma interpretação associada a mudança de estado mental, emocional ou psicológico.

10.2 Entidades físicas

10.2.1 Abrir_carteira

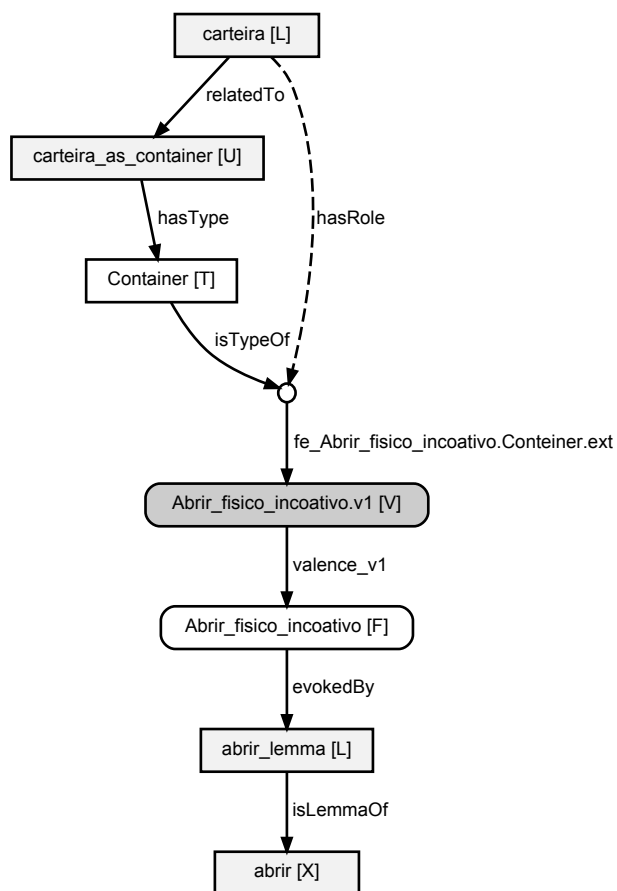
Motivação: Testar a inferência com entidades físicas conceptualizadas como containers.

Sentença: (...) a carteira abria e lá dentro a gente botava coisas (...)

Grafo: Figura 45

Comentários: Este experimento mostra a desambiguação relativa a alternância causativo-incoativo. Na valência escolhida via inferência, o *Elemento de Frame* CONTEINER tem a função gramatical *Externo*.

Figura 45 – Abrir_carteira



10.2.2 Abrir_porta

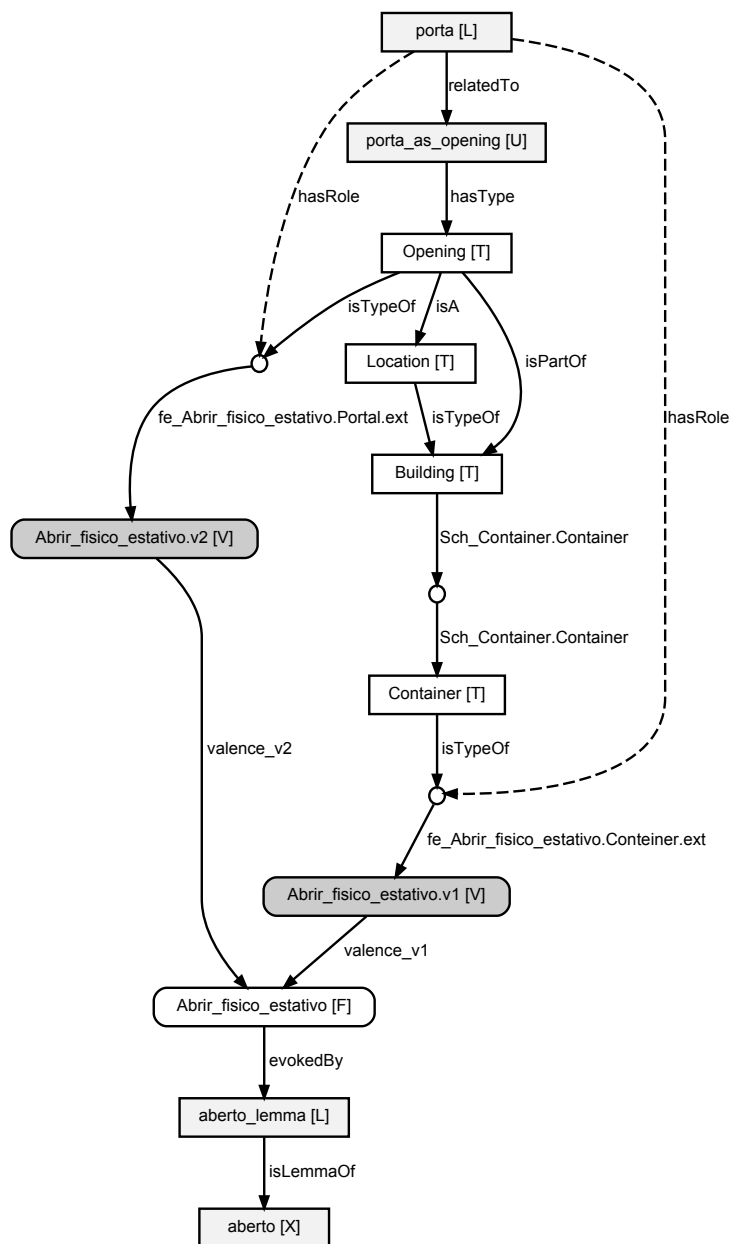
Motivação: Testar a inferência como entidades físicas conceptualizadas como containers, distinguindo as situações causativas e estativas.

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

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

Grafo: [Figura 46](#)

Figura 47 – Porta_aberta



10.2.3 Abrir_jornal

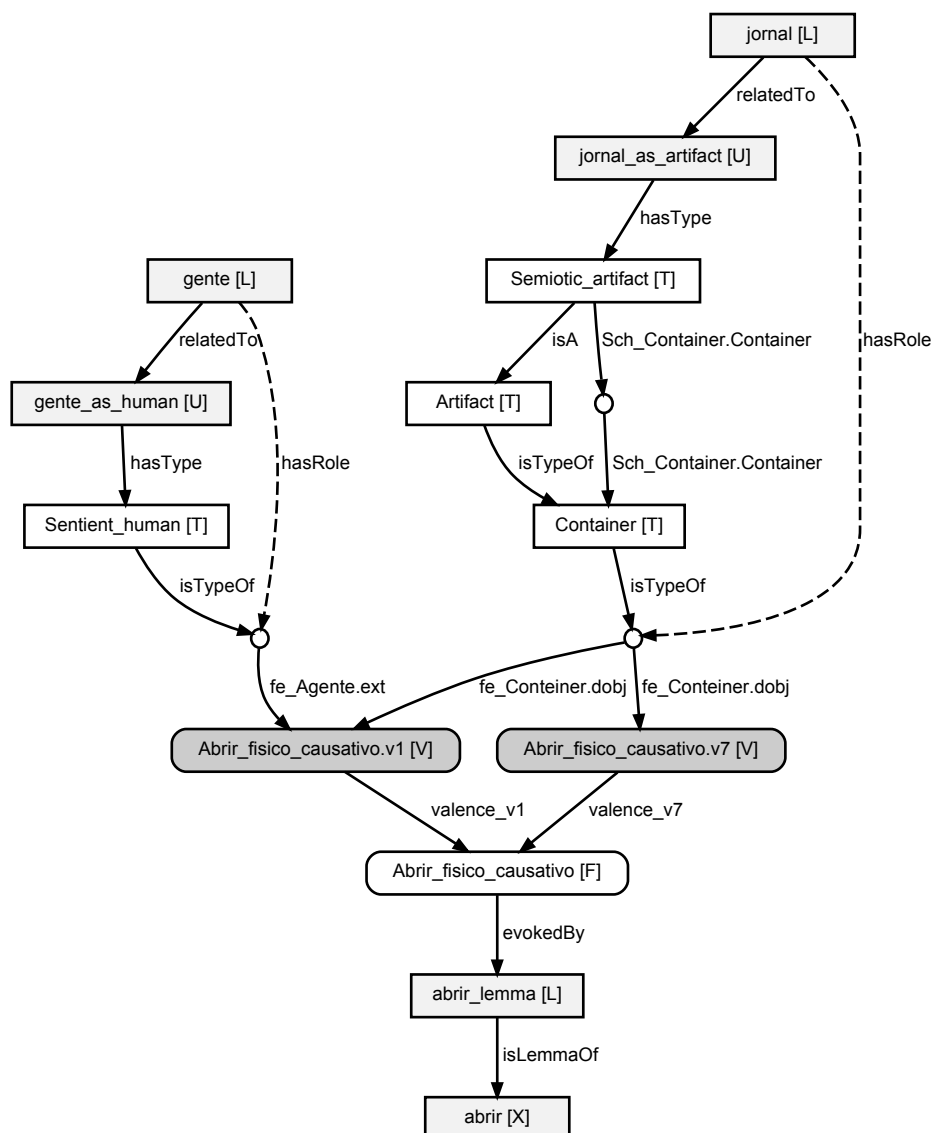
Motivação: Testar o processo de inferência com entidades polissêmicas. No caso, <jornal> tem tanto o sentido de "informação" quanto de um "artefato".

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

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

Grafo: Figura 48

Figura 48 – Abrir_jornal



Comentários: Neste caso o processo de inferência foi praticamente direto, por dois motivos:

- Existe uma relação *qualia* entre `INFORMATION` e `SEMIOTIC_ARTIFACT`, estabelecendo que informações são geralmente contidas em artefatos semióticos. Esta relação permite ao processo de inferência reforçar a ativação de <jornal> como um artefato.
- Uma vez que geralmente contém informações, `SEMIOTIC_ARTIFACT` é conceptualizado como um container, e associado com o *Elemento de Frame* `ABRIR_ESQUEMA.CONTEINER`. Este é mais um exemplo da integração entre a ontologia e o recurso lexical.

10.3 Entidades físicas x Aspecto

10.3.1 Abrir_museu

Motivação: Neste experimento são usadas sentenças construídas, com o objetivo de testar o uso aspectual de <abrir.v>, comparado com o seu sentido de movimento físico, na situação incoativa.

Sentença: *O museu abre.*

Segmentos: abrir.target, museu.ext

Grafo: Figura 49

Figura 49 – Abrir_museu

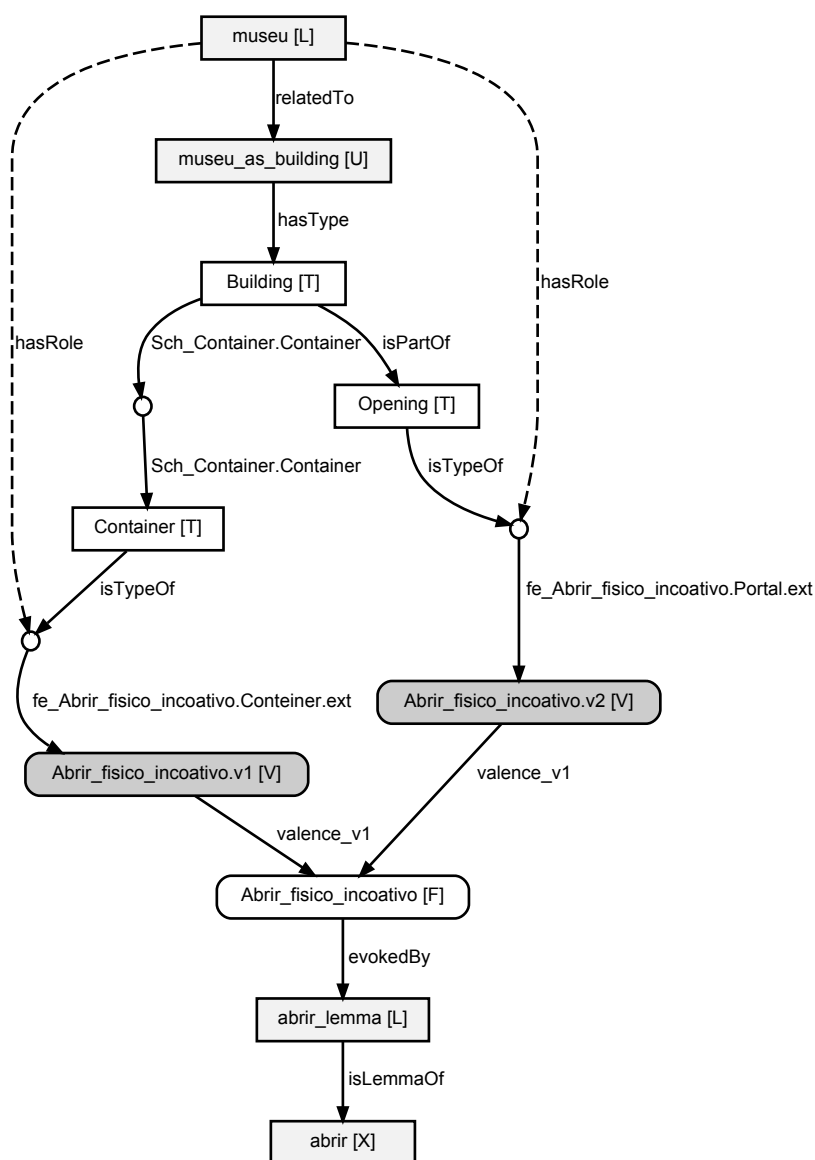
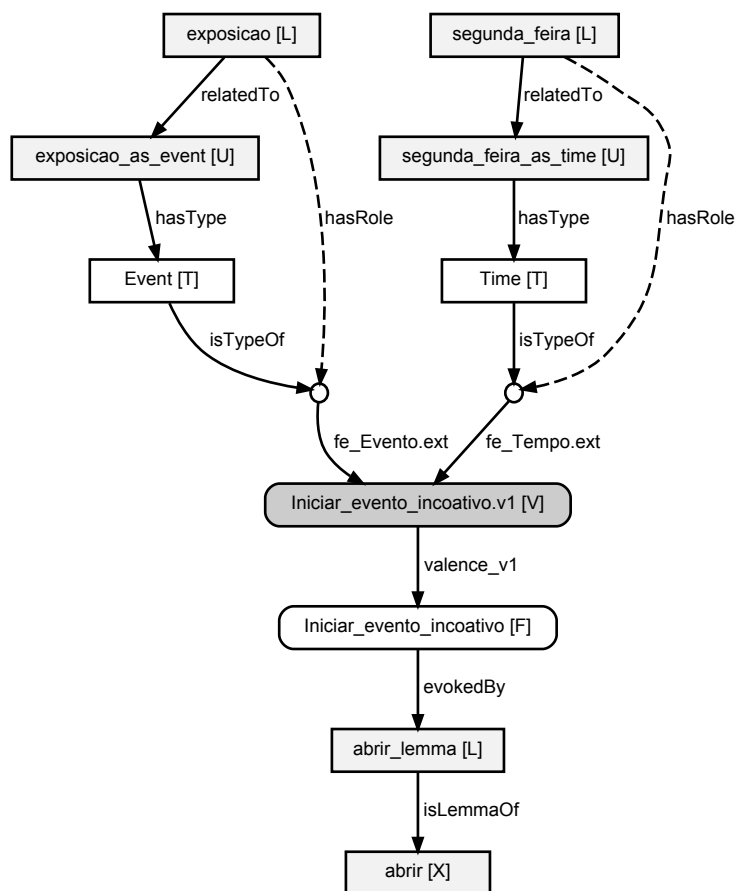


Figura 51 – Abrir_exposicao



Comentários: Novamente, com o auxílio do *Elemento de Frame* TEMPO, o processo de inferência determina o *Frame* IV.2.INICIAR_EVENTO_INCOATIVO como o mais provável para a interpretação da sentença.

10.3.3 Abrir_shopping

Motivação: Testar o processo de inferência com entidades de *Tipos Complexos*, simultaneamente com o uso aspectual de <abrir.v>. Neste caso, <shopping> tem tanto o sentido de "comércio" quanto de uma "construção" (ou "prédio").

Sentença: (...)Os shoppings na zona oeste abrem hoje (...)

Segmentos: abrir.target, shopping.ext, hoje.dep

Grafo: Figura 52

Comentários: Este exemplo é mais complexo e permite destacar vários fatos:

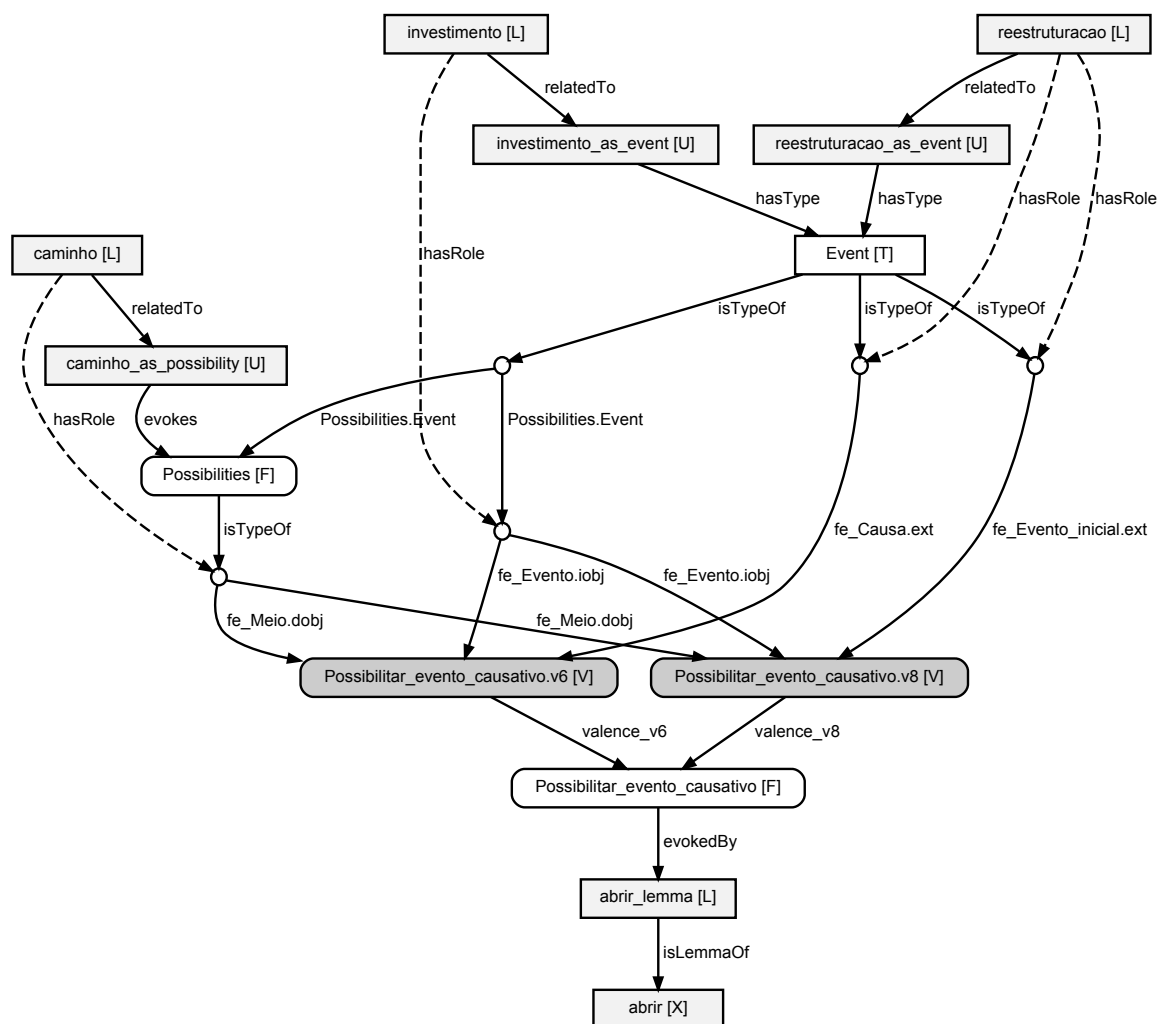
- a) Novamente, o *Frame* IV.2.INICIAR_EVENTO_INCOATIVO é ativado devido ao *Elemento de Frame* TEMPO;

Sentença: (...) programa de qualidade e reestruturação abrem caminho para investimentos de US\$ 100 milhões em três anos (...)

Segmentos: abrir.target, reestruturacao.ext, caminho.dobj, investimento.iobj

Grafo: Figura 54

Figura 54 – Abrir_caminho (possibilidade)



Comentários: A distinção do uso modal é possível porque a polissemia de <caminho> (ou como uma abertura ou como uma possibilidade) é resolvida através da evocação do *Frame* POSSIBILITIES, da FrameNet. Este experimento demonstra, mais uma vez, a importância da integração dos recursos lexicais na resolução das ambiguidades.

10.5 Metonímias

10.5.1 Abrir_cerveja

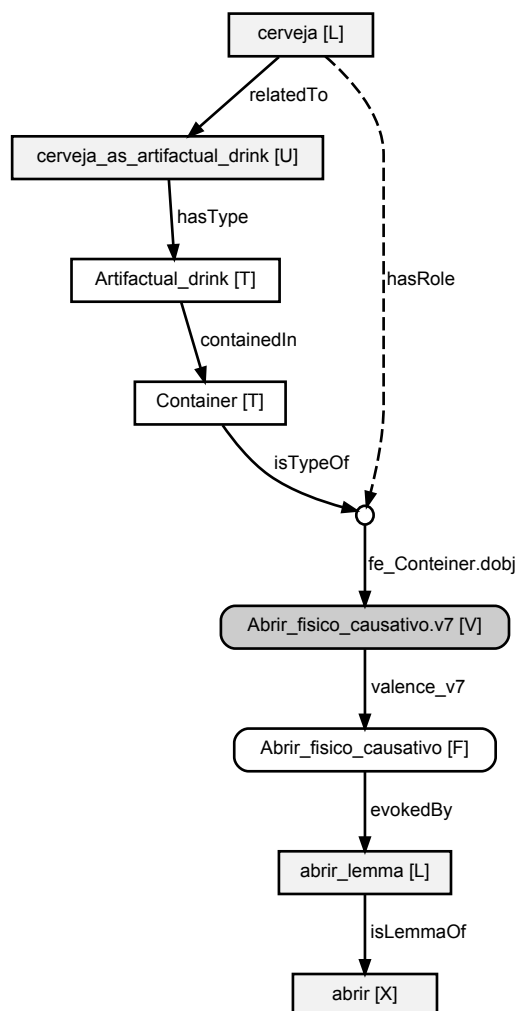
Motivação: Testar os algoritmos na resolução de uso metonímicos. Neste caso, o container é conceptualizado pelo seu conteúdo.

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

Segmentos: abrir.target, cerveja.dobj

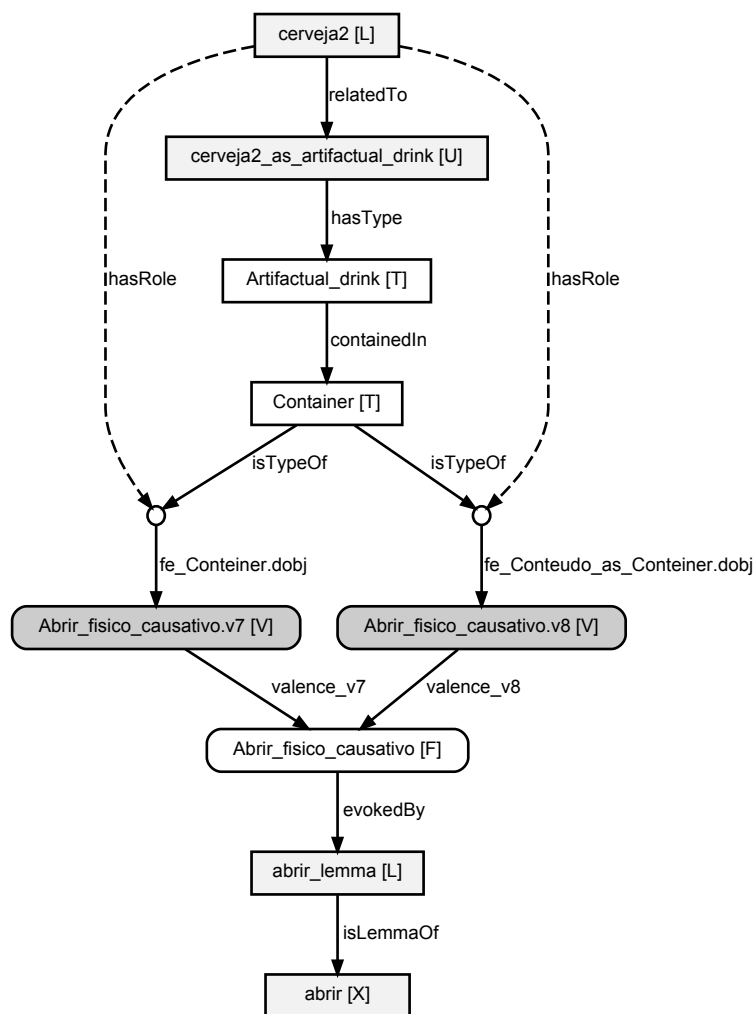
Grafos: [Figura 55](#), [Figura 56](#)

Figura 55 – Abrir_cerveja



Comentários: A [Figura 55](#) mostra que uma interpretação possível para a sentença é a consideração que <cerveja> é subentendida por um container (a garrafa de cerveja, por exemplo). Para a [Figura 56](#) foi criada uma valência (valence_V9) e um *Elemento de Frame* CONTEUDO_AS_CONTEINER, específicos para esta situação, a fim de deixar mais claro o processo metonímico que está ocorrendo.

Figura 56 – Abrir_cerveja (2)



10.5.2 Abrir_marcador

Motivação: Testar o processo metonímico em que uma entidade é conceptualizada como um evento.

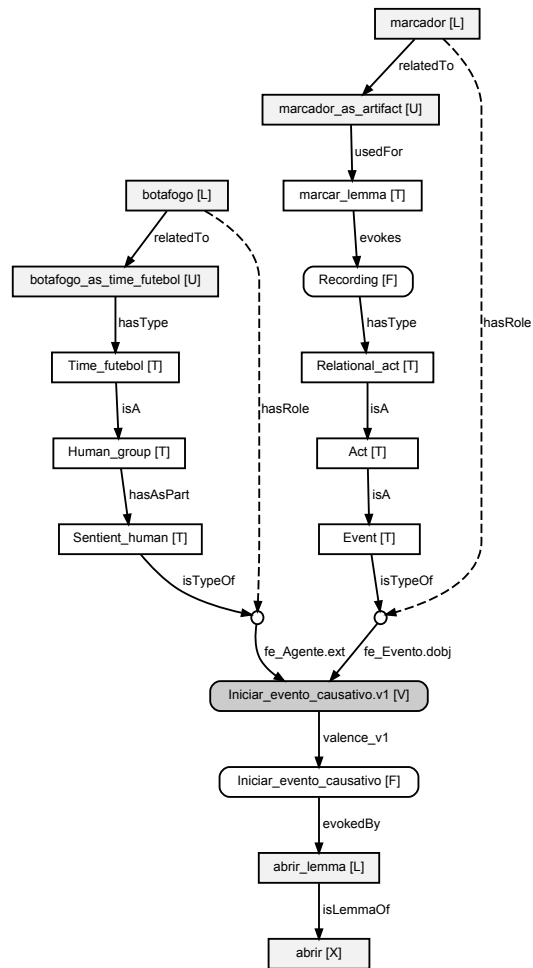
Sentença: o *Botafogo* comandou as ações no primeiro tempo e esteve perto de abrir o marcador.

Segmentos: abrir.target, botafogo.ext, marcador.dobj

Grafo: [Figura 57](#)

Comentários: Este experimento demonstra o uso das relações *qualia* entre uma *Unidade Semântica* e uma *Unidade Lexical* (relação télica *usedFor* entre <marcador_as_artifact> e <marcar_lemma>) e do uso do *Frame RECORDING*, a fim de que <abrir marcador> possa ser interpretado como "iniciar o evento de marcação" (*Frame* IV.1.INICIAR_EVENTO_CAUSATIVO). Na realidade, uma situação mais complexa é obtida se <marcador>

Figura 57 – Abrir_marcador



for considerado um SEMIOTIC_ARTIFACT (um artefato que contém informação), pois artefatos como estes são conceptualizados como containers nesta rede. Assim, o *Frame* III.1.ABRIR_FISICO_CAUSATIVO também seria evocado.

Conclusão

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

O significado tem diversos aspectos que precisam ser considerados. Para (JACKENDOFF, 2012), por exemplo, o significado tem uma função referencial, de associar as palavras às coisas do mundo; o significado tem uma função inferencial, que conecta os significados entre si; o significado de uma sentença é mais rico que o significado de suas palavras componentes; e, sobretudo, o significado é inconsciente e está muitas vezes oculto por trás das palavras. Assim, a questão do significado é difícil pois estes aspectos, entre outros, podem ser abordados de variadas maneiras e com variados objetivos. Realmente, são diversas as abordagens linguísticas que, generalizadas sob o nome de "Semântica", vão trazer o significado como seu objeto de estudo.

Neste trabalho, adotamos uma abordagem híbrida, reunindo contribuições da Semântica de Frames (para representação do significado considerando aspectos cognitivos) e da Teoria do Léxico Gerativo (para representação do significado considerando aspectos mais formais). O resultado foi uma aplicação computacional, chamada LUDI (*Lexical Unit Discovery through Inference*), que pode ser enquadrada na área de Compreensão da Língua Natural.

A Semântica de Frames é uma abordagem para estudo do significado lexical baseado nos trabalhos de Fillmore e seus colaboradores. Sua ideia central é que o significado de uma palavra deve ser descrito em relação a um **frame semântico**. Um *Frame*, neste contexto, é uma representação esquemática de estruturas conceituais e padrões de crença, práticas, instituições, etc. que provê a base para uma interação significativa em uma comunidade de fala. Um *Frame* representa um sistema de conceitos relacionados entre si de tal forma que, para compreender um deles é necessário compreender toda a estrutura onde ele se situa. Para a Semântica de Frames as palavras são categorizações das experiências e cada uma destas categorias é sustentada por uma situação motivadora, ocorrendo contra o cenário de conhecimento e experiência representado pelo *Frame*. O significado de uma expressão linguística é interpretado de acordo com o *Frame* de fundo que representa a cena no momento do enunciação.

Crucialmente, para a Semântica de Frames, a compreensão de uma palavra envolve a compreensão do *Frame* evocado por aquela palavra. Assim, a compreensão de itens lexicais polissêmicos demanda, antes de tudo, saber qual sentido está sendo usado em

dado contexto, ou seja, qual *Frame* está sendo evocado. Este é justamente o objetivo do *framework* LUDI.

Por outro lado, a Teoria do Léxico Gerativo (TLG), proposta por Pustejovsky, está apoiada na hipótese de que a natureza intrínseca, o uso pretendido ou possível, a constituição e a forma de geração (ou origem) de uma entidade (referenciada por um item lexical) permite predizer a participação do item lexical em um enunciado. Estes aspectos, que formam a chamada estrutura *qualia*, serviram de base para a construção de léxicos multicamadas e multifacetados como, por exemplo, o projeto SIMPLE-CLIPS, que originou a Ontologia SIMPLE, usada neste trabalho. No entanto, uma abordagem puramente gerativa deixa de lado um componente importante do significado lexical, que é a cena onde o item ocorre, capturada pela Semântica de Frames. Daí, nossa opção por alinhar as duas abordagens.

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

Para a Linguística Computacional, a contribuição é uma técnica de análise de sentenças baseada em um modelo conexionista, usando a técnica de *Ativação Propagada*. Os modelos encontrados na literatura, relativos à desambiguação de *Frames*, utilizam predominantemente técnicas baseadas em probabilidade, seja através de Máxima Entropia, seja através de Redes Bayesianas. No *framework* LUDI usamos um modelo conexionista estruturado, neurologicamente plausível, e um processo de inferência baseado em evidência, não em probabilidade. Consideramos que este tipo de inferência é mais adequado cognitivamente, pelo menos para uma certa classe de situações, em que experiências anteriores possibilitam o estabelecimento de evidências. A [Tabela 13](#) sintetiza as abordagens escolhidas neste trabalho, em comparação com as escolhas mais comuns em PLN.

Estas contribuições possuem valor por si mesmas. No entanto, julgamos que a maior contribuição é justamente a demonstração de que conceitos e ideias caros à Linguística Cognitiva podem ser usados em uma implementação computacional. A literatura de viés cognitivista, relacionada aos processos de categorização, conceptualização através de espaços mentais e abordagens gramaticais usando construções, por exemplo, apresenta, via de regra, um sentimento de resistência (quando não de rejeição) a qualquer tipo de

Tabela 13 – Abordagens escolhidas no *framework* LUDI.

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

formalização. Acreditamos que pesquisas como as realizadas pelo grupo NTL (*Neural Theory of Language*) ou esta, que levou à construção do *framework* LUDI, mostram que há um caminho viável de aproximação.

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

- a) Atualmente as funções gramaticais estão sendo fornecidas como dados de entrada; para uma automatização maior da análise, seria necessário o uso de um *parser* sintático que atribuísse estas funções;
- b) Estamos trabalhando apenas com palavras que evocam *Frames* (*Frame Evoking Elements*) e *Unidades Semânticas* associadas a tipos da ontologia SIMPLE; para uma análise completa da sentença, é necessário o cadastro dos *Frames* e/ou Esquemas associados às preposições, artigos, quantificadores, etc;
- c) Não foi feita nenhuma abordagem do ponto de vista construcional, mas apenas baseada em valências; o modelo, como está hoje, pode ser facilmente estendido para alguns casos simples de construções. Uma sugestão para trabalho futuro seria testar a aplicação com construções do Constructicon (FILLMORE; LEE-GOLDMAN; RHOMIEUX, 2012);
- d) Acreditamos que o modelo conexionista, como implementado, pode facilitar a resolução de anáforas e outras referências no contexto; estes testes, porém, não foram realizados nos experimentos realizados;
- e) A ontologia SIMPLE foi utilizada por ser uma ontologia já existente e baseada na TLG; no decorrer do trabalho ficou claro que, para uma aplicação mais genérica, ela precisa ser expandida e possivelmente associada ou alinhada com outras ontologias de topo, como a DOLCE (MASOLO et al., 2003);
- f) Apenas os itens lexicais necessários aos experimentos foram adicionados à ontologia; para uma aplicação mais genérica, seria necessário léxico mais completo

e abrangente. Uma sugestão é o alinhamento da ontologia SIMPLE com a ontologia DBPedia (LEHMANN et al., 2014);

- g) O modelo conexionista estruturado pode ser expandido de diversas formas, permitindo outros tipos de inferência e classificação. Uma opção é implementar alguns recursos semelhantes ao do modelo SHRUTI (SHASTRI, 1999).

À guisa de conclusão, acreditamos que a questão proposta para esta pesquisa, no capítulo 1, recebe uma resposta plausível com a implementação do *framework* LUDI. A Semântica de Frames, enriquecida com a contribuição da Teoria do Léxico Gerativo e o uso de Tipos Ontológicos, possibilita não só a desambiguação lexical, mas se constitui em uma promissora base de pesquisa na área da Compreensão da Língua Natural.

Referências

- AMARO, R. *Computation of verbal predicates in portuguese: relational network, lexical-conceptual structure and context*. Tese (Doutorado) — UNIVERSIDADE DE LISBOA, 2009. Citado na página 33.
- ASHER, N. *Lexical Meaning in Context: A Web of Words*. Cambridge: Cambridge University Press, 2011. (Lexical Meaning in Context: A Web of Words). ISBN 9781107005396. Citado na página 33.
- ATKINS, S. et al. Lexicographic relevance: selecting information from corpus evidence. *International Journal of Lexicography*, v. 16.3, p. 251–280, 2003. Citado 2 vezes nas páginas 48 e 51.
- BAKER, C.; FILLMORE, C.; CRONIN, B. The Structure of the FrameNet Database. *International Journal of Lexicography*, v. 16, n. 3, p. 281–296, 2003. Citado 2 vezes nas páginas 23 e 52.
- BAKER, C. F. FrameNet, current collaborations and future goals. *Language Resources and Evaluation*, v. 46, n. 2, p. 269–286, ago. 2012. ISSN 1574-020X. Citado 2 vezes nas páginas 51 e 52.
- BOAS, H. Frame Semantics as a framework for describing polysemy and syntactic structures of English and German motion verbs in contrastive computational lexicography. In: RAYSON, P. et al. (Ed.). *Proceedings of the Corpus Linguistics 2001 Conference*. Lancaster: University Centre for computer corpus research on language, 2001. p. 64–73. Citado na página 25.
- BORST, W. *Construction of Engineering Ontologies*. Tese (Doutorado) — University of Twente, 1997. Citado na página 60.
- BOS, J. A Survey of Computational Semantics: Representation, Inference and Knowledge in Wide-Coverage Text Understanding. *Language and Linguistics Compass*, v. 5, n. 6, p. 336–366, jun. 2011. ISSN 1749818X. Citado na página 18.
- BRUGMAN, C.; LAKOFF, G. Cognitive topology and lexical networks. *Lexical ambiguity resolution*, 1988. Citado na página 42.
- BRYANT, J. E. *Best-Fit Constructional Analysis*. Tese (Doutorado) — University of California at Berkeley, 2008. Citado 2 vezes nas páginas 30 e 75.
- CASTILHO, A. T. de. *Nova gramática do português brasileiro*. São Paulo: Contexto, 2010. ISBN 9788572444620. Citado na página 138.
- COPLEY, B.; WOLFF, P. Theories of causation should inform linguistic theory and vice versa. In: COPLEY, B.; MARTIN, F. (Ed.). *Causation in Grammatical Structures*. Oxford: Oxford University Press, 2013. cap. 2, p. 1–71. Citado 3 vezes nas páginas 184, 185 e 186.
- CROFT, W.; CRUSE, A. *Cognitive Linguistics*. Cambridge, MA: Cambridge University Press, 2004. Citado na página 43.

- CRUSE, A. *Meaning in Language: An Introduction to Semantics and Pragmatics*. 3a. ed. Oxford: Oxford University Press, 2010. Citado 4 vezes nas páginas 41, 43, 45 e 53.
- DIAS-DA-SILVA, B. C. O estudo Lingüístico-Computacional da Linguagem. *Letras de Hoje*, v. 41, n. 2, p. 103–138, 2006. Citado 2 vezes nas páginas 29 e 57.
- DIEDERICH, J. Spreading Activation and Connectionist Models for Natural Language Processing. In: GÄRTNER, H.-M. (Ed.). *Theoretical Linguistics*. Berlin: Mouton de Gruyter, 1990. v. 16, p. 25–64. ISBN 16134060. Citado na página 77.
- DODGE, E. *Constructional and Conceptual Composition*. Tese (Doutorado) — University of California at Berkeley, 2010. Disponível em: <<http://www.eric.ed.gov/ERICWebPortal/recordDetail?accno=ED527781>>. Citado 2 vezes nas páginas 54 e 55.
- DOLBEY, A.; ELLSWORTH, M.; SCHEFFCZYK, J. BioFrameNet: A Domain-Specific FrameNet Extension with Links to Biomedical Ontologies. *KR-MED*, p. 87–94, 2006. Citado na página 56.
- DURAN-MUNOZ, I. Managing polysemy in the adventure tourism discourse with Frame Semantics. In: SLODZIAN, M. et al. (Ed.). *Workshop Proceedings of the 9th International Conference on Terminology and Artificial Intelligence*. Paris, France: INALCO, 2011. p. 40–42. Citado na página 25.
- EVANS, V. *How words mean: Lexical concepts, cognitive models, and meaning construction*. Oxford: Oxford University Press, 2009. ISSN 0022-2267. ISBN 0022226710. Citado na página 33.
- FELDMAN, J.; BALLARD, D. Connectionist models and their properties. *Cognitive science*, v. 254, n. 1982, p. 205–254, 1982. Citado 2 vezes nas páginas 75 e 76.
- FELDMAN, J. A. *From Molecule to Metaphor - A Neural Theory of Language*. [S.l.]: MIT Press, 2006. ISBN 0262062534. Citado 5 vezes nas páginas 40, 46, 75, 76 e 85.
- FELLBAUM, C. *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press, 1998. Citado 2 vezes nas páginas 33 e 51.
- FERRARI, L. *Introdução à Linguística Cognitiva*. São Paulo: Contexto, 2011. Citado na página 38.
- FILLMORE, C. Frame Semantics. In: *Cognitive Linguistics: Basic Readings*. Berlin: Walter de Gruyter, 2006. cap. 10, p. 373–400. Citado 2 vezes nas páginas 22 e 44.
- FILLMORE, C. Valence issues in Framenet. In: GUTZ, T.; VOTTELER, K. (Ed.). *Trends in Linguistics, Studies and Monographs : Valency : Theoretical, Descriptive and Cognitive Issues*. Berlin: Mouton de Gruyter, 2008. p. 141–174. Citado 4 vezes nas páginas 48, 89, 136 e 139.
- FILLMORE, C.; BAKER, C. A frames approach to semantic analysis. *The Oxford Handbook of Linguistic Analysis*, n. 1, p. 313–340, 2010. Citado 7 vezes nas páginas 44, 45, 46, 47, 48, 50 e 100.
- FILLMORE, C.; JOHNSON, C. R.; PETRUCK, M. R. Background to Framenet. *International Journal of Lexicography*, v. 16, n. 3, p. 235–250, 2003. Citado 4 vezes nas páginas 22, 32, 47 e 51.

- FILLMORE, C.; LEE-GOLDMAN, R.; RHOMIEUX, R. The FrameNet Constructicon. In: BOAS, H.; SAG, I. A. (Ed.). *Signed-Based Construction Grammar*. Stanford: CSLI Publication, 2012. p. 309 – 372. Citado na página 163.
- FILLMORE, C. C. J. Frame semantics. *Linguistics in the morning calm*, Hanshin Publishing Co., Seoul, South Korea, p. 111–137, 1982. Citado 4 vezes nas páginas 22, 32, 44 e 45.
- FILLMORE, C. J. The case for case. In: _____. *Universals in Linguistic Theory*. New York: Holt, Rinehart & Winston, 1968. Citado na página 58.
- FILLMORE, C. J. The need for a frame semantics in linguistics. In: KARLGRÉN, H. (Ed.). *Statistical Methods in Linguistics*. [S.l.]: Scriptor, 1977. Citado na página 44.
- FILLMORE, C. J. Frames and the semantics of understanding. *Quaderni di Semantica*, v. 6, n. 2, p. 222–254, 1985. Citado 2 vezes nas páginas 22 e 44.
- GÄRDENFORS, P. *Conceptual spaces - the geometry of thought*. [S.l.]: MIT Press, 2000. I–X, 1–307 p. ISBN 978-0-262-07199-4. Citado na página 98.
- GÄRDENFORS, P. Semantics Based on Conceptual Spaces. In: *ICLA*. [S.l.: s.n.], 2011. p. 1–11. Citado na página 98.
- GAWRON, J. M. Frame Semantics. In: MAIENBORN, C.; HEUSINGER, K. von; PORTNER, P. (Ed.). *Semantics: An International Handbook of Natural Language Meaning*. Berlin: Mouton de Gruyter, 2008, (HSK Handbooks of Linguistics and Communication Science Series, v. 23). p. 1–35. Citado na página 44.
- GEERAERTS, D. *Theories of Lexical Semantics*. Oxford: Oxford University Press, 2009. 368 p. ISBN 9780198700302. Disponível em: <<http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780198700302.001.0001/acprof-9780198700302>>. Citado na página 24.
- GEERAERTS, D.; CUYCKENS, H. (Ed.). *The Oxford Handbook of Cognitive Linguistics*. Oxford: Oxford University Press, 2007. Citado na página 39.
- GÄRDENFORS, P. Conceptual spaces as a framework for knowledge representation. *Mind and Matter*, v. 2, p. 9–27, 2004. Citado na página 98.
- GRIES, S. Corpus-based methods and cognitive semantics: The many senses of to run. ... IN *LINGUISTICS STUDIES AND MONOGRAPHS*, 2006. Citado 2 vezes nas páginas 42 e 43.
- GRUBER, T. R. A Translation Approach to Portable Ontology Specification. In: *Proceedings of Japanese Knowledge Acquisition Workshop (JKAW92)*. [S.l.: s.n.], 1993. Citado na página 60.
- GRUBER, T. R. Towards principles for the design of ontologies used for knowledge sharing. In: GUARINO, N.; POLI, R. (Ed.). *Formal Ontology in Conceptual Analysis and Knowledge Representation*. [S.l.]: Kluwer., 1994. v. 43, n. 5-6, p. 907–928. Citado na página 60.
- GUARINO, N. Formal Ontology and Information Systems. n. June, p. 3–15, 1998. Citado 2 vezes nas páginas 60 e 61.

- HAUSER, M. D.; CHOMSKY, N.; FITCH, W. T. The faculty of language: what is it, who has it, and how did it evolve? *Science (New York, N.Y.)*, v. 298, n. 5598, p. 1569–79, dez. 2002. ISSN 1095-9203. Citado na página 38.
- HIRST, G. Semantic interpretation and ambiguity. *Artificial Intelligence*, v. 34, n. 2, p. 131–177, mar. 1988. ISSN 00043702. Citado na página 77.
- HIRST, G. Ontology and the Lexicon. In: *Handbook on ontologies*. [S.l.: s.n.], 2009. p. 269–292. ISBN 978-3-540-70999-2. Citado na página 62.
- HOFFMANN, T.; TROUSDALE, G. *The Oxford Handbook of Construction Grammar*. OUP USA, 2013. (Oxford Handbooks Series). ISBN 9780195396683. Disponível em: <<http://books.google.de/books?id=Lahhtcy60LgC>>. Citado na página 58.
- HORROCKS, I. Reasoning with expressive description logics: Theory and practice. In: VORONKOV, A. (Ed.). *Proceedings of the 19th International Conference on Automated Deduction (CADE 2002)*. [S.l.]: Springer, 2002. Citado na página 68.
- HUANG, C.-R. et al. *Ontology and the Lexicon*. Cambridge, MA: Cambridge University Press, 2010. Citado 4 vezes nas páginas 62, 63, 64 e 65.
- JACKENDOFF, R. *A User's Guide to Thought and Meaning*. [S.l.]: OUP Oxford, 2012. ISBN 9780191620683. Citado na página 161.
- JEZEK, E.; LENCI, A. When GL meets the corpus types and coercion phenomena Research goals and methodology. 2007. Citado na página 103.
- JOHNSON, M. Image-schematic basis of meaning. v. 9, 1989. Citado na página 127.
- JURAFSKY, D.; MARTIN, J. H. *Speech and Language Processing*. 2. ed. [S.l.]: Prentice Hall, 2008. ISBN 0131873210. Citado na página 75.
- KANEIWA, K.; IWAZUME, M.; FUKUDA, K. An Upper Ontology for Event Classifications and Relations. In: *Proceedings of the 20th Australian Joint Conference on Advances in Artificial Intelligence*. Berlin, Heidelberg: Springer-Verlag, 2007. (AI'07), p. 394–403. ISBN 3-540-76926-9, 978-3-540-76926-2. Citado na página 95.
- KOKAR, M. M.; MATHEUS, C. J.; BACLAWSKI, K. Ontology-based situation awareness. *Information Fusion*, Elsevier B.V., v. 10, n. 1, p. 83–98, jan. 2009. ISSN 15662535. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S1566253507000218>>. Citado na página 65.
- LAKOFF, G. *Women, Fire, and Dangerous Things*. Chicago: University of Chicago Press, 1987. ISBN 9780226468044. Citado 3 vezes nas páginas 40, 41 e 42.
- LAKOFF, G. Some empirical results about the nature of concepts in mind and language. *Mind and Language*, 1990. Citado na página 44.
- LAKOFF, G. Explaining embodied cognition results. *Topics in cognitive science*, v. 4, n. 4, p. 773–85, out. 2012. ISSN 1756-8765. Citado na página 127.
- LAKOFF, G.; JOHNSON, M. *Metaphors we live by*. Chicago: University of Chicago Press, 1987. Citado 2 vezes nas páginas 43 e 122.

- LANGACKER, R. W. *Foundations of Cognitive Grammar, Vol. 1*. Stanford: Stanford University Press, 1987. Citado na página 43.
- LEHMANN, J. et al. DBpedia - A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia. *Semantic Web Journal*, 2014. Citado na página 164.
- LENCI, A. Building an ontology for the lexicon: Semantic types and word meaning. *Ontology-Based Interpretation of Noun Phrases: Proceedings of the First International OntoQuery Workshop*, 2001. Citado na página 70.
- LENCI, A. et al. SIMPLE _ Linguistic Specifications. n. March, p. 1–405, 2000. Citado 2 vezes nas páginas 70 e 72.
- LEWANDOWSKA-TOMASZCZYK, N. Polysemy, Prototypes and Radial Categories. In: *The Oxford Handbook of Cognitive Linguistics*. Oxford: Oxford University Press, 2007. cap. 6, p. 140–169. Citado na página 41.
- LIEN, C. A frame-based account of lexical polysemy in Taiwanese. *Language and Linguistics*, n. 1873, p. 119–138, 2000. Citado na página 25.
- LIU, M.-C.; CHIANG, T.-Y.; CHOU, M.-H. A frame-based approach to polysemous near-synonymy: the case with mandarin verbs of expression. *Journal of Chinese Language and Computing*, v. 15, n. 3, p. 137–148, 2005. Disponível em: <<http://dblp.uni-trier.de/db/journals/jclc/jclc15.html#LiuCC05>>. Citado na página 25.
- LUTZ, C. *Description Logics Home Page*. 2006. Disponível em: <<http://dl.kr.org>>. Citado na página 66.
- LYONS, J. *Introdução à Linguística teórica*. São Paulo: Companhia Editorial Nacional, 1979. Citado na página 41.
- MASOLO, C. et al. *Wonderweb deliverable18, ontology library (final)*. [S.l.], 2003. Citado 4 vezes nas páginas 33, 94, 95 e 163.
- MATSUBAYASHI, Y.; OKAZAKI, N.; TSUJII, J. A Comparative Study on Generalization of Semantic Roles in FrameNet. In: *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*. [S.l.]: Association for Computational Linguistics, 2009. p. 19–27. Citado na página 82.
- MOREIRA, A.; SALOMAO, M. M. M. Análise ontológica aplicada ao desenvolvimento de Frames. *Alfa : Revista de Linguística (São José do Rio Preto)*, v. 56, n. 2, p. 491–521, dez. 2012. ISSN 1981-5794. Citado 3 vezes nas páginas 25, 47 e 56.
- MOURA, H. M. d. M.; PEREIRA, J. S. d. V. A interface léxico-enciclopédia no léxico gerativo: um estudo do verbo preparar. *Revista Anpoll*, v. 1, n. 16, p. 57–73, 2004. Citado na página 36.
- NARDI, D.; BRACHMAN, R. J. An Introduction to Description Logics. In: *The description logic handbook: Theory, implementation and applications2003*. [S.l.: s.n.], 2003. p. 5–41. ISBN 0521781760. Citado na página 66.

- NUZZOLESE, A. G.; GANGEMI, A.; PRESUTTI, V. Gathering lexical linked data and knowledge patterns from FrameNet. *Proceedings of the sixth international conference on Knowledge capture - K-CAP '11*, ACM Press, New York, New York, USA, p. 41, 2011. Citado na página 82.
- OSSWALD, R.; Van Valin Jr, R. D. FrameNet, frame structure, and the syntax-semantics interface. *Frames and Concept Types*, 2014. Citado 2 vezes nas páginas 25 e 51.
- OVCHINNIKOVA, E. *Integration of World Knowledge for Natural Language Understanding*. Amsterdam: Atlantis Press, 2012. Citado 4 vezes nas páginas 25, 56, 59 e 60.
- PETRUCK, M. R. L. Frame Semantics. In: VERSCHUEREN, J. et al. (Ed.). *Handbook of Pragmatics*. [S.l.]: John Benjamins, 1996. Citado na página 44.
- PRÉVOT, L.; BORGIO, S.; OLTRAMARI, R. A.: Interfacing ontologies and lexical resources. p. 1–12, 2005. Citado na página 80.
- PUSTEJOVSKY, J. The Generative Lexicon. *Computational Linguistics*, The MIT Press, Cambridge (USA), v. 17, n. 4, p. 409–441, 1991. Citado 3 vezes nas páginas 33, 34 e 103.
- PUSTEJOVSKY, J. *The Generative Lexicon*. Cambridge, MA: MIT Press, 1995. Citado 7 vezes nas páginas 26, 33, 35, 36, 37, 105 e 113.
- PUSTEJOVSKY, J. *Generativity and Explanation in Semantics: A Reply to Fodor and Lepore*. 1998. 289–311 p. Citado na página 37.
- PUSTEJOVSKY, J. The Semantics of Lexical Underspecification. *Folia Linguistica*, v. 32, n. 3-4, p. 1–28, 1998. ISSN 0165-4004. Citado 2 vezes nas páginas 33 e 109.
- PUSTEJOVSKY, J. Type theory and lexical decomposition. *Journal of Cognitive Science*, p. 1–38, 2006. Citado 4 vezes nas páginas 102, 103, 104 e 111.
- PUSTEJOVSKY, J. et al. Towards a Generative Lexical Resource: The Brandeis Semantic Ontology. In: *Proceedings of the Fifth Language Resource and Evaluation Conference*. [S.l.: s.n.], 2006. Citado na página 94.
- RAIMOND, Y.; ABDALLAH, S. *The Event Ontology*. 2007. Citado na página 95.
- ROSA, J. a. L. G. *Fundamentos da Inteligência Artificial*. Rio de Janeiro: LTC, 2011. Citado 2 vezes nas páginas 57 e 58.
- RUIMY, N.; MONACHINI, M.; DISTANTE, R. CLIPS, a Multi-level Italian Computational Lexicon: a Glimpse to Data. *LREC*, p. 792–799, 2002. Citado na página 70.
- RUPPENHOFER, J. et al. *FrameNet II: Extended Theory and Practice*. Berkeley, California: International Computer Science Institute, 2010. Citado 9 vezes nas páginas 48, 50, 52, 55, 81, 90, 121, 130 e 136.
- RUSSELL, S. J.; NORVIG, P. *Artificial Intelligence: a Modern Approach*. 3a.. ed. New Jersey: Pearson, 2010. ISBN 9780136042594. Citado na página 67.
- SALOMÃO, M. et al. *Copa 2014 Framenet Brasil*. Juiz de Fora, 2011. Citado 4 vezes nas páginas 22, 28, 32 e 49.

- SALOMÃO, M. M. FrameNet Brasil: um trabalho em progresso. *Calidoscópico*, v. 7, n. 3, p. 171–182, dez. 2009. ISSN 16798740. Citado 6 vezes nas páginas [25](#), [45](#), [46](#), [47](#), [104](#) e [137](#).
- SALOMÃO, M. M. M. Gramática das construções: a questão da integração entre sintaxe e léxico. *Revista Veredas*, v. 6, n. 1, p. 63–74, 2002. Citado na página [38](#).
- SCHÄFER, F. The Causative Alternation. *Language and Linguistics Compass*, v. 3, n. 2, p. 641–681, mar. 2009. ISSN 1749818X. Citado na página [120](#).
- SCHEFFCZYK, J.; BAKER, C. F.; NARAYANAN, S. Reasoning over Natural Language Text by Means of FrameNet and Ontologies. In: _____. *Ontology and the Lexicon: A Natural Language Processing Perspective*. Cambridge, UK: Cambridge University Press, 2010. (Studies in Natural Language Processing), cap. 4, p. 53–71. Citado 2 vezes nas páginas [56](#) e [82](#).
- SCHEFFCZYK, J.; PEASE, A.; ELLSWORTH, M. Linking framenet to the suggested upper merged ontology. *Formal Ontology in Information Systems: ...*, 2006. Citado na página [56](#).
- SCHERP, A. et al. F—a Model of Events Based on the Foundational Ontology Dolce+DnS Ultralight. In: *Proceedings of the Fifth International Conference on Knowledge Capture*. New York, NY, USA: ACM, 2009. (K-CAP '09), p. 137–144. ISBN 978-1-60558-658-8. Citado na página [95](#).
- SEARLE, J. R. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press, 1983. (Cambridge paperback library). ISBN 9780521273022. Citado na página [21](#).
- SHASTRI, L. Structured Connectionist Models of Semantic Networks. *Computers & Mathematics with Applications*, v. 23, n. 2, p. 293–328, 1992. Citado 2 vezes nas páginas [76](#) e [88](#).
- SHASTRI, L. Advances in Shruti— A neurally motivated model of relational knowledge representation and rapid inference using temporal synchrony. *Applied Intelligence*, v. 11, p. 79–108, 1999. Citado na página [164](#).
- SHASTRI, L.; AJJANAGADDE, V. A connectionist system for rule based reasoning with multi-place predicates and variables. 1989. Citado na página [76](#).
- SHUTOVA, E. et al. A Computational Model of Logical Metonymy. *ACM Trans. Speech Lang. Process.*, ACM, New York, NY, USA, v. 10, n. 3, p. 11:1—11:28, 2013. ISSN 1550-4875. Citado na página [122](#).
- SOWA, J. F. Processes and Participants. In: EKLUND, P. W.; ELLIS, G.; MANN, G. (Ed.). *ICCS*. [S.l.]: Springer, 1996. (Lecture Notes in Computer Science, v. 1115), p. 1–22. ISBN 3-540-61534-2. Citado 2 vezes nas páginas [105](#) e [106](#).
- SOWA, J. F. *Knowledge Representation: Logical, Philosophical and Computational Foundations*. Pacific Grove, CA, USA: Brooks/Cole Publishing Co., 2000. ISBN 0-534-94965-7. Citado 2 vezes nas páginas [105](#) e [106](#).
- TALMY, L. Force dynamics in language and cognition. *Cognitive science*, 1988. Citado na página [185](#).

- TALMY, L. *Toward a Cognitive Semantics, Vol 1: Concept Structuring Systems*. [S.l.]: MIT Press, 2000. Citado 14 vezes nas páginas 27, 28, 52, 53, 185, 186, 188, 189, 190, 191, 192, 193, 194 e 195.
- TALMY, L. Foreword. In: GONZALEZ-MARQUEZ, M. et al. (Ed.). *Methods in Cognitive Linguistics*. Amsterdam: John Benjamins Publishing Company, 2006. cap. Foreword, p. xi–xxi. Citado na página 28.
- TONELLI, S. et al. Investigating the semantics of frame elements. In: TEIJE, A. et al. (Ed.). *Knowledge Engineering and Knowledge Management*. [S.l.]: Springer Berlin Heidelberg, 2012. p. 130–143. Citado 2 vezes nas páginas 81 e 82.
- TORAL, A.; MONACHINI, M. SIMPLE-OWL: a generative lexicon ontology for nlp and the semantic web. In: *Workshop of Cooperative Construction of Linguistic Knowledge Bases*. [S.l.: s.n.], 2007. Citado 2 vezes nas páginas 73 e 74.
- TORRENT, T. T.; ELLSWORTH, M. Behind the Labels: Criteria for defining analytical categories in FrameNet Brasil. *Revista veredas*, v. 17, n. 1, p. 44–65, 2013. Citado na página 137.
- TSATSARONIS, G.; VAZIRGIANNIS, M.; ANDROUTSOPOULOS, I. Word Sense Disambiguation with Spreading Activation Networks Generated from Thesauri. *IJCAI 2007*, p. 1725–1730, 2007. Citado 2 vezes nas páginas 77 e 87.
- VERONIS, J.; IDE, N. Word sense disambiguation with very large neural networks extracted from machine readable dictionaries. *Proceedings of the 13th conference on . . .*, p. 389–394, 1990. Citado na página 77.
- W3C. *OWL2*. 2012. Disponível em: <<http://www.w3.org/TR/owl2-overview/>>. Citado na página 68.

Apêndices

APÊNDICE A – Modelo Conexionista

A.1 Ponderação dos links

No Modelo Conexionista do *framework* LUDI, o peso atribuído a cada *link* é sempre 1.0, exceto para os *links* em que o destino é um tipo semântico da ontologia SIMPLE. Neste caso, o peso do *link* é calculado em função da posição do tipo na hierarquia da ontologia. A função logística adaptada (A.1) é usada para este cálculo.

$$(A.1) \quad \text{Peso}(\text{nível}) = \frac{1 + (1 * \exp(-1 * \text{nível}))}{1 + (50 * \exp(-1 * \text{nível}))}$$

A Tabela 14 mostra os Tipos Semânticos, organizados por nível na hierarquia e o peso associado ao *link*.

Tabela 14 – Peso do *link* por Tipo Semântico SIMPLE.

| Nível | Tipos Semânticos | Peso |
|-------|---|--------|
| 1 | Top | 0,0705 |
| 2 | Agentive, AgentiveType, Constitutive, ConstitutiveType, Entity, Telic, TelicType, Type | 0,1461 |
| 3 | Abstract_entity, Amount, Cause, Concrete_entity, Event, Group, Part, Quality, Representation, Substance_phase | 0,3008 |
| 4 | Abstract_location, Act, Artifact, Aspectual, Body_part, Change_state, Cognitive_fact, Convention, Domain, Drink, Food, Human_group, Information, Institution, Language, Liquid, Living_entity, Location, Material, Metalanguage, Moral_standard, Movement_of_thought, Number, Organic_entity, Phenomenon, Physical_entity, Property_region, Psychological_event, Quality_dimension, Sign, Solid, State, Substance, Time, Unit_of_measurement | 0,5315 |
| 5 | Area, Artifact_food, Artifactual_area, Artifactual_drink, Artifactual_material, Artwork, Attribute_value, Building, Clothing, Cognitive_event, Container, D_3_location, Disease, Event_quality, Existence, Experience_event, Flavouring, Furniture, Geopolitical_location, Instrument, Micro-organism, Modal_event, Money, Motion, Natural_substance, Non_relational_act, Opening, Perception_event, Physical_quality, Psychological_quality, Relational_act, Relational_state, Semiotic_artifact, Sentient, Social_quality, Speech_act, Stimulus, Substance_food, Time_futebol, Vegetal_entity, Vehicle, Weather | 0,7530 |
| 6 | Animal, Configuration_state, Constitutive_state, Flower, Fruit, Human, Identity_state, Location_state, Plant, Possession_state, Time_state, Transfer, Vegetable | 0,8919 |
| 7 | Agent_of_persistent_activity, Agent_of_temporary_activity, Air_animal, Earth_animal, Patient_of_event, People, Profession, Role, Water_animal | 0,9572 |
| 8 | Ideo, Kinship, Social_status, X, Y, Z | 0,9838 |

A.2 Redes de <abrir.v>

As listagens a seguir apresentam exemplos comentados de algumas definições de nós e links ¹, usadas para construir a rede dos experimentos. Foram definidas entradas para o esquema imagético, para o esquema básico e para os vários *Frames*, *Elementos de Frame* e *Valências* apresentados no capítulo 9, além dos lemas, *Unidades Lexicais*, *Unidades Semânticas* e *Tipos Ontológicos* adicionais.

Existem dois formatos para a entrada de dados: a definição de um nó e a definição de um *link*. A definição de um nó segue o formato

N,Tipo_nó,Id_nó

enquanto a definição de um *link* tem o formato

Tipo_link, Nó_origem, Label_link, Nó_destino [,Função_gramatical | Restrição] [,Tipo_de_papel_temático] [,Instanciação_nula]

A.2.1 Definição do *Frame* Esquemático

```
#####
# Frames - Schema
#
# Definicao do Frame
#
N,F,Sch_Container
#
# Definicao dos Elementos de Frame
#
N,E,fe_Sch_Container.Container
N,E,fe_Sch_Container.Boundary
N,E,fe_Sch_Container.Portal
N,E,fe_Sch_Container.Interior
N,E,fe_Sch_Container.Exterior
#
# Links Frame-Elementos de Frame
#
FE,Sch_Container,Sch_Container.Boundary,fe_Sch_Container.Boundary
FE,Sch_Container,Sch_Container.Container,fe_Sch_Container.Container
FE,Sch_Container,Sch_Container.Portal,fe_Sch_Container.Portal
FE,Sch_Container,Sch_Container.Interior,fe_Sch_Container.Interior
FE,Sch_Container,Sch_Container.Exterior,fe_Sch_Container.Exterior
#
# Links Lexema-Lema
#
XL,abrir,usedByLemma,abrir_lemma
```

¹ A definição completa da rede pode ser encontrada no site do projeto: <<http://ludi.ematos.info>>

```
XL,aberto,usedByLemma,aberto_lemma
#####
```

A.2.2 Definição do Esquema Básico

```
#####
# Frames - I.Abrir_esquema
#
# Definicao do Frame
#
N,F,Abrir_esquema
#
# Definicao dos Elementos de Frame
#
N,E,fe_Abrir_esquema.Container
N,E,fe_Abrir_esquema.Fronteira
N,E,fe_Abrir_esquema.Portal
N,E,fe_Abrir_esquema.Conteudo
N,E,fe_Abrir_esquema.Interior
N,E,fe_Abrir_esquema.Exterior
N,E,fe_Abrir_esquema.Agente
N,E,fe_Abrir_esquema.Causa
N,E,fe_Abrir_esquema.Instrumento
#
# Links Frame-Elementos de Frame
#
FE,Abrir_esquema,Abrir_esquema.Container,fe_Abrir_esquema.Container
FE,Abrir_esquema,Abrir_esquema.Fronteira,fe_Abrir_esquema.Fronteira
FE,Abrir_esquema,Abrir_esquema.Portal,fe_Abrir_esquema.Portal
FE,Abrir_esquema,Abrir_esquema.Conteudo,fe_Abrir_esquema.Conteudo
FE,Abrir_esquema,Abrir_esquema.Interior,fe_Abrir_esquema.Interior
FE,Abrir_esquema,Abrir_esquema.Exterior,fe_Abrir_esquema.Exterior
FE,Abrir_esquema,Abrir_esquema.Agente,fe_Abrir_esquema.Agente
FE,Abrir_esquema,Abrir_esquema.Causa,fe_Abrir_esquema.Causa
FE,Abrir_esquema,Abrir_esquema.Instrumento,fe_Abrir_esquema.Instrumento
#
# Links EF-EF com os Elementos de Frame do Frame Esquematico
#
EE,fe_Abrir_esquema.Container,Sch_Container.Container,fe_Sch_Container.
  Container
EE,fe_Abrir_esquema.Fronteira,Sch_Container.Boundary,fe_Sch_Container.
  Boundary
EE,fe_Abrir_esquema.Portal,Sch_Container.Portal,fe_Sch_Container.Portal
EE,fe_Abrir_esquema.Conteudo,Sch_Container.Interior,fe_Sch_Container.
  Interior
EE,fe_Abrir_esquema.Interior,Sch_Container.Interior,fe_Sch_Container.
  Interior
```

```

EE,fe_Abrir_esquema.Exterior,Sch_Container.Exterior,fe_Sch_Container.
  Exterior
#
# Associa{\c c}{\~a}o de Tipos Ontologicos aos Elementos de Frame
#
ET,fe_Abrir_esquema.Agente,hasType,AgentiveType
ET,fe_Abrir_esquema.Causa,hasType,AgentiveType
ET,fe_Abrir_esquema.Instrumento,hasType,AgentiveType
#####

```

A.2.3 Definição do Cluster II.Movimento_corporal

```

#####
# Cluster - II.Movimento_corporal
#
# Definicao dos Frames
#
N,F,Movimento_corporal_causativo
N,F,Movimento_corporal_incoativo
N,F,Movimento_corporal_estativo
#
# Definicao dos Elementos de Frame
#
N,E,fe_Movimento_corporal_causativo.Agente
N,E,fe_Movimento_corporal_causativo.Causa
N,E,fe_Movimento_corporal_causativo.Instrumento
N,E,fe_Movimento_corporal_causativo.Parte_corporal
N,E,fe_Movimento_corporal_causativo.Experenciador
N,E,fe_Movimento_corporal_incoativo.Parte_corporal
N,E,fe_Movimento_corporal_incoativo.Experenciador
N,E,fe_Movimento_corporal_estativo.Parte_corporal
N,E,fe_Movimento_corporal_estativo.Experenciador
#
# Definicao das Valencias (relativas a <abrir.v>)
#
FV,Movimento_corporal_causativo,valence_v1,Movimento_corporal_causativo.
  v1
FV,Movimento_corporal_causativo,valence_v2,Movimento_corporal_causativo.
  v2
FV,Movimento_corporal_causativo,valence_v3,Movimento_corporal_causativo.
  v3
FV,Movimento_corporal_incoativo,valence_v1,Movimento_corporal_incoativo.
  v1
FV,Movimento_corporal_estativo,valence_v1,Movimento_corporal_estativo.v1
#
# Definicao dos EF de cada Valencia; para cada EF sao definidos:
# - Funcao gramatical: ext, dep, dobj, iobj

```

```
# - Tipo de papel tematico: agentiveRole ou telicRole
# - se permite omissao ou nao (ni)
#
VE,Movimento_corporal_causativo.v1,fe_Movimento_corporal_causativo.
  Agente.ext,fe_Movimento_corporal_causativo.Agente,ext,agentiveRole
VE,Movimento_corporal_causativo.v1,fe_Movimento_corporal_causativo.
  Parte_corporal.dobj,fe_Movimento_corporal_causativo.Parte_corporal,
  dobj,telicRole
VE,Movimento_corporal_causativo.v1,fe_Movimento_corporal_causativo.
  Experenciador.dep,fe_Movimento_corporal_causativo.Experenciador,dep,
  telicRole,ni
#
VE,Movimento_corporal_causativo.v2,fe_Movimento_corporal_causativo.Causa
  .ext,fe_Movimento_corporal_causativo.Causa,ext,agentiveRole
VE,Movimento_corporal_causativo.v2,fe_Movimento_corporal_causativo.
  Parte_corporal.dobj,fe_Movimento_corporal_causativo.Parte_corporal,
  dobj,telicRole
VE,Movimento_corporal_causativo.v2,fe_Movimento_corporal_causativo.
  Experenciador.dep,fe_Movimento_corporal_causativo.Experenciador,dep,
  telicRole
#
VE,Movimento_corporal_causativo.v3,fe_Movimento_corporal_causativo.
  Instrumento.ext,fe_Movimento_corporal_causativo.Instrumento,ext,
  agentiveRole
VE,Movimento_corporal_causativo.v3,fe_Movimento_corporal_causativo.
  Parte_corporal.dobj,fe_Movimento_corporal_causativo.Parte_corporal,
  dobj,telicRole
VE,Movimento_corporal_causativo.v3,fe_Movimento_corporal_causativo.
  Experenciador.dep,fe_Movimento_corporal_causativo.Experenciador,dep,
  telicRole
#
VE,Movimento_corporal_incoativo.v1,fe_Movimento_corporal_incoativo.
  Parte_corporal.dobj,fe_Movimento_corporal_incoativo.Parte_corporal,
  dobj,telicRole
VE,Movimento_corporal_incoativo.v1,fe_Movimento_corporal_incoativo.
  Experenciador.dep,fe_Movimento_corporal_incoativo.Experenciador,dep,
  telicRole
#
VE,Movimento_corporal_estativo.v1,fe_Movimento_corporal_estativo.
  Parte_corporal.dobj,fe_Movimento_corporal_estativo.Parte_corporal,
  dobj,telicRole
VE,Movimento_corporal_estativo.v1,fe_Movimento_corporal_estativo.
  Experenciador.dep,fe_Movimento_corporal_estativo.Experenciador,dep,
  telicRole
#
# Associa{\c c}{\~a}o de Tipos Ontologicos aos Elementos de Frame
#
```

```
ET,fe_Movimento_corporal_causativo.Agente,hasType,Sentient,e
ET,fe_Movimento_corporal_causativo.Causa,hasType,Event,e
ET,fe_Movimento_corporal_causativo.Instrumento,hasType,Instrument,e
ET,fe_Movimento_corporal_causativo.Parte_corporal,hasType,Body_part,e
ET,fe_Movimento_corporal_causativo.Experenciador,hasType,Sentient
ET,fe_Movimento_corporal_incoativo.Parte_corporal,hasType,Body_part,e
ET,fe_Movimento_corporal_incoativo.Experenciador,hasType,Sentient
ET,fe_Movimento_corporal_estativo.Parte_corporal,hasType,Body_part,e
ET,fe_Movimento_corporal_estativo.Experenciador,hasType,Sentient
#
# Links inibitorios para tratamento da alternancia causativo-incoativo
# - a ocorrencia do frame causativo inibe a ocorrencia dos frames
#   incoativo e estativo
#
VE,Movimento_corporal_incoativo.v1,inhibitedBy,
  fe_Movimento_corporal_causativo.Agente,ext,1
VE,Movimento_corporal_estativo.v1,inhibitedBy,
  fe_Movimento_corporal_causativo.Agente,ext,1
#
# Links inibitorios para tratamento do CoreSet
#
EE,fe_Movimento_corporal_causativo.Agente,inhibits,
  fe_Movimento_corporal_causativo.Causa,ext,1
EE,fe_Movimento_corporal_causativo.Agente,inhibits,
  fe_Movimento_corporal_causativo.Instrumento,ext,1
#
# Links EF-EF com o Esquema Basico
#
EE,fe_Movimento_corporal_causativo.Agente,Abrir_esquema.Agente,
  fe_Abrir_esquema.Agente
EE,fe_Movimento_corporal_causativo.Causa,Abrir_esquema.Causa,
  fe_Abrir_esquema.Causa
EE,fe_Movimento_corporal_causativo.Instrumento,Abrir_esquema.Instrumento,
  fe_Abrir_esquema.Instrumento
EE,fe_Movimento_corporal_causativo.Parte_corporal,Abrir_esquema.Portal,
  fe_Abrir_esquema.Portal
EE,fe_Movimento_corporal_causativo.Experenciador,Abrir_esquema.Container,
  fe_Abrir_esquema.Container
EE,fe_Movimento_corporal_incoativo.Parte_corporal,Abrir_esquema.Portal,
  fe_Abrir_esquema.Portal
EE,fe_Movimento_corporal_incoativo.Experenciador,Abrir_esquema.Container,
  fe_Abrir_esquema.Container
EE,fe_Movimento_corporal_estativo.Parte_corporal,Abrir_esquema.Portal,
  fe_Abrir_esquema.Portal
EE,fe_Movimento_corporal_estativo.Experenciador,Abrir_esquema.Container,
  fe_Abrir_esquema.Container
#
```

```
# Evocacao dos Frames pelos Lemas
#
LF,abrir_lemma,evokes,Movimento_corporal_causativo
LF,abrir_lemma,evokes,Movimento_corporal_incoativo
LF,abrir_lemma,evokes,Movimento_corporal_estativo
#####
```

A.2.4 Exemplos de Unidades da Ontologia

Estes exemplos mostram como é realizada a integração da FrameNet com a Ontologia SIMPLE.

```
#####
#
# Ontologia
#
# Associacao entre Tipos Ontologicos e o Frame Esquematico;
#   indicam a conceptualizacao associada a um Tipo.
#
TE,Container,Sch_Container.Container,fe_Sch_Container.Container
TE,Building,Sch_Container.Container,fe_Sch_Container.Container
TE,Opening,Sch_Container.Portal,fe_Sch_Container.Portal
TE,Semiotic_artifact,Sch_Container.Container,fe_Sch_Container.Container
#
# Associacoes Lema-Unidade Semantica e Unidade Semantica-Tipo Ontologico
#
LU,porta,relatedTo,porta_as_opening
UT,porta_as_opening,hasType,Opening
#
LU,empregada,relatedTo,empregada_as_profession
UT,empregada_as_profession,hasType,Profession
#
LU,carteira,relatedTo,carteira_as_container
UT,carteira_as_container,hasType,Container
#
LU,boca,relatedTo,boca_as_body_part
UT,boca_as_body_part,hasType,Body_part
#
LU,mao,relatedTo,mao_as_body_part
UT,mao_as_body_part,hasType,Body_part
#
LU,olho,relatedTo,olho_as_body_part
UT,olho_as_body_part,hasType,Body_part
#
LU,eu,relatedTo,eu_as_human
UT,eu_as_human,hasType,Human
#
LU,gente,relatedTo,gente_as_human
```

```
UT, gente_as_human , hasType , Human
#
LU, jornal , relatedTo , jornal_as_information
UT, jornal_as_information , hasType , Information
#
LU, jornal , relatedTo , jornal_as_artifact
UT, jornal_as_artifact , hasType , Semiotic_artifact
#
LU, exposicao , relatedTo , exposicao_as_event
UT, exposicao_as_event , hasType , Event
#
LU, museu , relatedTo , museu_as_building
UT, museu_as_building , hasType , Building
#
LU, domingo , relatedTo , domingo_as_time
UT, domingo_as_time , hasType , Time
#
LU, segunda_feira , relatedTo , segunda_feira_as_time
UT, segunda_feira_as_time , hasType , Time
#
LU, hoje , relatedTo , hoje_as_time
UT, hoje_as_time , hasType , Time
#
LU, shopping , relatedTo , shopping_as_building
UT, shopping_as_building , hasType , Building
#
LU, shopping , relatedTo , shopping_as_seller
UF, shopping_as_seller , evokes , Commerce_sell
#
# Atribuicao de um Tipo Ontologico a um Frame
#
FT, Commerce_sell , isA , Relational_act
#
LU, posicao , relatedTo , posicao_as_cognitive_fact
UT, posicao_as_cognitive_fact , hasType , Cognitive_fact
#
LU, empresario , relatedTo , empresario_as_profession
UT, empresario_as_profession , hasType , Profession
#
LU, vaqueiro , relatedTo , vaqueiro_as_profession
UT, vaqueiro_as_profession , hasType , Profession
#
LU, faca , relatedTo , faca_as_instrument
UT, faca_as_instrument , hasType , Instrument
#
LU, caminho , relatedTo , caminho_as_opening
UT, caminho_as_opening , hasType , Opening
```

```
LU,caminho,relatedTo,caminho_as_possibility
#
# Evocacao de um Frame por uma Unidade Semantica
#
UF,caminho_as_possibility,evokes,Possibilities
#
# Links inibitorios entre Frames
#
FF,Possibilities,inhibits,Abrir_fisico_causativo
#
LU,reestruturacao,relatedTo,reestruturacao_as_event
UT,reestruturacao_as_event,hasType,Event
#
LU,investimento,relatedTo,investimento_as_event
UT,investimento_as_event,hasType,Event
#
LU,informatica,relatedTo,informatica_as_profession
UT,informatica_as_profession,hasType,Profession
LU,informatica,relatedTo,informatica_as_subject
#
# Associa{\c c}{\~a}o de uma Unidade Semantica a um Elemento de Frame
#
UE,informatica_as_subject,relatedTo,fe_Education_teaching.Subject
#
LU,curso,relatedTo,curso_as_event
UE,curso_as_event,Education_teaching.Course,fe_Education_teaching.Course
#
LU,garrafa,relatedTo,garrafa_as_container
UT,garrafa_as_container,hasType,Container
UT,garrafa_as_container,contains,Liquid
#
LU,cerveja,relatedTo,cerveja_as_artifactual_drink
UT,cerveja_as_artifactual_drink,hasType,Artifactual_drink
#
# Relacao qualia entre Tipos Ontologicos
#
TT,Drink,madeOf,Liquid
TT,Artifactual_drink,containedIn,Container
UT,cerveja_as_artifactual_drink,madeOf,Liquid
#
LU,cerveja2,relatedTo,cerveja2_as_artifactual_drink
UT,cerveja2_as_artifactual_drink,hasType,Artifactual_drink
TT,Drink,madeOf,Liquid
TT,Artifactual_drink,containedIn,Container
UT,cerveja2_as_artifactual_drink,madeOf,Liquid
#
LU,marcador,relatedTo,marcador_as_artifact
```

```
UT,marcador_as_artifact,hasType,Artifact
UT,marcador_as_artifact,usedFor,marcar_lemma
LF,marcar_lemma,evokes,Recording
#
LU,botafogo,relatedTo,botafogo_as_time_futebol
UT,botafogo_as_time_futebol,hasType,Time_futebol
#####
```

APÊNDICE B – Causação

Os linguistas têm reconhecido, já a muito tempo, que a causação possui um papel importante no estudo do significado. Muito da estrutura sintática das sentenças é organizado em torno de eventos (ou de situações), bem como de suas relações causais. Tradicionalmente, porém, a causação tem tomado a forma de uma relação primitiva CAUSE, com pouca elucidação a seu respeito. A causação é discutida aqui como base para o estudo da alternância causativo-incoativo, na seção 9.1.1.

Copley e Wolff (2013), analisando o tema, acreditam que "desempacotar" a noção de causação poderia proporcionar uma compreensão mais profunda de alguns fenômenos linguísticos e de sua sustentação na estrutura conceptual. Eles classificam então as teorias de causação em dois grandes grupos: teorias da dependência e teorias da produção.

B.1 Teorias da dependência

As Teorias da Dependência têm em comum a ideia de que a causação pode ser explicada por meio de uma dependência entre a causa (C) e o efeito (E). Nesta teoria a causação pode ser reduzida à correlação ou regularidade, se as condições forem ajustadas e as variáveis, potencialmente confusas, forem fixadas corretamente. Dentre as teorias analisadas por Copley e Wolff (2013) estão: dependência lógica, dependência contrafactual, probabilidade e modelos Bayesianos.

A **Teoria da Dependência Lógica** se apoia no apelo intuitivo de que a causação pode ser definida em termos de condições necessárias e/ou suficientes. Uma causa é um fator sem o qual o efeito nunca teria ocorrido, ou seja, a causa é necessária para o efeito. Versões mais brandas da teoria consideram a causação apenas como uma condição suficiente.

A **Teoria da Dependência Contrafactual** é baseada na proposição contrafactual de que E poderia não ter ocorrido sem C. Na realidade esta teoria pode ser pensada como uma paráfrase da proposição de que C é necessário para E.

A **Teoria da Regularidade** assume que, se é verdade que um evento C causa um evento E, é verdade que eventos similares a C são invariavelmente seguidos de eventos similares a E. Uma das dificuldades bem conhecidas desta teoria é a de que existem causas que não são invariavelmente seguidas por seus efeitos. Esta observação motivou o surgimento de teorias de **Dependência Probabilística**. Em termos básicos, a ideia é que uma variável C aumenta a probabilidade de uma variável E se a probabilidade de E, dado C, for maior que a probabilidade de E, na ausência de C. Nesta teoria, por exemplo, se fumar causa cancer, a probabilidade de um fumante ter cancer é maior que a probabilidade

de um não-fumante ter câncer.

Finalmente, **Redes Causais Bayesianas** podem ser usadas para representar a estrutura causal de um domínio e sua distribuição de probabilidades subjacente. A estrutura causal é representada por um grafo acíclico direto com nós e setas, em que a distribuição de probabilidade consiste de probabilidades condicionais e incondicionais associadas a cada nó. Usando a teoria da probabilidade é possível fazer previsões sobre as relações causais.

B.2 Teorias da produção

As Teorias da Produção caracterizam a causação em termos de conceitos tais como força e energia, vendo a causação como uma produção ou um processo. Esta produção pode envolver a transmissão de quantidades conservadas de energia, pode ser vista em termos de um "poder causal" que habilita as entidades a receberem ou transmitirem energia ou ainda pode ser considerada em termos de forças sendo comunicadas de um agente para um paciente. [Copley e Wolff \(2013\)](#) fazem a análise de algumas propostas associadas a estas teorias, resumidas a seguir.

Uma primeira proposta é compreender a causação primariamente como um processo, ao invés de ser uma relação entre eventos. Um processo causal é visto como a transmissão de uma "marca causal", ou seja, a propagação de uma modificação local na estrutura. Um processo causal poderia ser instanciado se, por exemplo, alguém põe um pedaço de vidro vermelho em frente a uma fonte de luz. Neste caso, o pedaço de vidro comunicaria uma marca ao processo, que transmitiria a marca para um local diferente, por exemplo, uma parede.

Outra proposta considera que a causação envolve a transferência de uma quantidade conservada. Neste caso, haveria dois tipos principais de causação: persistência e interações. As interações causais ocorreriam quando as trajetórias de dois objetos se interceptam e há uma troca das quantidades conservadas.

Finalmente, é proposto ainda que a causação seja especificada em termos de configurações de forças, avaliadas em relação a um vetor de estado final. Diferentes configurações de forças são definidas de acordo com a tendência do paciente ao estado final, a concordância do agente e do paciente e a força resultante agindo no paciente. Estas diferentes configurações possibilitam a existência de diferentes categorias de relações causais, incluindo as categorias de impedimento e causa. Este tipo de teoria tem um paralelo na Linguística Cognitiva, em especial nos modelos desenvolvidos por Talmy (([TALMY, 1988](#)),([TALMY, 2000](#))), apresentados a seguir.

B.3 Semântica da Causação

Talmy (2000) apresenta um estudo em que a semântica da causação é analisada conforme geralmente é representada na linguagem. Ele diferencia diversos tipos de situações causativas e mostra como estes tipos podem ser combinados. Talmy também destaca, como Copley e Wolff (2013), que os estudos sobre causação mostram que não existe uma noção situacional única, como muitos tratamentos linguísticos tem considerado, em que possa ser usado apenas o verbo CAUSE. Ao contrário, os diversos tipos causativos podem ser associados a outros verbos que perfitem, por exemplo, o resultado, o evento em si, o instrumento usado, a autoria ou a agentividade. Nesta seção são apresentados os principais tipos analisados, com algumas sentenças ilustrativas.

B.3.1 Eventos autônomos

O termo "causativo", conforme a análise semântica feita por Talmy, deve ser entendido de maneira diferente da noção científica de causa, no mundo físico. Cientificamente, a totalidade de um fenômeno constitui um *continuum* causal, do qual uma porção delimitada conceitualmente (um evento) é compreendido como estando relacionado a uma causa externa, além de conter relações causais em si mesmo. O evento da água se derramando de um tanque, por exemplo, é compreendido como a ação da força gravitacional, exercida pela Terra sobre a água. No entanto, linguisticamente, uma sentença pode especificar um evento que é percebido como acontecendo por si mesmo, sem uma relação causal. Por exemplo, a sentença (B.0) é dita não-causativa e especifica um evento considerado autônomo.

(B.0) A água derramou do tanque.

B.3.2 Situação causativa básica

A situação causativa básica considerada por Talmy consiste de três elementos: um evento simples (ou seja, um evento que poderia ser considerado autônomo), "alguma coisa" que imediatamente causa o evento e a relação causal entre os dois. Considerando que "alguma coisa" seja também vista como um evento, esta composição poderia ser representada sintaticamente por uma das duas estruturas (B.1)a ou (B.1)b.

(B.1) (a) [Evento-A] causa [evento-B]
(b) [Evento-B] resulta de [evento-A]

Na terminologia adotada por Talmy, o Evento-A é chamado de **evento causador** (*causing event*) e o evento-B de **evento causado** (*caused event*). Nas sentenças (B.2)a e (B.2)b, por exemplo, o impacto da bola é o evento causador e a quebra da janela é o evento causado.

- (B.2) (a) O impacto da bola quebrou a janela. (O impacto da bola causou a quebra da janela)
- (b) A janela quebrou com o impacto da bola. (A quebra da janela resultou do impacto da bola)

Uma observação importante, quando se considera a situação causativa básica como um todo, é em relação ao elemento que funciona como Figura (na relação Figura-Fundo) no evento causador. Este elemento funciona como o **Instrumento** para a situação causativa inteira. No exemplos (B.2)a e (B.2)b a bola é considerada como o instrumento.

Outra questão de interesse está relacionada à "oposição dinâmica". Nesta situação parece haver uma oposição entre as tendências de repouso ou movimento das figuras associadas ao evento causador e ao evento causado. Por exemplo, na sentença (B.3) o instrumento do evento causador (o vento) se opõe à tendência de repouso do elemento do evento causado (a bola).

- (B.3) A bola rolou pelo campo levada pelo vento.

B.3.3 Situações causativas complexas

Talmy também analisa situações causativas mais complexas, como arranjos e composições das duas situações anteriores (eventos autônomos e situação causativa básica). Um terceiro elemento (a intencionalidade) pode também ser considerado e é analisado na seção seguinte. Dentre as situações que não envolvem a intencionalidade, pode-se destacar de forma resumida as seguintes:

- a) Destaque (*foregrounding*) de um elemento: são as situações em que parece haver um destaque de um elemento (por exemplo, o Instrumento) e a relação dele com a situação como um todo.
- (B.4) A bola quebrou o vaso caindo sobre ele.
- b) Causação inicial (*onset causation*): são as situações em que parece haver um ponto inicial marcado para o evento.
- (B.5) A caixa deslizou com o meu empurrão.
- c) Causação serial (*serial causation*): são as situações em que o segundo evento causa um terceiro, e o elemento atuando como instrumento do evento anterior pode ser também o instrumento do evento seguinte.
- (B.6) O vento forte derrubou as folhas que cobriram completamente o chão.
- d) Causação habilitada (*enabling causation*): são as situações em que a ocorrência de um evento possibilita a ocorrência de outro.

(B.7) Ele esvaziou o tanque tirando a tampa do ralo.

B.3.4 Agentividade

Talmy (2000, p. 513) caracteriza a noção de agentividade usando três critérios: uma entidade com um corpo (ou partes), que possui vontade (volição) e age com intenção. O corpo (ou as partes) respondem à volição. A intenção se aplica a estas respostas e, opcionalmente, ao demais eventos consequentes. Um exemplo típico da expressão da agentividade é apresentado na sentença (B.8).

(B.8) Eu matei o caracol batendo nele com minha mão.

Deve-se observar que uma entidade consciente, representada sintaticamente como o sujeito de uma construção causativa, pode ou não intencionar o evento causado. Esta diferença distingue o conceito de "agente" do conceito de "autor", ou seja, na autoria não há intencionalidade. Por exemplo, compare-se a sentença (B.8) com a sentença (B.9).

(B.9) Eu matei o caracol quando minha mão bateu nele.

As noções de "agente" e "autor" também devem ser distintas da noção de "vítima" (*undergoer*). Da mesma forma que o "autor", a "vítima" não intenciona o evento mencionado, mas também não possui agentividade sobre as ações que culminam naquele evento. Os eventos são autônomos e "acontecem" com a "vítima".

(B.10) Eu perdi minha caneta em algum lugar da cozinha.

Uma outra perspectiva sobre a agentividade seria a da "auto-agentividade", quando a ação se dá sobre o próprio agente, como ilustrado na sentença (B.11).

(B.11) O homem saltou do penhasco.

Finalmente, é possível ainda considerar a "agentividade causada", quando o exercício da agentividade é visto como um evento (especificamente um evento cognitivo) que pode, ele mesmo, ser causado (sentença (B.12)).

(B.12) As notícias da morte dele me entristeceram.

APÊNDICE C – Dinâmica de Forças

Para Talmy (2000), uma característica fundamental da linguagem é que ela possui dois subsistemas: o subsistema gramatical e o subsistema lexical. Enquanto o subsistema gramatical é restrito por regras e representa a "estrutura" conceptual, o subsistema lexical é aberto, não restrito e é responsável pelo "conteúdo" conceptual.

O modelo *force dynamics* (Dinâmica de Forças - DF) de Talmy faz parte deste modelo cognitivo mais abrangente, que representa a organização dos conceitos especificados na estrutura gramatical. O modelo DF reconhece as motivações morfossintáticas e semânticas das construções causativas como provenientes do subsistema gramatical. Desta forma, a causação é colocada em paralelo com outras categorias nocionais, como aspecto/tempo, negação, número e gênero, que também tem uma representação gramatical na linguagem (TALMY, 2000, p. 410-411). Este nível de análise permite abstrações da noção fundamental de causação no mundo real para outros domínios, através de vários processos cognitivos.

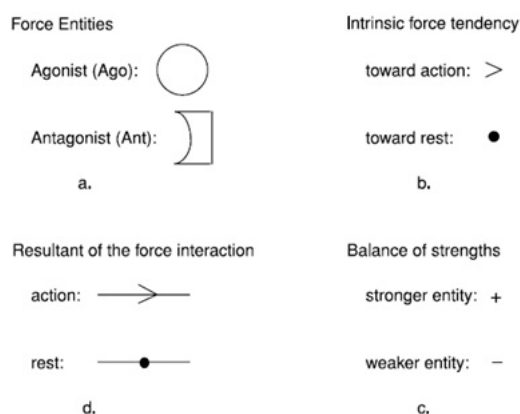
A DF está relacionada a como as entidades interagem com respeito à força, o que inclui o exercício da força, a resistência a tal força, a superação de tal resistência, o bloqueio da expressão da força e a remoção de tal bloqueio. Assim, a DF compreende a expressão da força como a oposição entre entidades e os resultados de tais interações com o tempo. Isto significa que a causação é essencialmente um conceito de dinâmica de forças e a DF é uma generalização da noção tradicional de "causativo" (TALMY, 2000, p. 409). O modelo DF decompõe a noção de "causativo" em "primitivas", para incluir as noções de "permissão", "habilitação", "ajuda" e "impedimento".

C.1 Notações

O modelo DF emprega notações especializadas (Figura 58) para denotar os conceitos básicos na representação das interações na dinâmica de forças.

Na DF, duas entidades, o Agonista (AGO) e o Antagonista (ANT), interagem em relação à força. Considera-se que cada uma destas entidades tem uma tendência de força intrínseca em um dado tempo, ou em direção ao movimento ou em direção ao repouso. No entanto, para que a interação de forças seja possível, as duas entidades devem apresentar tendências opostas. O AGO e o ANT não possuem a mesma intensidade de força: um deve ser mais fraco e o outro mais forte, em um dado tempo. Na DF, o resultado da interação de forças é colocado em termos do estado do AGO em relação à força depois da interação. Em outras palavras, cada interação é iniciada com a intenção (ou o potencial) de influenciar o

Figura 58 – Notações em DF.



Fonte: Talmy (2000, p. 414)

AGO, para que ele manifeste uma tendência de força diferente. Este objetivo é o ponto focal na expressão das interações na dinâmica de forças.

C.2 Padrões básicos de DF

A natureza do choque entre as entidades engajadas na interação de forças pode ser representada de formas distintas na DF. Os padrões demonstram, em cada caso, os estados resultantes possíveis e também fornecem acesso as informações de fundo, exibidas nas construções causativas.

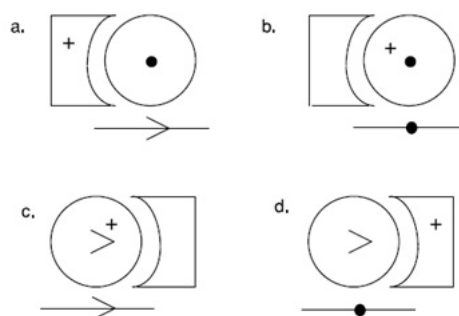
C.2.1 Padrões *Steady-State*

O padrão *steady-state* é subjacente a todos os padrões complexos da DF na linguagem (TALMY, 2000, p. 413). A interação entre o AGO e o ANT produz um dos quatro resultados mostrados na Figura 59.

Os padrões na Figura 59(a) e na Figura 59(d) demonstram o que pode ser referido como "causação prototípica". Isto envolve um "choque positivo" do ANT sobre o AGO (porque o ANT possui maior intensidade de força), o que leva o AGO, eventualmente, a um estado oposto ao seu estado original. Uma vez que a interação envolvida na Figura 59(a) e na Figura 59(d) é estável (ela continua durante uma parcela considerável de tempo), ela ilustra a chamada "causação estendida". Assim, a situação causativa, como geralmente é definida, poderia ser representada como na Figura 59(a) e na Figura 59(d), como várias características adicionais quando necessário.

Os padrões na Figura 59(b) e na Figura 59(c), no entanto, demonstram que nem todas as interações entre o AGO e o ANT resultam em uma mudança de estado do AGO.

Figura 59 – Padrões steady-state.



Fonte: Talmy (2000, p. 415)

Tais padrões ilustram o que Talmy (2000, p. 415) chama de "*despite category*", onde um AGO mais forte resiste com sucesso à força do ANT e mantém seu estado original. Assim, mesmo que esta situação não seja geralmente analisada como causativa, ela está fortemente associada como a causação (especificamente a "causação falha"). As sentenças a seguir fornecem exemplos destes padrões:

- (C.0) (a) CAUSAÇÃO: A pipa continua subindo porque o vento sopra forte.
 (b) RESISTÊNCIA: O telhado continua intacto apesar do forte vento.
 (c) SUPERAÇÃO: A bola continua rolando apesar da grama alta.
 (d) BLOQUEIO: Os manifestantes não entraram por causa da grade de ferro cercando o prédio.

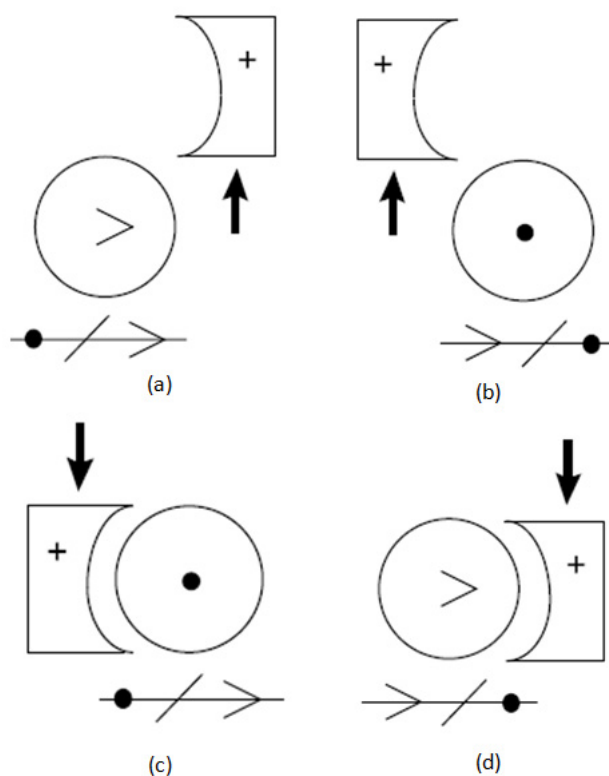
C.2.2 Padrões *Shifting*

A dinâmica das forças também pode ser encontrada em padrões de *shifting* (deslocamento). Nestes padrões, a interação entre o ANT e o AGO não é estável, mas se desloca com o passar do tempo para produzir um resultado diferente. Estes padrões são de dois tipos, ilustrados na Figura 60: separação do ANT e do AGO ((a) e (b)) e introdução repentina da força do ANT no AGO ((c) e (d)).

As setas na Figura 60(a-d) indicam os padrões de deslocamento ou movimento do ANT; as setas apontando para cima indicam a remoção da força do ANT no AGO ((a) e (b)), enquanto aquelas apontando para baixo indicam a introdução da força do ANT no AGO ((c) e (d)). Além disso, nas figuras (a) e (d), cada configuração de DF tem dois estados resultantes separados pela barra: o estado resultante na esquerda representa o resultado da interação entre o ANT e o AGO antes do deslocamento, enquanto o estado resultante na direita mostra o resultado da interação depois do deslocamento.

As figuras 60(a) e 60(b) ilustram a situação de *letting* (deixar), que envolve a

Figura 60 – Padrões de deslocamento.



Fonte: Talmy (2000, p. 418)

remoção da força do ANT sobre o AGO, resultando que o AGO, eventualmente, manifeste seu estado original ou atividade (que em (a) é em direção ao movimento e em (b) em direção ao repouso). Assim, a Figura 60(a) representa *onset letting of motion* (início da permissão de movimento), enquanto a Figura 60(b) representa *onset letting of rest* (início da permissão de repouso). Deve ser ressaltado, no entanto, que apesar da remoção da força do ANT sobre o AGO, o ANT continua sendo a entidade mais forte; o estado resultante exibido pelo AGO não é necessariamente em virtude do força do AGO sozinho. Em outras palavras, em 60(a) e 60(b), o AGO pode continuar a mostrar seu estado original somente com a permissão do ANT; assim, o ANT pode ser considerado responsável pelo eventual estado do AGO.

Por outro lado, a figura Figura 60(c) e a figura Figura 60(d) apresentam um cenário relativamente diferente. Em (c), um ANT como maior força repentinamente age sobre um AGO que está em repouso, causando o movimento do AGO. Em (d), um AGO pode ser repentinamente bloqueado por um ANT mais forte, causando uma mudança de estado do AGO. Talmy (2000, p. 418) nota que o padrão em (c) é mais associado com a categoria de causação, onde um objeto (comumente) em repouso sofre repentinamente a ação de outro entidade, resultando em uma mudança de estado do último. O padrão em (c) é referido

como *onset causation* (início de causação), enfatizando a natureza repentina da ação do ANT. Intuitivamente, todos os padrões de causação, incluindo os padrões *steady-state*, devem começar como *onset causation*.

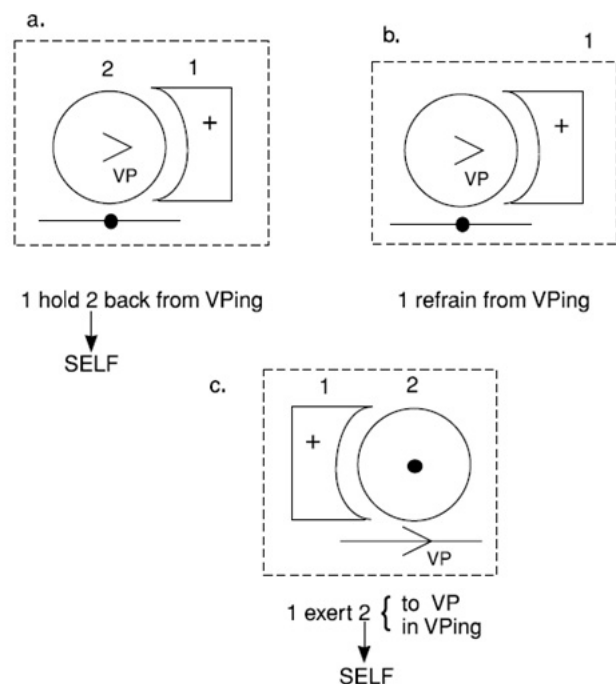
C.3 Domínios da DF

O modelo DF estende as interações de forças do físico para o abstrato. Talmy identifica os domínios intra-psicológico e inter-psicológico como áreas onde a dinâmica de força pode operar.

C.3.1 Domínio intra-psicológico

De acordo com Talmy (2000, p. 430), na linguagem as noções que são usualmente associadas com interações físicas, tais como "empurrar", "bloquear" e "permitir" podem ser estendidas para os domínios psicológicos através de metáforas sistemáticas. O modelo DF torna possível representar as expressões de interação de forças ocorrendo dentro de uma única psique. Este domínio envolve um "self dividido" em duas entidades, as quais estão engajadas em interações de força: uma parte de "desejo" (AGO) e uma parte de "bloqueio" ou "estímulo" (ANT). As duas entidades, com quantidade de forças opostas, são parte da mesma psique.

Figura 61 – O "self dividido".



Fonte: Talmy (2000, p. 431)

Como ilustrado na [Figura 61\(a\)](#) e na [Figura 61\(b\)](#), a expressão da causação intra-psicológica pode envolver três padrões diferentes. Em (a), a parte "desejo" do self dividido (AGO) é bloqueado por um ANT mais forte, que cerceia a atividade do AGO. Este padrão pode se exemplificado pela sentença (C.1).

(C.1) Ele impediu a si mesmo de responder.

Aqui, um ANT mais forte (a parte "bloqueio" do self) é expresso como o sujeito impedindo com sucesso a manifestação do AGO (expresso com o reflexivo "si mesmo"). Na [Figura 61\(c\)](#), no entanto, nota-se que um ANT mais forte pode agir como um estímulo sobre o AGO que está em repouso, provocando uma ação. O padrão em (b) demonstra que a causação intra-psicológica pode ser lexicalizada; as duas entidades em oposição na interação não podem ser individualmente especificadas em uma sentença, uma vez que a psique inteira deve ser expressa como sujeito. Neste padrão, as interações entre os elementos do "self dividido" são reunidas em um único item lexical. Esta é a razão de porque em (b) o AGO e o ANT não são numerados individualmente, mas a psique inteira é expressa como o ator.

[Talmy \(2000, p. 433\)](#) propõe que a dinâmica de força psicológicas é uma propriedade subjacente à todas as entidades sencientes. Ele afirma que enquanto parece que as manifestações de força das entidades (sencientes ou não sencientes) emanam da força do seus corpos físicos, em essência as entidades sencientes têm uma fonte de força renovável, a psique. Este fato é materializado em situações onde a força física do AGO, quando comparada com a força física do ANT, é muito mais fraca, mas o AGO resiste com sucesso a pressão do ANT. Por exemplo, na sentença (C.2) quando as forças do AGO (o estudante) e do ANT (os colegas) são medidas baseando-se apenas na quantidade física, o AGO é mais fraco. Assim, é possível dizer que a resistência do AGO se deve a uma fonte que é não física.

(C.2) O estudante resistiu a pressão dos colegas para fumar um cigarro.

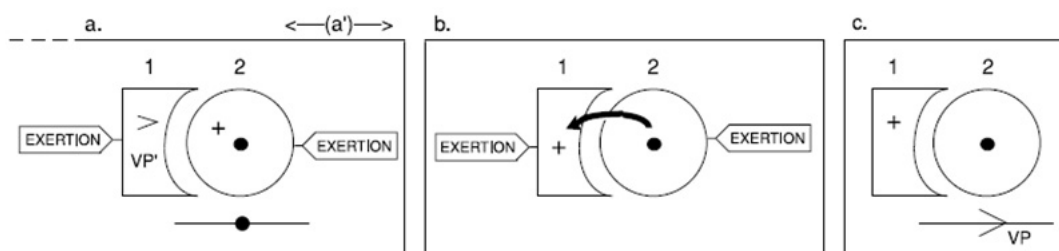
C.3.2 Domínio inter-psicológico

A dinâmica de forças pode ser expressa entre duas psique distintas através da extensão metafórica dos conceitos de "empurrar" ou o exercício de "pressão" de uma entidade sobre outra. Em interações sociais, mesmo se a expressão do exercício de força por uma entidade (senciente) sobre outra não é física, a interação é ainda transmitida por itens lexicais que expressam força física, sugerindo uma relação metafórica entre os dois domínios. Na sentença (C.3), o sujeito (AGO) é concebido como estando sob a influência de uma entidade não expressa (ANT), para submeter-se a uma certa mudança de estado. Neste exemplo, a interação não é física, mas ocorre entre as psique do AGO e do ANT. Note-se, no entanto,

que a interação de forças entre psique pode envolver fases de negociação de forças (exercício, manutenção ou produção) entre as entidades em oposição. Assim, a sentença acima pode apresentar as fases ilustradas na [Figura 62](#).

(C.3) Ele está sob muita pressão para testemunhar na corte.

Figura 62 – Padrões de fases no domínio inter-psicológico.



Fonte: [Talmy \(2000, p. 439\)](#)

As entidades sencientes engajadas em uma interação de forças dependem, para o suprimento e manutenção de força, de uma psique que gerencie o gasto de força. Na [Figura 62\(a\)](#), observa-se um ANT com uma tendência de força em direção a uma ação, que age sobre um AGO em repouso. Entretanto, uma vez que o AGO tem um equilíbrio de forças mais forte, para o ANT ter sucesso na mudança de estado do AGO, o AGO deve abandonar sua força para haver um deslocamento de grande intensidade do AGO para o ANT ([Figura 62\(b\)](#)). É somente depois do deslocamento da intensidade de força que o ANT pode derrubar o estado do AGO. Assim, na sentença já apresentada (C.3), o sujeito (ele), que está no controle de sua ação, pode reagir a pressão da entidade não-expressa através da decisão de testemunhar na corte. Da mesma forma, o AGO pode manter sua tendência de força durante a interação, resistindo com sucesso a pressão do ANT. A natureza de cada fase da ação é informada pelo conhecimento factivo do falante.

C.4 DF e causação

Após uma análise mais detalhada da interação de forças entre as entidades AGO e ANT, [Talmy \(2000, p. 428\)](#) apresenta a DF como um *framework* capaz de generalizar a noção básica de causação, que tradicionalmente é estudada com base apenas em uma noção uniforme e atômica, representada por "CAUSE". Resumidamente, o *framework* da DF permite considerar:

- a) não apenas a noção de "causar", mas também a de "permitir";
- b) não apenas os casos prototípicos de causar/permitir, mas também:

- início da provocação de ação (Figura 60(c))
 - início da provocação de repouso (Figura 60(d))
 - causação estendida da ação (Figura 59(a))
 - causação estendida do repouso (Figura 59(d))
 - permissão prototípica: início da permissão de ação (Figura 60(a))
 - outros tipos de permissão (p.ex. Figura 60(b))
- c) não apenas os casos em que o ANT é mais forte, mas também os casos em que o ANT é mais fraco.
- d) não apenas os casos em que o resultado é nomeado, mas também os casos em que a tendência é nomeada.
- e) não apenas a entidade que afeta (ANT) como sujeito, mas também a entidade afetada (AGO) como sujeito.

Anexos

ANEXO A – Ontologia SIMPLE

A.1 Ontologia SIMPLE

Figura 63 – Ontologia SIMPLE - Topo

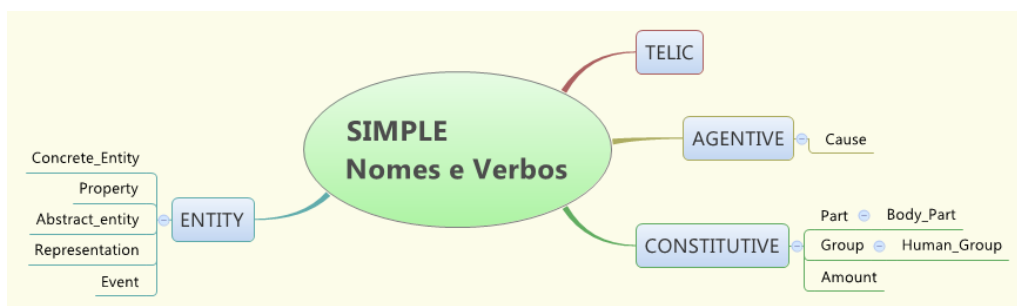


Figura 65 – Ontologia SIMPLE - Relações

