

UNIVERSIDADE DO VALE DO RIO DOS SINOS - UNISINOS
UNIDADE ACADÊMICA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM LINGUÍSTICA APLICADA
DOUTORADO EM LINGUÍSTICA APLICADA

Carolina Müller

**PRINCÍPIOS METODOLÓGICOS PARA A CONSTRUÇÃO DE UMA ONTOLOGIA
BASEADA NA SEMÂNTICA DE FRAMES**

São Leopoldo
2015

UNIVERSIDADE DO VALE DO RIO DOS SINOS - UNISINOS
UNIDADE ACADÊMICA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM LINGUÍSTICA APLICADA
DOUTORADO EM LINGUÍSTICA APLICADA

Carolina Müller

**PRINCÍPIOS METODOLÓGICOS PARA A CONSTRUÇÃO DE UMA ONTOLOGIA
BASEADA NA SEMÂNTICA DE FRAMES**

Tese apresentada como requisito parcial para a obtenção do título de Doutora, pelo Programa de Pós-Graduação em Linguística Aplicada da Universidade do Vale do Rio dos Sinos – UNISINOS.

Orientadora: Prof^ª. Dr^ª. Rove Chishman

São Leopoldo

2015

FICHA CATALOGRÁFICA

M958p Müller, Carolina
Princípios metodológicos para a construção de uma ontologia baseada na semântica de *frames* / Carolina Müller. – 2015.
177 f. : il.

Tese (doutorado) – Universidade do Vale do Rio dos Sinos, Programa de Pós-graduação em Linguística Aplicada, São Leopoldo, RS, 2015.
“Orientadora: Profa. Dra. Rove Chishman”

1. Linguística Aplicada. 2. Ontologia. 3. Domínio jurídico. 4. Semântica de *frames*. I. Título.

CDU 81'33

Catlogação na Fonte:

Bibliotecária Mariana Dornelles Vargas – CRB 10/2145

*Ao meu pai (in memoriam) e minha
mãe por sempre me incentivarem e
valorizarem meu estudo.*

*Aos meus amores, Den
pela compreensão, paciência e amor.*

AGRADECIMENTOS

É tempo de agradecer...

...agradecer a todos que estiveram ao meu lado e que, de uma forma ou outra, contribuíram para a realização deste trabalho.

De modo especial, agradeço:

À professora Rove Chishman por ter acreditado em mim e me permitido fazer parte do seu grupo de pesquisa; pela paciência e incentivo; pela compreensão às minhas limitações e dificuldades; por ser uma orientadora sempre atuante e por me apresentar de forma tão apaixonada a Linguística.

Ao meu marido Denis, pelo incentivo, ajuda e, sobretudo, paciência e compreensão.

Ao meu amado Daniel pela compreensão nos momentos em que estive afastada para estudar.

Aos professores do PPGLA pelo incentivo e ensinamentos, me permitindo conhecer cada vez mais a área da Linguística Aplicada.

Ao Programa de Pós-Graduação em Linguística Aplicada da Universidade do Vale do Rio dos Sinos, pela oportunidade de realizar esta pesquisa.

Ao Programa de Bolsas Capes-CNJ, por ter me concedido a bolsa de estudos.

Ao meu querido amigo de infância o Excelentíssimo Senhor Dr. Carlos Fernando Noschang Júnior, Juiz de Direito da 2ª Vara Cível da Comarca de Campo Bom, pela ajuda especializada.

Aos meus queridos colegas de doutorado pelo apoio, pelas conversas, pela ajuda, pelas trocas. Especialmente, agradeço à Thais por toda a ajuda em relação aos frames e principalmente pela força no decorrer dos últimos anos.

Aos meus queridos colegas da IENH pela compreensão nos períodos de ausência.

À equipe diretiva da IENH por todo o apoio recebido no decorrer do doutorado, permitindo-me a participação em eventos e o afastamento quando necessário.

Aos amigos por compreenderem as minhas angústias e me apoiarem sempre que necessitei.

Aos meus queridos familiares que acompanharam e incentivaram esta minha caminhada.

Simplesmente agradeço!

*O êxito da vida não se mede pelo
caminho que você conquistou,
mas sim pelas dificuldades
que superou no caminho.*

Abraham Lincoln

RESUMO

Esta tese apresenta uma metodologia fundamentada na teoria da Semântica de Frames (FILLMORE, 1982) para a construção de ontologias de domínio. A metodologia proposta segue passos reinterpretados a partir de metodologias vigentes, porém embasados no conceito de frames semânticos (cenas com seus participantes e diferentes papéis) para a descrição dos significados. O principal objetivo desta pesquisa reside em refletir sobre as convergências entre ontologias (GRUBER, 1996; GUARINO, 1998) e frames, de modo a acreditar-se que a combinação destas duas formas de organizar o conhecimento garantem a representação mais precisa de conceitos do domínio jurídico. A proposta assume que a inserção da informação dos frames, especialmente a relacionada aos diferentes papéis que os participantes podem exercer, é capaz de prover uma descrição detalhada dos conceitos do domínio e, assim, prover mais significado à ontologia. A metodologia prevê cinco etapas que correspondem ao ciclo de vida da ontologia, porém têm como foco o estágio de conceitualização, no qual são descritos os conceitos do domínio através do uso dos frames. Realizou-se a construção de uma ontologia piloto tomando como base os Juizados Especiais Criminais, cujos processos tramitam em primeira instância e correspondem a crimes de menor potencial ofensivo. O estudo resultou em uma metodologia linguística para o desenvolvimento de ontologias de domínio capaz de ampliar a descrição dos conceitos, provendo detalhamentos sobre os eventos, seus participantes e os diferentes papéis por eles exercidos, de modo a prover mais significado à estrutura ontológica, contribuindo com a recuperação da informação. A etapa de avaliação da ontologia foi realizada com base no cálculo da abrangência, obtendo um resultado de 84,4% de informações representáveis na estrutura ontológica, demonstrando a proficiência da metodologia proposta no que se refere à identificação de informações relevantes advindas dos documentos jurídicos.

Palavras-chave: metodologia – ontologia de domínio – domínio jurídico – Semântica de Frames

ABSTRACT

This thesis presents a methodology for the construction of domain ontologies based on the theory of Semantic Frames (FILLMORE, 1982). The proposed methodology reinterprets existing methodologies, but grounded in the concept of semantic frames (scenes of its participants and different roles) for the description of meanings. The main objective of this research lies in reflecting on the similarities between ontologies (GRUBER, 1996; GUARINO, 1998) and frames in order to believe that the combination of these two ways of organizing knowledge ensures the most accurate representation of legal domain concepts. The proposal assumes the inclusion of information of frames, especially those related to the different roles that scene participants may assume, is capable of providing a detailed description of domain concepts and thus provide more meaning to the ontology. The methodology provides five steps that correspond to the ontology life cycle, but are focused on the conceptualization stage, in which describes the domain concepts through the use of frames. A pilot ontology was built based on the Special Criminal Courts context, whose cases proceed through at first instance and account for crimes of lesser offensive potential. The study resulted in a linguistic methodology to develop domain ontologies able to expand the description of the concepts, providing detailing of events, its participants and the different roles they assume in order to provide more meaning to the ontological structure, aiming at contributing specially in information retrieval of documents. The evaluation stage of ontology was based on calculation of coverage (recall), getting a score of 84.4% of representable information on the ontological structure, demonstrating the usefulness of the proposed methodology with regard to the identification of relevant information from legal documents.

Keywords: methodology - domain ontology - legal field - Semantic frames

Lista de figuras

Figura 1: Ontologia – área de estudo interdisciplinar	38
Figura 2: Visão geral da LRI-Core.	45
Figura 3: Representação parcial da estrutura da LKIF-Core.....	47
Figura 4: Hierarquia dos ramos jurídicos na OntoJuris.....	49
Figura 5: Esquema síntese do capítulo 2	56
Figura 6: Frame <i>CRIMINAL_INVESTIGATION</i>	70
Figura 7: Relação de herança entre <i>frames</i>	71
Figura 8: Relação de perspectiva entre <i>frames</i>	72
Figura 9: Relação de usa entre <i>frames</i>	73
Figura 10: Relação entre <i>frames</i> e <i>subframes</i>	73
Figura 11: Esquema síntese do capítulo 3	78
Figura 12: Etapas metodológicas propostas	99
Figura 13: Etapas metodológicas propostas – foco na conceitualização.....	102
Figura 14: Esquema síntese do capítulo 4	108
Figura 15: Organograma do Poder Judiciário.	111
Figura 16: Mapa conceitual do Procedimento Sumaríssimo.	121
Figura 17: Relações entre <i>frames</i>	130
Figura 18: Geração da <i>stoplist</i> no AntConc.....	134
Figura 19: Resultado concordanciador – inquirição.....	135
Figura 20: Resultado concordanciador – alegações.	135
Figura 21: Resultado concordanciador – juiz.	135
Figura 22: Hierarquia de classes - Protégé.	144
Figura 23: Relação parte-de – classe EventosLegais - Protégé.	145
Figura 24: Inclusão de comentário para documentação no Protégé.	146
Figura 25: Inclusão da relação de equivalência – Conciliação, Composição e Acordo.....	147
Figura 26: Propriedade de objeto <i>participa</i>	148
Figura 27: Propriedade de objeto <i>ouvidaEm</i>	148
Figura 28: Relações não taxonômicas classe <i>Vítima</i>	149
Figura 29: Classe <i>FrameProcSum</i> para inclusão dos <i>frames</i> como instâncias.....	149
Figura 30: Classe <i>Papeis</i> para inclusão dos papéis como instâncias	150
Figura 31: Representação das relações n-árias.	151
Figura 32: Relações n-árias entre Autor, frame e papel.	152
Figura 33: Inclusão das unidades lexicais.	153
Figura 34: Unidades lexicais – Frame <i>Debate_Oral</i>	153
Figura 35: Interface para criação das propriedades de objeto.	154
Figura 36: Propriedades de dados.....	155
Figura 37: Resultado avaliação automática – OOPS!	156
Figura 38: Avaliação manual - documentos	157
Figura 39: <i>OntoFrameJur</i>	159

Lista de quadros

Quadro 1: Resumo de aspectos históricos acerca das ontologias	35
Quadro 2: Constituição do <i>corpus</i> de pesquisa	115
Quadro 3: <i>Frame CRIME_MENOR_POTENCIAL_OFENSIVO</i>	121
Quadro 4: <i>Frame INÍCIO_PROCEDIMENTO</i>	122
Quadro 5: <i>Frame AUDIÊNCIA_PRELIMINAR</i>	122
Quadro 6: <i>Frame CONCILIAÇÃO</i>	123
Quadro 7: <i>Frame TRANSAÇÃO_PENAL</i>	124
Quadro 8: <i>Frame HOMOLOGAÇÃO</i>	124
Quadro 9: <i>Frame DENÚNCIA</i>	125
Quadro 10: <i>Frame AUDIÊNCIA_DE_INSTRUÇÃO_E_JULGAMENTO</i>	125
Quadro 11: <i>Frame OITIVA_TESTEMUNHAS</i>	126
Quadro 12: <i>Frame INTERROGATÓRIO</i>	126
Quadro 13: <i>Frame DEBATE_ORAL</i>	126
Quadro 14: <i>Frame PROLAÇÃO_SENTENÇA</i>	127
Quadro 15: <i>Frame RECURSO</i>	128
Quadro 16: Variantes e <i>frames</i>	134
Quadro 17: Estrutura de classes – categoria Documentos Legais	135
Quadro 18: Estrutura de classes – categoria Instituições Legais.....	136
Quadro 19: Estrutura de classes – categoria Participantes Legais.....	136
Quadro 20: Estrutura de classes – categoria Eventos Legais.....	137
Quadro 21: Relações não taxonômicas – papéis dos participantes – Acusação.....	138
Quadro 22: Relações não taxonômicas – papéis dos participantes – Defesa.....	138
Quadro 23: Relações não taxonômicas – papéis dos participantes – Vítima	138
Quadro 24: Relações não taxonômicas – papéis dos participantes – Autor do fato.....	138
Quadro 25: Relações não taxonômicas – papéis dos participantes – Conciliador.....	139
Quadro 26: Relações não taxonômicas – papéis dos participantes – Juiz.....	139
Quadro 27: Relações não taxonômicas – papéis dos participantes – Turma recursal.....	139
Quadro 28: Relações não taxonômicas – papéis dos participantes – Delegado de Polícia.....	139
Quadro 29: Relações não taxonômicas – papéis dos participantes – Testemunhas.....	139
Quadro 30: Resultado da avaliação manual	156

Lista de Siglas e abreviaturas

CI – Ciência da Informação
CNJ – Conselho Nacional de Justiça
DAML + OIL – Linguagem de marcação semântica
DOLCE – Descriptive Ontology of Linguistics and Cognitive Engineering
EF – Elemento de frame
FN – FrameNet
FOLaw – Functional Ontology of Law
IA – Inteligência Artificial
ICLA – International Cognitive Linguistics Association
ICT4Law – ICT Coverging on Law: Next Generation Services for Citizens, Enterprise, Public Administration and Polimarkers
ITTIG – Institute of Legal Information Theory and Techniquese
JECrim – Juizado Especial Criminal
KIF – Knowledge Interchange Format
LC – Linguística Cognitiva
LC – Linguística Cognitiva
LKFI-Core – Core Ontology of Basic Concept
LOIS – Lexical Ontologies for Legal Information Society
LRI-Core – Laboratorium voor Rechts Informatics
MFPFO – Metodologia multifacetada para construção de ontologias com anotação semântica
OIL – Ontology Interchange Language
OKBC – Open Knowledge Base Connectivity
OOPS! – Ontology Pitfall Scanner
OWL – Ontology Web Language
PLN – Processamento da Linguagem Natural
PROJUDI – Processo Judicial Digital
RDF – Resource Description Framework
RI – Recuperação da Informação
RI – Recuperação da Informação
SF – Semântica de Frames
STF – Superior Tribunal Federal
SUMO – Suggest Upper Merged Ontology
TJ-PR – Tribunal de Justiça do Paraná
TJ-RS – Tribunal de Justiça do Rio Grande do Sul
TJ-SC – Tribunal de Justiça de Santa Catarina
TOVE – Toronto Virtual Enterprise
UL – Unidade Lexical
W3C – Wordl Wide Web Consortium
XML – Extensilble Markup Language

SUMÁRIO

1	INTRODUÇÃO	15
2	ONTOLOGIAS E A REPRESENTAÇÃO DO CONHECIMENTO	21
2.1	Ontologias: panorama interdisciplinar e histórico das acepções	23
2.1.1	Ontologias na Filosofia	24
2.1.2	Ontologias na Ciência da Computação	26
2.1.3	Ontologias na Ciência da Informação	31
2.1.4	Ontologias e sua relação com a Linguística	33
2.2	Ontologias na área jurídica	39
2.3	Estrutura das ontologias	51
2.4	Esquema síntese do capítulo	56
3	SITUANDO A SEMÂNTICA DE FRAMES	57
3.1	Linguística Cognitiva	59
3.2	Semântica de Frames como contributo teórico	61
3.3	Recurso lexical baseado em <i>frames: FrameNet</i>	69
3.4	Esquema síntese do capítulo	78
4	ETAPAS METODOLÓGICAS: UMA PROPOSTA COM BASE NA SF	79
4.1	Panorama de metodologias	80
4.1.1	Método proposto por Uschold e King (USCHOLD & KING, 1995)	82
4.1.2	TOVE – Toronto Virtual Enterprise (GRUNINGER & FOX, 1995)	84
4.1.3	Methontology (Fernández-López et al., 1997)	85
4.1.4	Método 101 (NOY & McGUINNESS, 2001)	88
4.1.5	Metodologia MFPFO (LIM, LIU & LEE, 2011)	91
4.1.6	Reflexão sobre as metodologias apresentadas	93
4.2	A proposta metodológica	96
4.2.1	Planejamento, especificação de requisitos e aquisição do conhecimento	100
4.2.2	Conceitualização	101
4.2.3	Implementação	104
4.2.4	Avaliação e documentação	105
4.3	Esquema síntese do capítulo	107
5	MODELAGEM E IMPLEMENTAÇÃO DA ONTOLOGIA	109
5.1	Planejamento e especificação de requisitos	109
5.2	Aquisição do conhecimento	115

5.3	Conceitualização	117
5.3.1	Mapa Conceitual	118
5.3.2	<i>Frames</i>	122
5.3.3	Lista de palavras	132
5.3.4	Classes e subclasses	137
5.3.5	Relações não taxonômicas	139
5.4	Implementação	143
5.5	Avaliação e documentação	155
5.6	Síntese do capítulo: a OntoFrameJur	159
6	CONSIDERAÇÕES FINAIS	160
	REFERÊNCIAS	163
	ANEXO A – Catálogo de armadilhas comuns – OOPS!	171
	ANEXO B – Termos de Audiência JECrim – etapa de avaliação	173

1 INTRODUÇÃO

Com a evolução constante dos sistemas informatizados e o advento exponencial da *Web*, a disseminação de conteúdo e conhecimento através das redes virtuais cresce a cada dia. Visando organizar todo este conteúdo disponibilizado em formato digital, os sistemas de recuperação e organização da informação tornaram-se fundamentais e passaram a figurar no rol de estudos de diferentes áreas, especialmente na Ciência da Computação (CC) e Ciência da Informação (CI).

O desafio neste contexto digital é organizar, armazenar e recuperar as informações com vistas a atender as necessidades dos usuários de diferentes comunidades e possibilitar o acesso a todo o conteúdo disponível. Em razão disso, diversos trabalhos e pesquisas têm sido realizados no sentido de desenvolver sistemas de recuperação da informação, dentre os quais podemos exemplificar: (a) a *WebSemântica* (BERNERS-LEE, HENDLER & LASSILA, 2001), cujo objetivo é a criação de metodologias e tecnologias usando padrões para aumentar o desempenho de atividades automatizadas; (b) a inserção de anotação semântica em documentos (SOUZA, 2005) e (c) o uso de estruturas que descrevam uma área através de conceitos e relações, tais como as **ontologias** (GRUBER, 1993).

Seguindo as orientações do *World Wide Web Consortium* (W3C)¹, as ontologias se apresentam como estruturas que organizam e descrevem o conhecimento de uma determinada área para a representação do conhecimento de diferentes domínios, tornando-se elementos eficazes e fundamentais para a recuperação da informação em sistemas informatizados. Atualmente, as ontologias têm sido amplamente utilizadas por informatas para refinamento de sistemas de busca na *Web* e criação de agentes que visam ao aprimoramento de *softwares* através do Processamento da Linguagem Natural (PLN).

Ontologia é um termo advindo da Filosofia e foi reutilizado por outras áreas, especialmente a CI e a CC, constituindo-se em um arcabouço de definições no que tange a cada área específica. Por ora, uma vez que a diversidade de denominações para ontologia será explorada no capítulo dois deste trabalho, podemos definir ontologia como estruturas que

¹ Disponível em <<http://www.w3.org/>> Acesso em 24/07/2013.

descrevem e organizam hierarquicamente um determinado domínio, tendo como base seus **conceitos e relações**.

Atualmente, as ontologias compreendem estudos interdisciplinares que envolvem a Lógica, a Inteligência Artificial (IA) e a Linguística promovendo a ligação entre a linguagem e a computação. Esta interface tem sido foco de diversos estudos, tanto na área da Linguística quanto na da Informática cujos principais objetivos são a Recuperação da Informação (RI) e o Processamento da Linguagem Natural (PLN) com vistas à criação de sistemas capazes de tornar a *Web* mais acessível e “inteligente”, assim facilitando o acesso aos usuários e permitindo buscas mais eficazes. Para cumprir com tais objetivos, é necessário valer-se de processos de representação da informação e do conhecimento de modo a fornecer suporte aos sistemas informáticos através de estruturas conceituais que descrevem diferentes domínios: as ontologias.

Entendemos que as ontologias têm se apresentado como uma alternativa para a resolução de problemas semânticos existentes nos sistemas computacionais, uma vez que estas permitem a compreensão compartilhada de um determinado domínio através de descrição de conceitos e suas relações, expressando o conhecimento de forma estruturada e organizada. Além disso, estas estruturas de organização do conhecimento também têm sido discutidas no âmbito da recuperação da informação jurídica, área na qual há uma grande quantidade de documentos que precisam ser organizados e facilmente recuperados a fim de agilizar os processos e facilitar a atuação de advogados e juízes.

Diante do exposto, esta tese discute o contexto das ontologias e sua aplicação para organização e recuperação da informação no domínio jurídico, estabelecendo uma estreita relação com a Semântica de Frames (FILLMORE, 1982) para a descrição dos conceitos da área, bem como propõe uma metodologia centrada em *frames* (FILLMORE, 1982) para a descrição ontológica do Processo Penal Brasileiro, cujos trâmites ocorram na 1ª instância do Juizado Especial Criminal, correspondendo ao procedimento sumaríssimo.

Esta pesquisa está vinculada ao projeto **Tecnologias Semânticas e Sistemas de Recuperação de Informação Jurídica** (Projeto CNJ-Acadêmico), desenvolvido pelo grupo SemanTec da Unisinos. Tal projeto se caracteriza pela interdisciplinaridade e ação conjunta das áreas do Direito, da Linguística, especialmente a Semântica Computacional e Semântica Cognitiva, e da Computação, com foco no Processamento da Linguagem Natural. A temática jurídica ligada ao contexto das ontologias com fins de recuperação da informação na era da

WebSemântica é foco de estudos do grupo SemanTec há mais de dez anos, tendo diversos trabalhos publicados nessa área.

As iniciativas e esforços do grupo no sentido de sistematizar a área jurídica devem-se ao fato deste campo caracterizar-se por conceitos complexos que se diferenciam por um alto grau de relacionamentos e níveis distintos de abstração. Tal complexidade pode ser compreendida levando em conta o próprio histórico da organização da informação jurídica.

A estruturação do sistema jurídico com vistas à recuperação da informação é também foco de outros grupos de pesquisa. Atualmente existem diferentes sistemas para a recuperação da informação jurídica, dentre eles podemos exemplificar com dois projetos desenvolvidos pelo *Institute of Legal Information Theory and Techniquese* – ITTIG – sob a responsabilidade de Daniela Tiscornia: o *JurWordNet, semantic web of legal terminology*² e o LOIS (*Lexical Ontologies for Legal Information Society*)³, que têm como finalidade integrar as informações jurídicas e facilitar as consultas ao sistema jurídico europeu.

As iniciativas no sentido de integrar informações jurídicas de diferentes sistemas, tais como as expostas anteriormente, necessitam levar em conta que o sistema jurídico apresenta um alto grau de complexidade podendo ser representado através de uma estrutura de links altamente intrincada, uma vez que as ideias e ações são fortemente interligadas (SCHEWEIGHOFER & LACHMAYER, 1997). Além disso, a área jurídica se caracteriza por sua organização em eventos inter-relacionados, nos quais os participantes desempenham diferentes papéis que se modificam de acordo com a situação e o contexto.

Tal assertiva acerca da estrutura complexa do domínio jurídico nos leva a tomar como base teórica a Semântica de Frames (FILLMORE, 1982), mais especificamente o conceito de *frame*. Um *frame* descreve uma situação típica, levando em consideração os aspectos culturais a ela relacionados e incluindo os participantes e suas condições. Cada *frame*, como uma categoria cognitiva, manifesta-se na língua por meio de palavras que o introduzem, isto é, *que o evocam*.

Acompanhando Tiscornia e Sagri (2012) reforçamos nossa escolha pela Semântica de Frames, uma vez que, na área jurídica, a relação unívoca entre termo e conceito é insuficiente para descrever a complexidade do domínio. Além disso, há uma estreita ligação entre o conhecimento jurídico e sua expressão linguística, estando intrinsecamente ligados ao

² Disponível em < <http://www.ittig.cnr.it/Ricerca/UnitaEng.php?Id=11&T=4> > Acesso em: 27/07/2013.

³ Disponível em < <http://www.ittig.cnr.it/Ricerca/UnitaEng.php?Id=70&T=E> > Acesso em: 27/07/2013.

contexto e condicionados à dimensão social. A abordagem da organização através dos *frames* permite descrever as diferentes nuances do domínio, tais como os conceitos e os termos relacionados a partir do contexto. Além disso, esta abordagem permite principalmente, a resolução da problemática em relação aos diferentes papéis que os participantes podem exercer no decorrer de um processo.

Seguindo esta linha, perguntamo-nos: *Considerando o propósito de prover maior significado à descrição na ontologia do domínio jurídico tomando como base a Semântica de Frames, quais são os passos metodológicos a serem adotados?*

A partir desta questão norteadora, estabelecemos como objetivo geral para esta tese *refletir sobre as convergências entre ontologias e frames, de modo a acreditar-se que a combinação destas duas formas de organizar o conhecimento garantem a representação mais precisa de conceitos e relações seguindo etapas metodológicas reinterpretadas à luz da Semântica de Frames.*

A fim de cumprir com nosso objetivo geral e dar resposta ao questionamento inicial, elencamos os seguintes objetivos específicos:

- Situar as pesquisas sobre ontologias, destacando sua aplicação nas áreas da CI, da CC, bem como sua aproximação com a área da Linguística.
- Ponderar sobre a teoria da Semântica de Frames e sua contribuição na inserção de conceitos e relações para descrever o procedimento sumaríssimo nos Juizados Especiais Criminais.
- Analisar as principais metodologias vigentes para a construção de ontologias a fim de verificar suas possíveis contribuições à proposta de etapas metodológicas que visem a inserção de conhecimento semântico baseado em *frames* na ontologia.
- Propor etapas metodológicas de desenvolvimento de uma ontologia baseada em *frames*.
- Apresentar uma ontologia piloto seguindo as etapas metodológicas propostas.

Um motivador desta tese é o fato de que uma ontologia do domínio jurídico possibilitará a ampliação da extração do conhecimento, permitindo o compartilhamento de informações relevantes aos usuários, sejam eles leigos ou especialistas. Através das ontologias as ferramentas de recuperação da informação podem ser aperfeiçoadas, permitindo maior exatidão no gerenciamento do conhecimento.

Outro elemento motivador para elencarmos este tema de estudo é o fato de pouco haver se tratado sobre metodologias para construção de ontologias que abarquem mais fortemente a questão linguística e a descrição dos significados. Promover esta discussão permitirá ampliar o escopo da Linguística na área de ontologias, bem como sua importância para a organização do conhecimento.

Esta pesquisa tem como principal justificativa o fato de que a conceitualização do sistema jurídico envolve uma grande complexidade de relacionamentos e eventos e se faz necessária a otimização destes processos através de sistemas de busca que permitam a recuperação da informação de forma automática.

Além disso, na área jurídica há um grande volume de documentos e informações relevantes armazenadas em diferentes bases de dados. Os documentos gerados por um processo jurídico podem servir de base para a produção de novos documentos; por esta razão, a área jurídica necessita de ferramentas capazes de permitir a rápida recuperação da informação.

Visando responder à questão formulada e atender aos objetivos previstos, este trabalho está organizado em seis capítulos. Considerando que este capítulo de introdução trouxe o contexto de pesquisa e sua justificativa, bem como a questão norteadora e os objetivos elencados, os demais capítulos cumprem com a função de descrever e sustentar as escolhas teóricas realizadas, bem como apresentar nossas reflexões para dar resposta ao questionamento inicial.

Cabe aqui uma observação relacionada à estrutura do trabalho. Organizamos o texto desta tese seguindo um critério de apresentar inicialmente os aspectos teóricos mais amplos, para depois trazer nossa proposta metodológica e de modelagem da ontologia. Assim, o texto segue uma progressão do geral para o aplicado, apresentando inicialmente os capítulos que envolvem uma descrição teórica e culminando naqueles em que modelamos a ontologia e expomos nossas reflexões. A escolha por esta estrutura se deu por duas razões: a primeira, pelo fato de adentrarmos a uma área interdisciplinar, na qual existem diferentes visões acerca das ontologias, sua construção e seu uso; e a segunda, por entendermos necessária a fundamentação teórica para a parte aplicada e o entendimento da proposta. Dito isso, passamos a apresentação do que tange a cada capítulo do trabalho.

O capítulo dois cumpre o papel de descrever a área das ontologias situadas como modelos de representação do conhecimento; para isso apresentamos um panorama acerca

destes artefatos considerando-os objeto de estudo de três áreas: a Filosofia, a Ciência da Computação, a Ciência da Informação e a sua relação com a Linguística⁴. O capítulo também cumpre com a função de apresentar as ontologias situadas na área jurídica exemplificando o arcabouço existente nesse domínio, além de descrever a estrutura e seus diferentes aspectos.

No capítulo três discorremos sobre a teoria da Semântica de Frames, situando-a na área da Linguística Cognitiva. Neste capítulo apresentamos considerações acerca do uso dos *frames* como elementos para a organização de uma ontologia, apontando para uma forma diferenciada de descrever conceitos e relações. Os capítulos dois e três compreendem o nosso referencial teórico centrado nas áreas supracitadas e embasam nossas reflexões para os capítulos seguintes.

O capítulo quatro apresenta um panorama acerca de metodologias para a construção de ontologias tomando como base as áreas da Ciência da Informação e Computação. Este panorama tem o intuito de justificar nossas escolhas para a construção da ontologia piloto do procedimento sumaríssimo e promover a reflexão para a nossa reinterpretação dos passos metodológicos tendo como base a Semântica de Frames. A discussão advinda permite encerrarmos apresentando a nossa proposta reinterpretada à luz da inserção de *frames* para a composição da ontologia.

Tendo apresentado nossas escolhas metodológicas, passamos ao capítulo cinco que tem o propósito de expor a parte aplicada deste trabalho. Neste capítulo analisam-se os passos metodológicos e descrevem-se as etapas de desenvolvimento da ontologia baseada em *frames*. O capítulo tem como principal propósito apresentar a modelagem e a validação da metodologia proposta.

Por fim, no capítulo seis tecemos considerações finais acerca da tese, apontando as contribuições para a área das ontologias e reiterando nossa escolha pela a Semântica de Frames como teoria suporte para nossa pesquisa. Além disso, destacamos possibilidades de uso da ontologia desenvolvida neste trabalho com vistas à integração em sistemas de recuperação da informação.

⁴ É importante esclarecer que as ontologias não são tema de estudos da Linguística, mas apresentam uma relação com a Semântica Computacional, uma área da Linguística Computacional.

2 ONTOLOGIAS E A REPRESENTAÇÃO DO CONHECIMENTO

O crescente aumento de informações armazenadas na *Web* tem conferido significativa importância às técnicas de organização de informação. Diferentes estruturas têm sido utilizadas para essa tarefa: estruturas que utilizam termos para sua organização, tais como **glossários**, **dicionários** e **léxicos computacionais**; estruturas que criam categorias e classificações, tais como as **taxonomias**; ou ainda, as estruturas que se organizam a partir de **conceitos** e seus relacionamentos, conhecidas por **ontologias**.

Por ser **conceito** um termo referido frequentemente neste trabalho e, a fim de evitar possíveis ambiguidades em relação ao seu uso, julgamos oportuno esclarecer que a nossa visão de **conceito** se diferencia das visões da Ciência da Informação (CI) e da Ciência da Computação (CC), uma vez que temos a Linguística Cognitiva (LC) como base teórica. Na CC, Gruber (1993) explica que a ontologia é organizada a partir de termos representacionais do domínio ou um conjunto de definições que associam os nomes às entidades: os conceitos. Já na CI, conforme Dahlberg (1978a), o conceito é formado por três elementos: o referente, suas características e sua forma verbal. No entanto, na nossa visão centrada na LC, entendemos **conceito** como a elaboração representativa do tratamento da **forma** como meio de acesso ao **significado**, interagindo com vários **processos cognitivos**. Considerando as diferentes acepções e o caráter interdisciplinar desta pesquisa, optamos por não fazer distinção no uso do termo, ainda que do ponto de vista teórico e empírico haja diferenças substanciais.

A organização da informação segue modelos de representação do conhecimento⁵ que permitem consistência aos sistemas e processos de análise. As ontologias, tesouros e taxonomias apresentam-se como os modelos de representação do conhecimento mais utilizados na atualidade e, por possuírem características comuns em um escopo conceitual, se relacionam por uma linha muito tênue que contribui para possíveis equívocos na definição e caracterização desses recursos.

⁵ O termo conhecimento é utilizado na área de Inteligência Artificial no desenvolvimento de sistemas baseados em conhecimento e sistemas especialistas. Para o desenvolvimento destas bases de conhecimento são utilizados formalismos de representação (NORVIG & RUSSELL, 2014).

Em função disso, este capítulo tem o propósito de apresentar um panorama acerca das ontologias como instrumentos de representação do conhecimento, considerando-as um artefato interdisciplinar cujas denominações variam conforme a área à qual estão relacionadas, desde a Filosofia até a sua ligação com a Linguística, passando pelas Ciências da Informação e da Computação. Nosso objetivo com este panorama é elucidar os diferentes entendimentos para o termo **ontologia** e chamar a atenção para a forma como a Linguística, em especial a Semântica, pode fazer parte deste campo interdisciplinar.

Ainda, por se tratar de um trabalho com interface com a área do Direito, buscamos apresentar neste capítulo exemplos de ontologias jurídicas que já estão sendo utilizadas em sistemas de representação do conhecimento e recuperação da informação, bem como destacar a importância destes recursos para a informática jurídica.

Outro ponto que cabe a este capítulo é a apresentação dos componentes de uma ontologia, seguindo as orientações de Gruber (1996). O autor define uma estrutura geral para representar o conhecimento através de ontologias. Esta estrutura defendida por Gruber (1996) é também a base para o desenvolvimento de diferentes metodologias para a construção de ontologias, foco a ser apresentado e discutido com maior profundidade no capítulo 4, no qual apresentamos nossa proposta metodológica.

A fim de abarcar tais objetivos, o capítulo se divide em quatro seções. A seção 2.1 cumpre a função de apresentar definições e tecer um panorama histórico acerca do uso de ontologias nas áreas da Filosofia, da Ciência da Computação e da Ciência da Informação, além apresentar ligação entre ontologias e Linguística através da Linguística Computacional, refletindo sobre a distinção entre ontologias formais e linguísticas. Optamos por apresentar cada área em uma subseção distinta apenas a fim de organização da leitura. Na seção 2.2 apresentamos a interface das ontologias com a área do Direito, exemplificando com as principais ontologias jurídicas já desenvolvidas, bem como tecemos considerações acerca da informática jurídica e o uso de ontologias para os sistemas de recuperação da informação. A estrutura que envolve uma ontologia, com base nos preceitos de Gruber (1996), é apresentada na seção 2.3 por se tratar de uma importante fundamentação para o nosso trabalho de proposição de uma metodologia que inclua relações semânticas baseadas em *frames*. E, por fim, na seção 2.4 apresentamos um esquema síntese com os aspectos abordados no capítulo.

2.1 Ontologias: panorama interdisciplinar e histórico das acepções

Definir ontologia é uma tarefa complexa, principalmente por se tratar de um conceito utilizado por diferentes áreas, cada uma com ênfase em seu objeto de estudo. O primeiro conceito de ontologia parte da Filosofia e vem fundamentar as demais conceituações nas diversas áreas que fazem uso dessa **ferramenta** ou **instrumento** de representação do conhecimento.

As ontologias, como estruturas para organização do conhecimento conforme sua origem filosófica, têm sido amplamente estudadas nos últimos anos por diversas áreas do conhecimento, entre elas a Ciência da Informação (CI) e a Ciência da Computação (CC).

Guarino (1998) propõe uma distinção para **Ontologia** e **ontologia**, que consideramos aqui importante destacar. **Ontologia** (com inicial maiúscula) se refere aos estudos ancorados na Filosofia, isto é, aos estudos referentes ao ramo da Filosofia que trata da natureza e da organização da realidade. **Ontologia** é um sistema de categorias que corresponde a uma determinada visão do mundo. Em contrapartida, **ontologia** compreende um estudo de um vocabulário específico que tem por objetivo descrever uma determinada realidade com fins computacionais. Os estudos aos quais nos dedicamos filiam-se ao termo **ontologia**, uma vez que temos o propósito de descrever a realidade do domínio jurídico através de conceitos e relações capazes de contribuir com a recuperação da informação. Consideramos necessário esclarecer que não faremos distinção na escrita do termo **ontologia**, usando-o sempre com inicial minúscula. Esta nossa escolha se deve ao fato de que o trabalho se dedica ao estudo de um domínio específico, procurando descrevê-lo e estruturá-lo e não à criação de uma estrutura capaz de “descrever o mundo” de forma genérica, tais como as ontologias fundacionais que apresentam uma ligação mais estreita com a Filosofia.

Por ser um conceito amplamente difundido e utilizado por áreas tão distintas, podem-se encontrar diversas definições para ontologias. Tal diversidade dificulta a escolha e utilização das técnicas disponíveis para a construção desse tipo de estrutura, sendo um ponto passível de esclarecimentos. Por esta razão, apresentamos a seguir as visões de cada área sobre a conceituação de ontologia e, por fim, buscamos organizar tais informações em um quadro síntese destacando as principais acepções.

2.1.1 Ontologias na Filosofia

Por ser a Filosofia a base do conceito de **ontologia**, julgamos oportuno apresentar brevemente o início filosófico que circunda as estruturas ontológicas atuais, principalmente para situar o assunto e trazer apontamentos sobre sua origem. Salientamos que esta seção tem o propósito de ser breve em relação à contextualização filosófica das ontologias uma vez que nosso trabalho não tem compromisso aprofundado com as bases relacionadas à Filosofia. A seção, portanto, cumpre com a função de situar as ontologias no contexto histórico, sem, no entanto, aprofundar tais aspectos, mas tendo em mente que a visão filosófica perpassa as demais uma vez que a discussão acerca das diferentes categorias do conhecimento parte da Filosofia.

O conceito de ontologia tem sua origem no grego, sendo a palavra composta por *onto* – ser e *logos* – estudo – portanto, **ontologia é a ciência que estuda o ser como tal**. Suas bases são fundamentadas na Filosofia, mais precisamente nas categorias criadas por Aristóteles. Categorias estas que servem como alicerce para classificar qualquer entidade, atribuindo propriedades que distinguem diferentes espécies do mesmo gênero. O filósofo propôs dez categorias básicas que servem para classificar qualquer objeto, a saber: **substância, quantidade, qualidade, relação, lugar, tempo, posição, estado, atividade e passividade**. A partir dessa noção inicial apresentada por Aristóteles, diversos filósofos fundamentam seus estudos sobre ontologias. A Filosofia define ontologia como **aquilo que existe**, estabelecendo uma relação com a metafísica das formas generalistas do século XVII. Esse termo tem sido empregado ao longo da história pela Filosofia e em estudos acerca das teorias sobre a natureza da existência.

De acordo com Borchert (2006), o termo **ontologia** teve sua origem no século XVII, tendo sido usado inicialmente por Jacob Lorhard na obra *Ogdoas Scholastica* (1606)⁶ e Rudolf Goclenius na obra *Lexicon Philosophicum* (1613)⁷. Porém, somente no século XVIII, com a publicação da obra *Philosophia Prima Sive Ontologia*, de Christian Wolff (1730)⁸ que o termo passou a ser utilizado pela comunidade de filósofos sendo considerado uma

⁶ Obra disponível em <http://goo.gl/bUOsI1> Acesso em 12/01/2015.

⁷ Obra disponível em <http://goo.gl/kSrA5E> Acesso em 12/01/2015.

⁸ Obra disponível em <http://goo.gl/NFq5HY> Acesso em 12/01/2015.

denominação similar a metafísica⁹ geral - *metaphysica generalis* - , a parte da metafísica que se ocupa de estudar as características dos seres em geral. Em sua obra, Wolff buscou estudar as ontologias, designando-as como **estruturas abstratas de caráter dedutivo**. Estas estruturas deveriam ser formadas a partir da análise de conceitos¹⁰, tais como: **ser, possibilidade, realidade, quantidade, qualidade, causa e efeito**.

René Descartes, Bento de Espinoza e Gottfried Wilhelm von Leibniz são representantes da Filosofia Moderna também envolvidos com estudos relacionados às ontologias. Tais pensadores discutiram a categoria **substância** proposta anteriormente por Aristóteles. Para Descartes, a substância deve ser dividida em espiritual e material, assim introduzindo uma dualidade para esta categoria proposta anteriormente. Espinoza reflete sobre a substância, proclamando que somente Deus pode ser visto nesta categoria. Já Leibniz apresenta a substância como sendo um conceito completo - *conceptus completus*, ou seja, dá conta de representar o ser por completo.

De acordo com Castro (2008), as reflexões acerca do conceito de substância realizadas pelos filósofos modernos influenciaram a sua modificação, contribuindo para que a noção proposta por Aristóteles (substância como individual e essência do ser) fosse revisitada e recriada. Tais reflexões contribuíram para uma nova perspectiva em relação ao avanço das correntes idealistas¹¹, fundamentando tentativas de proposição de novas ontologias de base idealista, tendo como destaques:

- A **Ontologia analítica crítica**, proposta por Nicolai Hartmann, que tinha como fundamento reconhecer o que a metafísica considerava insolúvel, buscando situar no seu lugar o que é racional e o que é irracional. Tinha, assim, uma proposta distinta das

⁹ Metafísica é uma das disciplinas fundamentais da Filosofia. A forma clássica dos sistemas metafísicos abordam os problemas primordiais da Filosofia teórica, ou seja, correspondem a ensaios que buscam descrever fundamentos, condições, leis, estruturas, causas e/ou princípios, o sentido e a finalidade da realidade do mundo ou dos seres em geral. A ontologia é o ramo central da metafísica, correspondendo à investigação das categorias básicas do ser/ente e como elas se relacionam entre si (REALE, 2002).

¹⁰ Por ser a Filosofia uma área que desenvolve atividades essencialmente conceituais, cabe um esclarecimento acerca do termo **conceito** a fim de evitar possíveis ambiguidades no uso do termo nas demais seções. Faria (2009), apresenta o termo a partir de sua etimologia, advindo do Latim *conceptus*, cujo significado corresponde a uma representação intelectual, complexa e abstrata do cerne de um objeto, ou seja, compreende o resultado entre um termo e a concepção que se tem dele. Logo, **conceito**, sob o viés filosófico, é o que atribui significado a um termo ou conjunto de termos, servindo como referência entre o pensamento e o objeto real.

¹¹ As correntes idealistas da Filosofia emergiram com o advento da modernidade, tomando a subjetividade como ponto central e considerando-a como fundamental para entendimento do mundo. Tais correntes originaram-se em Descartes, passando por Kant até chegar a Hegel, considerado um dos grandes idealistas da modernidade.

ideias subjetivas que pretendiam construir a essência a partir da existência (MORA, 1978).

- A **Ontologia fundamental**, proposta por Martin Heidegger, que buscava redefinir a Ontologia tradicionalmente descrita, apresentando uma nova compreensão da existência fundamentada *na finitude* permitindo o acesso à realidade. Foi uma proposta concebida visando a análise ontológica da compreensão do ser, da sua existência (SÁ, 2008).
- A **Ontologia transcendental**, de Edmund Husserl, que apresenta uma independização da Filosofia em relação às Ciências Naturais, destacando, assim, uma ligação entre a lógica formal e as ontologias. Husserl apresentou em sua proposta a concepção de **ontologia formal**, tornando-se o fundamento para as ontologias atuais e postulando que a materialidade particular deve ser descrita de maneira formal (HUSSERL, 1970; POLI, 2003).

Os aspectos filosóficos acerca das ontologias, especialmente a proposição da ontologia transcendental de Husserl, que cunhou a concepção de **ontologia formal**, embasam os estudos atuais sobre o assunto e fundamentam as definições nas demais áreas que estudam tais estruturas de representação do conhecimento. Nas seções seguintes, apresentaremos diferentes aspectos sobre as ontologias considerando as áreas da CC e da CI e sua intersecção com a Linguística, uma vez que os estudos semânticos têm-se apresentado fortemente necessários para a ampliação da descrição de domínios especialmente no que se refere aos estudos relacionados ao PLN.

2.1.2 Ontologias na Ciência da Computação

Historicamente, conforme Smith e Welty (2001), pode-se afirmar que a ligação entre Ciência da Computação e Filosofia deu-se pelo estudo realizado por Mealy em 1967, intitulado *Another Look at Data*¹². Neste estudo, o autor apresenta três esferas distintas no que se refere ao processamento de dados: (a) o mundo real; (b) o que existe na mente do homem; (c) símbolos armazenados em papel ou outro meio. No mesmo texto o autor destaca que, ao realizar o processamento dos dados, faz-se importante resgatar questões filosóficas sobre a existência das coisas no mundo, sem levar em conta suas formas de representação, e aponta

¹² Texto disponível em <http://www.computer.org/csdl/proceedings/afips/1967/5070/00/50700525.pdf> Acesso em 12/11/2014.

para as questões relacionadas às ontologias como estruturas para representar os dados. A partir daí, as ontologias passaram a despertar o interesse de pesquisadores da CC, destacando-se especialmente os estudos em IA, *WebSemântica*, Engenharia de *Software* e Arquitetura da Informação, todos voltados à representação do conhecimento.

De acordo com Moreira et al. (2004), o uso do termo **ontologia** tornou-se frequente na CC por volta dos anos 90, em projetos que buscavam a organização de grandes bases do conhecimento, tornando-as interoperáveis e melhor estruturadas. Esses projetos, desenvolvidos no contexto da *Web* e da IA, contribuíram para uma adaptação do significado da palavra ontologia.

Gruber (1992) justifica o uso do termo ontologia na área da CC comparando **o que existe** para a Filosofia e **o que existe** para a Inteligência Artificial (uma subárea da CC), ou seja, o autor afirma que, se para a Filosofia uma ontologia pode ser vista como **o que existe no mundo**, então, para a IA, uma ontologia deve ser considerada **o que existe para o sistema**, isto é, o conjunto de termos formais que representam um determinado conhecimento a ser inserido no sistema.

Mais tarde, Gruber (1993) apresenta uma definição para ontologia que é apontada na área da CC como a mais completa, sendo citada em vários artigos e textos sobre o assunto. O autor apregoa que uma ontologia corresponde a uma afirmação da lógica teórica, constituindo-se em **uma especificação formal e explícita de uma conceitualização**, ou seja, tudo o que existe pode ser representado de maneira formal e explícita.

Em relação à definição de Gruber (1993) cabem alguns esclarecimentos acerca dos termos empregados. Por **formal**, Gruber compreende que as ontologias devem ser processáveis por máquina, isto é, os computadores devem ser capazes de compreender o que está descrito, através de uma representação semântica, lógica e formal, favorecendo o raciocínio automático. Complementando **formal**, Gruber utiliza os termos **especificação** e **explícita**, que, para o autor, referem-se aos conceitos¹³, propriedades, relações, funções, restrições, axiomas que são definidos de forma explícita na modelagem dos dados.

Diante do exposto por Gruber (1993) podemos compreender que o termo **conceitualização** é utilizado para referir-se à descrição do significado de cada conceito

¹³ Gruber (1993) não define o significado do termo **conceito**, mas afirma que uma ontologia é definida por um conjunto de termos representacionais de um domínio, ou seja, um conjunto de definições que associam os nomes de entidades (classes, relações, etc) a termos – uma conceitualização. Assim, podemos depreender que Gruber compreende os conceitos como termos relacionados ao domínio.

incluindo suas relações semânticas com outros conceitos. Uma **conceitualização** pode ser considerada uma visão abstrata e simplificada daquilo que se pretende representar. Diante disso, todo tipo de agente de software ou sistema de recuperação da informação está comprometido com uma forma de **conceitualização** explícita ou implícita, sendo que compreende uma relação extensional, ou seja, um inventário da composição do domínio, uma listagem de todos os aspectos pertinentes àquele domínio, ainda que em nível de abstração.

Considerando a definição apontada por Gruber (1993), situamos este trabalho sob o viés da **conceitualização**, ou seja, nossa preocupação primordial na construção da ontologia será relacionada aos **conceitos** da área do Processo Penal Brasileiro e relações que se estabelecem entre eles (tópico a ser aprofundado no capítulo 4). Cabe apontar que, para a construção da ontologia, a formalização também corresponde a um fator essencial, porém, por se tratar de uma pesquisa fundamentalmente focada na semântica, apenas apresentaremos sugestões de formalização através do editor de ontologias, não tendo o compromisso com a escrita lógica de axiomas¹⁴.

Seguindo o propósito de conceituar ontologia, destacamos que a definição dada por Gruber (1993) é reformulada por Borst (1997), que também descreve a ontologia como uma **especificação formal e explícita de um conceito**, porém acrescenta a esta conceituação a palavra **compartilhada**, ou seja, corresponde a um conhecimento consensual dos fenômenos de determinado domínio. Assim, uma ontologia é uma descrição formal dos conceitos e relacionamentos que existem dentro de um domínio, o que implica que uma ontologia se relaciona com um vocabulário específico e com uma linguagem específica.

Novamente vislumbramos um aspecto relacionado à definição de ontologia em conformidade com este trabalho. O fato apontado por Borst (1997) acerca da necessidade de conceituação compartilhada, aludindo a um vocabulário e uma linguagem específicos relacionados a um domínio, corresponde a um aspecto característico do domínio jurídico, o qual apresenta termos e expressões peculiares que precisam ser expressos na ontologia de modo a caracterizar-se uma estrutura passível de compartilhamento.

Outra definição bastante presente em trabalhos sobre ontologias é dada por Guarino (1998), descrevendo-as como **artefatos** compostos por um conjunto de termos e palavras (um vocabulário) que têm a tarefa de descrever uma determinada realidade aliada aos significados

¹⁴ Na Filosofia, axiomas são verdades inquestionáveis universalmente válidas, muitas vezes utilizadas como princípios na construção de uma teoria ou como base para uma argumentação. Para a computação, os axiomas podem ser vistos como declarações definidas na ontologia utilizando a lógica de descrição para expressá-las (MATTAR, 2010).

das palavras e seus conceitos. Assim, sob essa visão, uma ontologia segue um caráter intensional¹⁵, tendo como base uma rede de relações. Ao referir-se a **uma rede de relações** a definição de Guarino reforça nossa proposta de inserção de relações semânticas ligadas aos *frames*¹⁶, uma vez que tornará possível a construção de uma ampla rede de relações entre os **conceitos** na ontologia do domínio jurídico.

Gruber e Guarino apresentam os conceitos mais utilizados para definir ontologias, sendo seus trabalhos representativos na área da CC. Conforme vimos, tais conceitos, ainda que apontem fortemente às questões formais relacionadas à construção destas estruturas, são altamente passíveis de uso para a fundamentação da nossa ontologia, principalmente no que se refere à conceitualização e compartilhamento.

Além da definição, julgamos oportuno tecer comentários acerca da classificação das ontologias. Na área da CC, Guarino (1998) sugere o desenvolvimento de diferentes tipos de ontologias de acordo com o nível de generalidade necessária. O autor propõe que as ontologias podem ser classificadas em: (a) ontologias de alto nível, (b) ontologias de domínio, (c) ontologias de tarefa e (d) ontologias de aplicação.

Conforme Guarino (1998), as ontologias de **alto nível** ou *top-level* (também podem ser empregados os termos **fundacionais**, **nucleares** ou *core* para definir este tipo de ontologia) são as que descrevem conceitos gerais, tais como **espaço**, **tempo**, **assunto**, **objeto**, **evento**, **ação**, sendo que são totalmente independentes de um problema particular ou domínio. Têm como principais características a abrangência e a reusabilidade, estando, algumas vezes, vinculadas a ontologias menores (de domínio, por exemplo) usadas por grandes comunidades. Tais ontologias, em nosso entendimento, são estreitamente ligadas ao conceito de ontologia cunhado pela Filosofia uma vez que buscam descrever o mundo e as coisas que nele existem.

¹⁵ Extensão e intensão são termos clássicos empregados no âmbito da Semântica. Para melhor explicar tais termos, apresentamos a definição do dicionário Oxford “A extensão de um predicado é a classe dos objetos que ele descreve: a extensão de ‘vermelho’ é a classe das coisas vermelhas. A intensão é o princípio pelo qual os objetos são escolhidos ou, em outras palavras, a condição que um objeto tem de satisfazer para ser corretamente descrito pelo predicado. Dois predicados (‘... é um animal racional’, ‘...é um bípede sem penas’) podem identificar uma mesma classe, mas o fazem por meio de uma condição diferente. [...] Um predicado ou qualquer outro termo ocorre num contexto extensional, numa frase, se puder ser substituído por outro predicado ou termo com a mesma extensão sem que o valor de verdade da frase sofra alteração: se João é um animal racional, e substituirmos ‘é um animal racional’ pelo predicado co-extensivo ‘é um bípede sem penas’, então João é um bípede sem penas. Outros contextos, como ‘Maria acredita que João é um animal racional’, podem não permitir a substituição, sendo denominados contextos intencionais”.

¹⁶ No capítulo 3 apresentaremos a noção de *frame* detalhadamente.

Podemos citar como exemplos deste tipo de ontologia as ontologias DOLCE - *Descriptive Ontology for Linguistics and Cognitive Engineering*¹⁷ (GANGEMI et al, 2002) e SUMO - *Suggested Upper Merged Ontology*¹⁸ (NILES & PEASE, 2001; SUMO, 2002), resultado de um trabalho cooperativo e interdisciplinar entre linguistas, filósofos, estudiosos das ciências cognitivas, entre outros. O objetivo das ontologias supracitadas é servir de base de conhecimento para ontologias mais simples, sendo compartilhadas por grandes comunidades de usuários.

As **ontologias de domínio** (GUARINO, 1998) têm como característica a descrição de um vocabulário relacionado através de um domínio genérico, buscando especializar os conceitos introduzidos nas ontologias de alto nível. Como exemplos, podemos citar as ontologias OntoJuris e OntoTrib (ARAÚJO et al, 2008): a primeira corresponde à organização das diferentes áreas do Direito e a segunda apresenta uma organização conceitual para o Direito Tributário, sendo ambas complementares. A ontologia que propomos neste trabalho também caracteriza-se como uma ontologia de domínio, uma vez que pretende descrever o Processo Penal Brasileiro nos juizados especiais.

As **ontologias de tarefa** (GUARINO, 1998) se caracterizam pela descrição de um vocabulário relacionado a uma tarefa ou atividade genérica, buscando especializar os conceitos introduzidos nas ontologias de alto nível. E as **ontologias de aplicação** são as mais específicas, pois são utilizadas em aplicações. É o tipo de ontologia que especializa os conceitos, tanto das ontologias de domínio quanto das de tarefa.

A área da CC também contribui com nosso propósito de construção de uma ontologia de domínio, dando subsídios para a estruturação da ontologia através da organização em classes, subclasses e relacionamentos. A estrutura de classes, subclasses e relacionamentos será apresentada de forma mais completa na seção 2.3.

Tendo apresentado o conceito e a classificação de ontologia para a área da CC tomando como base os teóricos mais representativos da área, julgamos conveniente trazer também a visão da CI, uma vez que se trata de uma área que tem como principal objetivo organizar e representar o conhecimento.

¹⁷ DOLCE é uma ontologia “fundacional” criada pelo *Instituto di Ontologia Applicada (Roma-Trento)*.

¹⁸ SUMO - Ontologia formal usada para mapear o WordNet. Mais informações em: <http://www.ontologyportal.org/>

2.1.3 Ontologias na Ciência da Informação

Um dos aspectos que nos leva a buscar subsídios também na área da CI para a construção da ontologia é o fato de que, nesta área, mais precisamente no que se refere à Organização e Representação do Conhecimento, uma ontologia tem o propósito de organizar e classificar os conceitos¹⁹ em categorias (retomando as bases filosóficas) levando em conta a perspectiva da linguagem do domínio e a forma de interação entre os sujeitos.

Soergel (1997) e Vickery (1997) destacam que a CI passou a fazer uso do termo **ontologia** no final dos anos 90, quando os métodos de classificação e organização do conhecimento passaram a suscitar interesse entre pesquisadores da CC, especialmente em função das demandas relacionadas à *Web*, ou seja, a necessidade de utilização de estruturas que permitissem um melhor acesso e recuperação da grande gama de informações disponíveis.

De acordo com Gilchrist (2003), para os profissionais da CI, uma ontologia constitui-se em um documento (ou arquivo) que define as relações entre termos e conceitos de maneira formal, apresentando similaridades com os vocabulários controlados²⁰ (muito difundidos na área da Biblioteconomia), tais como os tesouros²¹ e as taxonomias²². Tais similaridades encontram-se principalmente na forma de elaboração da estrutura desses instrumentos, uma vez que se faz necessária a organização dos conceitos em processos de categorização, classificação e definição das relações entre estes conceitos.

¹⁹ De acordo com Dahlberg (1978a) – que definiu o termo conceito na área da CI – o conceito é formado por três elementos: o referente (aquilo que se pretende conceituar), as características (soma dos enunciados verdadeiros sobre o referente) e a forma verbal – formando o **triângulo do conceito**. Assim, na área da CI que fundamenta seus estudos sobre ontologias na **Teoria do Conceito** elaborada por Dahlberg, o conceito passa de um artefato mental, abstrato, a um instrumento de pesquisa sendo manipulado para a construção de estruturas de organização do conhecimento.

²⁰ De acordo com a norma ISO 25964-1 um vocabulário controlado corresponde a uma lista prescrita de termos, títulos ou códigos, cada um representando um conceito referente a um domínio. São estruturas projetadas para aplicações em que é necessário identificar cada conceito com uma etiqueta consistente, por exemplo, indexando documentos com cabeçalhos para facilitar a localização da informação (ISO 25964-1:2011, p. 3).

²¹ Tesouros são apresentados pela ISO 25964-1 como vocabulários controlados estruturados, nos quais os conceitos são representados por termos e organizados de modo que as relações entre os conceitos são explicitados, além de apresentar termos e entradas para sinônimos ou quase-sinônimos (ISO 25964-1:2011, p. 12).

²² Taxonomia compreende um tipo de vocabulário controlado constituído por termos preferenciais de uma área organizados hierarquicamente. Nesta estrutura os termos são organizados em uma ou mais relações *é-um* ou *tipo-de* ou relações pai/filho, geral/específico (ANSI/NISO Z39.19:2005). Vale salientar que a taxonomia corresponde ao “esqueleto” da ontologia e, por esta razão, há uma linha muito tênue entre ambas nas definições da área da CI.

A área da CI busca referencial na discussão evidenciada por Smith (2004) acerca da diferenciação entre **Ciência** e **ontologia**²³, afirmando que a Ciência tem a função de **explicar** a natureza e a ontologia vem para **descrever, organizar e sistematizar** este conhecimento obtido através dos experimentos científicos. Tais preceitos tornam a CI a área responsável por uma vasta tradição no que se refere à organização dos domínios do conhecimento, aplicadas originalmente aos vocabulários controlados, que visem à organização documental. Esta área constitui um amplo arcabouço teórico e metodológico potencialmente aplicável ao desenvolvimento de ontologias, tendo como principais fontes os estudos de Otlet (1937), Ranganathan (1967), Dahlberg (1978a) e Hjørland (2002)²⁴.

De acordo com Sales, Campos e Gomes (2008), o estudo de ontologias em CI deve-se ao grande interesse pela área da Organização e Representação do Conhecimento, uma vez que as ontologias são estruturas que envolvem análise semântica visando otimizar a recuperação da informação, possibilitando aos programas de computadores uma “certa inteligência” no processamento das informações.

Levando em conta os preceitos da CI e da Organização do Conhecimento, as ontologias compreendem uma nova categoria de ferramentas para a representação do conhecimento, permitindo a organização do domínio através da descrição formal de relações e conceitos proporcionando melhorias na organização, descrição e recuperação da informação. Neste sentido, nosso trabalho tem este intuito: descrever e organizar o domínio do Processo Penal Brasileiro levando em conta conceitos e suas relações, estando, desta forma vinculado aos preceitos apontados pela CI.

No que se refere à classificação das ontologias, a área da CI apresenta uma organização diferente da apontada por Guarino (1998). Corcho, Gómez-Pérez & Fernández-Perez (2004) propõem que a classificação das ontologias deve estar concentrada no tipo de informação a ser modelada, ampliando a organização proposta por Guarino. Os autores defendem que as ontologias devem ser organizadas observando a seguinte divisão: (a) ontologias para a representação do conhecimento, (b) ontologias gerais e de uso comum, (c) ontologias de topo ou de nível superior (*upper ontologies*), (d) ontologias de domínio, (e) ontologias de tarefas, (f) ontologias de domínio-tarefa, (g) ontologia de métodos e (h) ontologias de aplicação.

²³ Ontologia como domínio de conhecimento preocupado com a natureza dos seres.

²⁴ Note-se que as referências advindas da área da CI são um tanto antigas, porém ainda muito pertinentes sendo citadas em diversos trabalhos atuais e representativos do assunto “ontologias” da mesma área.

Consideramos que os quatro tipos propostos por Guarino (1998) englobam em suas características todos os tipos propostos por Corcho, Gómez-Pérez & Fernández-Perez (2004), tornando a classificação mais abrangente, ainda que alguns tipos sejam semelhantes, apenas apresentando subdivisões. Neste trabalho, seguimos a visão de Guarino (1998) e propomo-nos ao desenvolvimento de uma ontologia de domínio, cujo propósito compreende descrever o Processo Penal Brasileiro utilizando a Semântica de Frames²⁵ como subsídio para o estabelecimento de relacionamentos mais detalhados entre eventos e papéis dos participantes.

2.1.4 Ontologias e sua relação com a Linguística

A área da Linguística vem se aproximando dos estudos sobre ontologias, sobretudo no concernente à forma de descrição dos conceitos, atuando nas áreas do PLN e Linguística Computacional. Cabe salientar que o estudo das ontologias não faz parte da tradição dos estudos linguísticos; observa-se, contudo, que as diferentes áreas que vêm se dedicando a este tema têm buscado subsídios na área da Linguística.

A aproximação entre ontologias e estudos linguísticos iniciou há algum tempo, conforme nos aponta Hutchins (1999). Segundo o autor, o vínculo entre as diferentes áreas teve como princípio a publicação do trabalho *Translation* de Weaver (WEAVER, 1955). Tal trabalho, ainda que não seja sobre ontologias especificamente, deu início a esses estudos, pois versava sobre uma proposta de utilização de técnicas estatísticas na área da organização da informação a fim de permitir que os computadores tivessem a capacidade de realizar traduções automáticas de um idioma para outro. O trabalho, considerado pioneiro no uso de computadores para fins de tradução, defendia a ideia de que um documento expresso em linguagem natural poderia ser considerado como um código e, desse modo, poderia ser processado através de outros códigos.

A partir de então, surgiram as primeiras pesquisas linguísticas, mais precisamente relacionadas aos estudos semânticos e vinculados à área do PLN, cujo objeto de estudo reside na geração e recepção automática de textos tomando como base as informações semânticas.

A Semântica Lexical já figura como ponto fundamental para a Linguística Computacional há algum tempo e, conforme Saint-Dizier e Viegas (1995), corresponde a uma

²⁵ No capítulo 3 apresentaremos a teoria da Semântica de Frames e sua usabilidade na nossa ontologia.

área capaz de cobrir uma ampla gama de problemas relacionados à representação do conhecimento em sistemas computacionais.

As ontologias, sendo consideradas artefatos computacionais, cujo principal propósito é contribuir com sistemas de recuperação da informação, fazem uso da Semântica Lexical através das relações de hiponímia, meronímia, sinonímia. Também utilizam-se da Semântica Lógica para estabelecer as regras lógicas e específicas relacionadas aos conceitos. Porém, o detalhamento do significado, capaz de prover informações acerca dos diferentes papéis que cada participante pode exercer em um determinado domínio, ainda é um desafio para estudiosos da área de ontologias²⁶ (seja na CC ou na CI). Este é o principal aspecto que nos leva a buscar a Semântica de Frames para incrementar os aspectos conceituais e representacionais na ontologia. É através dos diferentes eventos e das diferentes formas de participação relacionadas ao domínio jurídico descritas nos *frames* que a ontologia proverá a descrição da área e trará uma possibilidade de contribuição para minimizar o problema.

Vossen (1998) e Farrar & Bateman (2005) discutem um ponto interessante sobre ontologias em PLN que se refere à distinção feita entre **ontologias formais** ou **conceituais** e **ontologias linguísticas**²⁷. Segundo Weinstein (1998), as ontologias têm sido denominadas como estruturas de conceitos e, apesar de registrarem conceitos e relações, muitas delas não apresentam a definição de axiomas, tornando-se apenas grafos²⁸ e não árvores²⁹ de relações. Essas estruturas Weinstein denomina **ontologias informais** ou **linguísticas**. Para Guarino (1998b), a distinção faz-se importante, uma vez que o termo **ontologia** vem sendo utilizado para caracterizar qualquer estruturação de domínio ou apenas uma lista encadeada de termos relacionados a um determinado domínio.

Seguindo a linha de distinção apresentada acima, Vossen (1998) declara que as **ontologias linguísticas** têm como principal característica registrar conceitos lexicalizados, isto é, as palavras que compõem o léxico de uma determinada língua.

²⁶ Conforme afirmado por Giancarlo Guizzardi em palestra proferida no Seminário de Pesquisa em Ontologias no Brasil – Ontobras. Recife, 2012.

²⁷ Optamos por apresentar inicialmente os argumentos dos autores nos quais nos baseamos para depois trazer nossa posição em relação a esta distinção, salientamos que esta opção deve-se apenas ao fato de tornar o texto mais linear e explicativo.

²⁸ Grafo em computação corresponde a uma abstração que permite codificar relacionamentos entre pares de objetos. A Teoria dos Grafos teve sua origem com o problema das Pontes de Königsberg, em 1735. (FIGUEIREDO, 2011).

²⁹ Uma árvore é uma estrutura de dados bidimensional, não linear, que possui propriedades espaciais e admite muitas operações de conjuntos dinâmicos (DE OLIVEIRA, ALUÍSIO & DE BARCELOS ALMEIDA, 2006).

Segundo Prévot et al. (2010), as **ontologias linguísticas** são fundamentadas na linguagem humana e portanto linguisticamente convencionizadas, razão pela qual, muitas vezes, não oferecem estruturas conceituais precisas. De acordo com os autores, as **ontologias linguísticas** reconhecem a linguagem natural, considerando o senso comum como elementos importantes para a organização e análise do conhecimento.

Em contrapartida, as **ontologias conceituais** ou **formais** têm como característica a organização entre conceitos. Há a inserção de níveis artificiais para que se alcance uma estrutura controlada dos conceitos referentes a um determinado domínio. As **ontologias formais** aplicam técnicas puramente lógicas para a seleção e a descrição do conhecimento, sendo necessário pouco envolvimento humano, o que as torna maiores. Já as **ontologias linguísticas** descrevem a semântica das unidades lexicais com base em diferentes perspectivas de organização conceitual.

De acordo com Gruber (1993), as ontologias, quando relacionadas ao PLN, sempre serão ontologias formais uma vez que definem um vocabulário de uma determinada área usando a Lógica; desta forma, a sintaxe e a semântica dependem da forma como serão representadas logicamente a partir de uma linguagem de representação.

Ainda distinguindo ontologias formais e linguísticas, De Almeida Campos (2010) defende que as ontologias linguísticas correspondem a redes de palavras nas quais estão envolvidas as diferentes associações que podem ser feitas entre os variados termos de uma determinada área ou domínio, não tendo como base especificamente relações lógicas. De acordo com a autora, em uma ontologia linguística poder-se-ia considerar que “couro sintético” é um tipo de “couro”. Já em uma ontologia formal, levando-se em conta o conceito de couro como um produto de origem animal, a relação de subordinação citada anteriormente (entre “couro” e “couro sintético”) só poderia ser estabelecida com base em uma relação lógica. Assim, segundo as autoras, uma ontologia formal representa as redes conceituais pautadas em cadeias de relações lógicas, tais como generalização e agregação.

Autores como Guarino (1998b), Vossen (1998), De Almeida Campos (2010) e Prévot et al. (2010) **consideram como ontologias linguísticas os dicionários estruturados, ou WordNets**³⁰ e **FrameNets**, que são grandes bases lexicais disponíveis em várias línguas. Nessas bases de dados os substantivos, verbos, adjetivos e advérbios estão agrupados em

³⁰ A WordNet (Princeton) e a EuroWordNet são bases de dados que organizam os itens lexicais a partir de relações semânticas conceituais, não considerando a ordem alfabética, apresentam uma organização que poderia ser comparada à de um léxico mental. As relações paradigmáticas (sinonímia, antonímia, hiponímia e meronímia) também fazem parte do acervo das WordNets.

conjuntos de sinônimos (chamados de *synsets*), cada um expressando um conceito distinto. Estes *synsets* são agrupados por diferentes relacionamentos lexicais e semântico-conceituais e, por esta razão, as *WordNets* são consideradas como ontologias linguísticas.

Avaliamos que, apesar desta diferenciação entre ontologias linguísticas e formais apontada acima, ambas possuem um aspecto em comum, isto é, permitem o trabalho colaborativo entre máquinas e seres humanos através de uma equipe interdisciplinar composta por linguistas, informatas e especialistas do domínio ao qual a ontologia visa a descrever o conhecimento. Há, no entanto, um detalhe a ser observado: a visão sobre ontologias linguísticas discutida acima muito se aproxima dos **léxicos computacionais**.

Os léxicos computacionais são recursos tecnológicos com informações linguísticas – fonéticas, morfológicas, semânticas e as relações semânticas de sinonímia, antonímia, meronímia, hiponímia, troponímia e relações associativas – utilizados em sistemas de PLN, tendo como aplicações iniciais a tradução automática. Estes recursos estão vinculados a grandes bases de dados e permitem a interoperabilidade em sistemas de PLN.

Calzolari, Zampolli & Lenci, (2002) chamam atenção para a necessidade de haver uma estreita relação entre léxicos computacionais e ontologias, ressaltando que esta relação deve se dar através de um *continuum*, uma vez que ambos tratam de conteúdo relacionado às palavras. Guarino (1998) também destaca a importância da integração entre léxicos computacionais e ontologias, considerando isso uma condição essencial para a implementação da *Web Semântica*. Corroborando com Calzolari, Zampolli & Lenci (2002) e Guarino (1998), Agnoloni et al. (2009) defendem a necessidade de vincular os léxicos computacionais às ontologias a fim de tornar estas mais interoperáveis e lexicalizadas e aqueles mais estruturados contribuindo para o PLN.

Tais preceitos reforçam nossa proposta de embasar o estabelecimento das relações entre os conceitos na Semântica de Frames e buscar uma aproximação com a *FrameNet*, a qual consideramos um léxico computacional, mas que também pode ser entendido como uma ontologia linguística.

Concluindo esta seção, apresentamos um quadro resumindo as principais conceituações para o termo **ontologia**, considerando as áreas descritas anteriormente.

FILOSOFIA		
Século XVII	Origem do termo	<i>Jacob Lorhard</i> na obra <i>Ogdoas Scholastica</i> (1606) e <i>Rudolf Goclenius</i> na obra <i>Lexicon Philosophicum</i> (1613)
Século XVIII	Ontologia passa a ser considerada como similar a metafísica geral pela comunidade de filósofos	<i>Philosophia Prima Sive Ontologia</i> , de <i>Christian Wolff</i> (1730)
Filosofia Moderna	Discussão acerca da categoria “substância” de Aristóteles permitindo sua recriação – nova perspectiva e avanço das correntes idealistas	René Descartes, Bento de Espinoza e Gottfried Wilhelm von Leibniz
	Ontologia analítica crítica	Nicolai Hartmann
	Ontologia fundamental	Martin Heidegger
1970	Ontologia transcendental Surge o termo “ontologia formal” dando base para estudos da Ciência da Computação	Edmund Husserl
CIÊNCIA DA COMPUTAÇÃO		
1967	<i>Another Look at Data</i> – aponta a ligação da Filosofia à Computação – para processar os dados é necessário levar em conta suas formas de representação. Surge o interesse pelo tema nas áreas de Banco de Dados, Engenharia de Software e Inteligência Artificial.	Mealy
1992	Para a Filosofia ontologia representa “o que existe no mundo” e para a Inteligência Artificial a ontologia deve ser considerada o “que existe para o sistema”.	Gruber
1993	Ontologia é uma <i>especificação formal e explícita</i> de uma <i>conceitualização</i> , ou seja, tudo o que existe pode ser representado de maneira formal e explícita.	Gruber
1997	Ontologia é <i>especificação formal, explícita e compartilhada de um conceito</i> , ou seja, corresponde a um conhecimento consensual dos fenômenos de determinado domínio.	Borst
1998	Ontologia é um “artefato” composto por um conjunto de termos e palavras (um vocabulário) que têm a tarefa de descrever uma determinada realidade aliada aos significados das palavras e seus conceitos.	Guarino
1998	Sugere o desenvolvimento de diferentes tipos de ontologias de acordo com o nível de generalidade necessária: (a) ontologias de alto nível, (b) ontologias de domínio, (c) ontologias de tarefa e (d) ontologias de aplicação.	Guarino
2001	Criação da ontologia de alto nível (core) SUMO	Niles & Pease, 2001
2002	Criação da ontologia de alto nível (core) Dolce	Gangemi et al.
CIÊNCIA DA INFORMAÇÃO		
1997	O termo ontologia passou a ser usado frequentemente, quando os métodos de classificação e organização do conhecimento passaram a suscitar interesse entre pesquisadores da CC e da Linguística, especialmente em função das demandas relacionadas à <i>Web</i> , ou seja, a necessidade de utilização de estruturas que permitissem um melhor acesso e recuperação da grande gama de informações disponíveis	SOERGEL e VICKERY
2003	Uma ontologia constitui-se em um documento (ou arquivo) que define as relações entre termos e conceitos de maneira formal, apresentando similaridades com os vocabulários controlados (muito difundidos na área da Biblioteconomia), tais como os tesouros e as taxonomias.	GILCHRIST
2004	Propõem uma classificação das ontologias concentrada no tipo de informação a ser modelada: (a) ontologias para a representação do conhecimento, (b) ontologias gerais e de uso comum, (c) ontologias de topo ou de nível superior (<i>upper ontologies</i>), (d) ontologias de domínio, (e) ontologias de tarefas, (f) ontologias de domínio-tarefa, (g) ontologia de métodos e (h) ontologias de aplicação.	Gómez-Perez, Fernández-Perez e Corcho
2008	O estudo de ontologias em CI deve-se ao grande interesse pela área da Organização e Representação do Conhecimento uma vez que as ontologias são estruturas que envolvem análise semântica visando otimizar a recuperação da informação, possibilitando aos programas de computadores uma “certa inteligência” no processamento das informações	Sales, Campos e Gomes
LINGUÍSTICA		
1949	<i>Translation</i> – obra apresenta os primeiros passos para estudos linguísticos de recuperação da informação. Início do PLN.	Weaver
1998	Distinção entre ontologias formais e ontologias linguísticas. As WordNets são consideradas ontologias linguísticas.	Vossen
2010	As ontologias linguísticas são fundamentadas na linguagem humana e portanto linguisticamente convencionadas, razão pela qual, muitas vezes, não oferecem estruturas conceituais precisas. De acordo com os autores, as ontologias linguísticas reconhecem a linguagem natural, considerando o senso comum como elementos importantes para a organização e análise do conhecimento	Prévot et al.

Quadro 1: Resumo de aspectos históricos acerca das ontologias
Fonte: elaborado pela autora

A partir da síntese também podemos chamar a atenção para o que afirma García Marco (2007): as ontologias são estruturas de representação do conhecimento que influenciaram (ou ainda influenciam) uma nova área da pesquisa altamente transdisciplinar, que envolve inúmeros campos. García Marco apresenta esse pressuposto transdisciplinar de forma gráfica, conforme pode ser visualizado na figura 1.

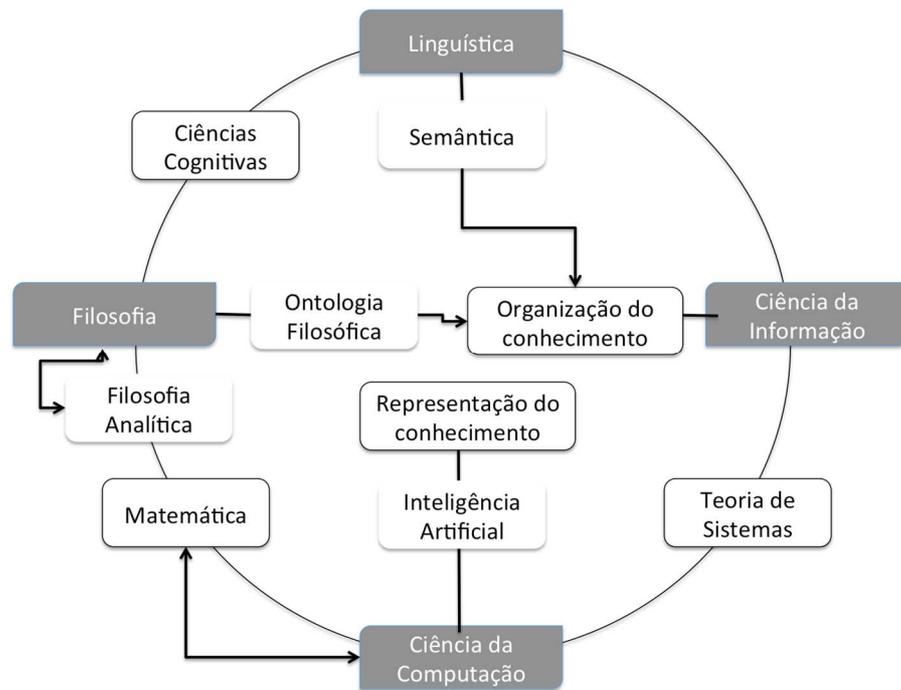


Figura 1: Ontologia – área de estudo interdisciplinar
Fonte: Adaptação de García Marco (2007, p. 548)

Como se vê na figura, as áreas da Filosofia, CC, CI e Linguística apresentam um inter-relação com foco na representação do conhecimento, tomando as ontologias como ponto comum nesta tarefa e, influenciando, cada uma sob a sua ótica, o desenvolvimento desta área considerada por García Marco (2007) como sendo transdisciplinar. Cabe novamente salientar que a Linguística não estuda ontologias propriamente, mas está engajada em aprimorar as soluções semânticas para estabelecimento de relações mais significativas entre os conceitos, caso ao qual nos dedicamos nesta pesquisa buscando demonstrar a potencialidade de Semântica de Frames para esta descrição.

Entendemos que as ontologias, tal como ferramentas de representação do conhecimento, receberam diferentes acepções, porém em todas as áreas cumprem o mesmo propósito. Contudo ainda não chegaram a ser um tema de estudos transdisciplinar conforme afirma García Marco (2007). No nosso entendimento, atualmente, pode-se considerar este tema como um estudo interdisciplinar apenas. Na nossa pesquisa, vamos focar o aspecto

semântico, principalmente no que se refere à estruturação dos conceitos e suas relações, tomando como base os preceitos das demais áreas para a estruturação formal da ontologia e teorias linguísticas para a determinação das relações entre tais conceitos.

Destas observações, pode-se concluir que a importância de uma ontologia é descrever e representar uma área do conhecimento. Determinado o domínio, sua *ontologia é o sistema de representação do conhecimento deste domínio*. Dessa forma, para se estruturar um sistema de representação de conhecimento eficiente e seu vocabulário é realizada uma análise ontológica formal do domínio.

Após a revisão da literatura embasada nas áreas apresentadas nesta seção, consideramos que *uma ontologia descreve formalmente os conceitos que estruturam um determinado domínio, bem como as relações semânticas existentes entre eles*. Entendemos *ontologia como um sistema que define e classifica o conhecimento, estruturando seus conceitos de forma clara e precisa e formalizando-os de modo a serem tratados por computadores*. Em suma, *uma ontologia deve ser construída com base em conceitos e nas relações existentes entre eles, levando em consideração o contexto do domínio*.

2.2 Ontologias na área jurídica

Diversas áreas valem-se das ontologias para melhorar o desempenho de sistemas de recuperação da informação. Na área jurídica tal prática mostra-se fundamental, uma vez que há um grande volume de documentos e informações relevantes e a necessidade de acessar de forma rápida e eficaz todo esse conteúdo.

De acordo com Mommers (2010), as ontologias legais apresentam alguns problemas. O primeiro se refere às variações relacionadas ao tempo e ao local, uma vez que os sistemas jurídicos podem sofrer modificações no decorrer do tempo e nos diferentes locais. Esta variedade de sistemas jurídicos implica em diferentes conceitos e relações, impossibilitando a criação de uma ontologia capaz de abarcar o todo da área jurídica, ou seja, raramente as ontologias seriam interoperáveis e reutilizáveis. Motivo este que nos leva a pensar uma ontologia específica para o Processo Penal Brasileiro baseada em *frames*.

O fato de estruturar o domínio tendo como base a Semântica de Frames pode, no nosso entendimento, minimizar o problema de interoperabilidade e reusabilidade previsto por Mommers uma vez que a descrição dos conceitos envolverá uma estrutura conceitual,

apresentando as cenas (*frames*) em representam o domínio e que os *frames* poderão ser readequados e reutilizados para a composição de outras ontologias jurídicas.

Outro problema encontrado no percurso da construção de ontologias jurídicas se refere ao uso da linguagem. A linguagem jurídica (também conhecida popularmente como “juridiquês”), ainda que de caráter protocolar e constituída por termos e expressões específicos da área (como por exemplo: “subiram os autos”, “por derradeiro transcrevo”, “juntada aos autos”, “vistos e relatados”, entre outros), corresponde à linguagem natural, ou seja, não é facilmente adaptável a uma linguagem formal como a usada por sistemas informatizados. Além disso, a estrutura sintática e semântica utilizada em textos jurídicos é um tanto peculiar e distante da normalmente usada pelas pessoas na comunicação. Em relação à linguagem jurídica, nossa preocupação neste trabalho recai especificamente sobre os termos equivalentes (ou variantes) que podem assumir diferentes sentidos de acordo com o contexto no qual se inserem. As fraseologias também são importantes no que se refere às relações entre os eventos e os participantes, sendo necessário atentar para elas, ainda que este não seja o foco da nossa pesquisa.

Outra dificuldade para a construção de ontologias jurídicas balizada por Mommers (2010) são as modificações na lei e as mudanças de significado de alguns termos, fazendo com que a ontologia tenha que ser aberta e adaptável, permitindo as alterações que forem surgindo no domínio com o passar do tempo. Neste trabalho, após a análise do *corpus*, temos o compromisso de elencar categorias e classes capazes de abarcar as informações sobre termos e legislação atual de modo a flexibilizar a estrutura possibilitando alterações futuras, caso seja necessário. Novamente o intuito de prover uma estrutura conceitual tendo como base os *frames* possibilita ajustes e readequações, uma vez que o foco da descrição estará justamente nos conceitos que, mesmo com mudanças na legislação ou nos termos da área, permanecerão inalterados em função da estrutura de *frames*.

Tais dificuldades na construção de ontologias jurídicas as diferenciam dos demais tipos de ontologias. Mommers (2010) apresenta uma distinção entre abordagens para a descrição de domínios legais, considerando três enfoques: (a) abordagens que se orientam pela interpretação semântica para a representação dos conceitos e relações; (b) abordagens epistemologicamente orientadas que focam o conhecimento do domínio; e (c) abordagens ontologicamente orientadas que salientam as entidades e relacionamentos constituintes do domínio. Vale salientar que estas abordagens tendem a se sobrepor no decorrer do desenvolvimento da ontologia.

Considerando os aspectos apontados por Mommers, reforçamos nossa proposta de construção de uma ontologia baseada na Semântica de Frames, uma vez que os conceitos serão descritos na ontologia tomando como base os *frames* específicos para o Processo Penal Brasileiro. Esta ação permitirá um detalhamento dos eventos e seus participantes, de modo a destacar os diferentes papéis que cada participante pode desempenhar no decorrer do processo.

Há registro de várias iniciativas no sentido de compor ontologias jurídicas e aprofundar estudos nesta área com vistas a ampliar a integração de dados e facilitar a recuperação da informação, porém o grande polo de estudos encontra-se na Europa, tendo como foco principal a Itália e a Holanda.

Na Itália destacam-se os projetos desenvolvidos pelo *Institute of Legal Information Theory and Techniquese* (ITTIG), cuja principal missão consiste em fazer uso da tecnologia para reinterpretar e integrar a informação jurídica no âmbito europeu, promovendo maior acesso e interoperabilidade de dados. É um instituto comprometido com pesquisas que abarcam diferentes setores, incluindo as metodologias e estratégias para a recuperação da informação jurídica, estratégias para a análise de textos legais, táticas para a conceituação e modelagem do domínio jurídico, além da investigação aprofundada da linguagem jurídica através de meios computacionais³¹.

O ITTIG faz parte de um projeto maior de trabalho com informação jurídica composto por diferentes grupos de pesquisa chamado *ICT4Law (ICT Converging on Law: Next Generation Services for Citizens, Enterprises, Public Administration and Policymakers)*³². Este projeto tem como compromisso desenvolver as tecnologias da informação no sentido de contribuir com a interoperabilidade e integração da área jurídica na Europa. Fazem parte deste projeto, além do ITTIG, sob responsabilidade de Daniela Tiscornia, a Universidade de Torino, através dos departamentos de Ciências Jurídicas, sob a responsabilidade de Gianmaria Ajani; Instituto de Economia, tendo como responsável Pietro Terna e o departamento de Informática, aos cuidados de Guido Boella. A Universidade de Piemonte também é parceira do projeto sob os cuidados dos professores Piercarlo Rossi e Laura Giordano.

³¹ Informações advindas do site do ITTIG, disponível em <http://www.ittig.cnr.it/Presentazione/Index.htm>, acesso em 28/04/15.

³² Mais informações sobre o projeto podem ser acessadas em <http://www.ict4law.org/>.

Além destes parceiros o projeto também conta com a participação de Nicola Guarino do Instituto de Ciências Cognitivas em Trento e de Monica Palmirani do Centro de Pesquisas em História do Direito e Informática Jurídica.³³

Citamos este projeto não só por ser representativo dos estudos em ontologias e recursos lexicais jurídicos, mas principalmente pelo foco ao qual se dedicam que consideramos relacionado a esta pesquisa. Os trabalhos desenvolvidos pelos grupos envolvidos com o *ICT4Law* têm uma preocupação em construir ontologias e recursos lexicais baseados em conceitos e relacionamentos semânticos, assim como nós temos a proposta de ampliar a descrição do Processo Penal Brasileiro através de uma ontologia cujo foco resida na descrição de eventos e papéis dos participantes tomando como base a Semântica de Frames.

Como trabalhos representativos deste grupo, podemos citar os projetos *LOIS - Lexical Ontologies for Legal Information Society* - (SAGRI e TISCORNIA, 2004)³⁴, *JurWordNet* (SAGRI, TISCORNIA & BERTAGNA, 2004) e o Projeto *Eunomos* (BOELLA et al., 2012). É interessante destacar que tais iniciativas são referidas pelos autores como léxicos computacionais ou ontologias linguísticas, novamente apresentando a distinção já mencionada anteriormente acerca da estrutura e organização em relação a termos e conceitos da área.

O *LOIS* (SAGRI e TISCORNIA, 2004) é um projeto co-financiado pela Comissão Europeia através do programa *e-Content* que visa à localização de *WordNets* descritivas do domínio jurídico para seis diferentes línguas europeias: italiano, inglês, alemão, tcheco, português e holandês. Estas *WordNets* são interligadas de acordo com o modelo da *EuroWordNet* (VOSSEN, 1998)³⁵, visando atingir a recuperação da informação nas diferentes línguas envolvidas no projeto. Com este projeto o cidadão ou o profissional do Direito tem acesso a uma base de dados com inúmeros documentos legais na sua linguagem, além de acessar tais documentos também em línguas diferentes.

Outro objetivo do projeto *LOIS* é prover a criação de um novo paradigma para acesso às informações e aos recursos jurídicos informatizados sob o ponto de vista multilíngue ultrapassando obstáculos terminológicos e de tradução.

³³ Informações advindas do site do projeto, disponível em www.ict4law.org, acesso em 28/04/15.

³⁴ Disponível em <<http://www.ittig.cnr.it/Ricerca/UnitaEng.php?Id=70&T=E>> Acesso em julho 2013.

³⁵ EuroWordNet é um banco de dados multilíngüe com wordnets para várias línguas europeias (holandês, italiano, espanhol, alemão, francês, tcheco e estoniano). Os wordnets são estruturados da mesma forma que o wordnet Americana para Inglês (Princeton WordNet, Miller et al 1990), em termos de synsets (conjuntos de palavras sinônimas) com relações semânticas básicas entre eles. Cada wordnet Representa um sistema de linguagem interna única de lexicalizações. Fonte: <<http://www.illc.uva.nl/EuroWordNet/>> Acesso em julho 2013.

Um projeto correlacionado ao *LOIS* é o *JurWordNet* (SAGRI, TISCORNIA, BERTAGNA, 2004), sendo que ambos se complementam pelo objetivo de representar o conhecimento jurídico e auxiliar na recuperação da informação. O *JurWordNet* é um léxico computacional (também pode ser chamado de ontologia linguística, se seguirmos a distinção descrita anteriormente) que abarca os termos jurídicos do italiano tendo como principal foco tratar a diversidade linguística presente na língua. Sua criação está fundamentada no *ItalWordNet* (SAGRI, TISCORNIA, BERTAGNA, 2004) e ambos baseiam-se na *WordNet* de Princeton e têm como principal objetivo prover soluções e recursos para a *WebSemântica*. É um recurso lexical jurídico que agrupa os itens lexicais em grupos de sinônimos para representar um mesmo conceito.

Na Universidade de Torino vem sendo desenvolvido o projeto *Eunomos* (BOELLA et al., 2012). O *Eunomos* consiste em um sistema de gestão de documentos legais representado em XML. Através desse sistema documentos do legislativo, as leis, por exemplo, podem ser recuperadas automaticamente a partir dos portais institucionais. Este sistema incorpora a *Syllabus Taxonomy* (AJANI et al., 2009), que foi desenvolvida para a construção de ontologias do domínio jurídico.

A ontologia *Eunomos* constitui-se em uma forte base de conhecimento para a análise e recuperação da legislação sendo concebida por profissionais da engenharia da informação, além de contar com o apoio de engenheiros do conhecimento legal que auxiliam na categorização e anotação das leis durante a construção do trabalho.

Boella et al. (2012) apresentam as principais categorias que compõem a *Eunomos*: cláusula deôntica, papel passivo, papel ativo, crime e sanção. Tais categorias são ampliadas e possuem subcategorias que correspondem às prescrições do Direito. A estrutura organiza as prescrições em conceitos e relações, permitindo ao usuário fazer buscas e aperfeiçoá-las, tornando-se um recurso útil para o trabalho oficial.

Outro ponto de referência no concernente à representação do conhecimento jurídico é a Universidade de Amsterdam, que desenvolve trabalhos relevantes em ontologia na área jurídica da União Europeia. André Valente e Joost Breuker, pesquisadores da universidade holandesa, foram os pioneiros (1996) na proposição de uma ontologia jurídica com categorias de base ligadas a subcategorias através de relacionamentos seguindo uma visão funcional do sistema jurídico – a *FOLaw - Functional Ontology of Law* (VALENTE & BREUKER, 1996). A *FOLaw* descreve os aspectos centrais da teoria jurídica, ampliando uma simples divisão da área em classes e subclasses.

Valente e Breuker (1996) apontam que os sistemas representativos do domínio jurídico necessitam estabelecer seu **compromisso ontológico** limitando a cobertura da ontologia, isto é, estabelecendo os limites entre o que será descrito ou não. A vantagem do uso de ontologias nesses sistemas é o fato de elas explicitarem o funcionamento da área, possibilitando a comparação de diferentes formas de representação do conhecimento jurídico em seu aspectos basilares.

Antes de descrever as ontologias desenvolvidas pelos holandeses, julgamos importante esclarecer o conceito de **compromisso ontológico**³⁶. Trata-se de um conceito advindo da Lógica que corresponde a uma pressuposição sobre a natureza da realidade. No âmbito das ontologias, mais especificamente quando se trata da Engenharia do Conhecimento, Gruber (1993) propõe uma adaptação declarando que o compromisso ontológico corresponde aos **acordos** ou **decisões** relacionadas a descrição do domínio de modo que os conceitos sejam usados de forma **coerente** e **consistente** através de um vocabulário compartilhado, tornando-se a **base para as decisões a serem tomadas no decorrer da modelagem**.

Levando em conta a posição de Breuker, Valente & Winkels (1997), que apontam que o compromisso ontológico corresponde ao conjunto de decisões tomadas na escolha dos conceitos e na sua descrição em detrimento de outros, ou seja, as decisões tomadas fundamentam a modelagem, consideramos que o nosso compromisso ontológico está relacionado às decisões que tomamos para a construção da ontologia, quais sejam: o recorte do domínio, considerando somente o Processo Penal Brasileiro em casos passíveis de julgamento no juizado especial; o uso da Semântica de Frames para fundamentar e organizar os relacionamentos entre conceitos; a organização da área seguindo quatro categorias base (eventos, participantes, instituições e documentos); a escolha de conceitos relacionados às categorias eventos e participantes para descrever através da teoria semântica escolhida; a usabilidade da ontologia em recursos lexicais ou sistemas de recuperação da informação.

Tendo esclarecido nossa posição em relação aos compromissos ontológicos relacionados a esta pesquisa, passamos a analisar e descrever os trabalhos desenvolvidos pelo grupo holandês. Dentre os trabalhos desenvolvidos por este grupo de pesquisadores, destacamos: a ontologia *LRI-Core - Laboratorium voor Rechts Informatica* - (BREUKER e HOEKSTRA, 2004)³⁷ que deu origem à *LKIF-Core - Core Ontology of Basic Legal Concept* -

³⁶ É possível encontrar a expressão “comprometimento ontológico” como sinônima para “compromisso ontológico”.

³⁷ Disponível em <<http://www.leibnizcenter.org/previous-projects/lricore>> Acesso em julho 2013.

(HOEKSTRA et al, 2007)³⁸, ambas participantes do projeto *Estrella* (LUPO, 2007)³⁹, cuja finalidade é desenvolver uma base ontológica para interligação entre os sistemas jurídicos europeus.

A *LRI-Core* (BREUKER e HOEKSTRA, 2004), desenvolvida pelo departamento de informática jurídica da Universidade de Amsterdam, é uma ontologia do domínio legal que apresenta uma conexão evidente com o mundo físico, isto é, considera que muitos conceitos encontrados no meio legal incidem também no senso comum, tendo-os incluído e representado na sua estrutura. É uma ontologia que tem como objetivo organizar de forma indexada outras ontologias do domínio servindo de base para novas estruturas.

Trata-se de uma ontologia *core* que compreende as seguintes categorias: conceitos físicos, conceitos mentais, funções e conceitos abstratos. Há ainda uma quinta categoria que corresponde aos termos de ocorrências, aqueles usados em situações genéricas. É considerada uma ontologia *core* por abarcar os principais conceitos comuns a todos os domínios legais. Uma visão geral desta ontologia pode ser vista na figura 2.

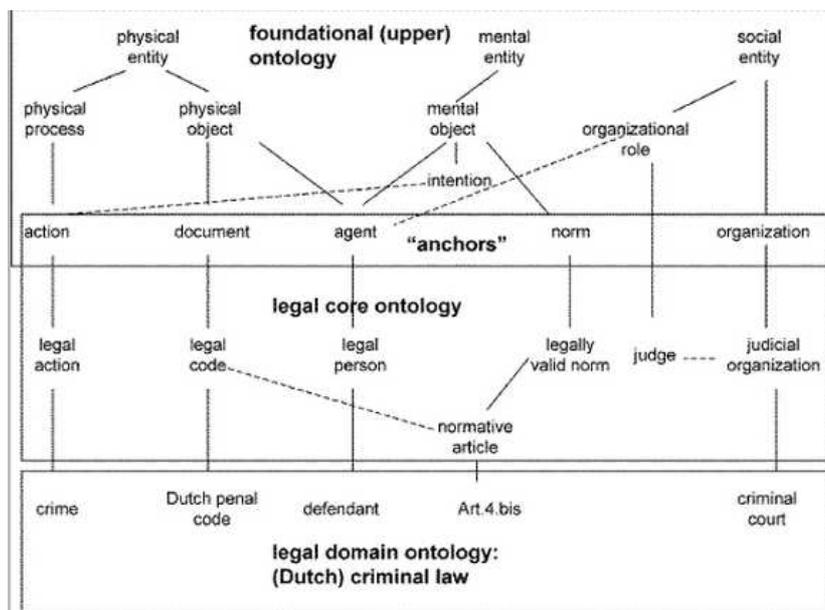


Figura 2: Visão geral da LRI-Core.
Fonte: BREUKER, 2004 in: eLEGI Workshop

Na figura 2 podemos visualizar as classes mais abstratas da LRI-Core: Conceito-Físico⁴⁰, Processo, Qualidade, Quantidade, Conceito-Mental, Papel,

³⁸ Disponível em <http://www.estrellaproject.org/?page_id=3> Acesso em julho 2013.

³⁹ Informações sobre o projeto Estrella em <<http://www.estrellaproject.org/lkif-core/>> Acesso em julho 2013.

⁴⁰ A fim de facilitar o entendimento do texto, destacaremos as palavras referentes aos nomes de classes nas ontologias usando a fonte Courier New. Esta notação será seguida em todos os capítulos do trabalho.

Conceito-Abstrato e Ocorrência. De acordo com Breuker e Hoekstra (2004), a ontologia faz uma distinção entre duas camadas, uma de topo e a outra nuclear. Esta organiza as classes Ação, Documento, Ator, Norma e Organização Social, estabelecendo uma ligação com os domínios legais, enquanto que aquela traz conceitos abstratos ligados ao domínio jurídico, contendo elementos do mundo físico e mental.

As classes Ação, Documento, Ator, Norma e Organização Social foram estudadas por Minguelli (2011) em sua dissertação de Mestrado, fundamentando as classes Eventos Legais, Documentos Legais, Instituições Legais e Participantes Legais atualmente elencadas pelo nosso grupo de pesquisa como sendo as principais categorias para a estruturação do domínio do Processo Penal Brasileiro.

Nesta pesquisa representaremos as referidas classes organizando-as de forma hierárquica para a construção da ontologia, mas nosso foco reside nas classes Participantes Legais (Ator, na *LRI-Core*) e Eventos Legais (Ação, na *LRI-Core*), uma vez que pretendemos ampliar a representação semântica dos **conceitos** através dos diferentes papéis que cada participante pode desenvolver em relação aos diferentes eventos do domínio. As outras duas classes são importantes para a estruturação do domínio, mas não interferem na representação semântica, levando em conta a Semântica de Frames.

A *LRI-Core* serviu de base para a construção de uma outra ontologia, a *LKIF-Core* desenvolvida no âmbito do Projeto *Estrella* do Centro Leibniz de Direito. *Estrella* é um projeto europeu para a representação transparente e padronizada a fim de estender a acessibilidade legal (LUPO et.al, IST-2004-027655). Seu principal objetivo é o desenvolvimento de uma plataforma baseada em padrões capaz de permitir a implementação de soluções legais para a gestão do conhecimento, visando apoiar a gestão dos documentos legais entre outras funções.

A ontologia *LKIF-Core* (HOEKSTRA et al, 2007) compreende o principal objetivo técnico do projeto *Estrella*. Esta ontologia segue as principais categorias da *LRI-Core*, porém é desenvolvida nos padrões XML (*Extensible Markup Language*) recomendados pelo W3C para soluções em *Web Semântica*, incluindo as linguagens RDF (*Resource Description Framework*) e OWL (*Ontology Web Language*), além de prever interfaces de aplicações para a interação com o conhecimento jurídico⁴¹.

⁴¹ Informações advindas do site do Projeto Estrella. Disponível em <<http://www.estrellaproject.org/>> Acesso em: 17/07/2013

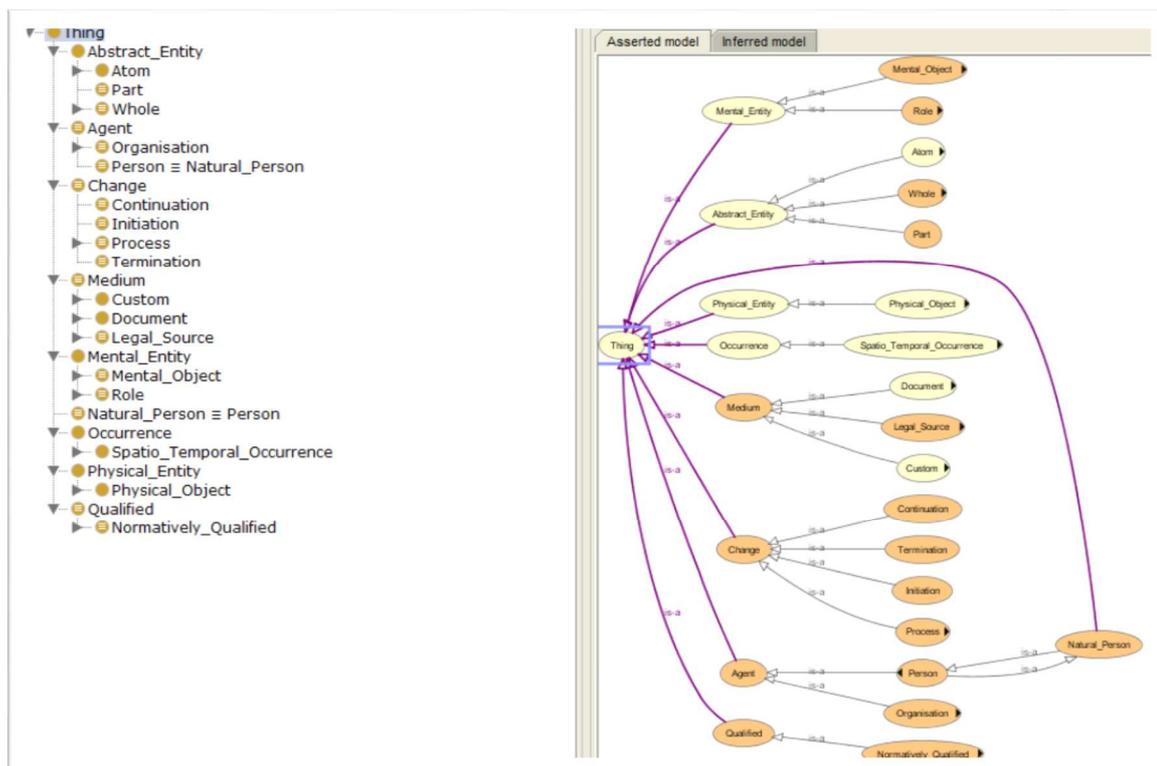


Figura 3: Representação parcial da estrutura da LKIF-Core
 Fonte: <http://www.estrellaproject.org/lkif-core/>. Acesso em 08/09/2013.

Esta ontologia é composta por quinze classes/módulos/categorias (como se pode ver parcialmente na figura 3), cada um responsável pela descrição de um conjunto de conceitos relacionados aos domínios legais e do senso comum. Assim como a *LRI-Core*, trata-se de uma biblioteca de ontologias relevantes para o domínio jurídico. Cinco classes correspondem a conceitos mais abstratos, estando estreitamente relacionados, são eles: Topo, Lugar, Mereologia, Tempo e Espaço-tempo. Os conceitos de nível básico foram acomodados em quatro classes: Processo, Papel, Ação e Expressão. Os agrupamentos básicos são estendidos para mais três classes que formam a ontologia legal: Ação legal, Papel legal e Norma. Além dessas classes, há ainda outras duas que cumprem o objetivo de cobrir o vocabulário básico: Modificação e Regras. E, finalmente, estas quatorze classes estão integradas ao módulo LKIF-Núcleo, que não fornece informações adicionais, mas cumpre o papel de ponto de acesso à biblioteca da ontologia⁴².

A *LKFI-Core*, pela forma como está estruturada e pelas classes que acomoda em seu nível básico, possibilita o reuso nesta pesquisa. Podemos fazer uso das classes consideradas de nível básico: Processo, Papel e Ação, bem como as subdivisões Ação legal,

⁴² Informações advindas do site sobre a ontologia LKIF. Disponível em <http://www.estrellaproject.org/?page_id=3> Acesso em: 17/07/2013.

Papel legal e Norma. Consideramos que as classes abstratas (Lugar, Mereologia, Tempo e Espaço-tempo) inclusas na LKFI-Core, também podem acomodar o domínio previsto para a construção da ontologia que aqui propomos. Tais classes podem prover a estrutura núcleo da nossa ontologia, permitindo sua ampliação através das classes advindas do estudo de *corpus*. Por se tratar de uma ontologia *core* (de topo ou fundacional), a *LKFI-Core* não descreve as classes de níveis mais baixos, tais como diferentes papéis e eventos, mas propicia a interligação entre elas, servindo de base para a ontologia de domínio que temos como objetivo.

No Brasil, ainda que em outro nível de aprofundamento, até mesmo porque são ontologias de domínio, podemos trazer como exemplos as ontologias OntoJuris e OntoTrib (ARAÚJO et al, 2008) desenvolvidas para o direito brasileiro.

A OntoJuris e a OntoTrib são ontologias desenvolvidas em OWL seguindo os preceitos do *W3C* para aplicações para a *Web Semântica*. Originalmente foram concebidas para representar modelos e informações pertinentes à área jurídica. A OntoJuris descreve a estrutura das normas jurídicas levando em conta a legislação e a hierarquia dos diferentes ramos jurídicos do Brasil. Como uma extensão da primeira, a OntoTrib corresponde à descrição do Direito Tributário, representando os conceitos pertinentes a essa área do Direito (ARAÚJO et al, 2008; DRUMOND & GIRARDI, 2008).

A OntoJuris é uma ontologia criada para representar as normas e seus tipos, além de descrever a estrutura do direito brasileiro, isto é, representa os tipos de instrumentos normativos e seus conteúdos e as diferentes áreas que compõem o sistema jurídico do Brasil. Sua estrutura permite relacionar as normas ao ramo específico ao qual dizem respeito. A divisão tradicional do Direito: (a) **público**, que compreende as leis e normas relacionadas ao interesse coletivo e o Estado; e (b) **privado**, que regulamenta os interesses particulares também está representada na ontologia. A OntoJuris é composta pelas classes *Case_Generator*; *Date*; *Dispositive*; *Legal_Affinity*; *Legal_Branch*; *Legal_User*; *Normative_Instrument*; *Type* (ARAÚJO et al, 2008; DRUMOND & GIRARDI, 2008).

Na figura 4 está representada a estrutura hierárquica dos ramos do Direito expressa na OntoJuris, que foi criada com o propósito de utilização em sistemas de filtragem e recomendação de informações.

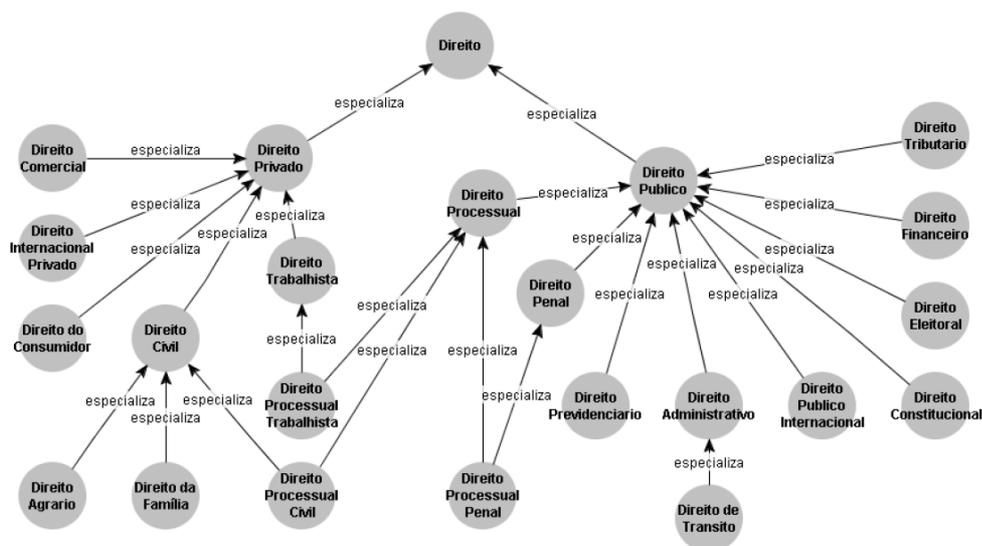


Figura 4: Hierarquia dos ramos jurídicos na OntoJuris
Fonte: ARAÚJO et al., 2008, p. 3

A OntoJuris organiza os diferentes ramos do Direito e nos auxilia nesta visão geral, situando-nos nesta área tão ampla. Na figura 4 percebe-se a gama de subdivisões do Direito Brasileiro e a nossa posição em relação a descrição do domínio.

A OntoTrib é uma ontologia para o domínio do Direito Tributário que foi desenvolvida a partir da estrutura da OntoJuris, ambas com o intuito de auxiliar usuários através de recomendações para o acesso às informações. Esta ontologia foi construída com base no conceito de tributo, enfatizando os conceitos do Direito Tributário e a classificação dos diferentes tipos de tributo existentes no país. A OntoTrib é composta pelas classes Tributo, Tributary_Application; Tributary_Concept; Tributary_Normative_Instrument e Tributary_Legal_User. (ARAÚJO et.al, 2008). Apresentamos brevemente esta ontologia apenas a título de exemplificação dos trabalhos realizados no Brasil, porém a OntoTrib não estabelece nenhuma relação com este trabalho por se tratar de um domínio específico do Direito.

Ainda cumpre-se citar o projeto desenvolvido por Aires Rover⁴³ (professor associado da Universidade Federal de Santa Catarina no cursos de Direito e Engenharia de Gestão do Conhecimento) denominado *Gestão do Conhecimento e TI no Poder Judiciário*, cujo objetivo principal é intensificar as reflexões e pesquisas relacionadas ao Poder Judiciário visando atender aos limites e desafios acerca da utilização da tecnologia da informação para o aprimoramento do Judiciário brasileiro. O projeto é composto por diferentes iniciativas no sentido de: verificar como a tecnologia da informação pode contribuir para a informatização

⁴³ Projeto relacionado ao edital 20/2010/CAPES/CNJ.

do sistema judiciário de forma mais efetiva e apresentar a efetividade da utilização do Modelo de Requisitos para Sistemas Informatizados de Gestão de Processos e Documentos do Judiciário – MoReq-Jus.

Complementando os estudos brasileiros, é necessário citar o pioneirismo do grupo SemanTec no sentido de explorar as formas de representação da informação jurídica sob um viés semântico.

As ontologias jurídicas, assim com as demais ontologias, quando vistas sob um viés social, facilitam o relacionamento homem-máquina em sistemas oferecidos na *Web*, tais como serviços de consultas a sites do governo. Se vista sob um olhar linguístico, as ontologias constituem-se em bases relacionais de informação léxico-semântica e lógico-conceituais capazes de auxiliar no processamento das línguas naturais (EVENS, 1988). E, na visão do PLN, de acordo com Fensel (2000), as ontologias configuram-se em modelos abstratos capazes de representar o mundo, sendo constituídas por conceitos e relacionamentos e processáveis por programas computacionais.

A Informática Jurídica é uma das áreas que muito se beneficia de estudos sobre as ontologias e sua aplicação na recuperação da informação (RI). De acordo com Rover (2010), a informática pode ser aplicada ao Direito sob dois enfoques: como **objeto** ou como **meio**. Se vista como **objeto**, caracteriza-se por tratar acerca da regulamentação e uso de computadores refletindo sobre práticas e consequências do uso de tecnologias, abrangendo as normas reguladoras para utilização de sistemas eletrônicos versando sobre privacidade, proteção ao software, liberdade, etc. Sob essa ótica pode ser chamada de **direito da informática**. Tendo como foco o **meio**, caracteriza-se como **informática jurídica**, versando acerca de metodologias e técnicas de processamento de informações através de sistemas informatizados.

O processo de informatização do Poder Judiciário teve início após a edição da Lei nº 11.280, de 16 de fevereiro de 2006, a qual determinou a inclusão de um parágrafo único⁴⁴ ao artigo 154 do Código do Processo Civil, determinando que os tribunais poderão proceder à comunicação oficial de processos através do meio eletrônico, desde que sejam atendidos os requisitos de autenticidade, integridade e validade jurídica permitindo que haja a interoperabilidade dos sistemas.

⁴⁴ “Art. 154 – Parágrafo único – Os tribunais, no âmbito da respectiva jurisdição, poderão disciplinar a prática e a comunicação oficial dos atos processuais por meios eletrônicos, atendidos os requisitos de autenticidade, integridade, validade jurídica e interoperabilidade da Infra-Estrutura de Chaves Públicas Brasileira (ICP) Brasil.”

De acordo com Paiva (2011), a partir daí surgem diversas questões referentes a organização, validação e autenticidade das informações apresentadas através dos sistemas, gerando a necessidade de criação de uma legislação específica sobre o assunto: é criada a Lei nº 11.280/2006 que tem o intuito de prescrever as normas acerca do uso de sistemas digitais para o judiciário. Essa Lei constitui-se o ponto de partida para o desenvolvimento de sistemas capazes de garantir a validade dos documentos digitais.

Diante da abertura prevista pela Lei nº 11.280/2006, diversos sistemas informatizados foram implantados no Brasil, porém tais programas ainda caracterizam-se pela limitação de automação de processos e pouca (ou nenhuma) integração com os demais, uma vez que cada comarca e cada estado desenvolve seu próprio recurso informatizado de acesso às informações.

A limitação na integração dos sistemas nas diferentes instâncias e o grande número de informações jurisprudenciais dificultam o trabalho dos magistrados no sentido de fundamentação de seus processos de acusação ou defesa, uma vez que estes realizam buscas para fundamentar seus argumentos. As ontologias podem contribuir para a minimização desta limitação permitindo a integração das informações judiciárias.

Dedicamos a seção seguinte para esclarecer, à luz dos preceitos dados por Gruber (1996), a estrutura necessária a uma ontologia independente de seu grau de formalidade ou domínio.

2.3 Estrutura das ontologias

A estrutura das ontologias, definindo os seus componentes, é também um aspecto bastante discutido na literatura tanto da área da CI quanto da CC. Autores como Gruber (1996), Gómez-Pérez (1999), Noy & McGuinness (2001) e Staab e Studer (2013) mencionam uma estrutura semelhante, porém o que os difere é a explicação dada para cada componente. Todos apresentam as **classes** como estruturantes da ontologia através de uma relação **taxonômica**, os **axiomas** como sendo sentenças verdadeiras em relação aos conceitos, a necessidade de inclusão de **relações** entre as diferentes classes e a importância das **instâncias** como representantes de elementos específicos para cada classe. Em razão da semelhança entre as definições e estruturas não vamos detalhar o posicionamento de cada autor, vamos, contudo, focar na estrutura apresentada por Gruber (1996) por considerá-la mais descritiva e detalhada.

Conforme Gruber (1996), os componentes básicos de uma ontologia são as **categorias**, **classes** e **subclasses**, organizadas em uma taxonomia; as **relações**, que representam o tipo de interação entre os conceitos de um domínio; as **propriedades** ou **atributos** referentes a cada entidade; os **axiomas**, usados para modelar sentenças sempre verdadeiras, e as **instâncias**, utilizadas para representar elementos específicos, ou seja, os próprios dados.

Para Gruber (1996), as **classes** correspondem aos conceitos referentes ao domínio. Tais conceitos podem ser concretos ou abstratos e representam um conjunto com características comuns sob uma única denominação. Por exemplo: a classe `Documentos Legais` denomina um grupo com características semelhantes, ou seja, agrupa todos os tipos de documentos relacionados ao domínio.

Já as **subclasses** são ocorrências particulares do objeto em relação à classe considerada. Uma subclasse também descreve conceitos de forma única, individualizada e concreta, fazendo referência a um objeto real (GRUBER, 1996). Por exemplo: a classe `Documentos Legais` pode ter como subclasses `Doutrina` e `Jurisprudência`, isto é, correspondem a tipos específicos de `Documentos Legais`.

As classes e subclasses são organizadas hierarquicamente por **relações semânticas** de **inclusão (é-um ou tipo-de)** e **parte-todo (parte-de)** relacionadas aos termos específicos do domínio, gerando a sua estrutura taxonômica. Além disso, são compostas de conceitos integrados através de relações semânticas, cuja natureza e número têm sido alvo de inúmeros estudos nas áreas da Computação, da Ciência da Informação, Recuperação da Informação e da Linguística.

A relação de **hiponímia** é conhecida como uma relação de inclusão, ou seja, a relação entre um termo mais específico em um mais geral (LYONS, 1987). Por exemplo: banana e maçã são **hipônimos** de fruta, logo fruta é **hiperônimo**, ou seja, inclui banana e maçã na categoria das frutas. A **hiponímia** é a relação mais importante para a construção da base da ontologia (a taxonomia), uma vez que é a mais frequente tanto entre os nominais quanto entre os verbos. Pode-se dizer que é uma relação lexical que corresponde à inclusão de uma classe em outra; tal inclusão depende do ponto de vista com que se analisará o significado.

Para identificarmos as relações de hiponímia entre unidades lexicais ou conceitos, podemos fazer uso dos axiomas *<X é-um Y>* ou *<X é-um-tipo-de Y>*. Por exemplo: *Petição Inicial é-um*⁴⁵ *Documento* ou *Habeas Corpus é-um-tipo-de Recurso*.

A **meronímia** é igualmente uma relação lexical caracterizada pelo traço de inclusão. No entanto, difere da hiponímia, pois estabelece uma relação hierárquica em que Y é parte de um outro objeto A. Dessa forma, trata-se de uma relação parte-todo, sendo caracterizada, dentro da hierarquia, pelo verbo **ter**. Para identificar a meronímia, pode-se fazer uso do axioma *<X é uma parte de Y; logo, Y tem um X>* (CRUSE, 1986). Por exemplo: *Decisão é-parte-de* um *Acórdão*. Um *Acórdão tem* *Decisão*.

Além das relações taxonômicas, as classes possuem características que são definidas pelos **atributos** ou **propriedades** que têm valor na subclasse; dessa forma, se estabelece uma relação hierárquica, de modo a compor uma taxonomia, estabelecendo, assim, uma estruturação na qual classes **pai** possuem classes **filho** (as subclasses), com atributos herdados dos pais. Por exemplo: considerando a classe *Participantes Legais* e suas propriedades – *ser humano, ter diferentes papéis, participar de eventos* – uma subclasse poderia ser *Autoridade Julgadora* que, automaticamente, herdaria as mesmas propriedades.

De acordo com Gruber (1996), as **propriedades** ou **atributos** são muito importantes na composição da ontologia, pois se referem à descrição de outros tipos de relações, sendo que estas correspondem a informações adicionais ligadas à estrutura hierárquica dada pelas classes e subclasses. É através das propriedades que ocorre a caracterização e a qualificação da categoria com a qual estão relacionadas. No caso do nosso trabalho as propriedades e atributos correspondem aos diferentes papéis assumidos pelos participantes nos eventos, sendo determinantes para a descrição do domínio.

Seguindo Gruber (1996), entendemos que os axiomas são sentenças verdadeiras em relação aos conceitos e servem como restrições relacionadas às classes. Como exemplo de axioma poderíamos dizer que *Maria é mulher* ou *Evento Legal gera Documento*. Outro exemplo seria dizer que *todo Participante Legal deve ser humano*, desta forma estamos usando um axioma para definir uma restrição, ou seja, caso seja vinculado um documento à classe *Participantes Legais*, o sistema acusaria um erro uma vez que documento não tem a propriedade de *ser humano*. Os axiomas ou restrições são necessários para evitar

⁴⁵ Visando facilitar a compreensão dos exemplos, as expressões referentes às propriedades das classes serão escritas usando a fonte *arial itálico*.

inconsistências na ontologia. Cabe aqui um esclarecimento: os axiomas, segundo Gruber (1996) devem ser definidos através de uma linguagem lógica, descrevendo cada situação seguindo os padrões lógicos. No caso do nosso trabalho, vamos prever as restrições, porém não temos o compromisso de escrevê-las através da linguagem lógica, este procedimento se dará diretamente no editor de ontologias (no capítulo 4 tecemos maiores esclarecimentos acerca desta questão).

Tais aspectos, por serem essenciais para a construção de ontologias, têm sido amplamente discutidos pelos autores das áreas da CC e da CI. Porém, não só a organização taxonômica e as propriedades/atributos são importantes para o desenvolvimento de sistemas capazes de depurar a linguagem natural, é fundamental levar-se em conta as questões referentes às **relações** existentes entre tais conceitos.

Os estudos sobre quais são as relações que devem compor uma ontologia são os mais variados e apresentam diferentes perspectivas de acordo com a área ao qual estão ligados. Segundo Hjørland (2007a), estudos que envolvem as relações semânticas são abordados de forma fragmentada no âmbito da CI. Para o autor, a área da CI deveria promover um debate teórico mais aprofundado acerca das questões semânticas e os pesquisadores desta área, mais precisamente os que trabalham com a representação do conhecimento, deveriam fundamentar seus trabalhos em teorias semânticas, favorecendo uma maior aproximação entre as áreas.

Também refletindo sobre as relações em ontologias, Sales, Campos e Gomes (2008) realizaram uma pesquisa na área da CI na qual estudaram as diferentes relações semânticas e propuseram um inventário de relações sistematizadas para serem usadas em ontologias. Neste trabalho, as autoras categorizaram as relações em **categoriais** e **formais**. Aprofundando, subdividiram-nas em relações **genéricas**, **partitivas** e **funcionais** e identificaram seis classes de relações: **categoriais genéricas**, **formais genéricas**, **categoriais partitivas**, **formais partitivas**, **categoriais funcionais** e **formais funcionais**. Com esta pesquisa, as autoras criaram um inventário de relações e, aparentemente, este problema acerca das ontologias estaria completamente solucionado, bastando escolher as relações adequadas a partir de uma lista.

Ocorre que, ao olhar para o domínio jurídico, percebe-se que apenas as relações **categoriais** e **formais** com suas subdivisões não dão conta de descrever adequadamente o domínio, ainda mais no que se refere aos diferentes eventos e seus participantes. Tais relações são adequadas para organizar a estrutura do domínio, mas não são úteis para representar as

peculiaridades da área, como, por exemplo, os diferentes papéis que cada participante pode exercer de acordo com o evento.

É justamente com base nesta discussão iniciada por Hjørland (2007a) e ampliada por Sales, Campos e Gomes (2008) que fundamentamos nosso propósito de apresentar a Semântica de Frames como teoria linguística capaz de ampliar a descrição dos significados incluindo diferentes relações entre as classes. Através da Semântica de Frames é possível estabelecer os diferentes papéis para cada participante nos eventos, permitindo descrever as especificidades do domínio. Por exemplo: em um processo uma mesma pessoa pode assumir diferentes papéis – a pessoa pode ser o **citado** no evento **Citação**, pode ser o **acusado**, no evento **Petição Inicial** e pode ser o **réu** no evento **juízo**. Todos esses papéis apresentam algumas propriedades comuns, mas se modificam em cada evento. É aqui que a Semântica de Frames tem a função basilar de distinguir os diferentes papéis. Vamos tratar os diferentes papéis como **relações não taxonômicas**.

Há consenso entre semanticistas e informatas de que as relações semânticas – taxonômicas ou não taxonômicas – são fundamentais na construção de ontologias. De acordo com Saint Dizier & Viegas (1995), as relações semânticas são um dos principais objetos de estudo da Semântica Lexical e interferem no plano do processamento da linguagem natural, uma vez que as ontologias são consideradas ferramentas para a representação do conhecimento, pois tais estruturas se caracterizam por um sistema classificatório bem descrito e marcado, permitindo que sejam passíveis de entendimento para a máquina.

Conforme já referimos, acreditamos na Semântica de Frames como teoria linguística capaz de cobrir as **relações não taxonômicas**, especificando os diferentes cenários (*frames*) em que cada participante está envolvido, bem como os diferentes papéis que podem ser desempenhados por eles. Através da inclusão destas relações vinculadas aos *frames* jurídicos⁴⁶, temos o intuito de prover à estrutura ontológica uma descrição mais exata e ampla do domínio, favorecendo o uso em recursos lexicais ou sistemas de recuperação da informação.

As relações taxonômicas são fáceis de incluir em uma ontologia, uma vez que são constituintes da estrutura hierárquica necessária para sua composição. Porém, as relações não taxonômicas que, de acordo com nossa proposta, são representadas através dos *frames* e seus elementos ainda constituem um desafio para os linguistas envolvidos com estudos voltados

⁴⁶ No capítulo 3 apresentamos uma discussão mais profunda acerca da Semântica de Frames explicitando sua importância para a descrição das classes e dos diferentes papéis dos participantes, considerando os verbos.

aos sistemas de recuperação da informação. Acerca destas relações vamos aprofundar no capítulo 3 no qual trataremos especificamente da noção de *frame* e da teoria da Semântica de Frames.

2.4 Esquema síntese do capítulo

A figura 5 apresenta um esquema síntese das ideias apresentadas no capítulo.

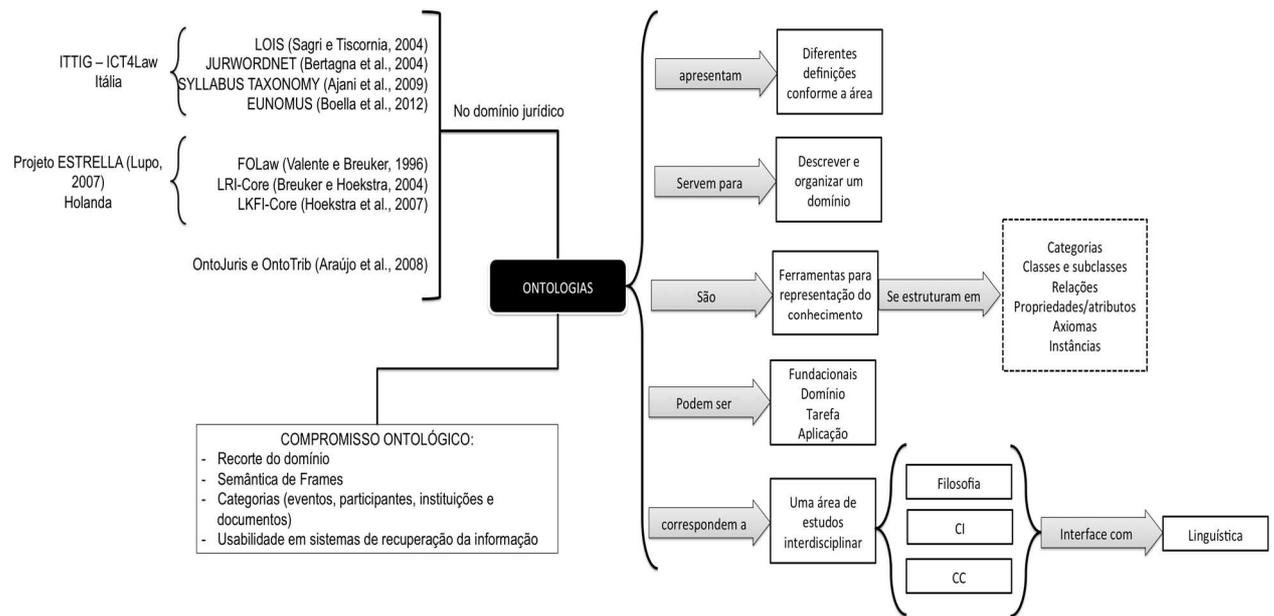


Figura 5: Esquema síntese do capítulo 2

Fonte: Elaborado pela autora

Nosso foco nesta tese é apresentar uma visão linguística na construção de ontologias, trazendo uma proposta metodológica que inclua a Semântica de Frames na descrição dos conceitos jurídicos, possibilitando um maior detalhamento dos eventos e seus participantes, por esta razão destacamos o capítulo seguinte para tratar da teoria linguística escolhida como base para nossa pesquisa.

3 SITUANDO A SEMÂNTICA DE FRAMES

Este capítulo cumpre com o papel de apresentar a teoria semântica à qual recorreremos para embasar nossa tese, esclarecendo fundamentos e conceitos relacionados. Tem como principal objetivo apresentar a Semântica de Frames (FILLMORE, 1982; 1985) como contributo para a descrição dos conceitos na ontologia e refletir sobre sua aplicabilidade nestas estruturas que visam à recuperação da informação. Para isso, apresentamos sua fundamentação na Linguística Cognitiva e sua aplicação computacional através da *FrameNet* (RUPPENHOFER et al., 2010).

A escolha por esta teoria semântica se deve ao fato de a área do Direito ser fundamentalmente organizada em diferentes eventos inter-relacionados, os quais podemos considerar como *cenas* ou *cenários* específicos; logo, entendemos que, organizar esta área através da Semântica de Frames, permite uma reflexão aprofundada em relação aos eventos, seus participantes e seus diferentes papéis para a descrição do domínio na ontologia.

A Semântica de Frames é uma das abordagens ligadas à Linguística Cognitiva (LC), uma corrente Linguística que corresponde a um conjunto de teorias que compartilham princípios relacionados à cognição humana, estabelecendo um novo paradigma na compreensão da linguagem. Esse novo paradigma advém de uma reflexão acerca do dualismo corpo-mente, compreendendo que ambos são constitutivos da composição dos significados partindo do uso da linguagem, opondo-se aos preceitos já estabelecidos apregoados pelo estruturalismo e pelo gerativismo.

Assim, a LC caracteriza-se por adotar uma perspectiva não modular da linguagem, considerando que os princípios cognitivos são compartilhados pelas capacidades cognitivas e pela linguagem. Seguindo esta linha, os significados correspondem à mediação entre linguagem e cognição. Ou seja, se trata de uma corrente linguística que considera que a linguagem é perspectivada em relação ao meio e a experiência humana.

De acordo com Salomão (2009), a LC filia-se às Ciências Cognitivas (Psicologia Cognitiva, Filosofia, Inteligência Artificial, entre outras) integrando-se a elas sob a perspectiva interdisciplinar de tratamento da cognição humana e considerando a intensa

relação da experiência corpórea com o entendimento do mundo, porém avançando no tratamento dos fenômenos linguísticos.

Conforme referimos, a Semântica de Frames (SF) é uma das teorias que fazem parte da LC. Tal abordagem foi desenvolvida por Fillmore e discutida em seus trabalhos de 1975, 1977, 1982 e 1985, tendo como foco principal a estrutura semântica dos itens lexicais e as construções gramaticais. Nessa perspectiva, o termo *frame* (Fillmore, 1982) corresponde a uma forma de estruturar o conhecimento através de cenas construídas na memória e organizadas através de um esquema baseado na experiência.

Os estudos relacionados à Semântica de Frames, no âmbito da LC, propiciaram a criação de um recurso lexical (ou léxico computacional) chamado *FrameNet*, desenvolvido em Berkeley, no *International Computer Science Institute*, sob a responsabilidade de Collin F. Baker, Charles J. Fillmore e John B. Lowe (BAKER; FILLMORE; LOWE, 1998). A *FrameNet* corresponde a um recurso lexical para o inglês que foi implementado seguindo os preceitos da Semântica de Frames com base em evidências colhidas em *corpora*. Foi a partir da implementação da *FrameNet*, no final da década de 90, que a SF passou a ser plenamente reconhecida como uma teoria.

Considerando os aspectos apresentados acima e o nosso objetivo com este capítulo, o organizamos em quatro seções. A seção 3.1 tem o intuito de apresentar a LC a fim de situar nossa pesquisa na área da Linguística, bem como apresentar as origens dessa abordagem cognitiva. Na seção 3.2 aprofundamos os preceitos envolvidos no nosso trabalho, destacando a Semântica de Frames e seu contributo teórico para a tese, refletindo sobre a aplicabilidade desta teoria para a descrição do domínio em uma ontologia jurídica. Por este ser um trabalho de interface entre Linguística e tecnologia, julgamos importante tecer esclarecimentos acerca do recurso lexical *FrameNet*. Para isso, destacamos a seção 3.3, a qual tem o objetivo de apresentar a *FrameNet*, sua origem, organização e sua relação com o domínio jurídico, bem como refletir sobre possibilidades de aproximação deste recurso lexical à ontologia que propomos neste trabalho. E, por fim, na seção 3.4, apresentamos um esquema síntese do capítulo.

3.1 Linguística Cognitiva

O crescente interesse pelo significado das palavras e os estudos psicolinguísticos de Eleanor Rosch sobre os protótipos e o processo de categorização (ROSCH 1999, ROSCH & MERVIS, 1975) foram os principais motivadores do surgimento da LC no final dos anos 70. Porém, sua institucionalização deu-se apenas por volta dos anos 90 com a criação da *International Cognitive Linguistics Association* - ICLA (Associação Internacional de Linguística Cognitiva)⁴⁷, a primeira publicação da revista *Cognitive Linguistics* (dirigida por Dirk Geeraerts) e a criação da coleção *Cognitive Linguistics Research* (editada René Dirven e Ronald Langacker e publicada por Mouton de Gruyter)⁴⁸ (SILVA, 1997).

Como principais representantes dessa vertente teórica podemos citar os norte-americanos George Lakoff, por suas contribuições acerca das metáforas (LAKOFF, 1987; LAKOFF & JOHNSON, 1980); Ronald Langacker, por seus trabalhos sobre a Gramática Cognitiva (LANGACKER, 1987, 1990, 1991); Leonard Talmy, em seus estudos sobre as estruturas linguísticas formais e as conexões entre tipologia semântica e universal (TALMY, 1988) e Charles Fillmore por seus estudos acerca da Semântica de Frames (FILLMORE, 1975, 1977, 1982, 1985). Também é importante citar os europeus Vyvyan Evans, por seus estudos sobre as estruturas lexicais e a representação do conhecimento (EVANS, 1988; EVANS & GREEN, 2006) e Dirk Geeraerts, que foi um dos primeiros europeus a explorar as possibilidades de um modelo teórico para a categorização (GEERAERTS, 1995, 2003, 2006; GEERAERTS & CUYCKENS, 2010).

A LC nega a tese de que a linguagem corresponde a um sistema autônomo com estrutura própria que independe de outros conhecimentos e que somente a forma linguística é portadora do significado, opondo-se ao estruturalismo e ao gerativismo. Diante disso, para a LC, o uso da linguagem envolve princípios cognitivos e experiências corporificadas, formando diferentes significações de acordo com o uso e o contexto.

Geeraerts (1995) aponta que a LC vê a linguagem integrada com as capacidades cognitivas humanas, afirmando que seus tópicos de maior interesse abarcam: as características estruturais da categorização na linguagem (tais como a prototipicidade, a polissemia, os

⁴⁷ Essa associação fomenta e promove pesquisas em LC desde então. Na página da associação é possível acessar importantes artigos, livros e coleções sobre o assunto, bem como conhecer mais sobre os representantes da área. Mais informações sobre a Associação podem ser acessadas em <http://www.cognitivelinguistics.org/en> (acesso em 27/01/2015).

modelos cognitivos, metáforas e imagens mentais); os princípios funcionais da organização linguística; a interface entre sintaxe e semântica; e a pragmática, como responsável por estabelecer a relação entre linguagem e pensamento através do uso.

Na visão da LC, a linguagem é vista como uma forma de organizar, processar e expressar informações com base no conhecimento de mundo. Tal assertiva fica mais clara com as palavras de Geeraerts & Cuyckens (2010, p. 5), ao afirmarem que a LC “é o estudo da linguagem em sua função cognitiva, onde cognitiva se refere ao papel crucial de intermediar estruturas informacionais em nossos encontros com o mundo.” Essa visão representa uma virada substancial em relação aos paradigmas anteriores.

Seguindo essa linha, Geeraerts (2006) sustenta que a LC ainda traz mais pontos importantes para a análise da linguagem: a excelência da semântica para a análise linguística, o significado linguístico visto sob enfoque enciclopédico, o significado como corporificado e o fato de que o significado linguístico apresenta-se sob diferentes perspectivas.

Na LC considera-se que o contexto orienta a construção do significado e a categorização linguística, opondo-se à ideia de que temos um léxico mental (conhecimento de dicionário) que contém todo o nosso conhecimento semântico dissociado dos outros saberes. Como afirma Langacker (1987), não há modo de fazer uma distinção entre semântica e pragmática, uma vez que isso resulta em um significado artificial; a concepção exequível para a semântica é aquela de natureza enciclopédica, que evita falsas dicotomias. O conhecimento enciclopédico é caracterizado pela LC como um sistema organizado e estruturado em forma de rede, levando em consideração os diferentes aspectos do conhecimento relacionados às palavras.

Geeraerts (2006) destaca ainda outro aspecto relevante na LC: a linguagem não pode ser vista sob uma visão objetivista, considerando que há uma única percepção correta para o significado, mas sim deve levar em conta a perspectiva da qual se vê o mundo e se constrói o significado, assim, a linguagem expressa uma visão do mundo que sofre influência da cultura, da vivência e das necessidades do indivíduo. Complementando essa ideia, Lakoff (1987) baliza que nossas experiências, vivências, cultura e posição no mundo estruturam nosso raciocínio e determinam a nossa compreensão e composição dos significados.

É através do nosso conhecimento enciclopédico que categorizamos as coisas no mundo e organizamos nosso léxico mental formando uma rede que nos permite compreender os significados das expressões linguísticas. Na LC, seguindo as ideias fundamentadas em

Roch (1999), a categorização corresponde a uma dimensão que constitui a estrutura da linguagem, na qual cada categoria é vista como um grupo de elementos que apresentam características similares, ou seja, são agrupados por semelhança, tendo características mais ou menos salientes, sendo que seus limites são frequentemente imprecisos. Essa proposta de categorização segue o rumo contrário da Semântica Lógica, que prevê um conjunto de condições necessárias e suficientes para que determinado elemento se encaixe em uma categoria.

A teoria do protótipo, como é conhecida esta concepção de categorização, teve início nos estudos psicolinguísticos de Eleanor Rosch e seus seguidores, quando estudaram as formas de categorizar as cores, aves e frutos e outras classes de entidades. Essa linha de investigação suscitou interesse no âmbito da Psicologia e da Inteligência Artificial (MEDIN & SMITH, 1984; NEISSER et al, 1996), cujas pesquisas seguiam a premissa de elaborar modelos formais para a memória conceptual humana.

Sintetizando os pontos principais da LC, podemos afirmar que se trata de uma coletânea de abordagens teóricas e metodológicas que, mesmo apresentando pontos divergentes em vários aspectos, se unem em um ponto central: a linguagem está ligada ao sistema cognitivo humano, não se caracterizando como uma faculdade modular e geneticamente determinada, mas contextualizada e integrada às demais faculdades cognitivas. Sendo assim, as habilidades cognitivas envolvidas na experiência humana também estão implicadas na linguagem e são os parâmetros para a sua estruturação.

Nesta seção apresentamos a LC e uma base de seus princípios objetivando ilustrar o escopo no qual se insere a Semântica de Frames, que compartilha a ideia central da miríade de paradigmas teóricos que a embasa, ou seja, a linguagem parte do conhecimento do falante tendo como base suas experiências e vivências, formando uma rede de significados através de modelos cognitivos ou *frames*. A Semântica de Frames é foco de nosso estudo, uma vez que a consideramos uma teoria capaz de abarcar a descrição dos significados do domínio jurídico permitindo uma melhor descrição e formalização da ontologia.

3.2 Semântica de Frames como contributo teórico

Nesta seção temos o intuito de tratar da teoria semântica que embasa esta tese, a Semântica de Frames (SF). Para isso, cumpre-se importante apresentar suas origens e ligação

com a Computação, bem como tecer esclarecimentos acerca do termo *frame*, usado como base na abordagem de Charles J. Fillmore (1975, 1977, 1982).

É oportuno rever brevemente o histórico dos estudos semânticos na área da Linguística para melhor compreender o lugar da SF neste arcabouço teórico. A **semântica** passou a ter relevância nos estudos linguísticos a partir da década de 70, pois, até então, as abordagens não apresentavam avanços em relação à descrição do significado das palavras. No estruturalismo, por exemplo, o significado era tratado pela análise componencial, ou seja, através de uma lista de condições que deveriam ser preenchidas para o estabelecimento do significado da palavra. Algumas linhas mais formalistas, consideravam o significado como uma informação intuitiva e que, por isso, não poderia ser descrito.

Abrindo um espaço de reflexão sobre a aproximação da semântica aos estudos sobre ontologias, cumpre-nos atentar ao fato de que os aspectos advindos do estruturalismo, como a análise componencial, ainda são vigentes em diversos trabalhos relacionados às ontologias nas áreas da CC e da CI, sem considerar o significado de forma totalitária, tendo a Semântica Formal como principal fonte de organização do significado. A semântica, sob um viés formal, é apresentada e tratada amplamente nos trabalhos relativos às ontologias nas áreas da CI e da CC, mas o significado propriamente dito é ainda um desafio para os ontologistas. Este é o ponto que nos faz buscar a SF como subsídio para propor uma forma de potencializar a descrição dos significados na estrutura ontológica.

O formalismo demasiado das abordagens semânticas instigou outras visões para tratar do significado, dentre elas, uma compatível com o gerativismo desenvolvida por Katz e Fodor em 1963. Mais tarde iniciou-se o movimento da Semântica Gerativa por Chomsky e seus alunos, trazendo uma oposição ao tratamento demasiadamente formal e sintético da linguagem. Complementando o cenário cultural, na psicologia, os estudos cognitivos também avançam tratando sobre a categorização e levando em conta os limites pouco definidos dos conceitos. E é justamente neste cenário de valorização e reflexão acerca do significado que, na década de 70, Charles J. Fillmore dá início a seus estudos sobre a Semântica de Frames, trazendo um novo paradigma para tratar do significado.

Fillmore buscou inspiração nos trabalhos de Minsky (1974) para adotar o termo *frame* como denominação para um sistema de conceitos relacionados de modo que o entendimento de um depende do entendimento de todo o sistema no qual ele está inserido. Cabe aqui, em função da interface computacional da pesquisa, apresentar o entendimento de Minsky acerca do termo *frame* tendo como objetivo mostrar a ligação entre as áreas, até mesmo porque, o

termo *frame* é também utilizado em trabalhos sobre ontologias e consideramos importante distinguir seus diferentes usos.

Na área da Computação, mais especificamente na IA, o termo *frame* é utilizado para referir-se à forma como dados se estruturam de modo a representar uma determinada situação, ou seja, compreendem um conjunto de informações sobre uma situação, que pode ser organizada através de propriedades (*slots*) que caracterizam cada circunstância. Ainda com base em Minsky (1974) e considerando os aspectos formais, a IA compreende *frame* como uma estrutura formada por **atributos, valores, relações e restrições** sobre os elementos envolvidos em cada situação. Para a Computação, em seus trabalhos voltados para a IA, os *frames* consistem em um conjunto de nós ou relações que representam características prototípicas de objetos ou situações, que podem ser alteradas de acordo com mudanças no mundo real, uma vez que eles contêm informações universais e particulares de uma comunidade.

Para Minsky (1974), que discorda da posição de representação do conhecimento através de fórmulas lógicas e defende que haja uma estrutura de dados para armazenar as informações relevantes sobre cada entidade, o significado deve ser organizado por meio de um conjunto de dados estruturados que representem situações típicas; segundo ele, tais estruturas correspondem aos *frames*⁴⁹. Os pressupostos do autor servem como base para diversos trabalhos na área da organização do conhecimento, sendo que o cerne de seus estudos concentra-se na hipótese de que a memória humana possui muitas propriedades típicas de situações ou objetos (estereótipos). Por exemplo: quando alguém fala em **estudar**, pode-se diretamente ligar esse item lexical a **aula, escola, professor** e até mesmo **livros**.

Minsky (1974) apresenta uma ligação dos *frames* ao PLN, uma vez que eles podem trazer mais informações sobre como deve ser utilizado (compreendido) certo conceito, prevendo uma série de ações ou expectativas, isto é, em um *frame* pode-se ter muitas informações que vão além da situação, advindas das representações do conhecimento do falante. Neste contexto, as palavras são componentes que compartilham *frames*, mas não estão necessariamente inter-relacionadas.

⁴⁹ Cabe aqui uma nota esclarecedora para o uso do termo *frame* no decorrer do trabalho, uma vez que ele será usado em várias situações, ora referindo-se aos *frames* relacionados à SF, ora aos referidos por Minsky. Por esta razão, consideramos importante adotar uma forma de diferenciação: sempre que fizermos referência ao termo *frame* na visão da SF usaremos o termo tão somente e, quando, a referência for de cunho computacional, acrescentaremos a referência a Minsky (1974) para evidenciar a diferença de significado.

Vemos aqui uma estreita ligação entre os *frames* propostos por Minsky (1974) e a estrutura das ontologias. Podemos afirmar que cada classe da ontologia com seus atributos, propriedades e axiomas corresponde a um *frame* segundo a visão de Minsky (1974). No entanto esta visão de *frame*, ainda que muito adequada em nível computacional, não corresponde ao que buscamos como contributo teórico na SF. Nosso intuito é promover a ligação das estruturas computacionais (*frames* de Minsky) às estruturas cognitivas (*frames* de Fillmore) aprimorando a gama de relações para cada conceito.

Fillmore (1975, 1977, 1982) inspira-se nos pressupostos de Minsky (1974) e traz para a LC o conceito de *frame*, adaptando-o de modo a corroborar com a ideia principal da abordagem teórica: um modelo diferenciado para a construção dos significados levando em consideração os elementos sintáticos e situacionais de um enunciado, a Semântica de Frames.

Inicialmente Fillmore (1975) apresenta os termos **cena** e *frame*. O primeiro, para se referir ao conjunto de cenários definidos culturalmente, relacionados a experiências e transações interpessoais, imagens corporais, crenças, experiências humanas e imagens, indo além do aspecto visual – uma estrutura cognitiva. Já o segundo, correspondendo ao conjunto de palavras, regras ou aspectos gramaticais relacionadas a uma determinada cena típica – uma estrutura linguística. Desta forma, o fato de conhecer a cena permite ao falante a escolha adequada das palavras e o conhecimento sobre estas palavras possibilita acessar a estrutura cognitiva construindo o significado. Com esta primeira proposição, Fillmore (1975) assinala que há uma diferença entre a estrutura cognitiva e a linguística, porém baliza que ambas são interdependentes para a construção do significado.

Seguindo seus estudos, Fillmore (1976) considera o **contexto** como parte da construção do significado. Assim define que o significado da palavra depende também do contexto no qual o enunciado está inserido, devendo este aspecto ser parte da teoria semântica. Esta constatação levou o autor a considerar que existem *frames interacionais* e *frames cognitivos*. Estes representam o conhecimento do falante em relação ao evento, seus participantes e diferentes papéis envolvidos (na cena), sendo que as escolhas linguísticas e o entendimento do significado dependem do conhecimento da cena. Enquanto aqueles correspondem às palavras escolhidas pelo falante adequadas a determinada situação ou cena. Por exemplo: perante um juiz ou em um documento jurídico o advogado usará o termo **aresto**, no entanto, ao conversar com seu cliente leigo poderia usar o termo **decisão do Tribunal**.

Mais tarde, em Fillmore (1982) define-se a Semântica de Frames. O autor a considera uma abordagem de pesquisa em linguística que segue uma metodologia empírica e que apresenta uma estrutura descritiva para o registro dos resultados da pesquisa. Segundo Fillmore a “Semântica de Frames oferece uma forma particular de olhar para o significado das palavras, bem como uma forma de caracterizar princípios para a criação de novos sintagmas e palavras” e, o autor complementa dizendo que desta forma é possível “adicionar novos sentidos às palavras e reunir os significados dos elementos em um texto no significado total do texto” (FILLMORE, 1982, p. 111).

A partir destas reflexões, Fillmore define o *frame* como um “sistema de conceitos relacionados de tal forma que para entender qualquer um deles você tem que entender a toda a estrutura na qual ele se encaixa” (FILLMORE, 1982, p. 111). Segundo o autor, quando há introdução de um determinado conceito em um ato de fala, um texto ou enunciado seguindo a estrutura dos *frames*, todos os outros conceitos relacionados serão ativados para a construção do significado. Seguindo nesta linha, a divisão entre cenas – como estruturas cognitivas – e *frames* – como estruturas linguísticas – é abandonada, uma vez que a concepção de *frame* é tratada como uma estrutura de **conceitos inter-relacionados** que abrange tanto o cognitivo quanto o linguístico. Nesta visão as palavras correspondem à categorias cognitivas, aproximando a teoria da SF às vertentes cognitivas da Linguística.

De acordo com Fillmore (1985), a pesquisa em SF é declaradamente empírica uma vez que busca encontrar a maneira como as expressões se encaixam em cada situação, considerando o contexto. Este aspecto relacionado à metodologia da SF, considerando-a empírica e não formal, nos faz refletir acerca da aproximação que pretendemos às ontologias, uma vez que estas são explicitamente formais. No nosso entendimento a organização da estrutura de *frames* através de uma metodologia empírica não prejudica a organização formal da ontologia, uma vez que ambas estruturas se complementam, ou seja, os *frames* construídos de forma empírica potencializarão os significados na estrutura formal da ontologia e também terão seu conhecimento formalizado a fim de suprir interesses relacionados à recuperação da informação.

O objetivo da SF é esclarecer os diferentes sentidos (variações) e as valências (funções sintáticas e semânticas) de itens lexicais, destacando a forma como cada conceito é organizado e representado na mente dos sujeitos. Reforçando essa ideia, fazemos uso das palavras de Croft (2009, p. 1) “a SF é uma teoria de como os conceitos são organizados e representados na mente”.

A SF leva em conta fatores culturais e situacionais para descrever a estrutura cognitiva de um evento, ou seja, considera o chamado conhecimento enciclopédico. Tem como base a multiplicidade de significados (SILVA, 1997) atribuídos a uma palavra, de acordo com a experiência humana. Segundo Croft e Cruse, “Fillmore utiliza-se de uma ampla gama de exemplos para demonstrar que existem fenômenos significativos da semântica, os quais não podem ser facilmente capturados em um modelo de semântica estrutural e traços semânticos” (CROFT e CRUSE, 2004, p. 8).

De acordo com Petrucci (1996), a SF ressalta o *continuum* entre linguagem e experiência considerando o *frame* como um sistema de conceitos inter-relacionados no qual, para compreender o significado de um conceito é necessário conhecer o contexto em que ele se insere. Ohara e Nikiforidou (2009) reforçam esta premissa, afirmando que a dicotomia existente entre pragmática e semântica é descartada pela SF uma vez que o significado é analisado conforme os aspectos sociais e do discurso estabelecendo um *continuum* para a construção do significado.

Segundo Fillmore (2006), na SF a compreensão de um **conceito** não se dá de forma isolada, mas sim a partir da sua posição em uma rede de conceitos (um sistema de conceitos), ou seja, para saber o significado de uma palavra implica saber, evocar e cena a que esta palavra remete. Seguindo esta abordagem, as palavras **evocam** o conhecimento organizado através da estrutura cognitiva do *frame*. Esta estrutura que permite a organização dos conceitos em um **sistema de conceitos inter-relacionados** possibilita à ontologia uma forma diferenciada de descrever o domínio, ou seja, permite que as ligações entre os **conceitos** incluam na estrutura informações sobre a cena à qual remetem, delimitando os papéis dos participantes.

Nessa abordagem, *frame* é um conceito fundamental que corresponde ao arcabouço conceitual presente na memória advindo de nossa interação com o mundo e da forma como solidificamos nossas vivências e conhecimentos. Fillmore (1982) considera *frame* uma construção intuitiva que possibilita a formalização de relações entre semântica e sintaxe como decorrência de uma análise lexical. Também poderíamos dizer que um *frame* é uma estrutura de conhecimento que ocorre a partir das interações refletidas pela linguagem.

Na SF o *frame* tem um papel fundamental da definição do significado, ou seja, o significado das palavras está diretamente ligado ao *frame* ao qual elas estão relacionadas. Assim, a interpretação de uma palavra ou enunciado requer que sejam acionadas estruturas de conhecimento que relacionam os **conceitos** com base nas diferentes cenas da experiência

humana. O exemplo clássico apresentado por Fillmore (1982) é relacionado aos verbos **vender, comprar, pagar, gastar, custar e cobrar**. Segundo o autor, para compreender o significado destes verbos é necessário ter conhecimento do *frame* *TRANSAÇÃO COMERCIAL*⁵⁰. O exemplo também serve para explicar o ponto de vista perspectivado, outro princípio da LC, ou seja, a transação comercial pode ser vista sob a perspectiva do comprador (que quer comprar algo) ou do vendedor (que quer vender algo).

Moreira e Salomão (2012) destacam outro aspecto relevante em relação à SF que é a ampliação dos papéis temáticos tradicionais (agente, paciente, tema, etc), os quais assumem funções temáticas específicas de acordo com o *frame*. Assim, cada situação descrita por um *frame* compreende um conjunto diferenciado de funções temáticas que melhor descrevem a cena, como pode ser exemplificado através do *frame* *TRANSAÇÃO COMERCIAL*, que tem como papéis centrais o **comprador** e o **bem**, enquanto que, na abordagem de papéis temáticos, os papéis seriam **agente** e **tema**. Observa-se que, com esta ampliação de papéis de acordo com a situação, o significado pode ser melhor explicitado uma vez que **agente** e **tema** são papéis genéricos e que se aplicam a diferentes participantes, enquanto que **comprador** e **bem** explicitam a função de cada um na cena.

Seguindo esta premissa, podemos usar o exemplo do evento **juízo**, no qual o juiz assume diferentes papéis de acordo com o contexto: ele pode assumir o papel de **juiz**, de **sentenciador**, de **analizador**, entre outros. Todos estes papéis, seguindo a visão dos papéis temáticos, se enquadrariam na categoria **agente**; porém, como se pode ver, o juiz desempenha diferentes papéis no decorrer do evento **juízo** e não há como categorizá-lo somente como **agente**, já que cada um dos papéis assumidos corresponde a diferentes formas de participação no evento.

No caso da nossa pesquisa, o conceito de *frame* tem fundamental relevância, uma vez que permitirá uma descrição ampla das cenas jurídicas envolvidas no Procedimento Sumaríssimo. Além disso, a noção de *frame* também será útil na descrição das diferenças de significado relacionadas ao domínio social de uso dos **conceitos**. Podemos citar como exemplo as palavras **inocente** e **culpado**, que, no contexto jurídico, estarão relacionadas ao *frame* *JULGAMENTO*, porém, no contexto social, apenas relacionam-se a alguém que cometeu um determinado delito.

⁵⁰ Para destacar a nomenclatura dos frames, faremos uso da notação *CAIXA ALTA ITÁLICO*.

A inclusão dos *frames* para a determinação dos significados jurídicos será de grande valia para a descrição do domínio na ontologia, pois, no Direito, assim como em outras áreas especializadas, se desenvolve uma linguagem particular e específica, na qual várias expressões da linguagem comum assumem diferentes acepções, adquirindo novos significados dentro do contexto jurídico. Nesse contexto, os profissionais fazem uso de diversas palavras e expressões polissêmicas, como é o caso do termo **citação**, que se refere tanto ao **evento** de citar alguém acerca de um processo, como ao **documento** propriamente dito.

O *frame* não só descreve uma cena, como também destaca seus participantes, que, de acordo com a teoria, são denominados **elementos de frame**. Os elementos de frame contribuem na descrição da estrutura do *frame*, representando os papéis que podem ser assumidos – situacionais ou semânticos – e constituindo-se como unidades básicas para o entendimento do *frame*. É justamente neste aspecto que há uma importante contribuição da SF para com a ontologia do domínio jurídico que nos propomos a desenvolver: os elementos de *frame* correspondem aos diferentes papéis que os participantes assumem nos eventos, por esta razão, serão úteis na descrição do conhecimento jurídico na ontologia.

O próprio Fillmore (1982), ao propor uma ampliação dos papéis semânticos visando à descrição dos *frames*, tomou como exemplo a área jurídica, destacando que, nesta área, não há apenas um conjunto de termos relacionados, mas sim um vocabulário especializado que implica o entendimento das palavras em um contexto, ou seja, o entendimento das cenas requer o conhecimento acerca dos envolvidos e do contexto no qual estão inseridos. O exemplo dado pelo autor é em relação à descrição dos verbos **culpar**, **acusar** e **criticar**, todos relacionados ao evento **juízo**. Sua análise demonstrou que os verbos demandam uma representação distinta, ainda que todos se relacionem a um mesmo evento.

A base teórica da SF deu origem a um recurso lexical que descreve e organiza os diferentes *frames*, destacamos a seção seguinte para apresentar este léxico computacional, trazendo conceituações e a sua estruturação a fim de complementar o arcabouço teórico uma vez que poderemos aproveitar estruturas já existentes para a organização da ontologia.

3.3 Recurso lexical baseado em *frames*: *FrameNet*

Estudos sobre Semântica de Frames inspiraram a criação do léxico computacional (ou recurso lexical) *FrameNet* (RUPPENHOFER et al., 2010), um projeto sediado no *Instituto Internacional da Ciência da Computação de Berkeley*, que consiste em um repositório com mais de 1200 *frames*, sendo interligados por aproximadamente 1675 relações e contando com um acervo de 13000 unidades lexicais, documentadas com cerca de 200.000 anotações manuais (ICSI, 2012; BAKER, 2014). Conforme Baker (2014), a FN foi lançada em 1997 como um projeto que buscava produzir uma base lexical para o inglês que fosse legível por máquinas e baseada na teoria da SF, considerada uma aplicação para essa teoria. Sua construção foi amparada em *corpus*, estabelecendo uma relação entre a abordagem de *frames* e a Lexicografia, sendo fruto do trabalho conjunto entre Fillmore e Atkins.

No contexto da FN, de acordo com Ruppenhofer et al. (2006), o conceito de *frame* corresponde a uma estrutura conceitual que descreve uma determinada situação, objeto ou evento reunindo seus participantes e demais elementos. Os *frames* correspondem à generalizações relacionadas a palavras agrupadas por temas semelhantes, as quais podem compartilhar papéis semânticos e até mesmo sintáticos.

Baker (2014) apresenta a terminologia envolvida na FN, salientando os principais aspectos. Segundo o autor, os diferentes papéis relacionados aos *frames*, são chamados de **elementos de frame** (EF) e as palavras que evocam cada *frame* são conhecidas como **unidades lexicais** (UL's). Baker aponta que cada UL corresponde aos **signos** nas palavras de Saussure, sendo a associação entre o **significante** e o **significado**. A forma está relacionada aos lemas, enquanto que o significado se estabelece através dos *frames*. Cada UL na FN equivale a um único sentido de uma palavra e, caso um lema tenha mais de um sentido, estará associado a mais de um *frame*.

De acordo com Ruppenhofer et al. (2010), uma UL é a associação de uma palavra com seu significado, sendo que cada palavra polissêmica pertence a um *frame* diferente. Desta forma, cada significado corresponde a uma estrutura distinta de informações para descrever uma determinada situação de acordo com a estrutura dos *frames*, esta estrutura envolve um objeto ou evento e seus elementos participantes.

A FN tem como objetivo principal identificar padrões de valência para verbos, substantivos, adjetivos, advérbios e preposições (somente na língua inglesa), através de

anotações e da exemplificação com sentenças reais. Além disso, tendo como base as noções de *frame* discutidas por Charles Fillmore, a FN também objetiva: (a) descrever as UL's relacionando-as aos *frames* que evocam, sendo estes também descritos com base nos EF; (b) utilizar um *corpus* para autenticação, validação e exemplificação das descrições efetuadas; (c) anotar as camadas identificando os EF, seus tipos sintagmáticos e suas funções gramaticais visando estabelecer possíveis combinações entre as UL's; (d) representar padrões de valência de cada UL; e (e) verificar e determinar as relações entre os diferentes *frames*.

Neste recurso lexical, os **elementos de *frame*** das situações podem ser organizados em nucleares (*Core*) e não nucleares (*Non-Core*), noções estas que estão fortemente ligadas à visão dos papéis semânticos que ocupam posições argumentais (participantes) ou não (não participantes). Os nucleares são os **elementos de *frame*** fundamentais para a caracterização do *frame* e se manifestam até mesmo na estrutura argumental traduzida pelo verbo em questão (predicador). Já os não nucleares não são tão importantes na conceituação, uma vez que não caracterizam o *frame* sozinhos. Na figura 6 podemos visualizar o *frame* *CRIMINAL_INVESTIGATION* e seus elementos de *frame*, bem como os tipos semânticos relacionados a cada elemento.

Criminal_investigation

[Lexical Unit Index](#)

Definition:

This frame describes the process that involves the determination by an authority, the **Investigator**, of the circumstances surrounding an **Incident** by means of inquiry.
 The police are **INVESTIGATING** the bank's reported loss of several thousand dollars from their downtown branch's top security vault.

Authorities are **INVESTIGATING** several murders.

FEs:

Core:

Incident [Inc]

The **Investigator** attempts to determine the exact circumstances surrounding the **Incident**, which is typically a criminal act (murder, robbery) or the result of such an act (death, loss of money).

Investigator [Inv]

The **Investigator** is the person or agency that investigates a person of a crime.

Semantic Type: Sentient

Suspect [Sus]

This FE denotes the **Suspect** that is investigated by the **Investigator**.

Semantic Type: Goal

Requires: Incident

The police are **INVESTIGATING** Dr. Phil.

Non-Core:

Duration [Dur]

The length of the investigation.

Semantic Type: Duration

The police carried out a **two-year INVESTIGATION** into terrorist kidnappings.

Time [Time]

When the **Investigator** investigates the **Incident**,

Semantic Type: Time

Our guys are **currently INVESTIGATING** her disappearance.

Figura 6: Frame *CRIMINAL_INVESTIGATION*.

Fonte: Disponível em:

<https://framenet2.icsi.berkeley.edu/fnReports/data/frameIndex.xml?frame=Criminal_investigation> Acesso em: 04/05/2015.

A FN organiza os diferentes *frames* através das relações entre *frames*, a saber: *inherits_from* (herda), *is_inherited_by* (é herdada por), *perspective_on* (perspectiva sobre), *is_perspectivized_in* (é perspectivada por), *uses* (usa), *is_used_by* (é usado por), *subframe_of* (subframe de), *has_subframe/s* (tem subframe/s), *precedes* (precede), *is_preceded_by* (é precedido por), e *see_also* (veja também).

O *FrameNet II: Extended theory and practice*⁵¹, livro que documenta a construção da FN, apresenta ainda outras duas relações entre *frames*: *is_inchoative_of* (é incoativa de) e *is_causative_of* (é causador de), estas referem-se a estados e causas, porém ainda não foram implementadas consistentemente na FN (RUPPENHOFER et al., 2010).

Conforme explicam Fillmore, Johnson e Petruck (2003), a relação de **herança** (*inherits_from*, *is_inherited_by*) implica que *frames* mais específicos herdam todas as características de um *frame* mais geral. Um *frame* pode herdar propriedades de outro *frame*, acrescentando-lhe detalhes. Assim como na relação **é-um** em que a classe-filho herda características da classe-pai, mas tem suas peculiaridades, os *frames* também podem herdar algumas características e incluir novas na sua descrição. Um *frame* mais simples pode herdar particularidades de um *frame* mais complexo e vice e versa.

A figura 7 ilustra a relação de herança (marcada pelas setas mais largas) entre os *frames* *EVENT*, *INTENTIONALLY_ACT*, *RULIN_LEGALLY* e *VEREDICT*, sendo que o último herda características de todos os demais; assim, para que seja um *VEREDICT* tem que ter as características de evento *INTENTIONALLY_ACT*⁵².

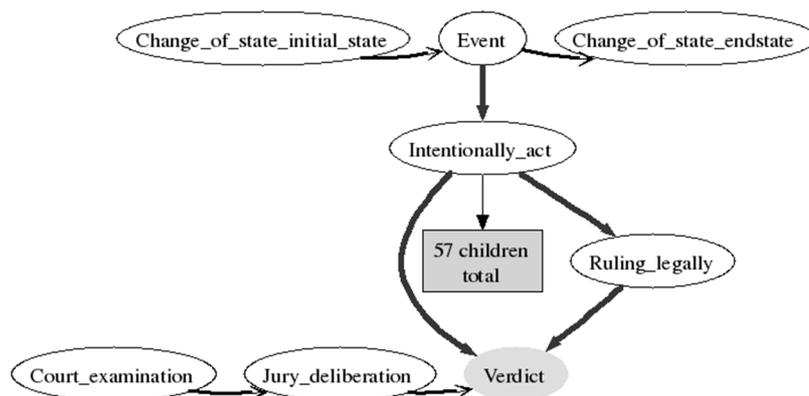


Figura 7: Relação de **herança** entre *frames*.

⁵¹ Também conhecido como *The Book*.

⁵² É importante esclarecer que a FrameNet descreve a linguagem geral, não a especializada como é o caso do Direito. A FN descreve apenas o frame *CRIMINAL_PROCESS* e, por esta razão apresentaremos os exemplos nesse escopo.

Fonte: Disponível em: <<https://framenet.icsi.berkeley.edu/fndrupal/FrameGrapher>> Acesso em: 19/08/2013.

A relação de herança entre *frame* e *subframe* equivale à relação de herança entre classes e subclasses na ontologia, sendo fundamental na organização hierárquica (taxonômica) das estruturas, tanto na FN quanto nas ontologias. Esta é uma relação que será incluída na ontologia de forma natural ao criar-se a estrutura taxonômica.

A relação de **perspectiva** (*perspective_on*, *is_perspectivized_in*) está relacionada a uma visão particular em um determinado *frame*. Um exemplo clássico é o *frame* de Transação comercial – pode ser visto pela perspectiva do vendedor ou do comprador. Ambas as cenas se referem ao *frame* *TRANSAÇÃO_COMERCIAL*, porém sob diferentes perspectivas.

Na figura 8 temos o exemplo da relação de perspectiva (marcada pela seta cinza) entre os *frames* *TRIAL* e *TRY_DEFENDANT*, ou seja, o réu tem uma perspectiva sobre o julgamento e julgar um réu exige que haja um julgamento. Assim, percebe-se que a relação de perspectiva estabelece uma inter-relação entre os eventos o que é muito comum na área jurídica, uma vez que os eventos se inter-relacionam e podem ser vistos de acordo com a perspectiva de cada participante.

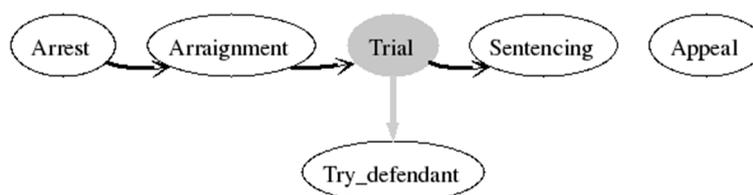


Figura 8: Relação de **perspectiva** entre *frames*.

Fonte: Disponível em: <<https://framenet.icsi.berkeley.edu/fndrupal/FrameGrapher>> Acesso em: 19/08/2013.

A relação de **usa** ou **é usado por** (*uses*, *is_used_by*) entre os *frames* pressupõe um *frame-pai* como plano de fundo para um *frame-filho*, ou seja, corresponde à cena maior na qual o *frame-filho* está inserido. A figura 9 elucida a ideia de um *frame usar* outro *frame*: para que haja um veredito é necessário que haja uma comunicação, logo o primeiro **usa** o segundo (a relação é representada pelas setas pontilhadas). Neste caso o *frame* *VEREDICT* usa algumas características do *frame* *COMMUNICATION*, mas não herda todas as suas características. Eles estão apenas relacionados por terem alguma semelhança.

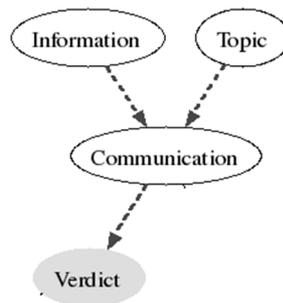


Figura 9: Relação de **usa** entre *frames*.

Fonte: Disponível em: <<https://framenet.icsi.berkeley.edu/fndrupal/FrameGrapher>> Acesso em: 19/08/2013.

Na relação de **subframe** (*subframe_of, has_subframe/s*), o frame-filho é um subevento do frame-pai, ou seja, é um tipo de evento representado pelo *frame* mais complexo. Essa relação assemelha-se à relação **é-um** ou classe/subclasse das ontologias. A relação de subframes pode ser observada na figura 10 pela representação das setas pontilhadas.

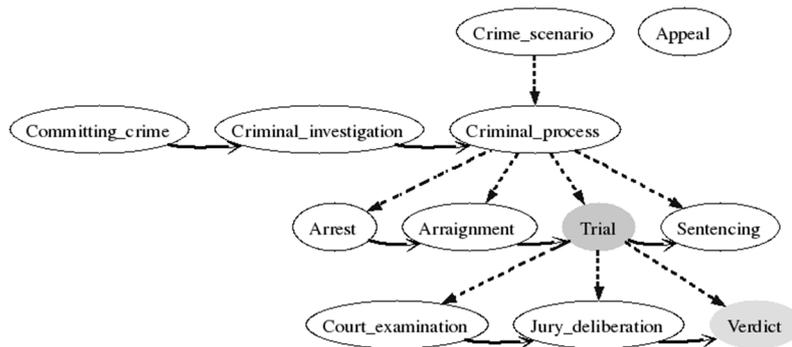


Figura 10: Relação entre *frames* e *subframes*.

Fonte: Disponível em: <<https://framenet.icsi.berkeley.edu/fndrupal/FrameGrapher>> Acesso em: 19/08/2013.

Como pode ser visto na figura 10, o *frame* **CRIMINAL_PROCESS** é um *subframe* de **CRIME_SCENARIO**. **ARREST**, **ARRAIGNMENT**, **TRIAL** e **SENTENCING** são *subframes* de **CRIMINAL_PROCESS** e **TRIAL** também tem *subframes*. Dessa forma há uma organização hierárquica entre os *frames*, novamente em consonância com a estrutura das ontologias.

Na figura 10 também está expressa a relação de **precedência** (*precedes, is_preceded_by*) através das setas contínuas. Conforme Petruck e De Melo (2012), essa relação acontece apenas entre *frames* pertencentes a um mesmo quadro, especificando a sequência em que os eventos ocorrem, ou seja, há uma ordem de acontecimentos que precisa

ser seguida. Como se pode ver na figura 10 antes de haver a **sentença** é necessário que haja o **juízo**.

A relação de **precedência** necessariamente precisa ser tratada pela ontologia, pois corresponde a uma importante organização da área jurídica, uma vez que os eventos ocorrem em uma sequência e interdependência.

A relação **veja também** (*see_also*) é usada em casos em que há grupos de *frames* semelhantes, mas devem ser diferenciados (comparados ou contrastados) de forma criteriosa e cuidadosa, ou seja, haverá uma definição no *frame* mais representativo que deverá contrastar com o outro, estabelecendo os limites pretendidos por eles. Essa relação não implica nenhuma relação direta entre os elementos de *frame*.

A FN apresenta diferentes aplicações e usos, conforme Baker (2014). Desde situações para melhorar seu desempenho e acesso até aplicações para ampliar seu potencial. Listamos, a seguir, alguns exemplos apontados por Baker (2014) de ampliação e aplicabilidade da FN em outras pesquisas:

- Trabalhos de rotulagem automática de papéis semânticos (ASRL) desenvolvidos inicialmente por Gildea e Jurafsky (2002). Posteriormente outros sistemas de rotulagem automática também foram desenvolvidos, como é o caso dos trabalhos de Erk e Pado (2006), Johansson e Nugues (2007) e Das et al. (2013).
- Desenvolvimento de softwares para a inclusão de novas UL's em *frames* já existentes (GREEN, 2004).
- Aplicativos que permitem extração de informação através de perguntas e respostas em trabalhos desenvolvidos por Mohit e Narayanan (2003) e Sinha (2008).
- Os dados da FN permitiram melhorias nos sistemas de recuperação da informação, em trabalho desenvolvido por Burchardt (2008).

Além disso, Baker (2014) salienta que a estrutura da FN também possibilitou melhorias nos estudos relacionados à análise de sentimentos, na tradução automática e no reconhecimento de eventos.

A FN não representa vocabulário especializado, conforme já referimos em nota, e tais descrições são destinadas a especialistas dos domínios, assim também os nomes próprios não fazem parte do arcabouço da FN (tais nomes são conhecidos pelo PLN como entidades nomeadas). Porém, existem iniciativas de uso da FN e da SF para a criação de léxicos

especializados, como é o caso do trabalho desenvolvido por Thomas Schmidt que criou um recurso computacional baseado na SF para o domínio do futebol – o Kictionary, em inglês, alemão e francês (Schmidt, 2008) e dos estudos para a criação de um léxico jurídico para o italiano, desenvolvido por Venturi et al. (2009).

Realizando uma busca simples no Google Acadêmico retornam centenas de trabalhos que relacionam a FN às ontologias ou abordam o uso da FN em sistemas de recuperação da informação, dentre os quais, destacamos alguns para exemplificar a gama de possibilidades de trabalhos relacionados a estas estruturas de organização do conhecimento e procuramos refletir acerca de semelhanças e diferenças em relação à esta pesquisa.

Um dos trabalhos de destaque foi iniciado por BĀRZDIŅŠ et al. (2008), no qual os autores propõem uma maneira de formalizar a FN através de uma ontologia multidimensional⁵³, levando em conta a herança entre os *frames* e os tipos semânticos permitindo o enriquecimento da anotação de textos e a identificação das anáforas. Através desta estrutura multidimensional, segundo os autores, é possível incluir as valências, as ligações entre as UL's e EF em frases anotadas na FN seguindo a Lógica de Primeira Ordem.

Esta é, no nosso entendimento, uma abordagem que transforma a FN e suas diferentes camadas em uma ontologia. Esta abordagem difere da nossa em todos os aspectos. Não faremos uso da FN propriamente dita para a organização da ontologia, apenas seguiremos seus pressupostos. A organização que prevemos para as classes e subclasses, bem como os relacionamento não será expressa em Lógica de Primeira Ordem em função de estarmos trabalhando na interface linguística da ontologia. E, por fim, não pretendemos incluir todas as camadas da FN – valências, tipos semânticos, anotação – na nossa ontologia, mas sim buscar subsídios neste recurso lexical.

Outro trabalho relacionado à FN que consideramos oportuno destacar é a pesquisa realizada por Venturi (2011). Ainda que não seja relacionado às ontologias, trata-se de uma iniciativa que visa a anotação de textos jurídicos tendo como base a FN. De acordo com a autora, o estudo objetiva mostrar a eficácia da FN a fim de propor novas perspectivas nos estudos relacionados à linguagem jurídica e recuperação da informação. Além disso, a partir de análises estatísticas tendo como base a anotação realizada através da FN, pode-se ampliar o

⁵³ Ontologia Multidimensional é uma abordagem representação do conhecimento espacial. Ontologias comuns (formalizadas em OWL) usam lógica de primeira ordem para definir as classes de indivíduos (objetos) e suas relações binárias (propriedades). Ao introduzir o conceito de uma ontologia multidimensional, estamos argumentando que sejam usadas médias geométricas em um espaço multidimensional através de formalismos baseados em lógica de primeira ordem para definição das classes e suas relações abarcando as camadas da FN (BĀRZDIŅŠ et al., 2008).

processamento automático de textos legais. Novamente trata-se de uma abordagem que diverge da nossa, pois não temos o intuito de anotar *frames* na FN tendo como base os textos jurídicos. No entanto, em caso de haver um sistema semelhante para o domínio jurídico brasileiro, entendemos que esta, junto da ontologia que pretendemos desenvolver, seja uma solução eficaz para o processamento da linguagem jurídica.

Moreira e Salomão (2012) também descrevem um trabalho que relaciona ontologias à estrutura da FN, porém novamente com um viés diferenciado da nossa proposta. As autoras propõem uma análise ontológica para verificar o enquadre de sentenças em diferentes *frames*. Sua proposta sustenta a modificação da metodologia de construção de *frames* na FN, levando em conta a estrutura das ontologias. O trabalho faz uso de três ontologias, considerando que uma apenas não seria capaz de capturar a natureza de todos os objetos expresso pela linguagem.

Até aqui apresentamos os preceitos que embasam a FN a partir da SF e proporcionamos uma resenha acerca de suas estruturas. Cumpre, no entanto, tecer reflexões pertinentes à tese a fim de justificar e elucidar nossas escolhas e pontos de vista.

Inicialmente, reforçamos nossa escolha teórica pela SF para a descrição do conhecimento na ontologia jurídica. Vislumbramos a contribuição da SF na organização do conhecimento jurídico, possibilitando a construção de uma ontologia capaz de abarcar mais fortemente o significado dos termos, fraseologias e itens lexicais. Seguindo a base teórica selecionada para sustentar esta tese, ponderamos que a inclusão de *frames* na estrutura ontológica poderá ser útil permitindo inferências acerca do domínio, estabelecendo maior grau de significação para os conceitos da área e possibilitando uma descrição ampla dos cenários e papéis dos participantes de cada evento, isto porque a área jurídica é fundamentalmente composta por eventos.

A SF abarca todos os aspectos que se relacionam à forma como os seres humanos conceptualizam as informações na sua mente, incluindo uma estrutura ontológica dos conceitos percebidos pelo homem em sua relação com o mundo, ou seja, torna-se uma importante ferramenta para o estabelecimento de relações em ontologias que pretendem descrever um determinado domínio levando em conta informações linguísticas mais aprofundadas, ampliando a significação de cada elemento e, assim, contribuindo fortemente no processamento da linguagem natural e no aprimoramento de sistemas de recuperação da informação.

A FN do inglês, a qual tomamos como base para nossa reflexão, apresenta um *frame* relacionado ao domínio jurídico, ainda que não tome como base a linguagem especializada: o *CRIMINAL_PROCESS*, cuja descrição envolve um evento no qual um suspeito é preso pelas autoridades, acusado como culpado e passa a ser processado como réu. Este *frame* possui diversos *subframes* que foram estudados por Bertoldi (2011) em sua tese de doutorado.

Conforme Bertoldi (2011) proceder a tradução automática dos *frames* do inglês para o português é uma situação delicada e problemática, especialmente porque suas bases advêm de sistemas distintos: a *Common Law*, para o sistema americano e a *Civil Law*, para o brasileiro. Além disso, há uma dificuldade também em relação às diferentes leis existentes em cada país. Contudo, o estudo de Bertoldi (2011) aponta alguns *frames* da FN que correspondem a *subframes* do *CRIMINAL_PROCESS* que podem ser considerados análogos e utilizados em uma *FrameNet* jurídica brasileira, vislumbrando uma possibilidade de tradução.

Se não avaliássemos a premissa da SF de considerar o contexto como elemento constitutivo do *frame*, poderíamos supor o uso dos *frames* elencados na pesquisa de Bertoldi como base para a inclusão na ontologia, apenas ampliando a sua descrição. No entanto, o fato de sua tradução ser possível, não significa que eles correspondam exatamente ao processo Penal Brasileiro relacionado aos juizados especiais. Diante disso, entendemos que a estrutura da FN será útil no sentido de balizar a construção dos *frames* do domínio, levando em conta o arcabouço teórico que a envolve. Neste trabalho, conhecer a FN foi importante para sustentar a organização dos *frames* levando em conta os elementos de *frame*, as unidades lexicais evocadoras e as relações entre *frames*. Salientamos que a construção de *frames* não é um compromisso desta pesquisa, porém é um passo necessário para a organização da ontologia de domínio uma vez que não podemos dispor da FN.

Outro aspecto relacionado à FN que contribui para esta tese são as relações entre *frames*. Dentre as citadas, destacamos como necessárias na descrição do conhecimento jurídico, as seguintes relações entre *frames*: herança, subframe e precedência. As duas primeiras assemelham-se às relações hierárquicas que estruturam a ontologia, correspondendo à organização em classes e subclasses. Enquanto que a outra corresponde a relações não taxonômicas e estarão relacionadas a diferentes *frames*, constituindo-se um desafio para a parte aplicada, uma vez que teremos de encontrar uma forma de formalizá-las e relacioná-las aos respectivos *frames*. A relação de perspectiva, conforme já referido, é também importante na representação do domínio, no entanto, não traz importantes contribuições no que tange à recuperação da informação e, por esta razão, não será analisada e incluída na ontologia.

Concluindo nossas reflexões acerca de Semântica de Frames e da FN reforçamos nossa tese de incluir na ontologia relações não taxonômicas baseadas em *frames* a fim de potencializar a descrição dos conceitos relacionados ao domínio do Processo Penal Brasileiro, provendo, assim, uma estrutura capaz de descrever a referida área de forma mais ampla, buscando minimizar os problemas relacionados às questões de significado para sistemas de recuperação da informação, atentando especialmente à forma de representação dos diferentes papéis de cada participante.

3.4 Esquema síntese do capítulo

A figura 11 apresenta um esquema síntese das ideias apresentadas no capítulo.

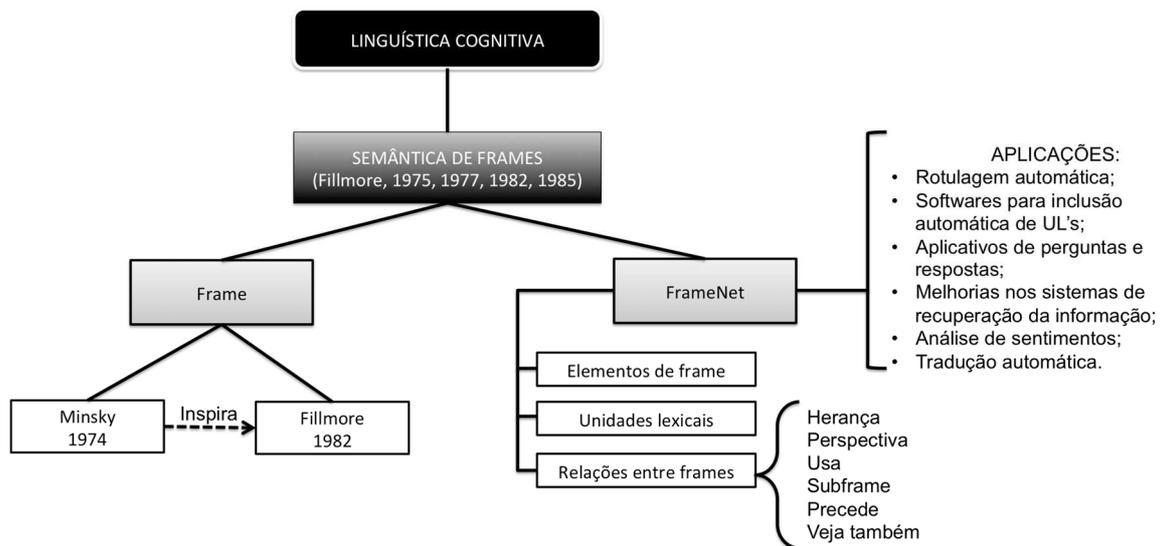


Figura 11: Esquema síntese do capítulo 3
Fonte: Elaborado pela autora.

O arcabouço teórico apresentado até aqui tem como função fundamental nossas escolhas metodológicas a serem detalhadas no capítulo seguinte.

4 ETAPAS METODOLÓGICAS: UMA PROPOSTA COM BASE NA SF

Os estudos sobre as ontologias na área da CI e da CC surgiram com maior intensidade na década de 80, através de propostas que viabilizassem meios de representar o conhecimento, recuperar informações e aprimorar o processamento da linguagem natural. Atualmente, conforme discutem Mendonça e Almeida (2014)⁵⁴, as ontologias têm sido aplicadas para diversos fins e em diferentes ramos, como por exemplo: (a) como base para a extração da informação utilizando-se do conteúdo disponível na *Web*; (b) usos para a tradução automática; (c) integração automática de vocabulários ou dicionários de dados; (d) integração de bancos de dados a softwares; (e) recuperação da informação em diferentes domínios; (f) aplicações práticas para a ampliação de recursos visando a *WebSemântica*; entre muitas outras aplicações. Nosso trabalho vem ao encontro das três últimas, podendo colaborar na integração entre sistemas e bancos de dados, na recuperação da informação jurídica e na colaboração em aplicações para *WebSemântica*.

Mesmo com todo este tempo de estudos e com toda a gama de usos e aplicações, ainda não há, conforme Mendonça e Almeida (2014), um consenso entre especialistas acerca das metodologias adotadas na construção das ontologias; segundo os autores, não há um modelo ou uma metodologia que sejam completamente aceitos ou que sejam capazes de cobrir diferentes domínios.

Pesquisando sobre metodologias para construção de ontologias deparamo-nos exatamente com esta grande variedade citada por Mendonça e Almeida (2014). Além disso, percebemos que as metodologias vigentes têm como foco a implementação de etapas, sem muito detalhar como cada uma deve ser desenvolvida e sem ater-se profundamente às relações entre classes e subclasses que extrapolem a hierarquia, isto é, que privilegiem outros tipos de relações.

Tais apontamentos nos fazem refletir sobre a possibilidade de implementação de uma metodologia que privilegie os aspectos de significação, tomando a SF como principal aliada

⁵⁴ Representantes da área da Ciência da Informação

para a criação de relações capazes de descrever o domínio jurídico focalizando os eventos e seus respectivos participantes, bem como seus diferentes papéis em cada evento.

Este capítulo, portanto, cumpre com o objetivo de apresentar e justificar nossa proposta metodológica baseada em *frames* para a construção da ontologia jurídica respondendo ao questionamento inicial: *Considerando o propósito de prover maior significado à descrição na ontologia do domínio jurídico tomando como base a SF, quais são os passos metodológicos a serem adotados?*. Antes, contudo, consideramos pertinente expor aspectos relevantes acerca do processo de construção de ontologias e apresentar as principais metodologias para seu desenvolvimento, bem como refletir sobre suas etapas de implementação. Adotamos esta estratégia, pois julgamos necessário apresentar o maquinário relacionado às metodologias a fim de sustentar nossa proposta.

Cumprase alertar que a seção que inicia o capítulo apresenta aspectos gerais acerca das metodologias já testadas e validadas na comunidade da CI. O objetivo de apresentar este consenso relacionado às metodologias é justamente proporcionar uma visão geral e sustentar nossas decisões em relação à proposta baseada em *frames*.

Para isso, o capítulo está organizado em três seções. A seção 4.1 corresponde ao panorama atual acerca das principais metodologias já desenvolvidas para a construção de ontologias e busca-se refletir sobre a forma como implementam as relações e a organização geral da ontologia. A seção 4.2 apresenta o detalhamento da metodologia que propomos para o desenvolvimento da ontologia jurídica baseada na SF. Nossa proposta leva em conta os princípios descritos pelos autores da área da CI e metodologias apresentadas que serviram de ponto de partida para a proposição das nossas etapas metodológicas. A partir dos aspectos apresentados pelos autores da CI e vislumbrados nas diferentes metodologias tornou-se possível refletir acerca da proposta de desenvolvimento de ontologia baseada em *frames*, atentando para as possibilidades de reuso ou as necessidades de adaptação e proposição de novas etapas. Concluindo, na seção 4.3, ilustramos as ideias centrais discutidas no capítulo, através de um esquema síntese.

4.1 Panorama de metodologias

Alguns dos principais estudiosos do tema (Fernández-López et al.,1997); Uschold & Gruninger, 1996 e Noy & McGuinness, 2001) defendem que o processo de construção de

ontologias corresponde a um **ciclo de vida** da ontologia e que tal processo envolve alguns estágios, a saber: (a) planejamento; (b) especificação de requisitos; (c) aquisição de conhecimento; (d) conceitualização; (e) formalização; (f) integração; (g) implementação; (h) avaliação; e (i) documentação.

A partir destes princípios fundamentam-se várias metodologias para a construção de ontologias, cada qual com seus focos e aplicações. Nos trabalhos sobre ontologias (mais frequentes na CI e na IA), são descritas diferentes metodologias que seguem um certo padrão, porém se diferem conforme a aplicação e o domínio da ontologia. Diante disso, nesta seção, temos por objetivo apresentar as principais metodologias⁵⁵, atentando para seus aspectos e fundamentos basilares, especialmente relacionados ao estágio de **conceitualização** e buscando estabelecer os princípios para os passos metodológicos que fundamentarão a construção da nossa ontologia jurídica.

Conforme destacam Mendonça e Almeida (2014), no final da década de 90 até o início dos anos 2000 foi o período em que se desenvolveram mais metodologias para a construção de ontologias, sendo que nos anos subsequentes o foco maior foi no seu aprimoramento tendo poucas novas metodologias desenvolvidas. Os autores ainda destacam que as metodologias existentes atualmente apresentam muitas semelhanças em relação às etapas previstas para o desenvolvimento de ontologias, porém algumas diferenças podem ser percebidas principalmente quando são metodologias desenvolvidas para diferentes aplicações e domínios.

Consideramos importante apresentar brevemente as principais metodologias tendo como foco suas etapas de desenvolvimento, principalmente as relacionadas ao estágio de **conceitualização**, por este ser o foco da nossa pesquisa.

Elencamos as seguintes metodologias selecionadas na literatura a partir da síntese elaborada por Mendonça e Almeida (2014): **Método de Uschold e King** (USCHOLD & KING, 1995), **TOVE** – *Toronto Virtual Enterprise* (GRUNINGER & FOX, 1995), **Methontology** (Fernández-López et al., 1997), **Método 101** (NOY & McGUINNESS, 2001), e **MFPFO** (LIM, LIU & LEE, 2011). Passamos, pois, à explanação destas metodologias, separando-as em blocos.

⁵⁵ Vale destacar que a literatura acerca de metodologias para o desenvolvimento de ontologias é vasta e variada, havendo possibilidades de encontrar mais de 20 metodologias válidas e preparadas para diferentes fins. Para a escolha das metodologias destacadas nesta seção, tomamos como base a pesquisa realizada por Mendonça e Almeida (2014). Realizamos uma busca na *web* e optamos por apresentar as metodologias mais citadas em trabalhos representativos sobre ontologias e que pudessem trazer contribuições para nossa pesquisa.

4.1.1 Método proposto por Uschold e King (USCHOLD & KING, 1995)

Com base no desenvolvimento da *Enterprise Ontology*, integrante do projeto *Enterprise* do Instituto de Aplicações em Inteligência Artificial da Universidade de Edimburgo em parceria com a IBM, Unilever e outros, Mike Uschold e Martin King idealizaram um método para o desenvolvimento de ontologias cujo processo de construção prevê como base **cenários de motivação**.

O método fundamenta-se na descrição de situações reais e na sequência de ações para que os objetivos sejam atingidos através dos cenários. Para os autores, os cenários de motivação auxiliam na descrição dos questionamentos que a ontologia pretende cobrir. A visão de construir a ontologia sob cenários de motivação condiz com nossa proposta de uma metodologia baseada em *frames* (RUPPENHOFER et al., 2006) que explicita os conceitos do domínio e suas relações através dos eventos, os participantes e seus respectivos papéis.

O método proposto por Uschold & King (1995) prevê quatro etapas:

- I. Identificação: nesta etapa é necessário definir os objetivos da ontologia (por que e para que ela será usada) e estabelecer as questões de competência, ou compromissos ontológicos. Segundo os autores, uma ontologia pode ser construída com a finalidade de compartilhar, reutilizar ou fazer parte de uma base de conhecimento. A proposta apresentada nesta pesquisa tem a finalidade de complementar uma base de conhecimento acerca do domínio jurídico.
- II. Construção: esta etapa subdivide-se em captura, codificação e integração.
 - a. Captura: é a fase de definição dos conceitos e relações de forma textual, ou seja, a criação de uma espécie de dicionário ou glossário com os conceitos e a forma como se relacionam.
 - b. Codificação: utilização de uma linguagem formal para codificar os conceitos e relacionamentos levantados na fase anterior. Nesta fase são definidas as classes, as relações e as instâncias.
 - c. Integração: verificar se há possibilidade de reuso de ontologias já existentes.
- III. Avaliação: utilizar critérios de ordem técnica para a avaliação, verificando se as especificações do início do projeto foram atendidas, validando as questões de competência.
- IV. Documentação: descrição de todo o processo de implementação, desde as finalidades da ontologia até a codificação estabelecida para a criação dos nomes das classes. A

etapa de documentação, como já vimos na seção anterior, é importante para o reuso da ontologia e para possíveis adaptações que sejam necessárias.

O estágio de **conceitualização** está previsto na metodologia proposta por Uschold & King (1995) na etapa de **construção** (II). Nesta etapa os autores sugerem a criação de um modelo do domínio através da seleção de conceitos, organizando-os em uma estrutura de classes e subclasses que, no nosso entendimento, corresponde à fase de modelagem conceitual prevista no estágio de conceitualização. Na nossa pesquisa, a etapa de **construção** prevista pelos autores corresponde à etapa de modelagem da ontologia, partindo da elaboração de um mapa conceitual do domínio que possibilitará a criação dos respectivos *frames*.

Nesta etapa de construção, Uschold & King (1995) atentam para o fato de observar as possíveis ambiguidades nas definições dos conceitos e relações, e afirmam que podem ocorrer vários termos correspondendo a uma única definição (sinonímia) ou um termo correspondendo a vários conceitos (polissemia ou homonímia). Os autores apresentam sugestões de tratar estes problemas na construção da ontologia, até mesmo propondo suprimir o termo. Entendemos que, na nossa abordagem de usar a SF para organizar as classes e subclasses e suas relações, podemos minimizar⁵⁶ este problema de ambiguidade sem a necessidade de exclusão de nenhum termo ou conceito. Isso se deve ao fato de estarmos trabalhando com *frames* (RUPPENHOFER et al., 2006), ou seja, os conceitos estarão organizados conforme os eventos aos quais estão relacionados, bem como os papéis dos participantes, possibilitando uma definição explícita e individual de cada conceito. O termo **defesa** é um bom exemplo para tratar sobre a ambiguidade: ora é um **evento**, no qual ocorrem os relatos das testemunhas de defesa e do próprio acusado; ora um **participante** que tem o papel de defender o réu. A estrutura de *frames* permite desambiguar o sentido do termo a partir das unidades lexicais evocadoras e dos papéis de cada participante, descrevendo o evento e relacionando seus participantes.

O problema que podemos enfrentar com esta abordagem é em relação à nomenclatura, uma vez que, por restrição da ferramenta de edição adotada, não podemos definir classes, subclasses e relações com um mesmo nome, porém este aspecto é facilmente resolvido definindo-se um padrão para a criação dos nomes incluindo uma sigla para identificar eventos, participantes, papéis e *frames*.

⁵⁶ Sugerimos que nossa abordagem apenas minimiza o problema de ambiguidade, pois temos ciência de que os *frames* não dão conta de identificar polissemia ou homonímia (um problema que ainda não foi solucionado na FN), apenas indicam que as unidades lexicais correspondem a *frames* distintos.

Inicialmente consideramos este método interessante por apresentar os **cenários de motivação** e relacionamos este aspecto ao nosso intuito de utilizar os *frames* (RUPPENHOFER et al., 2006) como aporte para a construção da ontologia; porém, ao atentarmos para suas etapas percebemos pouco detalhamento na fase de construção e nenhuma relação apontada aos cenários mencionados como motivadores do método. Além disso, o método não especifica a forma como os conceitos são elencados e nem mesmo a forma como são estabelecidas as relações, além das descrições formais.

4.1.2 TOVE – Toronto Virtual Enterprise (GRUNINGER & FOX, 1995)

Michael Gruninger e Mark Fox foram os idealizadores da metodologia baseada no projeto *Toronto Virtual Enterprise – TOVE*, tendo como base teórica a Inteligência Artificial e o desenvolvimento de ontologias para o domínio dos negócios corporativos. Os autores seguem Uschold (1995) e também utilizam os cenários de motivação para descrever os problemas referentes à ontologia.

Esta metodologia prevê seis etapas:

- I. **Elaboração e descrição dos cenários de motivação:** para os autores, os cenários de motivação correspondem aos problemas não descritos nas ontologias já existentes. Também chamados de cenário-problema, os autores assinalam a possibilidade de elaboração de um conjunto de soluções através de uma semântica informal dos objetos e suas relações que posteriormente possam ser incluídas na ontologia. Parece-nos, aqui, uma visão de cenários de motivação diferente da apresentada por Uschold & King (1995).
- II. **Formulação informal das questões de competência:** tendo como base os cenários motivacionais, são formuladas questões de competência, isto é, questionamentos sobre o que deverá ser representado, como deverá ser representado e o que a ontologia será capaz de responder ao ser desenvolvida.
- III. **Especificação dos termos através de linguagem formal:** são elencados os termos/conceitos referentes ao domínio com base nos cenários motivacionais e nas questões de competência que servirão de base para a estrutura formal da ontologia.

Após a escolha dos conceitos, é realizada a representação formal do conhecimento utilizando Lógica de Primeira Ordem ou *Knowledge Interchange Format (KIF)*⁵⁷.

- IV. Descrição formal das questões de competência: uso da linguagem formal para a descrição das questões de competência.
- V. Especificação formal dos axiomas: etapa de criação de regras em linguagem formal com o intuito de definir a semântica dos termos e as relações na ontologia.
- VI. Verificação da completude: nesta última etapa são estabelecidas condições para considerar a ontologia completa baseadas nas questões de competência elaboradas e formalizadas.

Como se vê pela descrição das etapas, esta metodologia é altamente formal, tendo como foco a lógica para estabelecimento de relações e até mesmo do significado dos conceitos. Segundo Corcho, Gómez-Pérez & Fernández-Perez (2004), esta metodologia pode ser usada como exemplo para a transformação de cenários informais em modelos legíveis por computadores, uma vez que expressa seu conteúdo através da lógica clássica.

Sob um ponto de vista computacional, que não vise à descrição semântica além de fórmulas lógicas com condições necessárias e suficientes, parece ser uma metodologia bastante adequada, ainda que partir somente dos cenários de motivação pareça-nos insuficiente para compor um domínio em sua completude.

4.1.3 Methontology (Fernández-López et al., 1997)

Asunción Gómez-Pérez, Mariano Fernández, Antônio J. De Vicente e Natalia Juristo, pesquisadores do laboratório de Inteligência Artificial da Universidade Politécnica de Madri, desenvolveram a metodologia, cujo principal objetivo é automatizar o processo de construção de ontologias.

De acordo com Fernández-López et al. (1997), a metodologia é composta por estágios de desenvolvimento, um ciclo de vida e técnicas para realizar planejamento, desenvolvimento e suporte das ações relacionadas à ontologia. O planejamento das atividades abrange a divisão de tarefas e supervisão sobre elas, o desenvolvimento é responsável pela construção da

⁵⁷ *Knowledge Interchange Format (KIF)* é uma linguagem projetada para utilização no intercâmbio de conhecimentos entre os diferentes sistemas de computador.

ontologia propriamente dita e o suporte corresponde ao processo de aquisição do conhecimento, documentação e avaliação, sendo que tais atividades co-ocorrem durante todo o ciclo de vida da ontologia.

No estágio de desenvolvimento a metodologia segue nove etapas:

- I. Especificação: etapa de elaboração de um documento em linguagem natural contemplando informações como: o objetivo da ontologia, seus propósitos e aplicações, quem serão seus usuários e qual o domínio que a ontologia pretende cobrir.
- II. Aquisição de conhecimento: etapa para busca de conhecimentos acerca do domínio através de contato com especialistas, documentos, livros, outras ontologias já existentes, enfim, é uma etapa de reconhecimento do domínio a ser descrito pela ontologia.
- III. Conceitualização: nesta etapa são elencados os termos da ontologia a partir do conhecimento adquirido nas etapas anteriores e organizados em um modelo conceitual. Vale salientar que modelo conceitual, sob o foco de Fernández-López et al. (1997), tem como função apresentar os conceitos importantes de um domínio, através das classes e subclasses, bem como suas relações, propriedades e atributos, sendo implementado através de uma estrutura.
- IV. Formalização: o modelo conceitual criado na etapa anterior é formalizado através de uma linguagem formal para a descrição de ontologias: lógica de descrições ou o modelo de *frames*. Cabe aqui esclarecer que este conceito de *frame* está relacionado às concepções de Minsky (1974) e corresponde ao conjunto de características lógicas de uma determinada entidade, não havendo relação com os *frames* de Fillmore (1982), os quais pretendemos integrar à ontologia para a ampliação das relações entre eventos, participantes e seus diferentes papéis.
- V. Integração: esta etapa prevê a integração com ontologias já existentes.
- VI. Implementação: nesta etapa os formalismos organizados nas etapas anteriores são implementados através de uma linguagem de representação de ontologias, as mais usadas e conhecidas são OIL⁵⁸, DAML+OIL⁵⁹ e OWL.

⁵⁸ Ontology Interchange Language - OIL (SOWA, 1999) é uma linguagem que se baseia em conceitos, classes e atributos. Permite a construção e representação de ontologias, através de semântica formal bem estabelecida tendo similaridades com as linguagens para a Web (FENSEL, 2000).

⁵⁹ DAML + OIL é uma linguagem de marcação semântica para recursos da Web. Baseia-se em padrões W3C anteriores, como RDF e RDF Schema. DAML + OIL fornece primitivas de modelagem comumente encontrados em linguagens baseadas em frames.

- VII. Avaliação: é uma etapa muito importante, pois tem o objetivo de garantir a qualidade de a adequação dos padrões descritos na ontologia verificando as possíveis inconsistências.
- VIII. Documentação: é uma etapa que visa facilitar a manutenção da ontologia e o seu reuso. Corresponde aos documentos com especificação de requisitos, resultados, convenções utilizadas, a forma como o modelo conceitual foi construído e os resultados obtidos após a avaliação.
- IX. Manutenção: etapa prevista caso haja ajustes a serem feitos na ontologia em função de adequações realizadas no domínio.

Atendo-se aos estágios previstos como princípios para construção de ontologias, dos quais destacamos o de conceitualização por considera-lo como o mais importante para o nosso objetivo, percebemos que a Methontology prevê uma etapa com o mesmo nome. Nesta etapa os autores enfatizam as estratégias de identificação de termos propostas por Uschold & Gruninger (1996) – *bottom-up*, *top-down* e *middle-out* – reforçando que a última é a mais profícua para a identificação dos termos primários da ontologia em um primeiro momento, para, depois, especializá-los ou generalizá-los. Esta técnica, segundo Gómez Pérez et al. (1997), permite uma identificação mais estável dos termos exigindo menos esforço dos desenvolvedores.

Na Methontology, após a extração dos termos, estes são agrupados em conceitos e verbos e, posteriormente, são verificadas as relações entre eles, além disso os verbos são direcionados aos conceitos aos quais estão relacionados. Esta abordagem se mostra eficiente para o nosso propósito de construção da ontologia usando a SF como aporte para a determinação das relações, em função da forma de agrupar os termos. Porém, na nossa abordagem temos que prever os EF's e as UL's (que não são somente verbos), ampliando as possibilidades de relações a serem estabelecidas a partir dos eventos e seus participantes.

Os autores da Methontology sugerem a criação de diagramas de árvores⁶⁰ para a classificação dos conceitos e organização dos verbos; contudo, nós entendemos que, para nosso fim, a melhor forma de fazer esta representação inicial é através de um mapa

⁶⁰ Em computação, diagrama de árvore é um modelo abstrato de uma estrutura hierárquica. Trata-se de uma estrutura não-linear constituída de nós com relações de parentesco - pai-filho (GOODRICH, TAMASSIA & MOUNT, 2011).

conceitual⁶¹ que englobe tanto conceitos, quanto palavras que os relacionam (sejam verbos ou nominais) ampliando as relações (não somente hierárquicas).

4.1.4 Método 101 (NOY & McGUINNESS, 2001)

Natalya F. Noy e Deborah L. McGuinness desenvolveram uma ontologia de vinhos e alimentos, utilizando o editor de ontologias Protégé-2000 e, a partir desta experiência, idealizaram o Método 101, cujo propósito é apresentar um guia simplificado de desenvolvimento de ontologias que possa ser útil a outros projetos. Neste método, as fases de modelagem conceitual e formalização ocorrem de modo simultâneo, uma vez que são executadas diretamente no editor de ontologias Protégé. A formalização neste método é baseada em *frames*, seguindo a visão de Minsky (1974).

O método compreende sete etapas:

- I. Determinar o domínio e o escopo da ontologia: nesta etapa as autoras sugerem que sejam respondidas quatro questões (que podem ser vistas como questões de competência). As questões sugeridas são: (a) Que domínio se pretende cobrir com a ontologia? (b) Qual o propósito de uso da ontologia? (c) Para que informações a ontologia deve prover respostas? (d) Quem vai usar e manter a ontologia? Segundo as autoras, estas perguntas servem para definir a ontologia e, ao final do processo, para verificar se a ontologia construída contém as informações necessárias para dar resposta às questões. Na nossa proposta metodológica, consideramos estas questões de competência para a definição do domínio e do escopo da ontologia, conforme será apresentado na seção seguinte.
- II. Considerar o reuso de outras ontologias: nesta etapa as autoras sugerem pesquisar outras ontologias relacionadas ao domínio para verificar a possibilidade de reuso de classes ou de modificação da estrutura para facilitar o desenvolvimento e promover uma maior integração de informações. Esta etapa elencada aqui é considerada importante por estudiosos de ontologias, tais como Guarino (1998) e Fernández-López et al. (1997); porém, o método 101 não esclarece de que modo este reuso deva ser implementado.

⁶¹ Na seção 4.3, ao apresentarmos nossa proposta metodológica aprofundaremos o conceito de mapa conceitual. Por ora, podemos definir mapa conceitual como sendo uma representação gráfica que indica os conceitos, palavras e suas relações para representar o conhecimento referente a um determinado assunto ou área.

- III. Enumerar termos: nesta etapa devem ser listados os termos importantes relacionados ao domínio para compor a ontologia. As autoras sugerem que sejam respondidas as seguintes questões: (a) Quais são os termos que estamos tratando? (b) Quais são as propriedades destes termos? (c) O que significam os termos? Parece-nos, aqui, que esta etapa é bastante intuitiva sendo possível para um domínio simples, como o dos vinhos descrito pelas autoras, porém inviável para um domínio complexo como o jurídico.
- IV. Definir classes e a hierarquia de classes: os termos definidos na etapa anterior devem ser selecionados de acordo com os objetos do domínio e agrupados por suas propriedades similares a fim de compor as classes e organizá-las em uma hierarquia (taxonomia). Segundo as autoras, não há uma forma única de determinar a hierarquia de classes de um domínio, mas sim a forma mais adequada em relação ao propósito da ontologia, ou seja, a forma de organizar as classes taxonomicamente dependerá do uso da ontologia, da sua aplicação e dos requisitos elencados inicialmente. Para esta organização, as autoras recomendam seguir as abordagens sugeridas por Uschold e Gruninger (1996) – *top-down*, *bottom-up* ou *middle-out*.
- V. Definir propriedades e restrições: a estrutura de classes definida na etapa anterior não é capaz de prover as informações necessárias para que a ontologia responda às questões de competência e, por isso, nesta etapa é necessário descrever a estrutura interna de cada classe, ou seja, determinar as propriedades das classes. Estas propriedades podem ser extraídas dos próprios termos relacionados ao domínio e servirão para descrever cada classe. Um conjunto de propriedades é chamado de *slot*.
- VI. Definir os valores das propriedades: as propriedades definidas para as classes podem assumir diferentes valores, dependendo da linguagem lógica com a qual a ontologia está sendo implementada. Um dos valores é chamado de **cardinalidade** – mínima ou máxima⁶²; única ou múltipla⁶³. Há ainda a possibilidade de definir se a propriedade é uma cadeia de caracteres (*string*), números (inteiros ou ponto flutuante), tipo lógico verdadeiro ou falso (*boolean*) ou lista de elementos.
- VII. Criar instâncias: as instâncias podem ser entendidas como “seres do mundo real” que preenchem a ontologia. Assim, nesta etapa devem ser elencadas as classes e preenchidas as propriedades definindo instâncias individuais.

⁶² Cardinalidade N se o slot deve ter ao menos N valores ou no máximo N valores. Cardinalidade 1 se o slot deve ter pelo menos um valor (requerido). Cardinalidade 0 se o slot é opcional.

⁶³ Cardinalidade única: permite no máximo 1 valor. Cardinalidade múltipla: permite qualquer número de valores.

Noy & McGuinness (2001) defendem que não há maneira correta para modelar um domínio, havendo sempre a possibilidade de propor diferentes alternativas. Segundo as autoras, a melhor solução dependerá do domínio e da aplicabilidade da ontologia sempre considerando que o processo de desenvolvimento não é linear e exige iterações e refinamentos durante todo o trabalho.

No Método 101, Noy & McGuinness (2001) apresentam uma lista de princípios⁶⁴ que devem ser observados para a criação da hierarquia de classes. Tais princípios são fundamentais para a criação de uma estrutura robusta e livre de ambiguidades e inconsistências. Ainda que todos os princípios apresentados pelas autoras sejam de grande valia para a construção de ontologias, consideramos que, para nosso propósito de desenvolvimento informal da ontologia, os seguintes princípios para a organização das classes devam ser seguidos:

- I. Relação **é-um** ou **tipo-de** (generalização): A tem uma subclasse B se todas as instâncias (ocorrência no mundo real) de B são também instâncias de A. Por exemplo: a classe Mamífero tem uma subclasse Gato e Sissy⁶⁵ é uma instância da subclasse Gato; então, por generalização, Sissy é um indivíduo da classe Mamífero. No domínio jurídico, podemos exemplificar através da classe AutoridadeJulgadora que tem uma subclasse Juiz e João da Silva é uma instância da subclasse Juiz, logo João da Silva é também uma instância da classe AutoridadeJulgadora.
- II. Transitividade da hierarquia de classes: A tem uma subclasse B, que tem uma subclasse C; logo, C é também uma subclasse de A, constituindo-se em uma relação **é-um** transitiva. Por exemplo: a classe Vinho tem uma subclasse VinhoBranco, que tem uma subclasse Chardonnay. A transitividade implica que a subclasse Chardonnay seja também uma subclasse de Vinho⁶⁶.
- III. Evolução da hierarquia de classes: é importante manter a consistência da estrutura, organizando as classes de forma genérica, pois podem ser necessárias manutenções gerando novas classes a serem incluídas na estrutura.

⁶⁴ A lista completa dos princípios pode ser acessada em http://liris.cnrs.fr/~amille/enseignements/Ecole_Centrale/What%20is%20an%20ontology%20and%20why%20we%20need%20it.htm. Acesso em 20/05/15.

⁶⁵ Para melhor visualização e entendimento dos exemplos, usaremos a fonte *arial* para destacar as instâncias.

⁶⁶ Exemplo extraído de Noy & McGuinness (2001).

- IV. Nomes das classes: os nomes das classes representam os conceitos do domínio, sendo que os nomes das classes poderão mudar, mas os termos relacionados aos conceitos sempre serão os mesmos. É importante seguir um padrão para nomear as classes e subclasses, evitando ambiguidades.
- V. Quando inserir (ou não) uma nova subclasse: para avaliar a necessidade de inclusão de uma nova subclasse é necessário atentar para as propriedades, restrições e relações. Se forem iguais, não há necessidade de criação de uma nova subclasse.

O Método 101, ainda que não apresente claramente aspectos relacionados às etapas de planejamento e concepção da ontologia, mostra de forma simplificada as etapas de formalização e implementação integradas, explicitando-as de modo a facilitar o entendimento do processo. Além disso, o método provê uma lista de princípios que elucidam a etapa de construção da taxonomia e auxiliam as escolhas do desenvolvedor. Por ser um método simplificado, as etapas de conceitualização e de implementação ocorrem de forma concomitante e diretamente no editor de ontologias Protégé, possibilitando maior interação no decorrer do processo de desenvolvimento.

4.1.5 Metodologia MFPFO (LIM, LIU & LEE, 2011)

Os autores conceberam a MFPFO – metodologia multifacetada para a construção de ontologias com anotação semântica – para a modelagem de estruturas relacionadas ao domínio de famílias de produtos visando reduzir o tempo e o esforço humanos no desenvolvimento de ontologias. O processo de construção da ontologia inicia-se com a coleta de informações sobre os produtos e suas variantes que são denominados “famílias de produtos”. Lim, Liu & Lee (2011) exemplificam uma família de produtos através dos computadores portáteis NC2XX, que têm em sua família os modelos NC200 e NC220 juntamente com todas as informações advindas dos fabricantes.

Esta metodologia apresenta seis etapas:

- I. Construção de uma taxonomia de família de produtos: o primeiro passo é organizar os produtos e suas variantes em uma taxonomia seguindo as características genéricas de cada produto a partir de uma lista de informações, tais como: número do produto, quantidade, descrição, montagem, etc.

- II. Extração de entidades⁶⁷: nesta etapa são elencadas as características específicas de cada produto da família de produtos.
- III. Identificação do conceito e geração da unidade com diferentes facetas⁶⁸: aqui os autores propõem que sejam definidas as possíveis facetas para cada entidade de cada produto, chamando-as de unidades semânticas. Esta etapa tem como objetivo verificar as associações possíveis entre os produtos permitindo o seu agrupamento.
- IV. Anotação semântica⁶⁹ e modelagem de cada faceta: o processo de anotação semântica consiste em associar as entidades com suas relativas facetas a partir dos agrupamentos realizados na etapa anterior. Nesta etapa são determinadas as inter-relações entre as entidades, conceitos e atributos de forma genérica visando à criação de uma ontologia semanticamente anotada.
- V. Estruturação da ontologia multifacetada e anotada semanticamente: este processo envolve a anotação das diferentes facetas tendo como foco as inter-relações. É considerada multifacetada, pois leva em consideração as diferentes facetas de cada produto, considerando as informações adicionais dos produtos e anotadas semanticamente porque estabelecem a relação entre as entidades através das facetas.
- VI. Avaliação e validação da ontologia: ao concluir a organização da ontologia é necessário testar e verificar inconsistências na sua modelagem.

Diante da descrição geral das metodologias apresentadas, poder-se-ia considerar que a mais adequada ao nosso propósito seria a **MFPFO** por apresentar etapas de construção de uma taxonomia, identificar conceitos e facetas, anotar semanticamente as diferentes facetas e relacioná-las a toda a estrutura. Porém, ao analisarmos atentamente o trabalho desenvolvido por Lim, Liu e Lee (2011), percebemos que a metodologia é fortemente ligada ao domínio “famílias de produtos”, isso implica um desenvolvimento restrito da estrutura em função da delimitação específica do escopo. Além disso, o significado de entidades, facetas e anotação semântica apresentado pelos autores é fortemente relacionado a estruturas lógicas e axiomas, sendo inapropriado para nosso intuito de ampliar o significado levando em conta os eventos, os participantes e seus diferentes papéis. Entendemos que os axiomas e propriedades são

⁶⁷ Para os autores da metodologia MFPFO o termo “entidade” se refere aos diferentes recursos que o produto possui.

⁶⁸ De acordo com Lim, Liu e Lee, o termo “facetas” corresponde às diferentes características que cada entidade possui. Cada consumidor pode estar interessado em diferentes facetas, por exemplo: o electricista está preocupado com a voltagem da luminária, enquanto que a dona de casa se interessa pela beleza.

⁶⁹ Para os autores da metodologia, “anotação semântica” corresponde às relações entre as facetas e as entidades, sendo que cada entidade pode ter diferentes facetas.

igualmente importantes para a estruturação da ontologia, porém este não é o foco do nosso trabalho. Para os autores, diferentemente da nossa posição, entidades estão estritamente ligados aos produtos, as facetas correspondem às diferentes características que cada produto apresenta e a anotação semântica traz apenas a relação entre as facetas e as entidades de uma forma simplificada e baseada em regras lógicas. Logo, esta metodologia não supre nosso intuito de ampliar as relações não taxonômicas através de conceitos da área jurídica e as relações existentes entre os diferentes eventos e seus participantes.

4.1.6 Reflexão sobre as metodologias apresentadas

Observando as principais metodologias descritas e suas etapas de desenvolvimento, se pode perceber que apresentam diferentes graus de aprofundamento e detalhamento dos processos, sendo algumas mais detalhadas, como a **Methontology** e a **TOVE**, e outras mais simples, mas que enfatizam as etapas de implementação, como por exemplo o **Método 101**.

Um ponto em comum que se observa nas metodologias descritas é o fato de referenciar como primeira etapa aspectos relacionados à definição e delimitação do domínio ao qual a ontologia se destina (com uma grande variação de nomenclatura, mas referindo-se ao mesmo aspecto), sendo, ao nosso ver, uma etapa necessária a todas as metodologias independente do domínio, uma vez que esta ação balizará todo o procedimento subsequente de estruturação da ontologia. Há ainda a referência sobre a determinação do escopo da ontologia, ou seja, qual será seu principal objetivo e aplicação, sendo outra etapa fundamental na implementação.

Outro ponto de interseção entre as metodologias descritas é a extração de termos, novamente apresentada com diferentes nomenclaturas, porém presente na maior parte dos exemplos citados. Na metodologia **TOVE** (GRUNINGER & FOX, 1995), por exemplo, temos a etapa (III), que se refere à concepção da terminologia formal, a qual, no nosso entendimento, corresponde a uma etapa de extração de termos do domínio. Já na **Methontology**, Gómez-Perez et al. (1997) estabelecem a etapa (II), denominando-a de etapa de aquisição do conhecimento, na qual são elencados os diferentes aspectos do domínio, sendo seguida da etapa (III), que corresponde à conceitualização. Novamente, relacionamos tais etapas à extração de termos, uma vez que correspondem aos princípios de organização da ontologia. No **Método 101** as etapas de extração de termos são bem evidentes e explícitas. Este é outro ponto que, no nosso entendimento, é fundamental figurar nas etapas de

desenvolvimento de ontologias a fim de cobrir o domínio através dos termos nele envolvidos. Vale salientar que a extração de termos é sempre associada a soluções computacionais, sendo utilizados diferentes recursos para a extração automática, privilegiando a quantidade de termos e os números de ocorrências. Na nossa proposta reinterpretamos a etapa de extração de termos. A etapa não tem o propósito de elencar termos para prover a estrutura de classes, uma vez que esta será dada à partir dos *frames*, mas cumpre com a função de auxiliar na determinação dos evocadores de *frame* e das unidades lexicais relacionadas aos *frames*. Na seção 4.3 elucidamos nossa proposta através do detalhamento da metodologia.

Nas metodologias citadas os termos extraídos são organizados em classes e subclasses, respeitando uma hierarquia, tal como no **Método 101**, que prevê uma etapa para a definição de classes através da organização dos termos extraídos. Na **Methontology**, a organização dos termos em classes e subclasses ocorre nas etapas de **integração** e **implementação**, na qual os conceitos são organizados hierarquicamente. Novamente identificamos uma variação na forma de nomear as etapas de definição de classes e subclasses, porém todas as metodologias apontam para a necessidade de organização do domínio através da hierarquia de classes, tomando como base os termos. O **Método 101** é o mais esclarecedor por apresentar de maneira simples e detalhada as etapas de implementação da hierarquia de classes, além de prover princípios para esta organização. Novamente, entendemos ser possível uma reinterpretação desta etapa: na nossa visão a estrutura de classes e subclasses será organizadas a partir dos frames.

Além desta parte que consideramos **estruturante** da ontologia (a definição do escopo, a extração de termos e o estabelecimento das classes e subclasses), atentamos para a forma como as metodologias descrevem a maneira de estabelecer as relações⁷⁰ entre as classes, uma vez que este ponto é um foco de estudo ao propor uma aproximação com a SF. Novamente não encontramos informações detalhadas sobre a forma de estabelecimento e implementação das relações, sendo que algumas metodologias sequer explicitam esta etapa. As metodologias que descrevem uma etapa para esta questão falam especificamente de relações lógicas. É o caso da **TOVE** (GRUNINGER & FOX, 1995), que, na etapa (V), remete à especificação de axiomas formais, e o **Método 101** (NOY & McGUINNESS, 2001), cuja etapa (VI) refere à necessidade de definição de propriedades e restrições. A **Methontology** (Fernández-López et al., 1997), em sua etapa de conceitualização, além de extrair os termos do domínio, também

⁷⁰ Não falamos aqui das relações taxonômicas, pois estas ocorrem naturalmente ao elencar classes e subclasses, mas das outras relações que ampliam o significado dos conceitos.

explicita as classes, subclasses e relações, sugerindo que estas podem ser organizadas de maneira formal através de Lógica de Primeira Ordem ou em modelos baseados em *frames* (MINSKY, 1974).

Em relação ao estágio de conceitualização, percebemos que o método de Uschold e King (1995) não explicita os passos necessários para a elaboração do modelo conceitual; no entanto, apresenta uma importante contribuição indicando as abordagens *bottom-up*, *top-down* e *middle-out* para a identificação dos termos relacionados ao domínio e à organização hierárquica da taxonomia.

No que tange à avaliação das ontologias, percebemos que somente o Método 101 não explicita uma etapa para esta atividade; as demais metodologias apresentam diferentes sugestões, porém não trazem detalhes sobre como efetivamente fazer a avaliação da estrutura ontológica.

Conforme já referimos, verificamos semelhanças nas metodologias apresentadas, porém a que consideramos mais completa e explícita em suas etapas de desenvolvimento é o **Método 101**, que pontua claramente todos os passos para a criação da ontologia, ainda que não apresente explicitamente a forma como devem ser estabelecidas as relações não taxonômicas. Segundo Mendonça e Almeida (2014), este método, junto com a **Methontology**, é destaque de uso em trabalhos sobre ontologias de domínio, porém não como métodos isolados, mas sim em combinação com outras metodologias a fim de prover suporte às limitações umas das outras.

Tal assertiva reforça nosso propósito de desenvolvimento de uma metodologia própria embasada nas já existentes, mas que privilegie a etapa de definição de classes, suas propriedades e restrições tendo como base a SF para o estabelecimento das relações não taxonômicas, não considerando tão fortemente o princípio de formalização lógica, tal como previsto por Fernández-López et al. (1997), Uschold & Gruninger (1996) e Noy & McGuinness (2001), ou seja, descrevendo a estrutura em Lógica de Primeira Ordem. Neste aspecto, seguiremos o Método 101, que propõe a formalização integrada à implementação diretamente no editor de ontologias.

As metodologias usadas atualmente apresentam alguns problemas no detalhamento dos passos, não explicitando-os claramente e apresentando poucos exemplos de implementação, sendo este mais um ponto que nos auxilia na decisão de tomar como base etapas das metodologias citadas, arrolando novos passos de inclusão de elementos semânticos

para garantir a descrição detalhada dos eventos jurídicos e seus participantes na nossa ontologia, ou seja, descrever uma etapa de estabelecimento de relações entre classes considerando a SF.

De acordo com Da Silva, Souza e Almeida (2008), a definição das etapas e métodos é descrita de forma muito superficial nos trabalhos que explicam as metodologias, pressupondo um conhecimento prévio do leitor para melhor entendimento. Não há, por exemplo, orientações que auxiliem o desenvolvedor nas decisões e procedimentos para a inclusão de conhecimento nas ontologias. Este foi um ponto que percebemos acerca das metodologias avaliadas, as etapas descritas de forma bastante sucinta, sem apresentar maiores detalhes.

Sayão (2001) reforça nosso intento de proposição de uma metodologia menos formal⁷¹, que privilegie o trabalho analítico e descritivo do desenvolvedor baseado na descrição dos conceitos através dos *frames*. O autor registra a necessidade de levar em conta as abstrações semânticas e as relações entre os conceitos linguísticos como forma de reflexão para as diferenças de significados entre os termos. Tal assertiva corrobora com nossa proposta de aprofundar a descrição dos conceitos e suas relações através dos *frames* no intuito de prover uma estrutura descritiva para a ontologia.

Conforme já referido, consideramos o estágio de conceitualização o mais importante na estruturação de uma ontologia que considere a SF como contributo teórico para sua implementação, porém propomos uma **reinterpretação** das etapas descritas nas metodologias citadas para a organização da estrutura taxonômica e estabelecimento das relações não taxonômicas. Reforçamos que as metodologias exemplificadas serviram como uma consistente base para a nossa reflexão e proposição de etapas diferentes no estágio de conceitualização. Assim, passamos à seção seguinte na qual apresentamos a nossa proposta metodológica para a construção da ontologia do domínio jurídico baseada na SF.

4.2 A proposta metodológica

Como vimos no capítulo 2, ontologia é um objeto de estudo interdisciplinar e muito difundido em diferentes iniciativas de prover conhecimento a sistemas de recuperação da informação, estudada especialmente pela CC e CI e com uma participação ainda tímida da Semântica Computacional. Porém, conforme apresentamos no início deste capítulo, é na área

⁷¹ Cumpre esclarecer que o termo formal se refere às etapas de formalização, estando fortemente ligado à organização do domínio através da Lógica de Primeira Ordem ou outra forma lógica de representação.

da CI que se desenvolveram as principais propostas metodológicas e os princípios a serem seguidos para a construção de ontologias, sendo, portanto, uma importante fonte para nossa reflexão.

Tendo em mente o nosso propósito de construção de uma ontologia baseada em *frames*, cujo objetivo é descrever o domínio de modo a detalhar seus eventos com os respectivos participantes e seus papéis, estudamos as metodologias descritas na seção anterior de visando analisar sua usabilidade neste trabalho. Nosso intuito, inicialmente, foi buscar uma metodologia capaz de abarcar a proposta desta tese. No entanto, deparamo-nos com metodologias herméticas que ressaltavam aspectos formais (lógicos), não possibilitando espaço para uma descrição de conceitos para além de regras e formas lógicas. Contudo, apesar desta constatação, entendemos que alguns passos são necessários levando em conta o propósito de descrição do domínio com vistas à recuperação da informação. Assim, tomamos alguns dos princípios apresentados nas seções anteriores e reinterpretemos à luz da nossa visão baseada na SF.

Conforme vimos na seção anterior, alguns aspectos são essenciais na construção de ontologias de domínio. Uma primeira questão que precisa ser considerada é a *delimitação clara do escopo da ontologia, explicitando detalhadamente o domínio*. O Método 101, assim como a Methontology e a TOVE apresentam este aspecto como uma etapa de especificação, na qual faz-se necessário identificar: o objetivo da ontologia, quem fará uso dela, qual será sua aplicação e que compromissos ontológicos pretende cobrir. Nesta etapa é fundamental esclarecer de forma explícita o domínio e os níveis de conhecimento que serão incluídos. No capítulo 5, que corresponde à parte empírica do trabalho, detalharemos nossas escolhas acerca do domínio e da definição do escopo da ontologia.

Embora não seja uma etapa elencada em todas as metodologias descritas no início do capítulo, consideramos, conforme Mendonça e Almeida (2014), que outro aspecto importante na implementação de uma metodologia é o estudo da possibilidade de reuso de ontologias de alto nível (ou ontologia de fundamentação) que descrevem os conceitos gerais não considerando um domínio específico (GUARINO, 1998), devendo estas ser usadas como ponto de partida. Nossa proposta não prevê exatamente o reuso de uma ontologia de fundamentação, mas a aproximação com a LRI-Core através das categorias **Participantes Legais**, **Eventos Legais**, **Documentos Legais** e **Instituições Legais**, conforme já foi mencionado no capítulo 2.

Outro aspecto que deve ser considerado na metodologia para a construção de ontologias é garantir uma descrição informal⁷² do domínio e uma maneira de formalização destas informações. Assim, Mendonça e Almeida (2014) afirmam que é importante garantir etapas que correspondam aos conceitos, a organização textual e gráfica das classes e subclasses, isto é, que descrevam textual e graficamente o domínio a ser modelado em forma de ontologia. Os autores também defendem a importância de prever etapas de implementação do conteúdo ontológico através de uma linguagem formal, como o OWL, por exemplo. Este é um ponto incluso em todas as metodologias citadas na seção anterior, normalmente associado à etapa de **implementação** e essencial para que a descrição do domínio seja ampla e legível por máquinas.

Nas metodologias descritas identificamos a distinção entre os aspectos informais (entendidos pelos autores da CI como a descrição em linguagem natural) e formais (descrição em linguagem lógica) no desenvolvimento de ontologias; porém novamente nos deparamos com informações genéricas e que não explicitam claramente a forma como cada tipo deve ser implementado. Assim, tomando como base as reflexões acerca das diferentes metodologias, a especificidade do domínio jurídico e o intuito de estabelecer relações não taxonômicas de forma mais detalhada, faz-se possível elencar etapas para a elaboração de princípios metodológicos próprios para a construção da ontologia proposta.

Seguindo a distinção entre etapas informais e formais, organizamos nossa proposta metodológica visando contemplar o **ciclo de vida** da ontologia, atentando aos estágios previstos por Fernández-López et al. (1997), Uschold & Gruninger (1996) e Noy & McGuinness (2001) e reinterpretando os passos principais da **Methontology** (Fernández-López et al., 1997), da Metodologia **TOVE** (GRUNINGER & FOX, 1995) e do **Método 101** (NOY & MCGUINNESS, 2001). Nossa proposta apresenta etapas inspiradas nas metodologias citadas, porém destaca o estágio de **conceitualização**, reforçando a organização das classes, subclasses e relações com base na SF.

É importante alertar que a base provida pelos autores da CI foi necessária para elencarmos a nossa proposta, e que algumas etapas serão respeitadas em função do objetivo de compor uma ontologia de domínio. Porém, a aproximação com a SF e a inclusão de *frames* na ontologia trouxeram a necessidade de elencarmos etapas diferenciadas para a estruturação da ontologia e criação das relações. Assim, nossa metodologia segue alguns dos princípios previstos pelos autores da CI, mas avança nos aspectos relacionados aos conceitos e suas

⁷² Descrição informal é aqui entendida como uma descrição textual, ou seja, em linguagem natural.

relações. As etapas elencadas por nós como proposta metodológica podem ser visualizadas na figura 12 e serão explicadas em seguida.

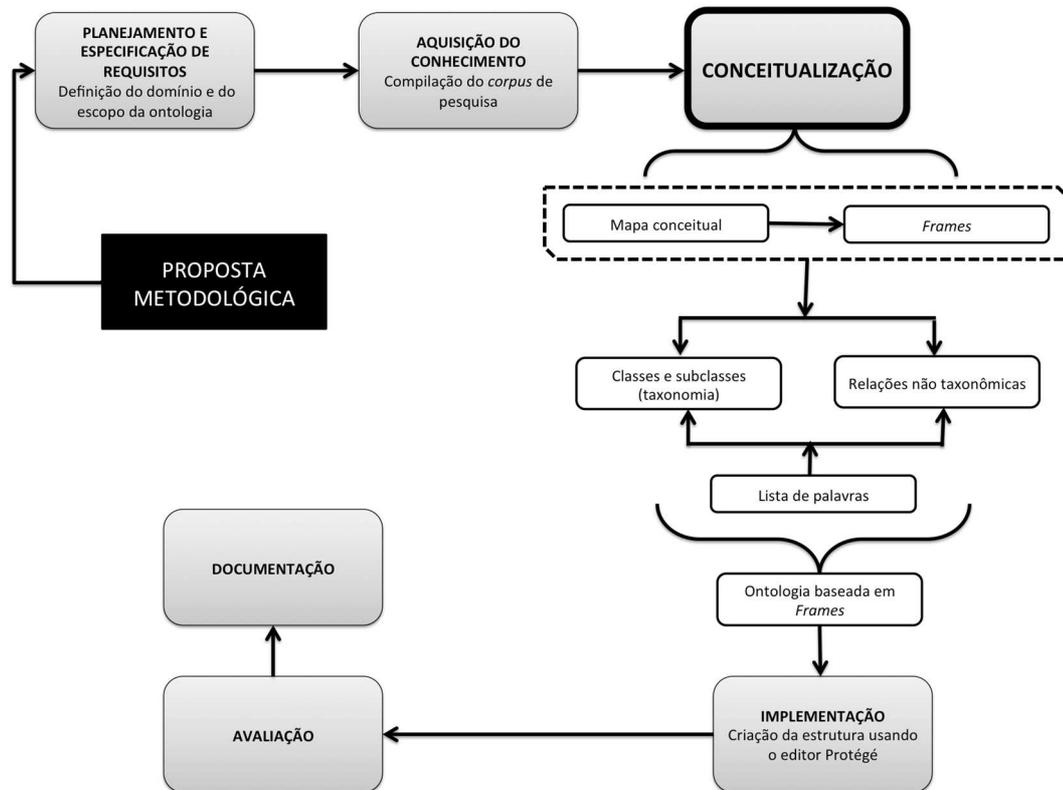


Figura 12: Etapas metodológicas propostas
Fonte: elaborado pela autora

A figura 12 mostra a sequência de passos metodológicos adotados para a construção da ontologia, tendo como foco principal a etapa de **conceitualização**, na qual são elencadas as classes, subclasses e relações. Nossa ênfase é neste estágio, justamente pelo fato de corresponder às etapas que atendem nosso objetivo principal de incluir mais significado à descrição dos conceitos através da SF.

Na sequência, apresentamos de forma sucinta uma explicação para cada estágio previsto nas etapas metodológicas, sem no entanto aprofundar aspectos relativos à construção da ontologia prevista neste trabalho, uma vez que este aprofundamento se dará na parte empírica, ou seja, no capítulo 5. Julgamos importante apresentar esta explicação para as etapas, a fim de reforçar aspectos essenciais já descritos anteriormente e apresentar nossa reinterpretação das etapas adaptadas ao nosso propósito. A escolha por organizar o trabalho desta forma corrobora com os trabalhos pesquisados, os quais apresentam primeiramente uma descrição geral, para depois exemplificar a metodologia descrita.

Optamos por apresentar a explicação das etapas em subseções a fim de organizar melhor o texto. Assim, na subseção 4.3.1 detalhamos os estágios de planejamento e especificação de requisitos e o estágio de aquisição do conhecimento. A subseção 4.3.2 apresenta o estágio de conceitualização, com a descrição das etapas e passos a serem seguidos. A subseção 4.3.3 traz o detalhamento da nossa proposta em relação ao estágio de implementação. E, finalizando a explicação acerca das etapas metodológicas, a subseção 4.3.4 apresenta as etapas de avaliação e documentação.

4.2.1 Planejamento, especificação de requisitos e aquisição do conhecimento

No primeiro estágio – **planejamento e especificação dos requisitos** – devem ser definidos os objetivos da ontologia, geral e específicos, planejando o andamento da atividade de desenvolvimento da estrutura ontológica. Neste estágio também deve-se definir o tipo de ontologia que será desenvolvida, que, de acordo com Guarino (1998) e Gómez-Pérez (2004) poderá ser de domínio ou de fundamentação. Para Fernández-López et al. (1997), é necessária a elaboração de requisitos que devem levar em conta: o propósito da ontologia, o grau de formalidade (mais ou menos formal) e o escopo. No caso do nosso trabalho, definimos a ontologia a ser construída como sendo de **domínio**, ou seja, será representativa de uma área específica e proverá a descrição baseada em *frames*.

Outro aspecto relativo a este estágio é a definição do domínio e do escopo a serem cobertos pela ontologia de forma bem explícita e definida, procurando responder às seguintes perguntas baseadas em Noy & McGuinness (2001): Que domínio a ontologia pretende cobrir? Para que a ontologia será utilizada?

Reinterpretando o estágio de **aquisição do conhecimento**, uma vez que o uso de *corpus* não é explicitado nas metodologias descritas, além da busca de informações acerca do domínio junto à especialistas, acrescentamos a necessidade de estabelecer um *corpus* para realização da pesquisa. Este *corpus* terá a função de auxiliar o desenvolvedor na seleção dos conceitos e construção das etapas futuras. Sugerimos a seleção de um ***corpus de consulta*** e de um ***corpus de pesquisa***.

O ***corpus de consulta*** tem a finalidade de prover ao desenvolvedor o material de base para a seleção dos **conceitos** referentes ao domínio e facilitar o entendimento da área para a construção da ontologia. Deve ser composto por documentos representativos da área, os quais

possam servir de base de consulta para o desenvolvedor no decorrer do trabalho. A partir do conhecimento da área devem ser relacionados os requisitos para a coleta do **corpus de pesquisa** considerando: representatividade da área, possibilidade de tratamento automático, quantidade e qualidade dos textos e confiabilidade das fontes de consulta. Após a coleta, o *corpus* deve ser compilado e gravado seguindo critérios pré-estabelecidos, como: formato dos arquivos, nome dos arquivos e pastas, preservação do arquivo original para posterior consulta (se necessário) e verificação da necessidade de manutenção dos textos, executando a análise individual a fim de retirar informações irrelevantes, tais como imagens e cabeçalhos.

4.2.2 Conceitualização

O estágio de **conceitualização**, considerado por nós o estágio mais importante no desenvolvimento da ontologia, corresponde à descrição do modelo conceitual (conjunto de conceitos e relações) a ser organizado para o domínio tomando como base os requisitos estabelecidos nos estágios anteriores. Este é o estágio que consideramos fundamentalmente uma reinterpretação das metodologias descritas na seção anterior. Justificamos esta assertiva pelo fato de não centrarmos a estruturação da ontologia em uma etapa de extração de termos (como visto nas demais metodologias) uma vez que propomos a criação da estrutura de classes, subclasses e relações tendo como base o conhecimento descrito através dos *frames*.

A figura 13 destaca o estágio de conceitualização e cada uma de suas etapas enumeradas a fim de auxiliar na sua descrição e facilitar a organização do texto. Na sequência, explicamos cada etapa distintamente seguindo a numeração da figura.

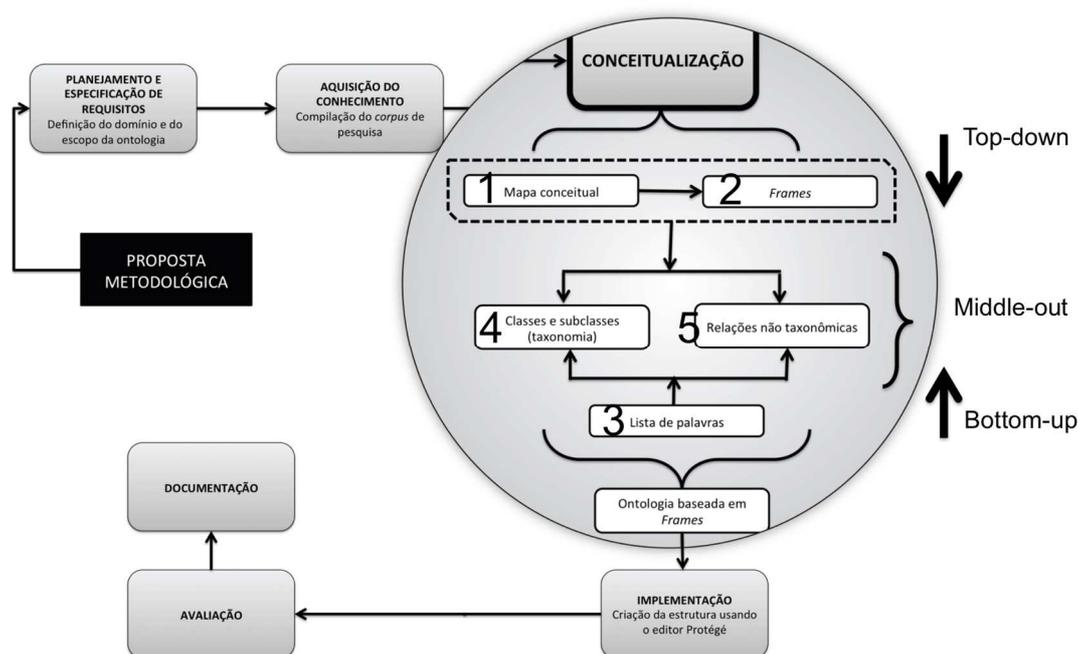


Figura 13: Etapas metodológicas propostas – foco na conceitualização

Fonte: elaborado pela autora

1. **Mapa conceitual:** nesta etapa o domínio deve ser representado através de um mapa conceitual após o estudo dos documentos elencados como *corpus* de consulta. Um mapa conceitual (NOVAK & GOWIN, 1996) é um esquema usado para representar um grupo de conceitos inseridos em uma rede de proposições, considerado um meio de estruturar o conhecimento através de uma forma visual para compartilhar significados, conceitos e relações. Segundo Novak & Gowin (1996), os mapas conceituais advêm da teoria da aprendizagem significativa desenvolvida por David Ausubel, que sugere a organização do conhecimento humano através de uma hierarquia de conceitos.
2. **Frames:** a partir da organização do conhecimento no mapa conceitual, é possível fazer uma primeira identificação dos *frames* relacionados ao domínio. A descrição dos *frames* envolve a proposição de uma glosa (definição), a identificação dos elementos de *frames* e a listagem das unidades lexicais evocadoras de cada *frame*. Vale reforçar que a organização do domínio em *frames* se dá de forma empírica e advém do conhecimento descrito no mapa conceitual. Após a identificação dos *frames* é possível, novamente com base nas informações do mapa, verificar quais são as relações entre os *frames* que devem ser destacadas. Ao finalizar as etapas 1 e 2 considera-se importante a consulta a especialistas da área no intuito de verificar se as descrições realizadas correspondem à realidade do domínio.

3. **Lista de palavras:** nesta etapa sugerimos o uso de uma ferramenta computacional de extração da informação linguística⁷³. A ferramenta atende a duas necessidades: listar de forma semiautomática as palavras referentes ao domínio e prover uma forma de visualização destas palavras em seu contexto. A lista de palavras será provida à partir do *corpus* de pesquisa. Esta lista é importante, pois permite visualizar exemplos concretos de uso dos conceitos contextualizados nos documentos ampliando a gama de relações não taxonômicas. Além disso, a lista será útil na identificação de unidades lexicais evocadoras de *frames* e de variantes (palavras equivalentes ou sinônimas), aspecto fundamental para a recuperação da informação e que precisa ser inserido na ontologia.
4. **Classes e subclasses:** partindo das categorias participantes legais, eventos legais, documentos legais e instituições legais, o mapa conceitual e os *frames* serão analisados de modo a auxiliar na organização da estrutura de classes e subclasses.
5. **Relações não taxonômicas:** a figura 13 também ilustra a abordagem *middle-out* para a estruturação das classes e subclasses na ontologia. É *middle-out* porque une as outras duas abordagens: ao elencarmos os conceitos organizando-os em um mapa conceitual do qual advém os *frames* do domínio, estamos usando a abordagem *top-down* – partindo dos conceitos mais amplos para especificá-los. E, ao listarmos as palavras relacionadas ao domínio e buscarmos suas relações com os conceitos, estamos seguindo a abordagem *bottom-up* – do mais específico (palavras) para o mais amplo (conceitos). Ambas, fornecerão subsídios para a organização da estrutura hierárquica e estabelecimento das relações.

Todas estas etapas permitiram a criação do modelo conceitual da **ontologia baseada em *frames*** através de uma definição minuciosa dos eventos e seus participantes de modo a prover uma estrutura capaz de cobrir conceitualmente o domínio. As informações advindas das etapas de análise conceitual, serão organizadas em quadros a fim de facilitar o estágio de implementação.

Concluído o estágio de conceitualização teremos a ontologia de domínio descrita de forma textual (*frames*) e gráfica. Esta representação será a base para a escolha das soluções na etapa de implementação no editor de ontologias.

⁷³ O termo “ferramenta computacional de extração da informação linguística” se refere aos sistemas capazes de processar automaticamente o *corpus* e extrair dele termos ou palavras representativos. No capítulo 5, ao detalharmos e exemplificarmos esta etapa, apresentaremos mais detalhes acerca das ferramentas de extração de informação linguística.

4.2.3 Implementação

Para a etapa de **implementação**, sugerimos seguir o Método 101 (NOY & McGUINNESS, 2001) e adotar o editor de ontologias **Protégé** em uma versão atualizada.

O Protégé 5.0⁷⁴ é uma ferramenta que permite construir ontologias de domínio, personalizar formulários de entrada de dados, inserir e editar dados, na qual possibilita-se a criação de bases do conhecimento guiadas por uma ontologia. É um editor de ontologias de código aberto e foi desenvolvido pela *Stanford Medical Informatics na Stanford University School of Medicine*.

O Protégé possui duas formas para o desenvolvimento de ontologias, via **Protégé-frames**⁷⁵ ou via **Protégé-OWL**, ambas com características comuns, porém diferentes linguagens de formalização.

No Protégé-frames, é possível construir e preencher ontologias com o protocolo OKBC (*Open Knowledge Base Connectivity*). Nesta interface, a ontologia é desenvolvida através de quadros baseados no domínio, adaptando a forma de entrada de dados e inserido diferentes dados de instanciação. O Protégé-Frames fundamenta-se nas ideias propostas por Minsky (1974), sendo que a noção de quadros/cenas deu origem às linguagens computacionais orientadas a objetos. Para o Protégé-Frames um quadro (*frame*) consiste em um conjunto de atributos (*slots*) que descrevem as características dos objetos representados através de seus valores. Os valores atribuídos a cada quadro também podem servir como ligação com outros quadros.

O Protégé-OWL é uma extensão do software que suporta a *Web Ontology Language* (OWL), especificada pela W3C como linguagem oficial de construção de ontologias para a *Web Semântica*. Uma ontologia desenvolvida em OWL pode incluir descrições de classes, propriedades e suas instâncias.

Optamos pela utilização do Protégé-OWL para a implementação da nossa ontologia, por se tratar de um editor que retorna um arquivo no formato OWL, suportado por aplicações para a *WebSemântica*, recomendado pela W3C. Temos ciência de que esta escolha nos

⁷⁴ As informações sobre o Protégé foram retiradas do Manual *A Practical Guide To Building OWL Ontologies Using Protégé 4 and CO-ODE Tools*, desenvolvido por Matthew Horridge, Simon Jupp, Georgina Moulton, Alan Rector, Robert Stevens, Chris Wroe, na University Of Manchester, 2007. Disponível em: <http://owl.cs.manchester.ac.uk/tutorials/protegeowltutorial/>. Acesso em: 10/01/2011

⁷⁵ O Protégé-frames é o editor que foi utilizado para a implementação do Método 101.

remeterá a buscar uma forma original de representar as relações entre *frames* na nossa ontologia, além de não termos a opção de utilizar o mesmo nome para diferentes entidades, uma vez que o Protégé-OWL não permite que duas unidades de diferentes nomes se refiram a um mesmo ente e nem dois nomes diferentes se refiram a um mesmo indivíduo.

Os esquemas quadros produzidos pela análise conceitual da etapa de conceitualização serão o ponto de partida para a implementação no editor de ontologias. Temos ciência de que a inclusão de classes e subclasses é uma atividade simples, porém para a implementação das relações, incluindo os papéis dos participantes, teremos o desafio de buscar alternativas compatíveis com o editor.

4.2.4 Avaliação e documentação

A etapa de **avaliação** incide em verificar se a estrutura ontológica proposta cumpre com a função à qual foi criada. Nos trabalhos das áreas de CI e da CC esta etapa corresponde a uma atividade efetivamente lógica, seguindo métricas para analisar os resultados.

Na literatura que trata de ontologias em PLN, Brewster et al. (2004) destacam uma forma de avaliação padrão que segue medidas de **precisão** e **abrangência** (também se pode encontrar o termo revocação). Estas medidas são usadas especialmente na área da RI.

Entende-se por precisão a proporção de documentos recuperados em uma busca e que foram considerados relevantes para o usuário e, por abrangência a proporção de documentos considerados relevantes que foram recuperados. Para a obtenção destes resultados, segue-se os seguintes cálculos: $P=AC/R$ (onde, P é precisão, AC é o número de acertos e R é a resposta da consulta) e $A=AC/C$ (onde, A é abrangência, AC é o número de acertos e C é o conjunto total de acertos). Para uso destas medidas, é necessário que se tenha um material previamente identificado (poderia ser chamado de modelo ou gabarito) com o qual serão feitas as comparações: um *corpus* etiquetado ou uma coleção de documentos relevantes.

No caso de avaliação das ontologias, no entanto, Brewster et al. (2004) defendem que a precisão corresponde a quantidade de conhecimento que foi corretamente identificado na ontologia em relação ao todo do conhecimento da ontologia. E a abrangência reflete o conhecimento que foi corretamente identificado com relação ao todo que deveria ser adquirido. A dificuldade, nesta abordagem, está em definir o todo do conhecimento que deve ser adquirido, uma vez que existem diferentes tipos de conhecimento. Assim, o problema

inerente à avaliação de ontologias reside no fato de que se precisa definir exatamente **o que** será avaliado e, para isto, ainda não há um consenso na comunidade acadêmica.

Outra abordagem de avaliação de ontologias sugerida na literatura (Hovy, 2002; Maedche e Staab, 2002) propõe que seja feita uma comparação entre a ontologia desenvolvida e uma outra considerada padrão (*Golden model*), analisando número de classes, propriedades e atributos. Esta abordagem, no entanto, nos parece inviável para avaliação de ontologias de domínio, pois, a menos que se tenha uma ontologia considerada padrão exatamente do mesmo domínio (o que dispensaria a construção de uma nova ontologia), a comparação seria insipiente.

Outra forma de avaliação disponível é uma ferramenta que avalia ontologias desenvolvidas em OWL disponível na *Web*, chamada OOPS! – *Ontology Pitfall Scanner*⁷⁶ (POVEDA-VILLALON et al., 2012). Esta ferramenta analisa a ontologia a fim de verificar se a modelagem está em conformidade com os padrões ou se apresenta algum tipo de inconsistência ou armadilha (*pitfalls*). É possível enviar o arquivo OWL para análise na página da OOPS!, e, usando o Jena API⁷⁷, a ferramenta faz uma varredura na estrutura ontológica e verifica automaticamente os problemas de acordo com um conjunto de erros. A listagem de erros⁷⁸ foi classificada pelas autoras em: erros de compreensão humana, consistência lógica, representação do mundo real e modelagem. Os erros encontrados são apresentados na tela, juntamente com uma explicação relacionada ao problema e sugestões de correção.

Considerando tais aspectos, vamos avaliar a ontologia seguindo duas diretrizes:

- 1- A submissão à ferramenta OOPS! para verificação das inconsistências (avaliação automática). A partir do resultado advindo da ferramenta, serão feitos os devidos ajustes.
- 2- O cálculo da abrangência (avaliação manual)
 - a. Escolha aleatória de quatro documentos do *corpus* de pesquisa;
 - b. Consulta ao especialista para marcação simples (sublinhando apenas) de quais são as informações relevantes presentes dos documentos;

⁷⁶ Disponível para acesso em <http://oops.linkeddata.es/>

⁷⁷ Jena API é um framework (biblioteca de aplicações) para desenvolvimento de sistemas para *WebSemântica* (<https://jena.apache.org/>).

⁷⁸ A listagem completa dos erros analisados pela ferramenta OOPS! pode ser lida no ANEXO A.

- c. Verificação da possibilidade de representação destas informações na ontologia;
- d. Cálculo do resultado seguindo a fórmula: $A=AC/C$.

Optamos pelo cálculo somente da abrangência em função da dificuldade de definir o todo do conhecimento que deveria ser adquirido pela ontologia, conforme apontado por Brewster et al. (2004)

Como última etapa do processo metodológico, seguindo os estágios elencados por Fernández-López et al. (1997), Uschold & Gruninger (1996) e Noy & McGuinness (2001), temos a **documentação**. Segundo os autores, esta etapa é fundamental para a manutenção da ontologia, permitindo que sejam recuperadas as informações essenciais acerca do seu desenvolvimento. Para isso, deve ser constituído um documento descrevendo passos de execução, escolhas, decisões tomadas no decorrer do percurso, padrões elencados, etc. No caso do nosso trabalho, entendemos que todas as informações acerca do desenvolvimento da ontologia estarão descritas no capítulo 5, portanto, constituindo-se na documentação da ontologia. No entanto, sugerimos que, em caso de reuso desta proposta metodológica, um documento com as informações de desenvolvimento seja redigido.

4.3 Esquema síntese do capítulo

As ideias desenvolvidas no capítulo podem ser visualizadas no esquema síntese (figura 14).

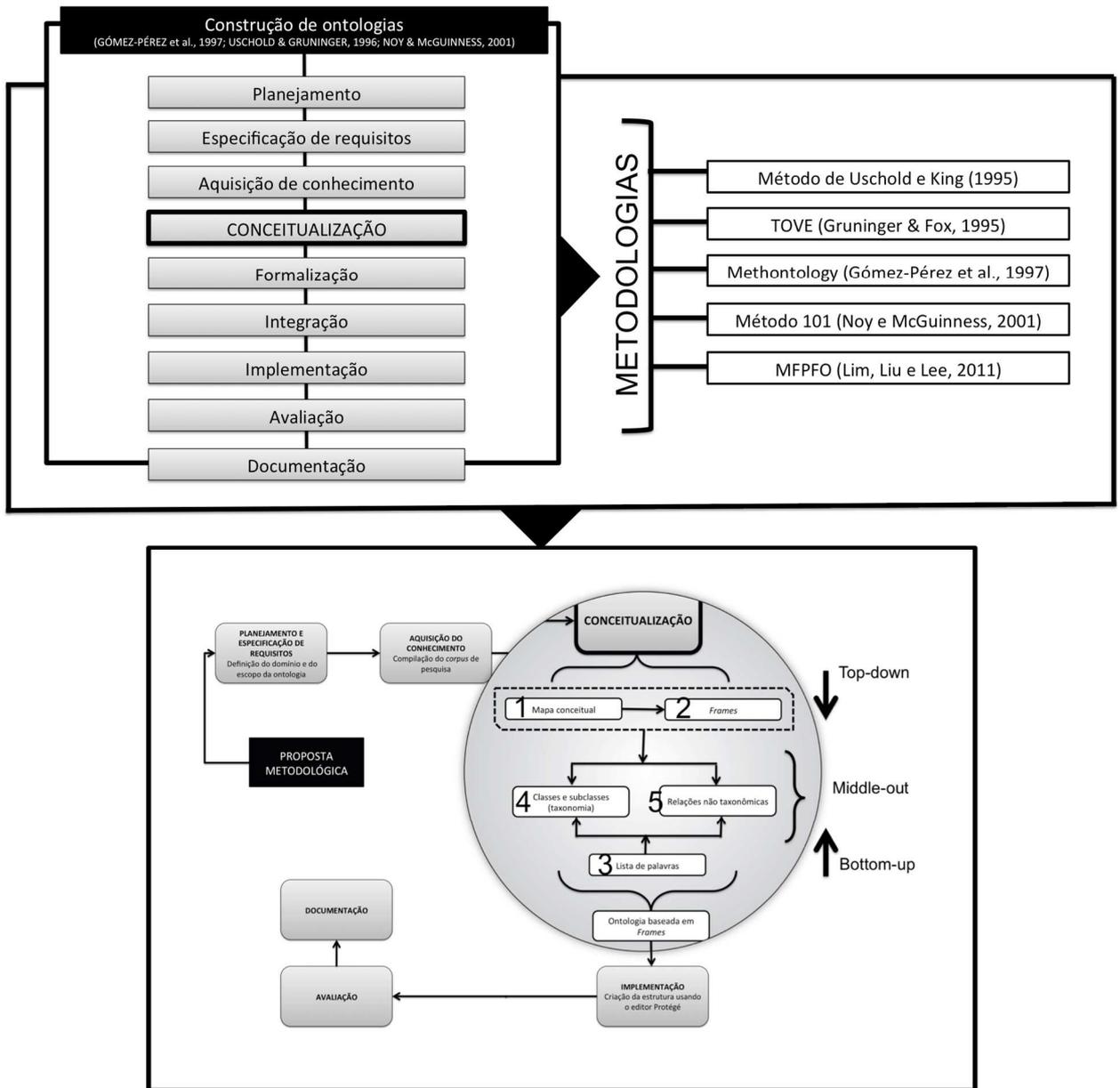


Figura 14: Esquema síntese do capítulo 4
Fonte: elaborado pela autora

No capítulo seguinte, apresentamos a modelagem da ontologia seguindo nossa proposta metodológica.

5 MODELAGEM E IMPLEMENTAÇÃO DA ONTOLOGIA

Nos capítulos anteriores apresentamos a fundamentação teórica que possibilitou a proposição de etapas metodológicas próprias para a construção da ontologia do domínio jurídico, bem como descrevemos a referida proposta de modo genérico através dos estágios previstos e suas principais ações.

Este capítulo tem como objetivo apresentar a modelagem da ontologia com base nas etapas metodológicas propostas, trazendo nossas reflexões acerca das informações e exemplificando cada etapa no decorrer do processo de desenvolvimento da ontologia.

Estruturamos o capítulo seguindo as etapas metodológicas descritas. Assim, o organizamos em seis seções a fim de apresentar o detalhamento de análise e desenvolvimento da ontologia. A seção 5.1 corresponde à etapa de planejamento e especificação de requisitos. Nesta seção são apresentados o domínio e o escopo da ontologia, situando-a na área do Direito Penal. Na seção 5.2 desenvolvemos a etapa de aquisição do conhecimento, na qual apresentamos nossa estratégia para conhecer o processo penal, bem como para elencar o *corpus* de pesquisa. A seção 5.3 oferece a descrição dos passos adotados para o cumprimento da etapa de conceitualização. Nesta seção são apresentadas as análises realizadas a partir do *corpus* e do conhecimento do domínio para a organização da estrutura de classes, subclasses e relações da ontologia à luz da SF. Seguindo as etapas previstas, na seção 5.4 descrevemos as soluções encontradas para a implementação dos dados através da ferramenta de edição de ontologias. Na seção 5.5 descrevemos as etapas adotadas para a avaliação da ontologia, bem como referenciamos a forma elencada como documentação de todos o processo de desenvolvimento. E, por fim, na seção 5.6 ilustramos a síntese deste capítulo com a visão parcial da ontologia piloto construída.

5.1 Planejamento e especificação de requisitos

Esta seção tem o objetivo de situar o estágio de planejamento e especificação de requisitos para o desenvolvimento da ontologia. Para isso, baseamo-nos nas perguntas

elencadas por Noy & Mc'Guinness (2001): *Que domínio a ontologia pretende cobrir? Para que a ontologia será utilizada?*. Organizamos a seção respondendo às questões.

Que domínio a ontologia pretende cobrir?

Esta pesquisa é vinculada ao projeto CNJ-Acadêmico, conforme já mencionado no capítulo de introdução deste trabalho, tendo como principal foco a área jurídica. Todavia, para a implementação da ontologia, uma vez que a área jurídica é ampla e possui diversas especificidades, é necessário estabelecermos alguns limites para o escopo da pesquisa e delimitação do *corpus*.

Para melhor entendimento dos limites estabelecidos como domínio e escopo deste trabalho, julgamos oportuno apresentar, ainda que brevemente, uma visão geral acerca do judiciário brasileiro. A partir desta visão geral será possível situar exatamente o escopo da pesquisa e definir claramente o domínio que pretendemos cobrir.

O Poder Judiciário é regido pela Constituição Brasileira de 1988 (BRASIL, 1988), tendo suas bases definidas no Capítulo III, artigos 92 a 126, ou seja, é a Constituição que demarca a estrutura e a organização do Poder Judiciário no Brasil, suas áreas e níveis de atuação, organizando-o em diferentes Tribunais e Justiças, cada um com suas competências.

O Judiciário Brasileiro apresenta uma estrutura hierárquica organizada em três instâncias, como se pode visualizar na figura 15. Além disso, a Justiça no Brasil se divide em Justiça Comum e Justiça Especializada. A primeira subdivide-se em **Estadual** – para solucionar casos entre pessoas, empresas, instituições ou crimes – e **Federal**, cuja função é julgar os casos de interesse da União, das autarquias ou das empresas públicas. Já a segunda, por ser especializada, tem a função de julgar casos referentes à Justiça do Trabalho, Justiça Eleitoral e Justiça Militar.



Figura 15: Organograma do Poder Judiciário.

Fonte: <http://www.guiadedireitos.org/images/stories/organograma-judiciario-inv-m.jpg>.
Acesso em 30/05/2015.

A **1ª instância** tem a função de analisar primeiramente os casos levados à Justiça sendo representada pelos Juízes dos fóruns e varas das variadas regiões do país que julgam os processos monocraticamente, isto é, individualmente. A **2ª instância** corresponde aos Tribunais de Justiça, onde são julgados os casos que sofreram apelação na instância anterior, sendo as decisões reavaliadas pelos Desembargadores, que podem ou não manter a decisão. Há ainda a **3ª instância**, que julga os casos controversos e que sofreram apelação na instância anterior, representada pelos Tribunais Superiores, nos quais os Ministros estipulam uma decisão final, não cabendo mais apelação. Aos Tribunais Superiores cumpre a função de garantir a interpretação unívoca da lei em todo o país.

O mais alto tribunal do Sistema Judiciário Brasileiro é o Supremo Tribunal Federal (STF), cuja principal função é garantir o respeito e cumprimento às leis federais ou estaduais

protegendo a Constituição Federal. É também responsabilidade deste tribunal o julgamento de políticos da esfera Federal, tais como o Presidente da República, os Senadores e Deputados.

A Justiça Comum se divide em Cível e Criminal, respeitando as esferas Nacional e Estadual. Enquanto que a Justiça Especializada corresponde a Justiça do Trabalho, Justiça Eleitoral e Justiça Militar, sendo estas responsáveis pelo julgamento de processos específicos de cada área e envolvendo seus servidores, como é o caso dos militares.

Na área Cível, a Justiça Comum julga casos relativos aos bens da população em geral – compra e venda e imóveis, veículos, crédito e transações comerciais, indenizações, etc – e questões de família – casamentos, divórcios, guarda e adoção de crianças, testamentos e heranças, etc. Os casos de menor complexidade, cujo valor do prejuízo não ultrapasse 40 salários mínimos podem ser julgados nos Juizados Especiais sendo mais rapidamente solucionados. Já os casos que ultrapassam o limite citado acima devem ser julgados nas Varas⁷⁹ Cíveis, sendo necessária a condução por um advogado representando as partes interessadas na ação.

Na área Criminal são julgados os crimes cometidos pelos cidadãos, desde que sejam caracterizados como crimes conforme o Código de Processo Penal. Os casos graves – roubo, agressão física, tráfico de drogas – são julgados pelos juízes das Varas Criminais, já os casos considerados gravíssimos – envolvendo mortes ou intenções de matar (crimes dolosos) – são julgados no Tribunal do Júri, composto de um júri popular que decide pela culpa ou não do acusado, cabendo ao juiz a definição da pena a ser cumprida.

Casos que envolvam crimes menos graves e que resultem em penas de até dois anos podem ser julgados nos Juizados Especiais Criminais, conhecidos como JECrim. O JECrim julga os casos com maior rapidez objetivando a conciliação entre as partes envolvidas a fim de prover a reparação dos danos causados ou a aplicação de penas que não envolvam a prisão.

Como se pode ver, ainda que em uma breve descrição, o Sistema Judiciário Brasileiro é complexo e possui várias ramificações, cada uma com suas especificidades. Esta visão

⁷⁹ “**Vara:** é área judicial em que o juiz de primeira instância exerce sua autoridade. Elas podem ser únicas (atendem a todo e qualquer processo), civis, criminais ou atender a outras especializações e são numeradas de acordo com o número de juízes da comarca. As varas especializadas são criadas de acordo com as demandas locais, sendo as mais comuns, além das já mencionadas: Vara da Família e das Sucessões, Vara da Infância e da Juventude, Vara de Acidentes do Trabalho e Vara da Fazenda Pública. Os Tribunais de Justiça não seguem essa divisão” (GUIA DE DIREITOS, disponível em http://www.guiadedireitos.org/index.php?option=com_content&view=article&id=672&Itemid=148. Acesso em 03/06/2015).

ampla do Sistema Judiciário Brasileiro foi implementada na OntoJuris (ARAÚJO et al, 2008) conforme mencionamos no capítulo 2 como exemplo de ontologia jurídica de domínio.

No nosso entendimento, cada área e subárea do Sistema Judiciário pode ser representada em uma ontologia de domínio em função das suas especificidades e características peculiares, pois, ainda que muitos eventos e participantes sejam semelhantes, estes exercem papéis diferentes em cada etapa do processo, em cada instância e em cada esfera do judiciário. Assim, entendemos que o domínio jurídico brasileiro precisa ser implementado em um conjunto de ontologias que atendam a diferentes subdomínios, ou seja, o domínio jurídico brasileiro deve ser implementado através de uma **rede de ontologias**⁸⁰.

Cumpre-se destacar que, neste trabalho, tomamos como foco uma destas ontologias e apenas sugerimos que a descrição ampla do domínio jurídico brasileiro seja realizada através de um conjunto de ontologias interligadas, ou seja, a rede de ontologias, não sendo nosso compromisso explicitar a forma como esta integração deva ser realizada. Contudo, vislumbramos que as etapas metodológicas propostas nesta pesquisa cumpram com o papel de embasar a construção das demais ontologias dos subdomínios jurídicos.

Diante do exposto, restringimos o domínio de cobertura da nossa ontologia na Justiça Comum Criminal, mais especificamente nos Juizados Especiais Criminais – JECrim – descrevendo o procedimento sumaríssimo. Os Juizados Especiais Criminais, previstos pela Lei nº 9.099/95, foram criados para tratar especificamente de *infrações penais de menor potencial ofensivo*, aquelas consideradas de menor gravidade. A lei considera que tais infrações correspondem a contravenções penais cuja pena atinja no máximo dois anos; como exemplos deste tipo de crime podemos citar: agressão física sem causar lesão, ameaças, lesão corporal leve, entre outros.

A criação destes juizados ocorreu em função das grandes demandas dos juizados criminais, visando desafogar os sistemas carcerário e judiciário. Para isso, são privilegiados os princípios da oralidade, da informalidade, da economia processual e da celeridade, resultando em penas menos punitivas e com maior cunho social que podem cominar em reparação de

⁸⁰ **Rede de ontologias** é um conceito desenvolvido na âmbito da Metodologia NeON (SUAREZ-FIGUEROA, GOMEZ-PEREZ & FERNANDEZ-LOPEZ, 2012), cujo foco é descrever modos de desenvolvimento de ontologias em um contexto multidisciplinar contemplando o uso de várias ontologias. Nesta metodologia o desenvolvimento se dá com base em cenários. Não apresentamos esta metodologia por ela envolver a integração de várias ontologias, não sendo o foco deste trabalho.

danos sofridos pela vítima ou aplicação de pena não privativa de liberdade (art. 62 da Lei nº 9.099/95).

Para que a ontologia será utilizada?

Atualmente há uma grande quantidade de informações no sistema judiciário que precisam ser organizadas e compiladas e fim de facilitar o acesso. Uma das iniciativas para facilitar o acesso às informações jurídicas é o processo eletrônico que foi instituído pelo Governo Federal através da Lei nº 11.280, de 16 de fevereiro de 2006, que dispõe sobre a informatização do processo judicial, alterando a Lei nº 5.869/1973 do Código de Processo Civil. Antes da lei, porém, alguns Tribunais do País já faziam uso de diferentes recursos eletrônicos para dar andamento aos processos, caracterizando-se como uma primeira fase (o início) do processo eletrônico.

Em 2006, o Conselho Nacional de Justiça (CNJ) tomou a iniciativa de lançar o projeto PROJUDI – Processo Judicial Digital⁸¹ – que consiste em um sistema de processo eletrônico que visa informatizar todos os tribunais do país para que todo o trâmite de um processo judicial se dê em meio informatizado, substituindo os atos processuais realizados em papel. A partir dessa iniciativa do CNJ, os Tribunais passaram a apresentar maior fluxo de informações internas e externas não só para os advogados e juízes, como também para a população. Atualmente vários Tribunais desenvolvem suas ferramentas para acompanhar a evolução digital dos processos.

Apesar de o cenário apontar para grandes melhorias no acompanhamento dos processos, ainda existem diversas dificuldades a serem ultrapassadas, tais como: limitações técnicas, limitações de estrutura de rede, a falta de interligação entre os diversos sistemas, entre outras. Alguns profissionais do direito afirmam que atualmente o que realmente ocorre é um *procedimento eletrônico*, diferenciando-se do que os especialistas chamam de *processo eletrônico* (FULLY e BERNARDES, 2011).

Corroborando com esta questão, Ruschel e Rover (2011) salientam que os atuais sistemas são limitados no que concerne à integração com outros sistemas, bem como nos processos de automatização para acesso à informação. No entanto, a proposição do edital CNJ n.020/2010, ao qual corresponde esta pesquisa, evidencia a preocupação por parte do Poder Judiciário de avançar na solução destas dificuldades.

⁸¹ Disponível em < <https://projudi.tjba.jus.br/projudi/> > Acesso em agosto de 2013.

Considerando tais aspectos, uma ontologia é útil no sentido de colaborar em sistemas que visem à integração informatizada dos tribunais, possibilitando a recuperação da informação e minimizando o trabalho manual de juízes e advogados nos sustento de suas teses de defesa ou de acusação, bem como na fundamentação das sentenças.

Além disso a ontologia pode ser utilizada para promover uma maior transparência do sistema judiciário, facilitando a sua compreensão. Mais especificamente, pensando em relação aos processos dos Juizados Criminais Especiais (nos quais, *à priori*, não se faz necessária a presença de um advogado), a ontologia, aliada a um sistema de recuperação da informação ou a um agente inteligente, pode ser útil para auxiliar o usuário leigo e trazer-lhe informações sobre como proceder ou até mesmo no andamento do seu processo.

5.2 Aquisição do conhecimento

Nesta pesquisa, o processo de aquisição do conhecimento iniciou-se através de reuniões e entrevistas com especialistas da área – o Juiz de Direito da 2ª Vara Cível de Campo Bom e um advogado criminalista de Novo Hamburgo – além da leitura de doutrina e leis, pesquisas na *web* sobre o assunto e busca de jurisprudências. Para melhor compreender o domínio foi estudado o *Manual de Procedimentos de Juizados Especiais Criminais*⁸² elaborado pelo Supremo Tribunal Federal e pelo Conselho Nacional de Justiça, o *Código do Processo Penal*⁸³ bem como sua atualização de 2011⁸⁴ e a lei que fundamenta a criação dos juizados especiais (Lei nº 9.099/95⁸⁵). A leitura dos documentos possibilitou vislumbrar a estrutura do domínio, seus principais eventos e respectivos participantes auxiliando na construção das etapas seguintes. É importante salientar que, por ser um domínio complexo, a aquisição do conhecimento perpassou todo o desenvolvimento do trabalho e, sempre que necessário, os especialistas foram convidados a auxiliar.

⁸² Documento disponível em <http://www.mpce.mp.br/orgaos/CAOCRIM/manuais/manualJERIM.pdf>. Acesso em dezembro de 2014.

⁸³ Documento disponível em http://www.planalto.gov.br/ccivil_03/decreto-lei/Del3689.htm. Acesso em dezembro de 2014.

⁸⁴ Documento disponível em http://www.planalto.gov.br/ccivil_03/ato2011-2014/2011/lei/112403.htm. Acesso em dezembro de 2014.

⁸⁵ Documento disponível em http://www.planalto.gov.br/ccivil_03/Leis/L9099.htm. Acesso em dezembro de 2014.

Tendo definido o domínio e o escopo da ontologia, passamos à definição do *corpus* de pesquisa o qual precisa ser representativo possibilitando a cobertura da área criminal. Para isso, elencamos um ***corpus de consulta***, composto pelos documentos listados acima e um ***corpus de pesquisa***.

O ***corpus de pesquisa*** tem a finalidade de prover uma lista de palavras a serem utilizadas na análise para a construção das classes e subclasses da ontologia, bem como analisar os *frames* relacionados e também a definição das relações existentes. Em função do domínio escolhido, encontramos algumas dificuldades para a coleta do *corpus* de pesquisa. Isso se deve ao fato de o JECrim ser julgado exclusivamente em 1ª instância, com isso as sentenças prolatadas nos processos não estão disponíveis para acesso livre *on-line*, tais como os acórdãos e as jurisprudências. Contudo, o contato com o Juiz de Direito nos possibilitou acesso a um conjunto de termos de audiência que poderão ser usadas na pesquisa desde que tenham as informações sigilosas (nomes, locais, etc) suprimidas.

Tentando solucionar o problema do *corpus* de pesquisa, acessamos o site do Tribunal de Justiça do Rio Grande do Sul (TJ-RS)⁸⁶ e percebemos que havia possibilidade de consulta a jurisprudências relacionadas a Ações Penais do Procedimento Sumaríssimo. Acessamos os documentos e, ao analisá-los, percebemos que se tratavam de ações instauradas nos Juizados Especiais, mas que correspondiam a crimes de menor potencial ofensivo contra a mulher ou relacionados a figuras da política e, portanto de competência de instância superior. Porém, ainda que as jurisprudências⁸⁷ relatassem casos que enviados para a 2ª instância, era possível identificar os eventos do procedimento sumaríssimo em 1ª instância.

Seguimos a mesma premissa para buscas nos sites do Tribunal de Justiça de Santa Catarina (TJ-SC)⁸⁸ e no Tribunal de Justiça do Paraná (TJ-PR)⁸⁹. No primeiro, encontramos a

⁸⁶ O site pode ser acessado em <http://www.tjrs.jus.br/site/>

⁸⁷ De acordo com Dimoulis (2011), a jurisprudência corresponde a um documento que institui um conjunto de decisões sobre determinada situação com base nas leis feitas pelos tribunais e que passam a ser usadas como fonte de consulta para novas decisões.

⁸⁸ O site pode ser acessado em <http://www.tjsc.jus.br/>

⁸⁹ O site pode ser acessado em <https://www.tjpr.jus.br/>

possibilidade de acessar os Termos Circunstanciados⁹⁰ e, no segundo, encontramos as Decisões Monocráticas⁹¹.

Consideramos que o *corpus* constituído é representativo, haja vista que inclui termos de audiências, jurisprudências, termos circunstanciados e decisões monocráticas; todos em formato digital e de acesso livre. Assim, determinamos a abrangência do *corpus*, limitando entre os três estados da Região Sul. O *corpus* de pesquisa corresponde a um total de 48384 palavras distribuídas em 70 documentos, conforme a organização do quadro 2.

Termos de Audiência	
Comarca de Campo Bom Vara criminal	12 documentos – período de 01/01/2013 até 01/01/2015
Jurisprudências	
TJ-RS	10 documentos – período 01/01/2013 até 01/01/2015
Termo Circunstanciado	
TJ-SC	40 documentos – período 08/04/2008 até 01/01/2015
Decisão monocrática	
TJ-PR	8 documentos - período 01/01/2013 até 01/01/2015
TOTAL DE DOCUMENTOS	70
TOTAL DE PALAVRAS	48.384

Quadro 2: Constituição do *corpus* de pesquisa.
Fonte: elaborado pela autora

Concluídas as etapas de planejamento, especificação de requisitos e aquisição do conhecimento temos todas as informações necessárias para o início da modelagem conceitual da ontologia.

5.3 Conceitualização

Entendemos o estágio de conceitualização como nossa principal contribuição para uma proposta metodológica de construção de ontologia baseada em *frames*. Nesta seção vamos

⁹⁰ O **termo circunstanciado** é um documento elaborado pela autoridade policial cuja finalidade é relatar um crime de menor potencial ofensivo para ser levado ao Juizado Especial Criminal.

⁹¹ **Decisão monocrática** é **decisão** proferida por um único magistrado, de qualquer instância ou tribunal. A **decisão monocrática**, em primeira instância, é a regra.

explicitar as etapas de análise seguindo a mesma ordem apresentada no capítulo anterior: mapa conceitual, *frames*, lista de palavras, classes e subclasses e relações não taxonômicas, culminando na organização da ontologia de forma textual através de quadros. Para melhor organização do texto optamos por dividir a seção em cinco subseções referentes a cada um dos passos supracitados.

5.3.1 Mapa Conceitual

Construir o mapa conceitual do procedimento sumaríssimo foi uma tarefa que exigiu um longo tempo e um conhecimento amplo da área. Visando conhecer detalhes acerca desse procedimento dedicamo-nos à leitura das leis relacionadas e também do referido manual. Para melhor entendimento do mapa consideramos oportuno explicar brevemente cada evento identificado:

- **Crime de menor potencial ofensivo:** considerados os crimes de menor gravidade e que resultam em penas de restrição de direitos ou multa, envolvendo o máximo de dois anos. Este crime deve ser registrado em uma delegacia de polícia, gerando um boletim de ocorrência.
- **Termo circunstanciado:** é lavrado pelo delegado de polícia a partir do boletim de ocorrência e é encaminhado para o Juizado Especial Criminal para dar início ao processo.
- **Audiência preliminar de conciliação:** nesta audiência o conciliador tenta um acordo entre a vítima e o autor do fato, visando o encerramento do processo.
- **Composição:** na audiência preliminar tenta-se a composição, ou seja, o conciliador tenta uma forma de acordo entre as partes que pode ser: composição de danos (o autor do fato repara os danos à vítima) ou transação penal (o autor do fato declara-se culpado e aceita uma das penas impostas). Caso haja composição, o juiz homologa e encerra-se o processo. Caso não seja aceita a composição, o processo segue.
- **Proposta de transação penal:** corresponde ao evento no qual o autor do fato declara-se culpado e aceita uma pena pecuniária: restrição de direitos ou multa, doações, serviços comunitários, etc. Sendo aceita a proposta de transação penal, o juiz homologa e encerra-se o caso.

- **Homologação:** nesta fase inicial do processo o juiz participa somente homologando o acordo a fim de encerrar o processo.
- **Denúncia:** no caso de não haver acordo na composição e nem na proposta de transação penal, o Ministério Público encaminha a denúncia, dando prosseguimento ao processo. No caso deste evento cabe uma ressalva: os documentos lidos referem-se à denúncia ou queixa. **Denúncia** corresponde aos casos típicos, quando o crime é considerado ação pública, sendo instaurado pelo Ministério Público. Já a **queixa** corresponde aos casos de menor ocorrência, relacionados a ações privadas, instaurados pela própria vítima e seu advogado. Em relação a este evento, optamos por seguir os casos típicos e de maior ocorrência, ou seja, a denúncia.
- **Audiência de instrução e julgamento:** nesta audiência há a participação do juiz de 1ª instância e são ouvidas as partes para esclarecimento dos fatos. São arroladas testemunhas para participarem desta audiência (tanto de defesa, quanto de acusação). Para esta audiência são levadas as provas.
- **Oitiva das testemunhas:** as testemunhas de acusação e de defesa são ouvidas pelo juiz.
- **Interrogatório:** após a oitiva das testemunhas, o acusado é interrogado pelo juiz.
- **Debates orais:** após a oitiva das testemunhas e o interrogatório do réu, a acusação e a defesa dispõem de vinte minutos (prorrogáveis por mais dez) para sustentarem oralmente suas teses. Em caso de impossibilidade de oitiva de todas as testemunhas ou alguma outra impossibilidade de conclusão da audiência no mesmo dia, os debates orais poderão ser substituídos por memoriais (documentos que relatam os depoimentos e os fatos).
- **Prolação da sentença:** concluída a audiência de instrução e julgamento o juiz profere a sentença. É gerado um termo de audiência no qual são registrados o resumo dos debates orais e a fundamentação da sentença.
- **Recurso:** pode ser instaurado após a homologação ou a prolação da sentença, caso uma das partes sinta-se prejudicada pelo resultado. O recurso é julgado por uma turma recursal composta por três juízes de 1ª instância. É interessante atentar que a turma recursal faz as vezes de 2ª instância, uma vez que há um recurso em relação ao resultado do processo.

Após a identificação dos eventos, passamos a analisá-los separadamente a fim de verificar os participantes envolvidos (sejam animados ou inanimados). Na primeira parte do processo, na qual se tenta o acordo entre as partes, o juiz tem apenas o papel de analisar o acordo e homologar a decisão com vistas ao encerramento do processo. Nesta etapa inicial há um conciliador, papel exercido por um bacharel em Direito (não necessariamente um advogado) que tem a função de tentar o acordo entre as partes. O juiz entra como autoridade julgadora somente quando não houver o acordo entre as partes, instaurando-se a audiência de instrução e julgamento. Estas informações foram também incluídas no mapa conceitual relacionando-as aos respectivos eventos.

Outro detalhe acerca dos participantes marcado no mapa foi em relação à participação do acusado no evento de proposta de transação penal. Como este evento visa a um acordo no qual o autor assume-se culpado, ele passa a ser chamado de **beneficiário**, pois a proposta de transação penal possibilita-lhe uma pena menor, beneficiando-o e evitando que seja condenado no final do processo.

E, concomitantemente, verificamos as relações entre os eventos, identificando que há uma sequência e uma dependência entre eles. Por exemplo: para que haja uma **audiência de conciliação** é necessário que tenha sido lavrado um **termo circunstanciado** em uma delegacia. Esta sequência e dependência ocorre entre vários eventos envolvidos no processo e precisou ser tratada no mapa a partir das relações.

Após esta análise acerca dos eventos, participantes e as relações passamos a criação do mapa conceitual efetivamente. Para construção do mapa usamos o software CmapTools⁹² que é uma ferramenta para elaboração de mapas conceituais, auxiliando na sua estruturação gráfica.

O mapa conceitual elaborado após a análise do procedimento sumaríssimo pode ser visualizado na figura 16.

⁹² É possível fazer o download gratuito do CmapTool em <http://www.baixaki.com.br/download/cmapttools.htm#ixzz3gd0omR5V>

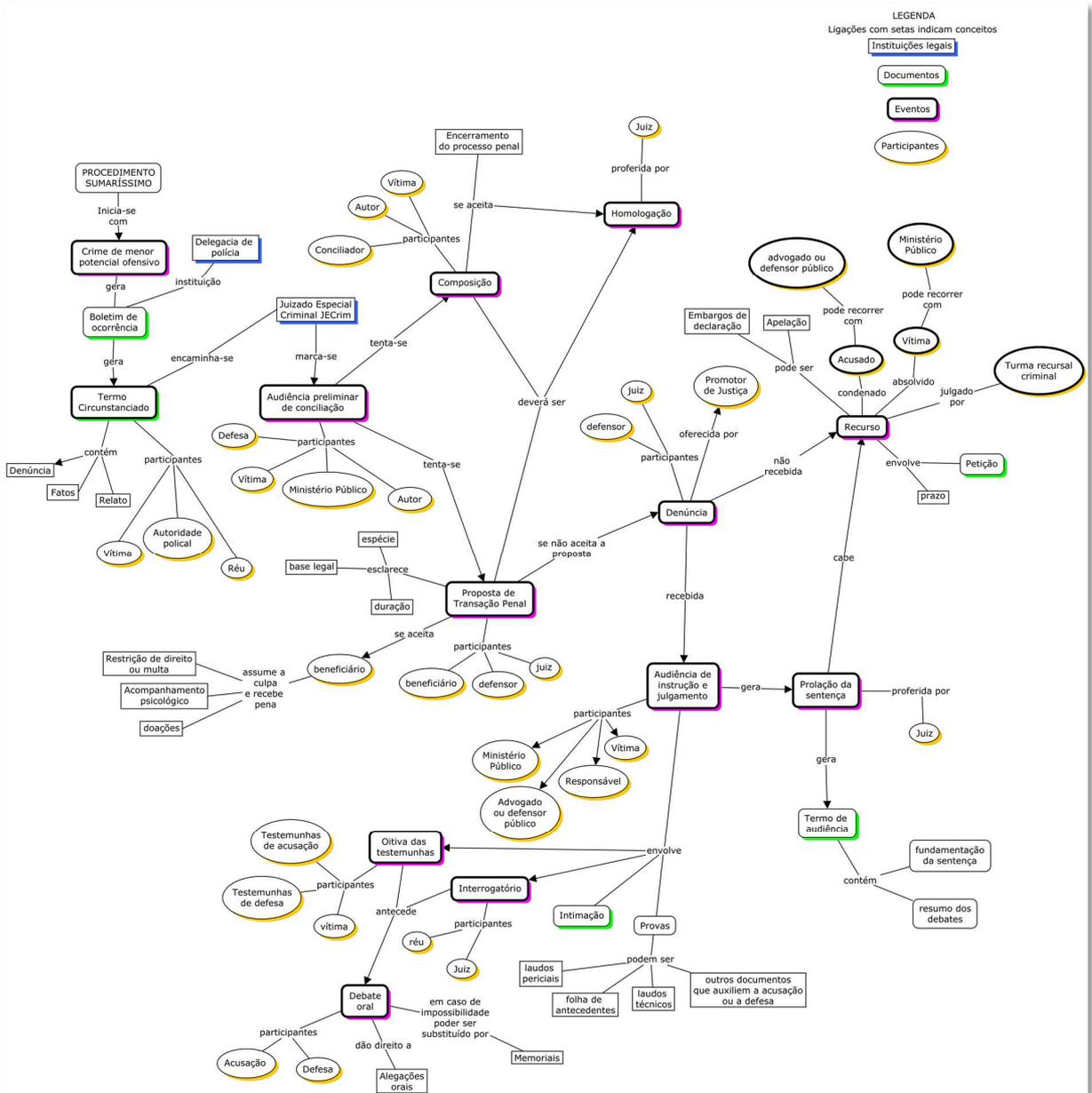


Figura 16: Mapa conceitual do Procedimento Sumaríssimo.

Fonte: elaborado pela autora.

A atividade final relacionada ao mapa conceitual foi buscar a identificação dos conceitos nele ilustrados a partir das categorias elencadas para a construção da ontologia – eventos legais, instituições legais, participantes legais e documentos legais. Como se pode ver na legenda, marcamos cada categoria com uma cor a fim de proporcionar uma melhor identificação.

Consideramos o mapa conceitual o ponto de partida para a construção dos *frames* do domínio, bem como a base para o estabelecimento das classes, subclasses e relações taxonômicas na ontologia; etapas estas que serão descritas na sequência.

5.3.2 *Frames*

De posse do mapa conceitual descritivo do procedimento sumaríssimo passamos a analisar quais seriam os *frames* relacionados a esse procedimento. Para a identificação dos *frames*, analisamos o mapa conceitual a fim de listar as cenas, todas representadas através de eventos. A partir desta primeira reflexão, elencamos treze *frames* a serem descritos para cobrir o domínio.

Para a descrição dos *frames* seguimos uma abordagem empírica descrevendo-os através do conhecimento da área adquirido nas etapas anteriores. Além disso, consultamos os documentos do *corpus* de consulta, o dicionário especializado de Plácido e Silva (2008) e, sempre que necessário, contamos com o apoio do especialista.

O primeiro passo para a construção de cada *frame* foi retomar as descrições dos eventos que embasaram o mapa conceitual, a fim de possibilitar a construção do *frame* através de uma glosa (definição). Entendemos que esta glosa precisa conter uma visão geral acerca da cena que o *frame* descreve, bem como apresentar o envolvimento dos participantes. É importante salientar que a construção da glosa segue os princípios da FN, e não da Lexicografia.

O próximo passo foi verificar no mapa conceitual quais são os participantes de cada evento, bem como as demais informações envolvidas a fim de delimitar os elementos de *frame*. Além disso, o conhecimento da dinâmica da área, adquirido e representado no mapa conceitual, sustentaram a definição dos elementos de *frame*. Em relação à definição dos elementos de *frame*, cumpre-se destacar que não seguimos a abordagem da FN, na qual são incluídas as noções de valência e tipologia definido elementos nucleares e periféricos através da anotação de sentenças. Justifica-se tal estratégia, levando-se em conta que as construções frasais que constam dos documentos jurídicos têm características semântico-sintáticas que nos impedem de seguir à risca a metodologia da FN de levar em conta predicados e argumentos.

Por fim, identificamos, a partir dos documentos do *corpus* de consulta, as unidades lexicais evocadoras dos *frames*. Note-se que aqui há uma interseção com a etapa de listagem de palavras que, após a análise, proverá outras unidades lexicais relacionadas aos *frames*, porém partindo do *corpus* de pesquisa. É necessário destacar que a proposição de *frames* não é o intento principal da metodologia; contudo, para ilustrar nossa proposta, precisamos partir de um certo número de *frames*.

O primeiro *frame* descrito foi *CRIME_MENOR_POTENCIAL_OFENSIVO*. Para a descrição deste *frame* identificamos na Lei nº 9.099/95 a caracterização deste tipo de crime e avaliamos quais seriam os elementos de *frame* (também através do texto da lei), bem como as unidades lexicais relacionadas. O *frame* pode ser visualizado no quadro 3.

FRAME: <i>CRIME_MENOR_POTENCIAL_OFENSIVO</i>	
Glosa:	<i>CRIME_MENOR_POTENCIAL_OFENSIVO</i> corresponde aos crimes de menor relevância, com ações julgadas e processadas nos ÓRGÃO JULGADOR. São definidos com base na LEI (nº 9.099/95). Correspondem aos processos penais de crimes com PENAS cominadas em até dois anos (TEMPO). Envolve um AUTOR que pratica o delito e uma VÍTIMA que sofre o delito.
Elementos de frame:	ÓRGÃO JULGADOR: juizados especiais criminais responsáveis pelo julgamento dos crimes de menor potencial ofensivo. LEI: base legal que define estes tipos de crimes. TEMPO: tempo máximo de pena prevista que caracteriza o crime de menor potencial ofensivo. PENAS: penalidades imputadas ao autor do crime. AUTOR: aquele que realizou a prática delituosa. VÍTIMA: quem sofreu a prática delituosa.
Unidades lexicais:	Crime, pena, potencial ofensivo

Quadro 3: *Frame CRIME_MENOR_POTENCIAL_OFENSIVO*

Fonte: Elaborado pela autora

Após a prática delituosa considerada um crime de menor potencial ofensivo há o registro do boletim de ocorrência em uma delegacia de polícia. A partir do boletim de ocorrência, lavra-se o termo circunstanciado que é enviado, juntamente com a vítima e o autor do fato, ao Juizado Especial Criminal, instaurando-se o início do procedimento sumaríssimo. Com o auxílio especializado do Juiz de Direito de 2ª Vara Cível de Campo Bom, identificamos que o boletim de ocorrência e a lavratura do Termo Circunstanciado não correspondem a *frames*, mas sim a documentos que estão incluídos no *frame* de *INÍCIO_PROCEDIMENTO*. Elencamos então, o segundo *frame*, que pode ser visualizado no quadro 4.

FRAME: INÍCIO_PROCEDIMENTO	
Glosa:	O início do procedimento se dá através do registro da ocorrência pela VÍTIMA em uma delegacia de polícia na qual a AUTORIDADE POLICIAL que lavra o Termo Circunstanciado (DOCUMENTO). Após a lavratura do termo, o AUTOR DO FATO e a VÍTIMA são encaminhados ao Juizado Especial Criminal (LOCAL) para o encaminhamento do processo.
Elementos de frame:	AUTORIDADE POLICIAL: é o delegado responsável pela lavratura do Termo Circunstanciado contendo o registro de ocorrência. AUTOR DO FATO: é o responsável pelo crime de menor potencial ofensivo. VÍTIMA: quem sofreu a prática delituosa. DOCUMENTO: termo circunstanciado lavrado na delegacia LOCAL: Juizado Especial Criminal que dará andamento ao processo.
Unidades lexicais:	Lavrar, lavratura, registro de ocorrência

Quadro 4: *Frame INÍCIO_PROCEDIMENTO*
 Fonte: Elaborado pela autora

Após o recebimento do termo circunstanciado em um Juizado Especial Criminal, é marcada uma audiência preliminar que tem por objetivo promover uma conciliação entre vítima e autor do fato. Nesta audiência há a presença de um conciliador (bacharel em Direito), que busca estabelecer um acordo entre as partes. Caso a audiência resulte em uma composição (termo técnico que se refere a conciliação entre as partes), o juiz homologa e encerra o caso. Para representar este evento, criamos o *frame AUDIÊNCIA_PRELIMINAR*, representado no quadro 5.

FRAME: AUDIÊNCIA_PRELIMINAR	
Glosa:	Com o termo circunstanciado ajuizado em um dos juizados especiais, ocorre a <i>AUDIÊNCIA_PRELIMINAR</i> que é realizada do modo informal com a presença da AUTORIDADE CONCILIADORA e DEFENSOR. Nesta buscam uma conciliação entre AUTOR DO FATO e VÍTIMA, resultando em uma COMPOSIÇÃO que deverá ser homologada pelo juiz.
Elementos de frame:	AUTORIDADE CONCILIADORA: é o participante designado para promover o debate entre as partes visando uma conciliação em relação ao fato quando a lei permitir (bacharel em Direito). AUTOR DO FATO: responsável pelo crime de menor potencial ofensivo realizado. VÍTIMA: quem sofreu o crime cometido ou se sente lesado por ele. DEFENSOR: advogados de ambas as partes, cuja função é defender os interesses dos envolvidos. Poderá ser designado um defensor público. DECISÃO: resultado da audiência que pode ser de conciliação ou composição de danos. Em caso de composição, encerra-se o processo. INFRAÇÃO PENAL: crime de menor potencial ofensivo que está sendo discutido na audiência.
Unidades lexicais:	Conciliação, conciliador, audiência preliminar

Quadro 5: *Frame AUDIÊNCIA_PRELIMINAR*
 Fonte: Elaborado pela autora

A audiência preliminar pode se encaminhar para dois eventos distintos, conforme o resultado do acordo entre vítima e autor. Pode-se decidir pela composição ou pela transação penal. A composição corresponde ao evento no qual se estabelece o acordo entre vítima e autor do fato, podendo resultar em uma conciliação (ambos decidem por encerrar o processo

com base no acordo efetivado) ou composição de danos (o autor do fato paga uma indenização ou ressarcir os danos da vítima). Havendo a composição, o juiz homologa a decisão e encerra-se o processo. Entendemos que composição corresponde a um subevento de audiência preliminar e criamos o *frame* **CONCILIAÇÃO** (quadro 6) para representá-lo. Cabe aqui um esclarecimento acerca do termo escolhido para designar o *frame*: o termo técnico para o evento é **composição**, porém optamos por usar **conciliação** para nomear o *frame*, objetivando um nome mais representativo semanticamente.

FRAME: CONCILIAÇÃO	
Glosa:	Corresponde a uma tentativa de que VÍTIMA e AUTOR DO FATO entrem em acordo em relação à INFRAÇÃO PENAL. A AUTORIDADE CONCILIADORA deve buscar as possibilidades de conciliação. Resulta em um ACORDO entre as partes envolvidas, o qual deverá ser homologado pelo juiz.
Elementos de frame:	VÍTIMA: quem sofreu dano AUTOR DO FATO: é o responsável pelo crime de menor potencial ofensivo. INFRAÇÃO PENAL: crime de menor potencial ofensivo que está sendo discutido na audiência. AUTORIDADE CONCILIADORA: é o participante designado para promover o debate entre as partes visando uma conciliação em relação ao fato. ACORDO: decisão tomada no ato de conciliação, podendo encerrar o processo, sugerir composição de danos.
Unidades lexicais:	Conciliar, conciliação, acordo

Quadro 6: *Frame* **CONCILIAÇÃO**
 Fonte: Elaborado pela autora

No caso de não haver a composição, há a possibilidade de transação penal. A transação penal é o evento no qual o autor do fato tem a oportunidade de ter sua pena reduzida caso assuma a culpa pela infração penal. Neste caso, além de ter a pena reduzida o autor do fato exime-se de ser considerado condenado no final do processo. A transação penal pode ser de pagamento de multa, restritiva de direitos, serviços comunitários, acompanhamento psicológico, entre outros. Neste evento o autor do fato passa a ser visto como beneficiário, uma vez que a decisão resultará em uma pena menor e o encerramento do processo. Havendo aceitação da transação penal por parte do beneficiário, a decisão é encaminhada ao juiz que analisará o caso e homologará a transação, encerrando o caso. Para representar este evento, criamos o *frame* **TRANSAÇÃO_PENAL**, visualizado no quadro 7.

FRAME: TRANSAÇÃO PENAL	
Glosa:	A AUTORIDADE ACUSADORA formulará proposta de <i>TRANSAÇÃO PENAL</i> que consistirá em medida restritiva de direitos: multa, prestação de serviços, doações ou acompanhamento psicológico (ESPÉCIE). Esta proposta deve esclarecer a ESPÉCIE e a DURAÇÃO da medida aplicada ao BENEFICIÁRIO. Aceita a proposta pelo BENEFICIÁRIO e por seu DEFENSOR, esta será submetida à avaliação da AUTORIDADE JULGADORA para homologação seguindo a BASE LEGAL.
Elementos de frame:	AUTORIDADE ACUSADORA: é o representante do Ministério Público ou da Promotoria de Justiça que formulará a proposta de transação penal. ESPÉCIE: tipo de medida a ser imputada ao autor como pena pelo delito. DURAÇÃO: período de tempo em que a medida será vigente. BENEFICIÁRIO: autor do delito que deverá cumprir a medida estabelecida a fim de não ser condenado culpado no final do processo. DEFENSOR: advogado designado para defender os direitos do autor do fato. AUTORIDADE JULGADORA: responsável pela homologação BASE LEGAL: leis nas quais a autoridade julgadora apoia a homologação.
Unidades lexicais:	Transação, proposta, medida aplicada

Quadro 7: *Frame TRANSAÇÃO PENAL*

Fonte: Elaborado pela autora

Com esta descrição, já é possível vislumbrar que *CONCILIAÇÃO* e *TRANSAÇÃO PENAL* são subframes de *AUDIÊNCIA PRELIMINAR* e que ambos antecedem a homologação realizada pelo juiz. Todo o processo até aqui somente envolve o juiz no evento de homologação, seja da composição ou da transação penal. Entendemos que o evento de homologação também seja um *frame* necessário para a descrição do domínio e o criamos com o mesmo nome. O *frame HOMOLOGAÇÃO* a ser visualizado no quadro 8.

FRAME: HOMOLOGAÇÃO	
Glosa:	<i>HOMOLOGAÇÃO</i> corresponde ao ato oficial no qual a AUTORIDADE JUDICIÁRIA aprova e reafirma as decisões advindas de um ACORDO entre as PARTES ou da aceitação da transação penal.
Elementos de frame:	AUTORIDADE JUDICIÁRIA: juiz responsável por presidir a sessão e dar como aceito o acordo. ACORDO: aceite das partes PARTES: participantes do processo que buscam um acordo em razão do dano – autor e vítima.
Unidades lexicais:	Homologar, aprovar, aceitar, analisar

Quadro 8: *Frame HOMOLOGAÇÃO*

Fonte: Elaborado pela autora

No caso de haver composição ou transação penal e estas terem sido homologadas pelo juiz, encerra-se o procedimento. Porém, caso não tenha se estabelecido acordo, o processo prossegue através do oferecimento de denúncia pelo Ministério Público. O *frame DENÚNCIA* está descrito no quadro 9.

FRAME: DENÚNCIA	
Glosa:	A <i>DENÚNCIA</i> é oferecida pela AUTORIDADE ACUSADORA, fazendo uma acusação ao AUTOR DO FATO, narrando o delito praticado (RELATO) e pedindo sua condenação com aplicação de pena correspondente. Resulta no aceite ou não da AUTORIDADE JULGADORA.
Elementos de frame:	AUTORIDADE ACUSADORA: representante do Ministério Público ou da Promotoria de Justiça responsável por oferecer a denúncia AUTOR DO FATO: o denunciado, praticante do delito. RELATO: documento narrando o delito e solicitando o arrolamento de testemunhas e aplicação de pena ao autor do fato. AUTORIDADE JULGADORA: juiz que recebe a denúncia e decide aceitando-a ou rejeitando-a com base nos fatos.
Unidades lexicais:	Oferecer denúncia, oferecer queixa, denunciar

Quadro 9: *Frame DENÚNCIA*
 Fonte: Elaborado pela autora

A partir deste momento o juiz passa a ter uma participação efetiva e, se aceita a denúncia, é marcada uma audiência de instrução e julgamento. Para esta audiência são arroladas testemunhas de acusação e de defesa que serão ouvidas pelo juiz, assim como o réu será interrogado prestando seu depoimento acerca da infração penal. O quadro 10 descreve o *frame AUDIÊNCIA_DE_INSTRUÇÃO_E_JULGAMENTO*.

FRAME: AUDIÊNCIA DE INSTRUÇÃO E JULGAMENTO	
Glosa:	<i>AUDIÊNCIA_DE_INSTRUÇÃO_E_JULGAMENTO</i> corresponde ao ato que ocorre após o aceite da denúncia, tendo data, hora (TEMPO) e LOCAL definidos. Neste ato, são ouvidos os relatos das partes (RESPONSÁVEL e VÍTIMA) e das TESTEMUNHAS, gerando a DECISÃO por parte da AUTORIDADE JULGADORA. Participam deste momento o representante da ACUSAÇÃO e o DEFENSOR do autor do crime.
Elementos de frame:	AUTORIDADE JULGADORA: aquele que dirige a audiência RESPONSÁVEL: aquele que se defende da acusação, sendo interrogado na audiência. VÍTIMA: aquele que sofreu dano e propôs a ação, sendo ouvido na audiência. TESTEMUNHAS: pessoas arroladas para auxiliar na acusação ou na defesa, sendo interrogadas na audiência. DECISÃO: resultado da apreciação da autoridade julgadora ACUSAÇÃO: representante do Ministério Público ou Promotoria de Justiça responsável pelo oferecimento da denúncia. DEFENSOR: advogado ou defensor público designado para defender os interesses do responsável pelo ato criminal. TEMPO: data e hora designados para a audiência. LOCAL: sede do juizado especial, na qual se realiza a audiência.
Unidades lexicais:	Audiência de instrução, audiência de instrução e julgamento, instrução, audiência

Quadro 10: *Frame AUDIÊNCIA_DE_INSTRUÇÃO_E_JULGAMENTO*
 Fonte: Elaborado pela autora (adaptado de Minghelli, 2015)

A audiência de instrução e julgamento ocorre seguindo uma sequência: primeiro são ouvidas as testemunhas, depois o réu é interrogado e, por fim, a acusação e a defesa dispõem de um tempo delimitado para expor oralmente suas teses. Para representar cada um dos

eventos, criamos os *frames* *OITIVA_TESTEMUNHAS*, *INTERROGATÓRIO* e *DEBATE_ORAL* que podem ser visualizados nos quadros 11, 12 e 13 respectivamente.

FRAME: OITIVA_TESTEMUNHAS	
Glosa:	Ato no qual a AUTORIDADE JULGADORA ouve o relato da VÍTIMA e das TESTEMUNHAS acerca da infração penal. Envolve uma INTIMAÇÃO para comparecimento à audiência e a juntada de PROVAS.
Elementos de frame:	AUTORIDADE JULGADORA: juiz responsável pela oitiva da vítima e das testemunhas. VÍTIMA: quem foi lesado pelo crime de menor potencial ofensivo. TESTEMUNHAS: pessoas arroladas para auxiliar a defesa e a acusação. INTIMAÇÃO: documento enviado para chamar as testemunhas para a audiência PROVAS: documentos que comprovem a acusação ou que auxiliem a defesa.
Unidades lexicais:	Ouvir, oitivas, testemunhas

Quadro 11: *Frame OITIVA_TESTEMUNHAS*

Fonte: Elaborado pela autora

FRAME: INTERROGATÓRIO	
Glosa:	O <i>INTERROGATÓRIO</i> corresponde ao ato no qual a AUTORIDADE INQUIRIDORA colhe o DEPOIMENTO (informações) do INTERROGADO (autor do ato/crime) sobre o CASO ouvindo suas versões sobre o fato e garantindo-lhe direito de defesa.
Elementos de frame:	AUTORIDADE INQUIRIDORA: pessoa responsável pelo interrogatório, questionando o INTERROGADO. DEPOIMENTO: fatos narrados pelo INTERROGADO a fim de sustentar sua defesa. INTERROGADO: a quem são dirigidas as perguntas da AUTORIDADE INQUIRIDORA. O autor do ato/crime. CASO: os fatos sobre os quais o depoimento versa TEMPO: quando o interrogatório ocorre (dia/hora) LOCAL: onde o interrogatório ocorre
Unidades lexicais:	Inquirir, interrogar, interrogatório

Quadro 12: *Frame INTERROGATÓRIO*

Fonte: Elaborado pela autora (adaptado de Minghelli, 2015)

FRAME: DEBATE_ORAL	
Glosa:	O <i>DEBATE ORAL</i> corresponde ao momento em que o a AUTORIDADE ACUSADORA e a AUTORIDADE DEFENSORA dispõem de tempo para análise crítica das PROVAS e exposição de argumentos visando o convencimento do juiz. Há um TEMPO limitado para cada participante expor suas alegações.
Elementos de frame:	AUTORIDADE ACUSADORA: representante do Ministério Público AUTORIDADE DEFENSORA: advogado ou defensor público PROVAS: documentos arrolados para argumentar em relação à defesa ou à acusação e os depoimentos colhidos das testemunhas. TEMPO: tempo que cada parte tem para expor seus argumentos perante o juiz (20 minutos para cada parte, prorrogáveis por mais 10)
Unidades lexicais:	Debater, provar, defender, argumentar

Quadro 13: *Frame DEBATE_ORAL*

Fonte: Elaborado pela autora

Observando estes *frames*, pode-se perceber que *OITIVA_TESTEMUNHAS*, *INTERROGATÓRIO* e *DEBATE_ORAL* são subframes de *AUDIÊNCIA_*

DE_INSTRUÇÃO_E_JULGAMENTO. Além disso, há uma precedência na sequência dos acontecimentos, ou seja, eles ocorrem seguindo uma ordem.

Concluída a audiência de instrução e julgamento, é realizada a prolação da sentença por parte do juiz. A prolação da sentença é registrada em um termo de audiência, no qual são relatados os depoimentos e a fundamentação baseada na lei. Encerra-se o processo, cabendo a possibilidade de recurso pela parte que sentir-se lesada pela decisão.

Entendemos que a prolação da sentença corresponde a um evento importante do processo e, criamos o *frame* *PROLAÇÃO_SENTENÇA* (quadro 14) para representá-lo.

FRAME: PROLAÇÃO_SENTENÇA	
Glosa:	<i>PROLAÇÃO_SENTENÇA</i> corresponde à decisão (<i>SENTENÇA</i>), fundamentada na lei (<i>FUNDAMENTAÇÃO</i>), proferida pela <i>AUTORIDADE JULGADORA</i> , na qual se decide sobre uma <i>INFRAÇÃO PENAL</i> e se impõe uma <i>PENA</i> . A sentença é registrada em um <i>TERMO DE AUDIÊNCIA</i> .
Elementos de frame:	<p><i>AUTORIDADE JULGADORA</i>: aquele a quem cabe decidir a questão.</p> <p><i>SENTENÇA</i>: decisão fundamentada na lei acerca do ato crime.</p> <p><i>FUNDAMENTAÇÃO</i>: as razões e motivos que sustentam a decisão com base legal.</p> <p><i>INFRAÇÃO PENAL</i>: crime de menor potencial ofensivo que configura conflito entre as partes.</p> <p><i>PENA</i>: sanção de natureza pecuniária que a lei impõe àqueles que infringem determinações legais ou administrativas. Poderá ser: multa, prestação de serviços comunitários, prestação pecuniária ou reparação de danos</p> <p><i>TERMO DE AUDIÊNCIA</i>: documento que registra a denúncia, o resumo dos debates e a fundamentação da sentença.</p>
Unidades lexicais:	Condenar, absolver

Quadro 14: *Frame* *PROLAÇÃO_SENTENÇA*
Fonte: Elaborado pela autora (adaptado de Minghelli, 2015)

Após a prolação da sentença ou da homologação do juiz é cabível um recurso, caso uma das partes sintam-se lesada pela decisão. Por ser um processo desenvolvido no Juizado Especial Criminal, este recurso é julgado por uma turma recursal composta por três juízes (todos de 1ª instância) que cumprem o papel de 2ª instância, isto porque os processos deste juizado permanecem somente na 1ª instância, diferindo dos demais cujos recursos são julgados nas instâncias superiores. Temos, então, o *frame* *RECURSO* representado no quadro 15.

FRAME: RECURSO	
Glosa:	O <i>RECURSO</i> poderá ser interposto contra a decisão da <i>AUTORIDADE JULGADORA</i> que rejeitar a denúncia, bem como da sentença absolutória ou condenatória ou da homologação da transação penal. O recurso poderá ser julgado por três juízes em exercício no primeiro grau de jurisdição (<i>TURMA RECURSAL</i>). <i>ACUSADO</i> pode recorrer via <i>DEFENSOR</i> e Ministério Público (<i>AUTORIDADE ACUSADORA</i>) também pode recorrer. Há um <i>PRAZO</i> que que seja interposto o recurso que deve ser feito através de uma <i>PETIÇÃO</i> .
Elementos de frame:	<i>AUTORIDADE JULGADORA</i> : juiz monocrático responsável pela decisão de rejeitar a denúncia. <i>TURMA RECURSAL</i> : grupo composto por três juízes em exercício no primeiro grau da jurisdição. <i>ACUSADO</i> : o responsável pelo crime de menor potencial ofensivo que se sentiu lesado pela decisão. <i>DEFENSOR</i> : advogado ou defensor público designado para defender o acusado. <i>VÍTIMA</i> : quem foi lesado pelo crime de menor potencial ofensivo. <i>AUTORIDADE ACUSADORA</i> : representante do Ministério Público <i>PRAZO</i> : tempo previsto para que se possa entrar com recurso contra a decisão monocrática. <i>PETIÇÃO</i> : documento de recurso sobre a decisão.
Unidades lexicais:	Recorrer, recurso, apelar

Quadro 15: *Frame RECURSO*
 Fonte: Elaborado pela autora

Outro aspecto a ser considerado é a relação entre os *frames*, pois será útil no momento de elencar a estrutura taxonômica da ontologia, constituindo as classes e subclasses. Analisando o mapa conceitual e o conhecimento adquirido para sua organização, foi possível estabelecer as relações entre os *frames* criados para o domínio. Escolhemos fazer uso das relações existentes na FN, porém percebemos (conforme vem sendo estudado por Minguelli, 2015) que o domínio exige o uso de relações diferenciadas em função de suas peculiaridades. Organizamos os *frames* e suas relações em uma figura seguindo os padrões de cores e setas da FN. A estrutura organizada está representada na figura 17.

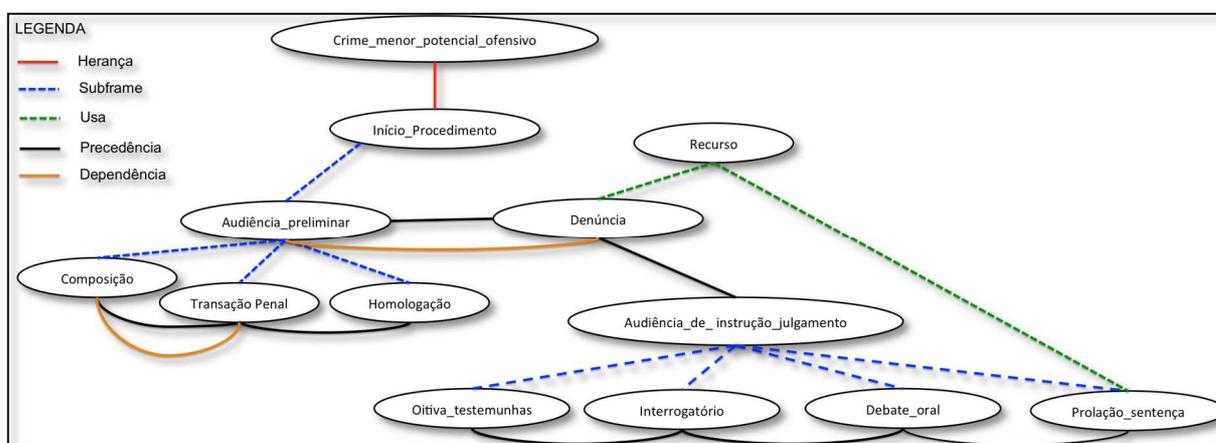


Figura 17: Relações entre frames.
 Fonte: elaborado pela autora.

As relações advindas da FN que são utilizadas na hierarquia de frames do procedimento sumaríssimo são: herança (marcada pela linha vermelha), subframe (pontilhado azul), usa (pontilhado verde) e precedência (linha preta).

Note-se, na imagem, identificada com a linha laranja, que há uma relação de **dependência**. Esta relação não faz parte do grupo de relações explicitados na FN. Neste trabalho não temos a pretensão de eleger novas relações entre *frames*, mas percebemos que nenhuma das relações disponíveis na FN cobria o aspecto relacionado ao fato de que a transação penal só ocorre no caso de não haver composição, ou seja, para existir a primeira é necessário que não tenha ocorrido acordo na segunda, gerando uma relação que consideramos de dependência entre os *frames*. O mesmo ocorre entre a audiência preliminar e a denúncia; a segunda só ocorre se a primeira não tiver sido homologada.

É importante alertar que, em função das peculiaridades do domínio jurídico, poderão existir várias relações diferentes das dispostas na FN. No entanto, no caso específico do procedimento sumaríssimo identificamos somente a dependência. Um aprofundamento sobre as relações entre os *frames* no domínio jurídico vem sendo desenvolvido por Minguelli em sua pesquisa de doutorado.

Cabe aqui um esclarecimento acerca da finalização da etapa de conceitualização: a descrição dos conceitos e dos *frames* precisava ser fiel à realidade e, para garantir esta necessidade, realizamos três formas de validação. A primeira, através da revisão da leitura dos documentos do *corpus* de consulta. A segunda, por meio de consulta a um advogado especialista da área criminal. E, por fim, uma terceira validação envolveu o parecer de um juiz conhecedor do procedimento sumaríssimo e dos Juizados Criminais Especiais. Esta revisão, sobretudo a que envolveu os especialistas da área, permitiu os ajustes necessários para a representação dos fatos.

Após a organização das relações entre os *frames*, concluímos as etapas que seguem a abordagem *top-down*. Na próxima etapa apresentamos os procedimentos relacionados à abordagem *bottom-up*, na qual elencamos a lista de palavras e as relacionamos aos conceitos.

5.3.3 Lista de palavras

Para a obtenção de uma lista de palavras relacionadas ao domínio, realizamos o tratamento do *corpus* de pesquisa em uma ferramenta computacional. Normalmente estas ferramentas provêm uma lista de palavras e seu número de ocorrências no *corpus*, um gerador de *n-gramas*⁹³, geração de palavras-chave (*keywords*)⁹⁴, gerador de *clusters*⁹⁵, colocados (*collocates*)⁹⁶ e o concordanciador⁹⁷. Tais recursos permitem um estudo aprofundado dos conceitos do domínio em seu contexto específico e facilitam ao desenvolvedor a identificação de palavras ou termos (simples e complexos), além de permitir a visualização das palavras em seu contexto no *corpus*. No nosso caso, as ferramentas úteis são a geração de palavras-chave, pois proverá a lista de palavras, e o concordanciador, para verificar o contexto em que cada palavra aparece no *corpus*.

Bevilacqua (2013, p. 17) apresenta algumas ferramentas computacionais para a extração de informação linguística: (a) softwares como: *Wordsmith Tools*⁹⁸ e *AntConc*⁹⁹ (b) sites e plataformas: *WebCorp*¹⁰⁰, *e-Termos* (DE OLIVEIRA, ALUÍSIO & DE BARCELOS ALMEIDA, 2006)¹⁰¹ e *Corpógrafo* (SARMENTO et al., 2004)¹⁰²; (c) grupos brasileiros que desenvolvem projetos nesta área e que disponibilizam recursos deste tipo: CoMET (USP), GETerm (UFSCAR), NILC (USP-UFSCAR), Termisul (UFRGS), TermNeo (USP) e TextQuim (UFRGS).

⁹³ Corresponde a uma sequência de palavras que aparecem no *corpus*. Na maioria dos sistemas se pode definir a extensão desta sequência, permitindo a identificação de termos com mais de uma palavra, como por exemplo “habeas corpus”, “Tribunal do Juri”, “subiram os autos”.

⁹⁴ Permite a criação de uma lista de palavras-chave que servirão de base para confrontar a lista de termos com o *corpus* de referência.

⁹⁵ Permite a verificação do grupo de palavras que faz parte do contexto a partir da busca pelas palavras-chave.

⁹⁶ Possibilita a verificação de palavras que aparecem junto às palavras consultadas.

⁹⁷ Lista as ocorrências da palavra consultada em seu contexto.

⁹⁸ Disponível em <http://www.lexically.net/wordsmith/>. Acesso em 13/06/15.

⁹⁹ Disponível em <http://www.laurenceanthony.net/software.html>. Acesso em 20/12/2013.

¹⁰⁰ Disponível em <http://www.webcorp.org.uk/live/>. Acesso em 13/06/15.

¹⁰¹ Disponível em www.etermos.cnptia.embrapa.br. Acesso em 21/09/2009.

¹⁰² Disponível em <http://labclup.letras.up.pt/cgi-bin/corpografo/user/sagi/splash.pl>. Acesso em 13/06/15.

Nesta pesquisa usamos a plataforma **e-Termos** para a compilação¹⁰³ do *corpus* e o software **AntConc** para a obtenção da lista de palavras e o estudo do contexto de uso a fim de verificar as variantes e as possíveis relações que não apareceram no mapa conceitual e nos *frames*. O primeiro, por estar disponível com livre acesso na *Web* e facilitar o *upload* do *corpus*, permitindo que este procedimento seja feito em bloco através de um arquivo compactado (.zip) e o segundo, por apresentar mais opções de análise do *corpus*, principalmente a facilidade de análise das palavras em seu contexto no *corpus*.

Após a compilação do *corpus* na plataforma e-Termos lançamos o arquivo em formato .txt no software AntConc e usamos a ferramenta de palavras-chave para obter uma lista de palavras. A primeira tentativa resultou em uma lista de mais de 1000 palavras, pois foi processada sem nenhum tipo de filtro. Nossa primeira constatação foi que o programa não processa os sinais gráficos e, para resolver este problema, usamos o editor de texto para substituir todas as palavras acentuadas e com cedilha por suas semelhantes sem o sinal, gravando um novo arquivo .txt para ser processado. Ainda no primeiro processamento, verificamos que o programa considerou até mesmo os artigos definidos como palavras e isso resultou em uma lista não satisfatória.

Para resolver este problema, identificamos a possibilidade de criação de uma *stoplist*. A *stoplist* é uma lista de palavras que não devem ser consideradas na análise do todo. A criação da *stoplist* se deu de forma manual, observando os resultados da primeira listagem. Na figura 18 ilustramos o processo de criação da *stoplist*, no qual destacamos com um quadro pontilhado a *textbox* na qual foram digitadas as palavras consideradas fora do domínio ou não necessárias, tais como artigos, preposições, pronomes, etc. A seleção das palavras nesta primeira seleção deu-se de forma empírica e resultou em uma nova lista com 241 palavras.

¹⁰³ Compilação significa a compactação de todos os documentos selecionados como corpus de pesquisa em um único documento, facilitando seu processamento geral.

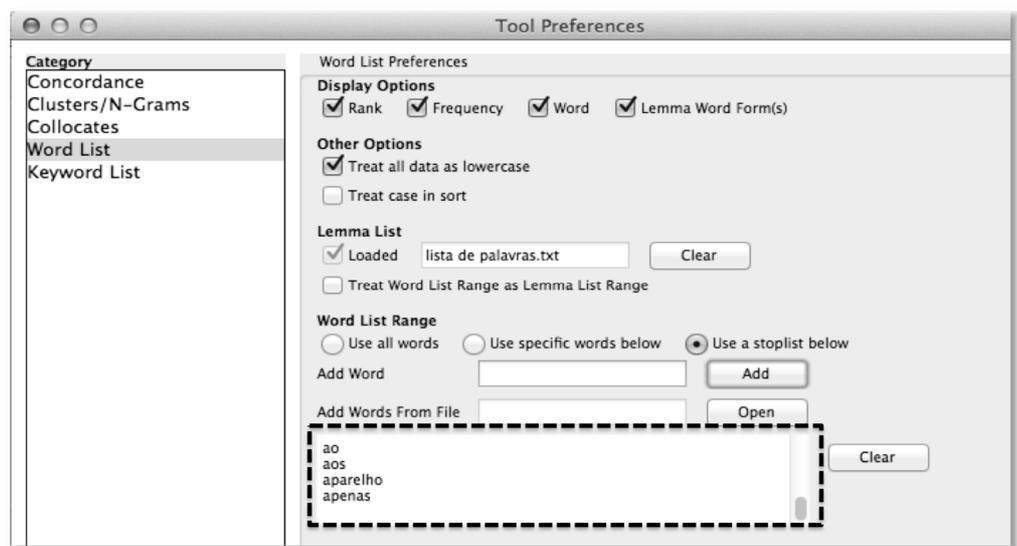


Figura 18: Geração da *stoplist* no AntConc.

Fonte: elaborado pela autora.

Para obtermos um resultado mais favorável, analisamos a segunda lista de palavras, ampliando a *stoplist*. Desta vez retiramos as variações verbais mantendo somente a forma no infinitivo e as variações de gênero e número, além de incluir na *stoplist* palavras que sabidamente não se constituíam como sendo do domínio, por exemplo mulher, homem, caracterizar, hipótese, suficientemente, porque, entre outras. Esta nova inclusão de palavras na *stoplist* resultou em uma lista com 98 palavras.

A partir da lista de palavras, passamos a analisar seu contexto através da ferramenta concordanciador do AntConc. Esta ferramenta permite visualizar a palavra no seu contexto, bem como mostra as diferentes ocorrências no *corpus*. Assim, é possível identificar se a palavra corresponde ou não a uma unidade lexical evocadora de *frame*, ampliando este aspecto em relação aos *frames* elaborados. Além disso, este processo permite a identificação de variantes um aspecto fundamental na recuperação da informação.

Para exemplificar a associação das palavras da lista aos *frames*, destacamos a palavra **inquirição**. A figura 19 mostra a palavra em dois contextos, os quais evocam o *frame* *AUDIÊNCIA_DE_INSTRUÇÃO_E_JULGAMENTO*, mais precisamente o *subframe* *OITIVA_TESTEMUNHAS*.

queixa, a audiencia prossegue com a inquiricao da vitima, das testemunhas de acusacao sede do juizo, para a sua inquiricao podera ser expedida carta precatória.

Figura 19: Resultado concordanciador – inquirição.

Fonte: elaborado pela autora.

Outro exemplo de palavra extraída da lista que foi identificada como unidade lexical evocadora de *frame* é **alegações**. Observando o resultado do concordanciador (figura 20) percebe-se que esta palavra evoca o *frame* *DEBATE_ORAL*. Esta análise nos possibilitou ampliar as unidades lexicais relacionadas a cada *frame*.

a defesa a oportunidade de oferecer alegacoes orais. O representante do Ministerio dos debates orais por memoriais ou alegacoes finais escritas, mesmo porque, conc o interrogatório do reu. Apresentadas as alegacoes finais sera prolatada a sentenca: :

Figura 20: Resultado concordanciador – alegações.

Fonte: elaborado pela autora.

Para exemplificar a identificação das diferentes funções exercidas pelos participantes através da lista de palavras, elencamos a palavra **juiz**. A análise realizada através do concordanciador permitiu ampliar a descrição das funções exercidas pelo juiz.

reduzida a escrito e homologada pelo juiz togado, mediante sentenca irrecorrivel, que t aplicada. A lei permite que o juiz reduza a multa ate metade, podendo , sera esta submetida a apreciacao do Juiz togado, caso nao esteja presidindo a sera feita pelo promotor ou pelo juiz togado. ou queixa - Frustrada a transacao, o juiz esclarecera sobre a possibilidade da concilia de provas - As provas que o juiz ouvira o defensor do reu, medida a prolocao de sentenca oral pelo juiz considerar excessivas, impertinentes ou prot oralidade e da identidade fisica do juiz togado. Em decorrenca da consagracao dos e dispensado o relatório e o juiz, o magistrado que presidir a audiencia de danos entre as partes, o juiz mencionara apenas os seus elementos de de danos entre as partes, o juiz homologa a composicao, registra-se a penal, proposta pelo promotor ou pelo Juiz homologa a composicao, registra-se a após o recebimento da denúncia pelo juiz. Havendo transacao, homologa-se por sentenc

Figura 21: Resultado concordanciador – juiz.

Fonte: elaborado pela autora.

Como se pode ver na figura 21, o juiz homologa a composição, pode reduzir a multa imputada, aprecia a denúncia, esclarece a possibilidade de conciliação, ouve o defensor do réu, preside a audiência. Todos estas funções do juiz, vislumbradas a partir da análise do *corpus* pelo concordanciador do AntConc, estão relacionadas aos diferentes *frames* e serão

incluídos na ontologia através de relações ou permissões que o juiz tem no decorrer do processo.

Identificamos também palavras como **data**, **local**, **endereço**, **nome** que são dados referentes aos indivíduos, ou seja, os atributos para os documentos (data, local, endereço) e para os participantes (nome, endereço). Estas informações também são importantes para a localização de informações em um sistema de busca e serão inseridas na ontologia como atributos (propriedades de dados no OWL).

A lista de palavras também possibilitou a verificação de variantes ou fraseologias relacionadas ao domínio e que se referem a um mesmo *frame*. Para uma melhor visualização, organizamos as palavras em um quadro indicando as variantes e o *frame* relacionado.

Palavra	Variantes	Frame
Acordo	Conciliação – composição	<i>CONCILIAÇÃO</i>
Ajuizamento	Ajuizar – marcar	<i>AUDIÊNCIA PRELIMINAR E AUDIÊNCIA DE INSTRUÇÃO E JULGAMENTO</i>
Apelação	Apelar – recorrer	<i>RECURSO</i>
Apreciação	Apreciar – avaliar	<i>HOMOLOGAÇÃO</i>
Arquivo	Arquivamento – arquivar – arquivar-se	<i>HOMOLOGAÇÃO</i>
Intimar	Intime-se – intimado – intimação	<i>OITIVAS E INTERROGATÓRIO</i>
Laudos	Provas	<i>AUDIÊNCIA DE INSTRUÇÃO E JULGAMENTO</i>
Proferir	Proferida – proferimento – prolatar – prolação	<i>PROLAÇÃO DA SENTENÇA</i>
Vítima	Ofendido – parte prejudicada	<i>CRIME DE MENOR POTENCIAL OFENSIVO – AUDIÊNCIA PRELIMINAR – CONCILIAÇÃO – AUDIÊNCIA DE INSTRUÇÃO E JULGAMENTO – RECURSO – OITIVA TESTEMUNHAS</i>
Defensor	Defesa – advogado – defensor público	<i>AUDIÊNCIA PRELIMINAR – TRANSAÇÃO PENAL – AUDIÊNCIA DE INSTRUÇÃO E JULGAMENTO – RECURSO – DEBATE ORAL</i>
Acusação	Ministério Público – promotoria de justiça – promotor – advogado de acusação	<i>TRANSAÇÃO PENAL – DENÚNCIA – AUDIÊNCIA DE INSTRUÇÃO E JULGAMENTO – RECURSO – DEBATE ORAL</i>
Autor	Autor do fato – acusado	<i>CRIME DE MENOR POTENCIAL OFENSIVO – INÍCIO DO PROCEDIMENTO – AUDIÊNCIA PRELIMINAR – CONCILIAÇÃO – RECURSO</i>
Autor	Beneficiário – beneficiado	<i>TRANSAÇÃO PENAL</i>
Autor	Réu – culpado – responsável	<i>AUDIÊNCIA DE INSTRUÇÃO E JULGAMENTO – INTERROGATÓRIO</i>
Autor	Interrogado	<i>INTERROGATÓRIO</i>

Quadro 16: Variantes e *frames*

Fonte: Elaborado pela autora

Note-se que as palavras relacionadas aos participantes (vítima, defensor, acusação e autor) não só apresentam as variantes, como também explicitam seus diferentes papéis em cada *frame*. Identificar as variantes é fundamental para o sistema de busca localizar as

informações, no entanto o grande desafio está na forma de explicitar os papéis que se modificam de acordo com a variante e o *frame* ao qual se relaciona.

Ao analisarmos as palavras em seus contextos, verificando as variantes e os *frames* aos quais se relacionam concluímos a etapa *bottom-up* para a organização da hierarquia de classes.

5.3.4 Classes e subclasses

Para a organização da estrutura hierárquica da ontologia, através das classes e subclasses, o mapa conceitual e os *frames* foram fundamentais. O primeiro para auxiliar na organização das categorias **Documentos Legais**, **Instituições Legais** e **Participantes Legais**. Já o segundo foi a base para a estruturação da categoria **Eventos Legais**.

Nesta etapa descritiva das classes e subclasses optamos pela representação em forma de quadros, nos quais apresentamos as categorias, as classes e as subclasses respectivamente. A primeira categoria a ser representada é Documentos Legais. A partir do mapa conceitual, elencamos os documentos que são gerados no decorrer do procedimento sumaríssimo. Note-se que estão representados somente os documentos referentes ao procedimento sumaríssimo, uma vez que nosso compromisso é com a descrição deste domínio.

CATEGORIA	CLASSES (relação é-um)
Documentos legais	Boletim de ocorrência
	Termo circunstanciado
	Termo de audiência
	Provas
	memoriais

Quadro 17: Estrutura de classes – categoria Documentos Legais
Fonte: Elaborado pela autora

A próxima categoria representada foi Instituições Legais. Novamente vale atentar para a nota referida acima: representamos somente as instituições envolvidas com o procedimento sumaríssimo, tendo ciência de que área jurídica envolve uma gama de outras instituições.

No caso do procedimento sumaríssimo temos uma peculiaridade: o processo se inicia através de um boletim de ocorrência que deve ser registrado em uma delegacia de polícia. Delegacia de polícia não corresponde exatamente a uma instituição relacionada ao sistema judiciário, mas sim ao sistema policial. No entanto, uma vez que é nesta instituição que se

instaura o procedimento sumaríssimo entendemos que, para fins de descrição ontológica, é necessária apontá-la como uma das instituições na ontologia.

CATEGORIA	CLASSES (relação é-um)	
Instituições Legais	Delegacia	
	Ministério Público	
	Fórum	SUBCLASSE (relação parte-de)
		Vara criminal
		SUBCLASSE JECrim

Quadro 18: Estrutura de classes – categoria Instituições Legais
Fonte: Elaborado pela autora

O mapa conceitual também proveu informações para a descrição da categoria Participante Legais. É importante salientar que nesta etapa estamos apenas elencando a estrutura taxonômica, uma vez que as relações serão estabelecidas da etapa seguinte.

CATEGORIA	CLASSES (relação é-um)	
Participantes Legais	Acusação	SUBCLASSES (relação parte-de)
		Promotor de justiça
		Advogado de acusação
	Defesa	SUBCLASSES (relação parte-de)
		Defensor público
		Advogado de defesa
	Partes	SUBCLASSES (relação parte-de)
		Vítima
		Autor do fato
	Autoridades	SUBCLASSES (relação parte-de)
		Conciliador
		Juiz
		Turma recursal
		Delegado de polícia
	Testemunhas	SUBCLASSES (relação parte-de)
Acusação		
	Defesa	

Quadro 19: Estrutura de classes – categoria Participantes Legais
Fonte: Elaborado pela autora

Para a estruturação das classes e subclasses da categoria Eventos Legais tomamos como base a organização dos *frames* do domínio observando as relações entre eles. Vale atentar para o fato de que as relações entre *frames* que são passíveis de inclusão na estrutura taxonômica são **herança** e **subframe**. As demais relações entre os *frames* serão tratadas como não taxonômicas.

CATEGORIA	CLASSES (relação é-um)	
Eventos Legais	Início do procedimento	
	Audiência Preliminar	
		SUBFRAMES (relação parte-de)
		Conciliação
		Transação penal
		Homologação
	Audiência de instrução e julgamento	
		SUBFRAMES (relação parte-de)
		Oitivas testemunhas
		Interrogatório
	Debate oral	
	Denúncia	
	Prolação da sentença	
	Recurso	

Quadro 20: Estrutura de classes – categoria Eventos Legais
Fonte: Elaborado pela autora

No quadros também identificamos as relações taxonômicas *é-um*, *tipo-de* e *parte-de* entre as classes e subclasses. A próxima etapa corresponde ao estabelecimento das relações não taxonômicas.

5.3.5 Relações não taxonômicas

As relações não taxonômicas foram estabelecidas a partir da análise dos *frames* do domínio. Esta análise permitiu descrever as relações entre os eventos e os participantes, visualizando a relação entre as classes e subclasses na ontologia. É justamente o estabelecimento das relações não taxonômicas que proverá a ampliação da descrição do significado na ontologia, sendo estas fundadas nos *frames*.

A primeira etapa de análise envolveu os participantes e os papéis exercidos por eles nos respectivos *frames*. Para determinar estas relações, partimos das informações constantes do quadro 16 e os *frames* desenvolvidos, bem como atentamos para informações acerca do envolvimento de cada participante advindas do *corpus* de consulta Organizamos as relações em quadros, apresentando cada participante separadamente.

Participante	Papéis	Frame
Acusação	autoridade acusadora	<i>DENÚNCIA</i>
	acusador	<i>AUDIÊNCIA_DE_INST._JUG</i>
	debatedor de acusação	<i>DEBATE_ORAL</i>
	defensor da vítima	<i>TRANSAÇÃO_PENAL</i>
	Funções/ações	
	pode entrar com	<i>RECURSO</i>
	oferece	<i>DENÚNCIA</i>
	arrola testemunhas	<i>AUDIÊNCIA_DE_INST._JUG</i>

Quadro 21: Relações não taxonômicas – papéis e funções dos participantes – Acusação
Fonte: Elaborado pela autora

Participante	Papéis	Frame
Defesa	autoridade defensora	<i>TRANSAÇÃO_PENAL</i>
	defensor do réu	<i>AUDIÊNCIA_INS._JULG.</i>
	debatedor de defesa	<i>DEBATE_ORAL</i>
	Funções/ações	
	pode entrar com	<i>RECURSO</i>
	arrola testemunhas	<i>AUDIÊNCIA_INS._JULG.</i>

Quadro 22: Relações não taxonômicas – papéis e funções dos participantes – Defesa
Fonte: Elaborado pela autora

Participante	Papéis	Frames
Vítima	prejudicado	<i>CRIME_MENOR_POTENCIAL_OFENSIVO</i>
	ofendido	<i>AUDIENCIA_PRELIMINAR</i>
	inquirido	<i>OITIVA_TESTEMUNHAS</i>
	lesado	<i>AUDIÊNCIA_INST_JULG</i>
	Funções/ações	
	registra BO	<i>INÍCIO_PROCEDIMENTO</i>
	participa	<i>AUDIÊNCIA_PRELIMINAR</i>
	participa	<i>AUDIÊNCIA_INST._E_JULG</i>
	é representado por MP	<i>DENÚNCIA</i>

Quadro 23: Relações não taxonômicas – papéis e funções dos participantes – Vítima
Fonte: Elaborado pela autora

Participante	Papéis	Frames
Autor	autor do fato	<i>CRIME_MENOR_POTENCIAL</i>
	denunciado	<i>INÍCIO_PROCEDIMENTO</i>
	beneficiário	<i>TRANSAÇÃO_PENAL</i>
	acusado	<i>DENÚNCIA</i>
	responsável civil	<i>AUDIÊNCIA_PRELIMINAR</i>
	citado	<i>AUDIÊNCIA_INST._JULG</i>
	sentenciado	<i>PROLAÇÃO_SENTENÇA</i>
	interrogado	<i>INTERROGATÓRIO</i>
	investigado	<i>INÍCIO_PROCEDIMENTO</i>
	Funções/Ações	
	pratica	<i>CRIME_MENOR_POTENCIAL</i>
	participa	<i>AUDIÊNCIA_PRELIMINAR</i>
	participa	<i>AUDIÊNCIA_INST_JULG.</i>
	representadoPor Defensor	<i>RECURSO</i>

Quadro 24: Relações não taxonômicas – papéis e funções dos participantes – Autor do fato
Fonte: Elaborado pela autora

Participante	Papéis	Frame
Conciliador	autoridade conciliadora	<i>AUDIÊNCIA_PRELIMINAR</i>
	conciliador	<i>CONCILIAÇÃO</i>
	Funções/ações	
	busca acordo entre as Partes	<i>CONCILIAÇÃO</i>

Quadro 25: Relações não taxonômicas – papéis e funções dos participantes – Conciliador
Fonte: Elaborado pela autora

Participante	Papéis	Frames
Juiz	autoridade julgadora	<i>AUDIÊNCIA_INST_JULG</i>
	autoridade inquiridora	<i>INTERROGATÓRIO</i>
	autoridade inquiridora	<i>OITIVA_TESTEMUNHAS</i>
	autoridade judiciária	<i>HOMOLOGAÇÃO</i>
	autoridade avaliadora	<i>DENÚNCIA</i>
	Funções/ações	
	analisa acordo	<i>TRANSAÇÃO_PENAL</i>
	homologa acordo	<i>TRANSAÇÃO_PENAL</i>
	analisa acordo	<i>CONCILIAÇÃO</i>
	homologa acordo	<i>CONCILIAÇÃO</i>
	ouve réu	<i>INTERROGATÓRIO</i>
	ouve testemunhas	<i>OITIVAS_TESTEMUNHAS</i>
	prolata sentença	<i>PROLAÇÃO_SENTENÇA</i>
decide pena	<i>PROLAÇÃO_SENTENÇA</i>	
intima Testemunhas	<i>AUDIÊNCIA_INST_JULG</i>	

Quadro 26: Relações não taxonômicas – papéis e funções dos participantes – Juiz
Fonte: Elaborado pela autora

Participante	Papéis	Frame
Turma recursal	Autoridade avaliadora	<i>RECURSO</i>
	Funções/ações	
	Avaliar	<i>RECURSO</i>

Quadro 27: Relações não taxonômicas – papéis e funções dos participantes – Turma recursal
Fonte: Elaborado pela autora

Participante	Papéis	Frame
Delegado de Polícia	Autoridade policial	<i>INÍCIO_PROCEDIMENTO</i>
	Funções/ações	
	Registrar BO	<i>INÍCIO_PROCEDIMENTO</i>
	Lavrar Termo Circunstanciado	<i>INÍCIO_PROCEDIMENTO</i>

Quadro 28: Relações não taxonômicas – papéis e funções dos participantes – Delegado de Polícia
Fonte: Elaborado pela autora

Participante	Papéis	Frame
Testemunhas	Prestar depoimento	<i>AUDIÊNCIA_INST_JULG.</i>
	inquirida	<i>OITIVAS_TESTEMUNHAS</i>
	Funções/ações	
	participam	<i>AUDIÊNCIA_INST_JULG.</i>
	São intimadas para	<i>AUDIÊNCIA_INST_JULG.</i>

Quadro 29: Relações não taxonômicas – papéis e funções dos participantes – Testemunhas
Fonte: Elaborado pela autora

Uma das problemáticas acerca da construção de ontologias reside justamente na dificuldade de representar os diferentes papéis que os participantes exercem no mundo real, conforme referenciado por Sowa (2012). Através da análise desta informação, baseada nos *frames* do domínio foi possível estabelecer a distinção entre os diferentes papéis assumidos por cada participante no decorrer do processo, minimizando o problema referido. As soluções de implementação na seção seguinte, sinalizam uma possibilidade de formalização para os papéis dos participantes em seus respectivos frames.

As relações entre os *frames* são igualmente importantes para a descrição do conhecimento do procedimento sumaríssimo na ontologia. Conforme referimos, as relações de **herança** e **subframe** já foram contempladas no passo de criação das classes e subclasses. No entanto, cumpre-se estabelecer a representação para as relações de **precedência** e de **dependência**. A relação de **usa** é marcada graficamente entre os *frames* na etapa anterior, mas não faz-se necessária no âmbito da recuperação da informação.

A **precedência** será marcada pela relação *precede* e a **dependência**, pela relação *dependeDe*, como se pode ver nos exemplos:

- [1] Oitivas *precede* Interrogatório
- [2] Interrogatório *precede* DebateOral
- [3] Denúncia *precede* AudiênciaInstruçãoJulgamento
- [4] TransaçãoPenal *dependeDe* não Composição
- [5] Homologação *dependeDe* Composição
- [6] Homologação *dependeDe* TransaçãoPenal

O mapa conceitual também provê a inclusão de relações não taxonômicas, tais como:

- [7] BoletimOcorrência *gera* TermoCircunstanciado
- [8] TermoCircunstanciado *éEncaminhadoPara* JeCrim
- [9] AudiênciaInstruçãoJulgamento *envolve* Provas
- [10] Memoriais *podeSubstituir* DebateOral
- [11] ProlaçãoSentença *gera* TermoAudiência

Concluída esta etapa de estabelecimento das relações, temos uma descrição linguístico-conceitual do domínio apresentando gráfica e textualmente toda a estruturação do procedimento sumaríssimo. Podemos afirmar que até aqui temos uma ontologia informal do domínio, isto porque, os dados estão ainda em linguagem natural e não formalizados no editor de ontologias.

5.4 Implementação

O estágio de implementação, conforme já referimos, tem como função a implementação dos dados linguístico-conceituais analisados no percurso de conceitualização. Para a implementação da ontologia, optamos pela ferramenta Protégé, uma vez que esta resulta em um arquivo OWL que poderá ser disponibilizado para sistemas de recuperação da informação. Vale salientar que, em função da extensão do trabalho, vamos trazer apenas exemplos de implementação, mostrando a solução encontrada para cada informação a ser inserida na ontologia.

O Protégé sugere um padrão para a nomenclatura das classes e das relações (propriedades). Na ferramenta, as classes devem ser nomeadas através do seguinte padrão **InicialMaiúscula**, ou seja, as palavras que compõem um nome de classe terão sempre a primeira letra maiúscula. E as relações (propriedades) inicial com minúscula e as palavras seguintes têm a inicial maiúscula, por exemplo, **propriedadeMaiúscula**. Em ambos os casos deve-se evitar acentuação e sinais gráficos. A fim de evitar futuros problemas no uso da ontologia, optamos por seguir a notação de nomenclatura adotada pelo Protégé.

O primeiro passo para a implementação é a organização hierárquica das classes e subclasses. No caso do Protégé, esta é a relação mais simples de ser modelada, uma vez que, ao criar uma classe e derivar suas subclasses, o próprio software já estabelece a relação *é-um* (hiponímia). Para a inserção das classes, sempre parte-se de uma classe inicial que compõe qualquer ontologia, mesmo quando está vazia; esta classe é chamada de **Thing**¹⁰⁴. Para o Protégé, uma classe pode ser explicada como um conjunto de indivíduos (ou conjunto de objetos), sendo **Thing** a classe raiz da qual todas as demais derivam, isto é, todas as classes serão consideradas subclasses de **Thing**.

¹⁰⁴ **Thing** é uma palavra reservada da linguagem OWL, isto é, faz parte de seu vocabulário específico, conforme padrão da W3C e, por esta razão, não realizamos a tradução do termo.

Outro aspecto importante a ser levado em consideração na criação da hierarquia de classes é o conceito de classes disjuntas (*Disjoint Classes*). As classes OWL podem se sobrepor e tornar a ontologia inconsistente, por isso é importante separar cada grupo de classes, criando-as como disjuntas, a fim de garantir que não ocorra a sobreposição. Isso garante que um indivíduo¹⁰⁵ instanciado como membro de uma classe não possa ser membro de nenhuma outra classe naquele grupo. Por exemplo, ao definir que *AutorFato* e *Vitima* são um grupo de classes disjuntas, estamos determinando que os indivíduos que pertencem a esta classe não podem pertencer a nenhuma outra daquele grupo. Na figura 22, apresentamos a hierarquia de classes estabelecida com base no estágio de conceitualização.



Figura 22: Hierarquia de classes - Protégé.
Fonte: elaborado pela autora.

¹⁰⁵ Indivíduo, membro e instância são considerados sinônimos no Protégé.

Ao incluir as classes e subclasse no editor de ontologias as relações taxonômicas são automáticas, evidenciadas pela relação *é-um*. Já a relação *parte-de* não funciona de forma tão simples. Identificamos a meronímia relacionada às classes *Jecrim*, *Vara* e *Forum*, pertencentes à categoria *IntituiçõesLegais*. Outra categoria que apresenta meronímia é *EventosLegais*, na qual *AudiênciaPreliminar* e *AudiênciaInstJulg* possuem classes merônimas. Ao incluir as classes, foi necessário atentar para este fato, pois precisamos criá-las como irmãs e não como filhas. Isto acarreta que a visão da taxonomia não evidencia a relação de parte-todo entre as classes e, além disso, exige que seja criada uma propriedade de objeto¹⁰⁶ *parteDe* para representar a meronímia.

Duas características do OWL nos levaram a buscar uma solução alternativa para a inclusão da relação de meronímia na ontologia: (1) no OWL as propriedades de objeto servem para estabelecer relações entre indivíduos e não entre classes; e (2) as limitações da Lógica de Descrição usada no OWL não permitem restringir os participantes das relações se eles não forem indivíduos. Com isso, se usássemos somente uma relação *parteDe*, não teríamos como definir individualmente os pares de relação.

Para resolver este problema criamos uma propriedade de objeto *parteDe* e oito subpropriedades *parteDe1*, *parteDe2*, e assim sucessivamente, nas quais podemos definir explicitamente os pares que se relacionam. As relações parte-de estabelecidas para as classes *AudiênciaPreliminar* e *AudiênciaInstJulg* podem ser visualizadas na figura 23. Usamos o mesmo recurso para estabelecer a relação de meronímia entre as classes *Forum*, *Vara* e *Jecrim*.

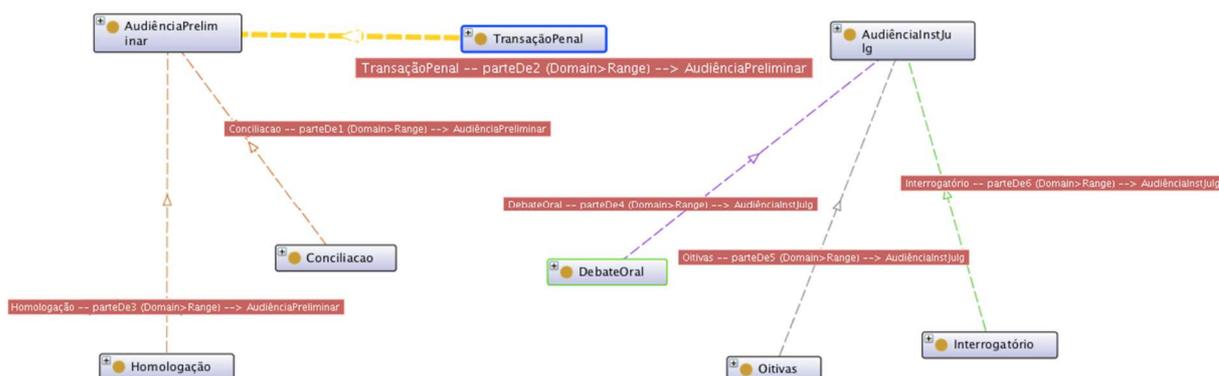


Figura 23: Relação parte-de – classe EventosLegais - Protégé.

Fonte: elaborado pela autora.

¹⁰⁶ Propriedade de objeto é um termo da linguagem OWL. As propriedades OWL representam relacionamentos. Existem dois principais tipos de propriedades: propriedades de objetos e propriedades de tipo de dados. As propriedades de objeto são relações entre dois indivíduos.

A expressão $\langle \text{Transação Penal} - \text{parteDe2 (Domain>Range)} \rightarrow \text{AudienciaPreliminar} \rangle$ corresponde à forma como o editor representa os pares de relação. Nesta relação, entendemos o *domain* como origem e o *range* como destino, ou seja, a relação tem uma origem em transação penal e o destino é audiência preliminar: transação penal é parte-de audiência preliminar. Desta forma foi possível definir todas as classes relacionadas através de relação parte-de.

Visando à documentação da ontologia, o Protégé permite a inclusão de anotações para cada classe. Estas anotações poderão ser úteis para desenvolvedores que queiram reutilizar a ontologia. No nosso caso, escolhemos usar o recurso de anotação para marcar os *frames*, que foram incluídos na estrutura taxonômica como classes, através de sua descrição. Entendemos importante deixar marcada a interface com os *frames* também através de uma explicação textual. Na figura 24 é possível visualizar um exemplo desta documentação dos *frames*, através da classe *Recurso*.

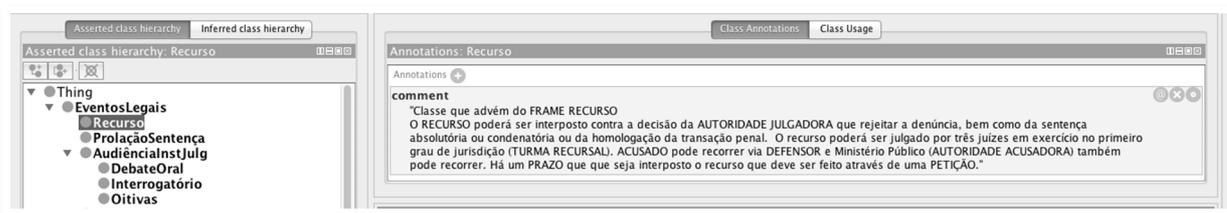


Figura 24: Inclusão de comentário para documentação no Protégé.
Fonte: elaborado pela autora.

Antes de buscar soluções para a inclusão das relações não taxonômicas na ontologia, vamos ampliar a estrutura de classes e subclasses inserindo informações sobre as variantes (analisadas e anotadas no quadro 16). Em um primeiro momento vamos apenas marcar que são classes equivalentes, pois esta é a denominação existente no Protégé para, mais tarde, buscar a forma mais adequada de relacioná-las aos respectivos *frames*. O Protégé dispõe da relação de equivalência (*Equivalent To*) que será usada para identificar as variantes.

Para incluir as variantes, é necessária a inserção das classes que serão consideradas equivalentes e, após, usar a relação de equivalência para fazer esta marcação. Exemplificamos este processo através das classes *Conciliacao*, *Composicao* e *Acordo*. A primeira já constava como uma classe na ontologia, enquanto que as outras duas foram incluídas para

serem marcadas como equivalentes. A inclusão da relação de equivalência pode ser visualizada na figura 25.

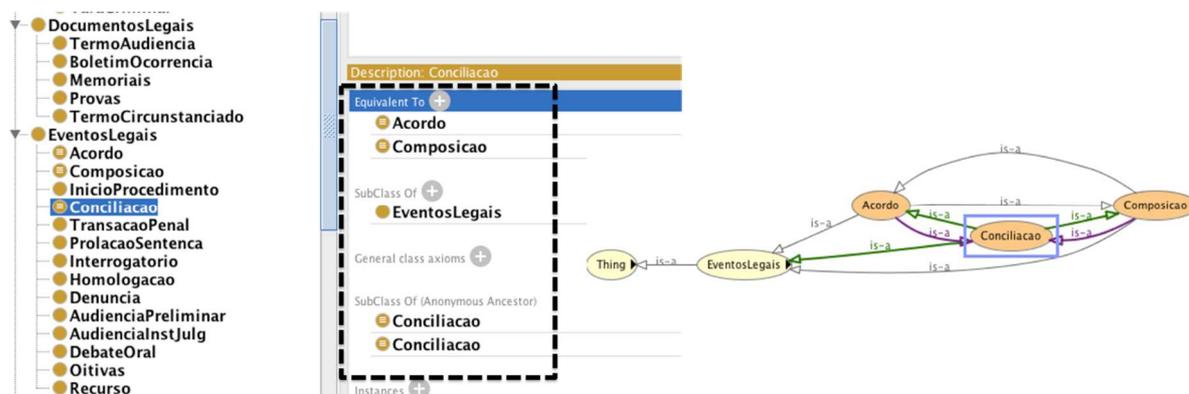


Figura 25: Inclusão da relação de equivalência – Conciliação, Composição e Acordo.
Fonte: elaborado pela autora.

A figura 25 mostra três formas de visualização da relação de equivalência: a primeira indicando na estrutura taxonômica (esquerda); a segunda, apresentando a descrição da classe (caixa pontilhada); e, a terceira, em formato gráfico ligando as classes por duas setas com a relação *é-um*. As outras variantes identificadas a partir do estudo dos *frames* e do mapa conceitual foram incluídas da mesma forma seguindo o disposto no quadro 16, com exceção das palavras relacionadas aos papéis dos participantes, uma vez que estas terão uma forma de inclusão diferenciada.

Passamos à análise das relações que correspondem aos papéis e funções dos participantes e nos deparamos com uma situação complexa, pois identificamos que temos relações **binárias** – de uma classe para outra classe, por exemplo: *Acusação oferece Denúncia* – e **n-árias** – de uma classe para várias outras, por exemplo: *AutorFato temPapelDe Beneficiário em TransaçãoPenal*. Esta situação é considerada complexa porque o OWL suporta somente as relações binárias. Para a descrição das n-árias vamos adotar a solução sugerida pelo W3C para estas situações, mais especificamente adotaremos o padrão 1 o qual soluciona esta questão introduzindo uma nova classe exclusivamente para permitir representar relações n-árias¹⁰⁷.

Exemplificamos inicialmente as relações binárias através do participante *Vítima*. As relações entre classes serão representadas através de propriedades de objeto. Conforme

¹⁰⁷ Solução encontrada no site do W3C. Disponível em www.w3.org/TR/swbp-n-aryRelations/#pattern1 Acesso em julho de 2015.

visualizado no quadro 23, a *Vítima* participa de Audiências, é ouvida em Oitivas e é representada pelo Ministério Público.

Representamos inicialmente a relação *participa* indicando que a *Vítima* *participa* Audiências. Como a participação da *Vítima* se dá tanto na Audiência Preliminar, quanto na de Instrução e Julgamento usamos o conector lógico **ou** (*or*). A criação da propriedade de objeto e pode ser visualizada na figura 26.

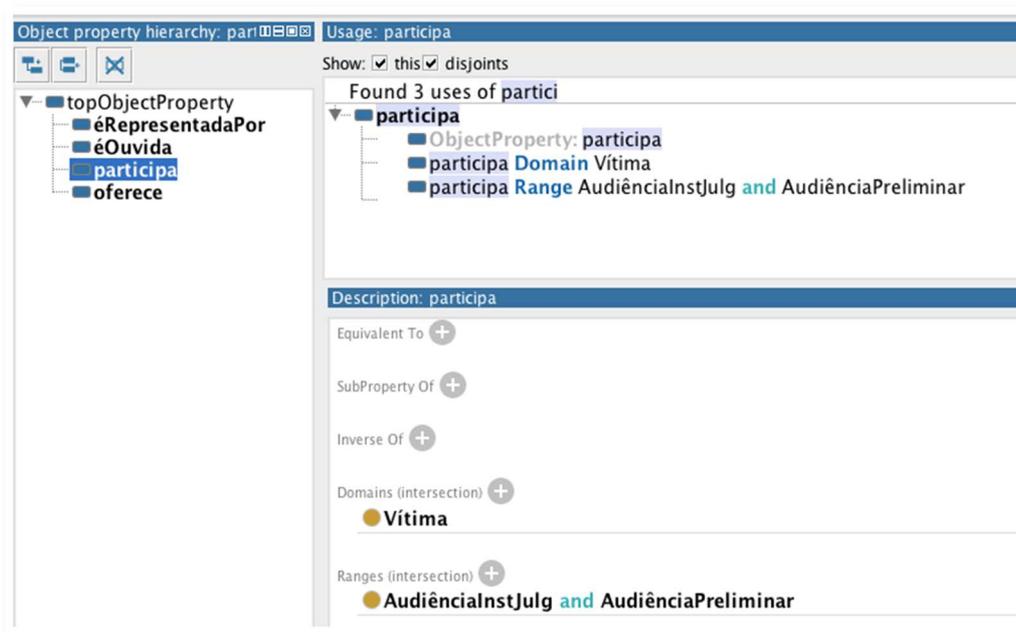


Figura 26: Propriedade de objeto *participa*.

Fonte: elaborado pela autora.

A classe *Vítima* também tem uma relação com a classe *Oitivas*, na qual ela é ouvida em seu depoimento. Para a representação desta relação, criamos a propriedade de objeto *ouvidaEm* que pode ser visualizada na figura 27. Esta figura apresenta um outro formato de visualização das relações disponível no Protégé, através de um esquema gráfico.

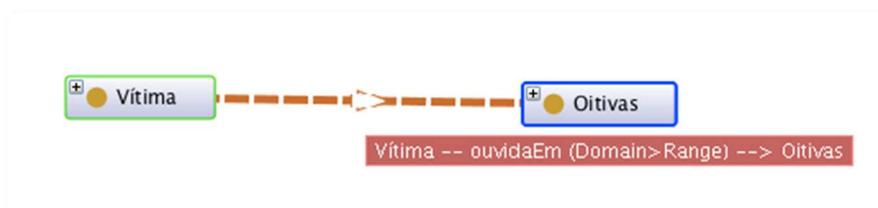


Figura 27: Propriedade de objeto *ouvidaEm*.

Fonte: elaborado pela autora.

Ainda referente à classe *Vítima*, temos a relação de ser representada pelo Ministério Público. Para representá-la criamos a propriedade de objeto *representadaPor* e incluímos na descrição da ontologia. A figura 28 mostra as relações incluídas para a classe *Vítima*.

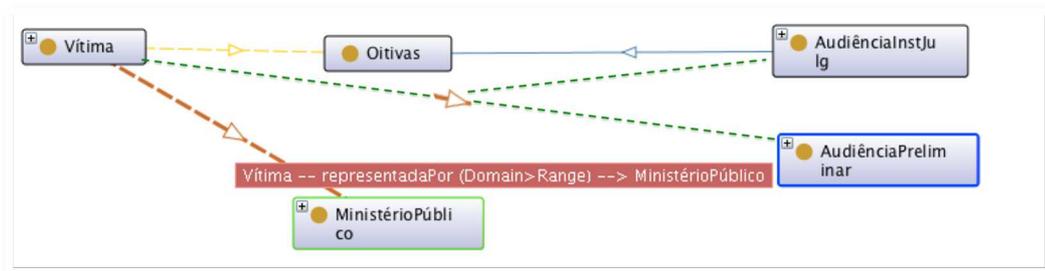


Figura 28: Relações não taxonômicas classe *Vítima*.
Fonte: elaborado pela autora.

A representação das relações n-árias, conforme já referimos, é um pouco mais complexa e seguiremos a sugestão do *W3C* para solucionar o problema. No entanto, não se pode apenas incluir uma classe para ligar todas as relações, conforme prescrito na solução do *W3C*, pois há uma série de ações que antecedem esta etapa.

A primeira ação para representar as relações n-árias foi criar uma classe para incluir todos os *frames*. Criamos a classe *FramesProcSum* e incluímos todos os *frames* como sendo indivíduos (*instances*) desta classe a fim de solucionar a representação. Na figura 29 apresentamos o resultado desta inclusão na ontologia.

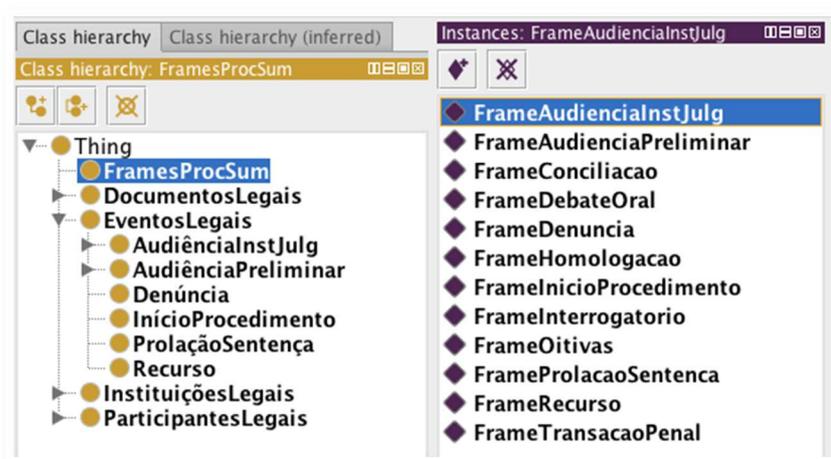


Figura 29: Classe *FrameProcSum* para inclusão dos frames como instâncias
Fonte: elaborado pela autora.

O segundo passo é a criação de uma classe `Papeis` e a inclusão de todos os papéis como indivíduos, seguindo a mesma abordagem já apresentada no primeiro passo. A figura 30 representa a classe `Papeis` e seus respectivos indivíduos. Salientamos que, até aqui, ainda não foram inseridas as relações; estes passos são apenas a preparação para sua inclusão.

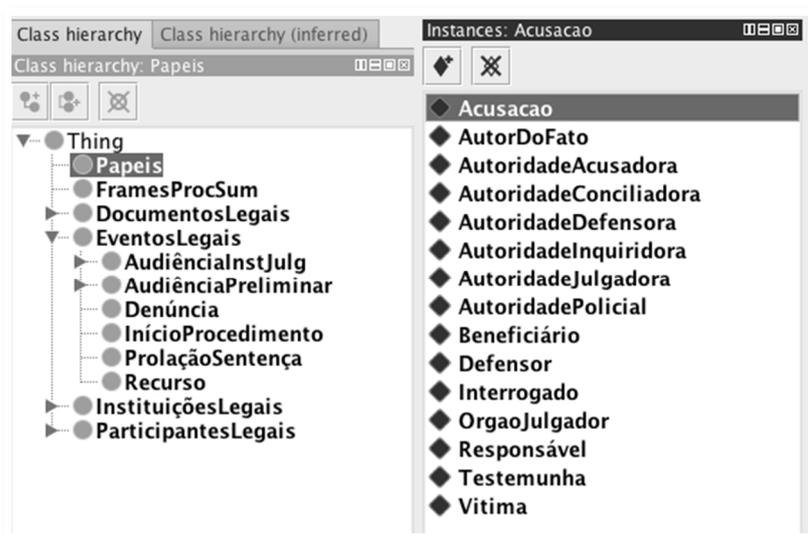


Figura 30: Classe `Papeis` para inclusão dos papéis como instâncias
Fonte: elaborado pela autora.

Após os passos de preparação para representação das relações n-árias, com a inclusão das novas classes e respectivos indivíduos, seguimos a abordagem sugerida pelo *W3C* e citada anteriormente, de inserir uma classe especificamente para representar a ligação entre os participantes e seus papéis nos diferentes *frames*. Para isso criamos a classe `Liga_Part_Papel_Frame`, cuja finalidade é exclusivamente representar a relação.

Criamos as propriedades de objeto *fazParteDoFrame*, *temPapel* e *temParticipante* e as relacionamos à classe `Liga_Part_Papel_Frame`. Esta ação permitirá estabelecer a relação n-ária entre Autor do fato e seu papel de Beneficiário no *frame* de Transação Penal, por exemplo (e todas as outras que seguirem este padrão). As relações criadas podem ser visualizadas na figura 31. A inclusão dos papéis dos demais participantes seguiu a mesma lógica exemplificada.

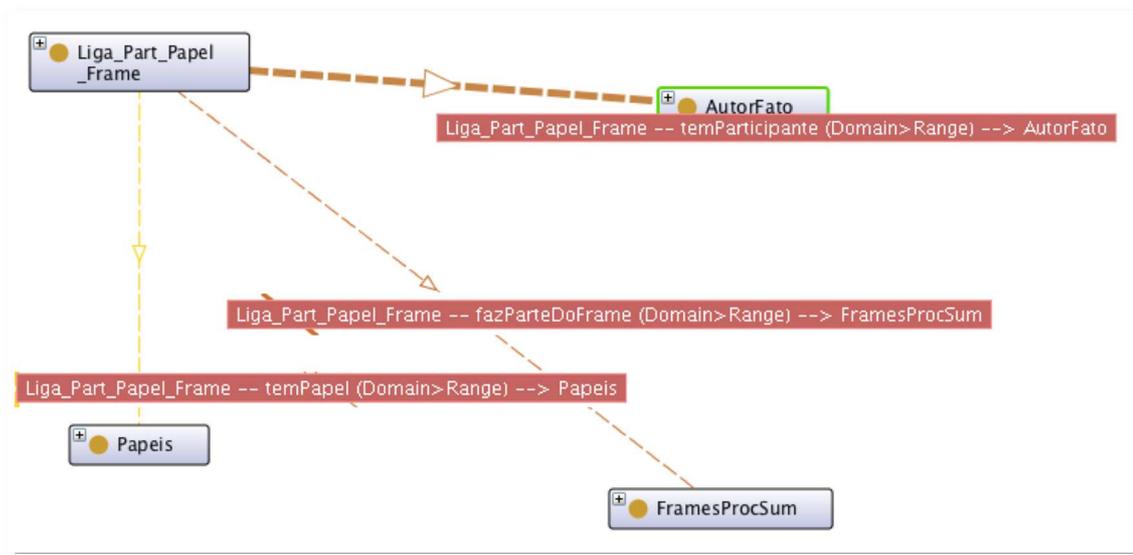


Figura 31: Representação das relações n-árias.
Fonte: elaborado pela autora.

Para exemplificar a relação de forma mais concreta, criamos um indivíduo na classe `AutorFato` (`Autor1108`) e estabelecemos as propriedades de objeto de modo a descrever que o *frame* `TRANSAÇÃO_PENAL` *temParticipante* `AutorFato`, que *temPapel* de `BENEFICIÁRIO`. A figura 32 ilustra, à esquerda a hierarquia de classes, no centro o indivíduo *liga* que é responsável pela ligação entre todas as relações e à direita a descrição através das propriedades de objeto. Desta forma conseguimos representar e descrever a relação entre o *frame*, o participante e seu papel no respectivo *frame*, sugerindo uma solução para resolver o problema relacionado aos diferentes papéis que os participantes podem desempenhar, uma vez que cada papel será explicitado e relacionado ao respectivo *frame*.

¹⁰⁸ Para identificar os indivíduos usaremos a fonte Arial Narrow.



Figura 32: Relações n-árias entre Autor, frame e papel.
Fonte: elaborado pela autora.

As relações entre os *frames* também são importantes para a descrição do domínio, sendo que **herança** e **subframe** são incluídas automaticamente através da estrutura hierárquica. Porém, para a inclusão das relações de **dependência** e **precedência** foi necessária a criação de propriedades de objeto: *dependeDe* e *precede*, respectivamente.

Em função das limitações da Lógica de Descrição não há como definir as regras que regem, no mundo real, as relações entre todos os eventos legais. Isto porque, as propriedades de objeto não tem, no OWL, a capacidade de restringir os participantes destas relações, mas sim de sugerir/indicar o tipo dos indivíduos. Por esta razão, o *domain* e o *range* da relação *precede* serão definidos como a superclasse *EventosLegais*. O mesmo vale para a relação *dependeDe*.

Outra informação advinda da análise conceitual que consideramos importante incluir na ontologia, são as unidades lexicais relativas aos *frames*. Para isso, o primeiro passo foi a inclusão de uma classe *UnidadeLexical* e suas instâncias: as unidades lexicais identificadas anteriormente (através da organização dos *frames* e pela análise da lista de palavras). A figura 33 ilustra este processo.

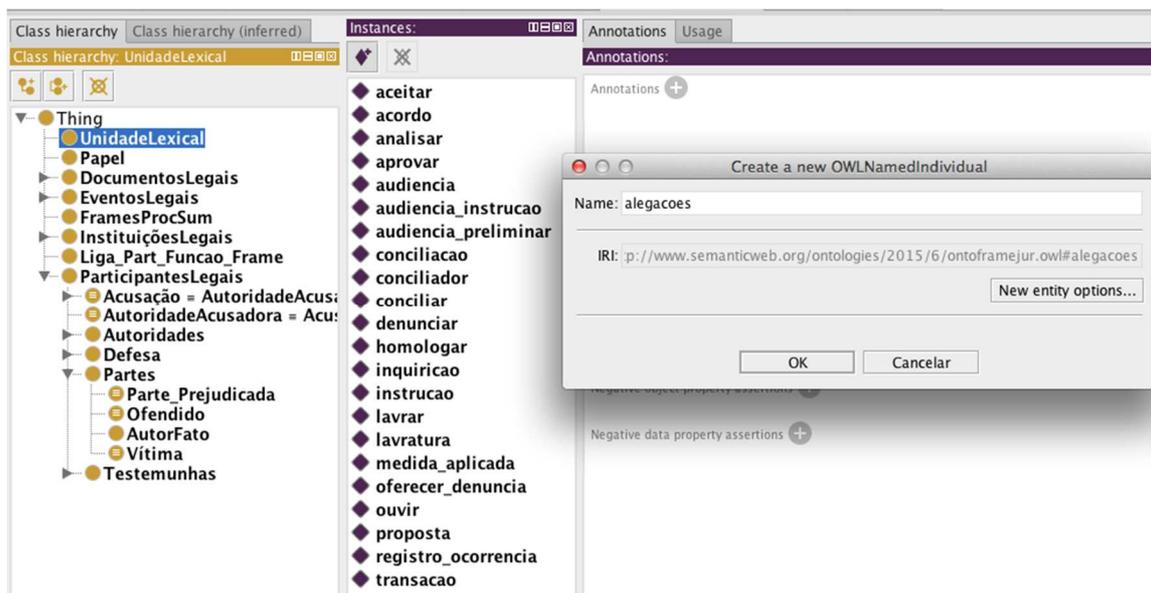


Figura 33: Inclusão das unidades lexicais.

Fonte: elaborado pela autora.

Para a inclusão das unidades lexicais, relacionando-as aos respectivos *frames*, criamos a propriedade de objeto *unidadeLexicalDe*. As relações foram estabelecidas entre os indivíduos das classes *FrameProcSum* e *UnidadeLexical*. Na figura 34 exemplificamos este processo através das unidades lexicais relacionadas ao *frame DEBATE_ORAL*, associando as UL's previstas na etapa de construção de *frames* (*debater*, *provar*, *defender*, *argumentar*) e também a UL *alegações*, que foi incluída a partir da análise da lista de palavras.

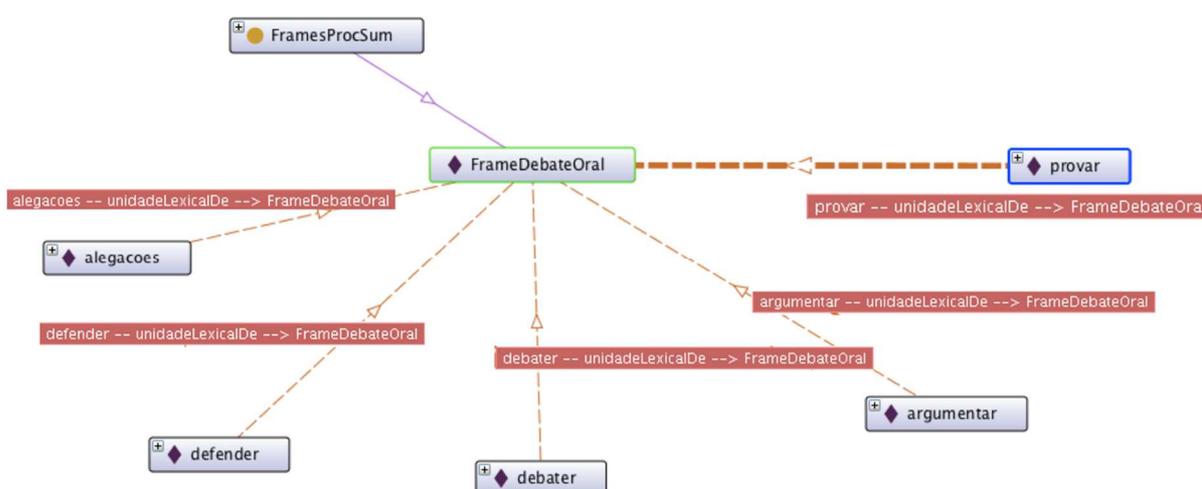


Figura 34: Unidades lexicais – Frame Debate_Oral.

Fonte: elaborado pela autora.

Seguindo com a implementação, incluímos as demais relações advindas do mapa conceitual e que foram listadas na seção anterior. Para representá-las criamos as propriedades de objeto *gera*, *eEncaminhadoPara*, *envolve* e *podeSubstituir*. A interface para a criação das propriedades de objeto poder ser visualizada na figura 35. A descrição destas relações segue os mesmos padrões já exemplificados quando apresentamos a solução de implementação para as relações binárias. A inclusão de todas as relações binárias seguiu este mesmo procedimento: criação da propriedade de objeto e estabelecimento do relacionamento.

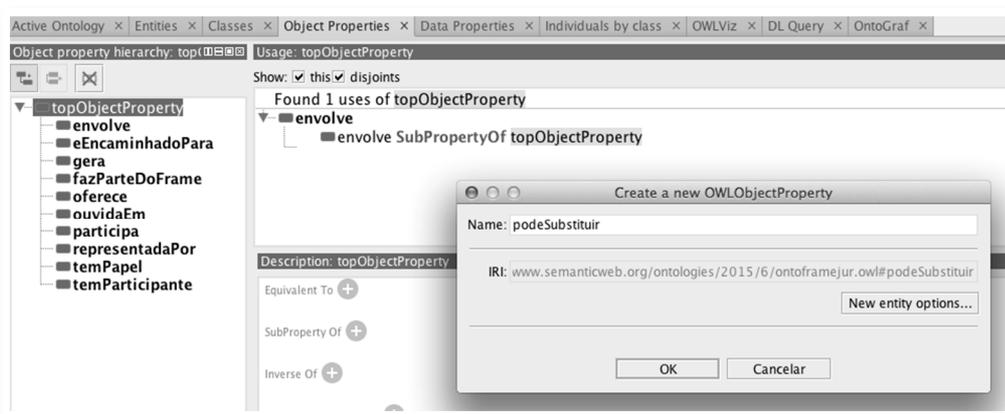


Figura 35: Interface para criação das propriedades de objeto.
Fonte: elaborado pela autora.

Ao final da implementação das relações entendemos ser necessário incluir na ontologia uma classe referente ao crime de menor potencial ofensivo. Esta classe não pertence às categorias básicas, mas corresponde a uma importante informação para o desenvolvimento do procedimento sumaríssimo. Criamos a classe *CrimeMenorPotencial* e a relacionamos com a classe *InicioProcedimento* através da propriedade de objeto *temInicioCom* estabelecendo a relação $\langle \text{InicioProcedimento} \text{ --} \rangle \text{ temInicioCom} (\text{domain} \rangle \text{range}) \text{ --} \rangle \text{ CrimeMenorPotencial}$.

Por fim, visando à recuperação da informação, consideramos a inclusão de atributos para os documentos e os participantes (nome, data, endereço – informação advinda da lista de palavras). Para a inclusão dos atributos, criamos propriedades de dados¹⁰⁹. Na figura 36 exemplificamos a inclusão das propriedades de dados para as classes *ParticipantesLegais* e *DocumentosLegais*.

¹⁰⁹ Propriedade de dados são relações definidas na linguagem OWL entre instâncias de classes e tipos de dados.

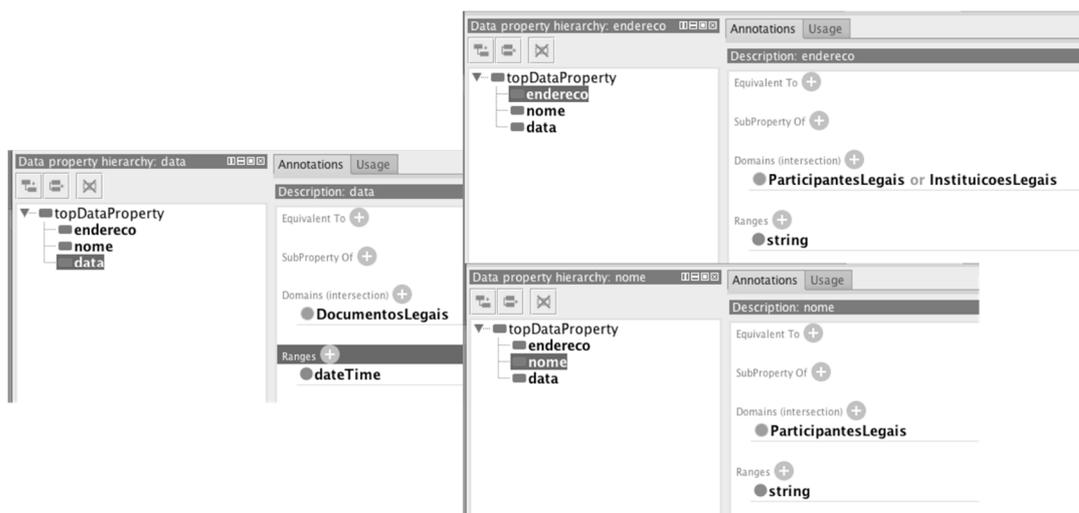


Figura 36: Propriedades de dados.

Fonte: elaborado pela autora.

As propriedades de dados estabelecem que todos os indivíduos das classes `ParticipantesLegais` e `InstituicoesLegais` têm um endereço, que todos os indivíduos da classe `ParticipantesLegais` têm nome e que todos os indivíduos da classe `DocumentosLegais` têm data. Nesta relação o *domain* é a classe e o *range* é o tipo de dado (*string* corresponde a uma sequência de letras e *dateTime*, ao formato da data)

Após a inclusão de todas as relações no editor, temos a ontologia piloto do procedimento sumaríssimo concluída e decidimos nomeá-la **OntoFrameJur**. A ontologia é composta por 51 classes e 38 propriedades de objeto definidas a partir das etapas de modelagem conceitual e nas necessidades de implementação. A `OntoFrameJur`, depois de concluída, passou pela etapa de avaliação, descrita na seção seguinte.

5.5 Avaliação e documentação

Seguindo o prescrito na nossa proposta metodológica, a avaliação da ontologia será realizada sob dois enfoques:

- 1- Verificação automática para identificação e ajustes de possíveis inconsistências através da ferramenta OOPS!
- 2- Verificação manual para identificação de informações relevantes na amostra do *corpus* que são passíveis de representação na ontologia. A realização desta verificação incide em:
 - a. Escolha aleatória de cinco documentos do *corpus* de pesquisa (amostra);

- b. Consulta ao especialista para marcação simples (sublinhando apenas) de quais são as informações relevantes presentes dos documentos;
- c. Verificação da possibilidade de representação destas informações na ontologia;
- d. Cálculo da abrangência do resultado seguindo a fórmula: $A=AC/C$

A ferramenta OOPS! avalia basicamente aspectos estruturais da ontologia, sendo válida para a verificação de erros ou inconsistências. A avaliação realizada segue os parâmetros descritos no ANEXO A.

A avaliação com a ferramenta foi satisfatória, identificando que a estrutura ontológica não apresentou nenhum erro considerado crítico. No entanto, retornou algumas sugestões para melhoria da estrutura. O retorno da ferramenta pode ser visualizado na figura 37.

Evaluation results	
It is obvious that not all the pitfalls are equally important; their impact in the ontology will depend on multiple factors. For this reason, each pitfall has an importance level attached indicating how important it is. We have identified three levels:	
<ul style="list-style-type: none"> ▪ Critical 🚫 : It is crucial to correct the pitfall. Otherwise, it could affect the ontology consistency, reasoning, applicability, etc. ▪ Important ⚠️ : Though not critical for ontology function, it is important to correct this type of pitfall. ▪ Minor 🟡 : It is not really a problem, but by correcting it we will make the ontology nicer. 	
[Expand All] [Collapse All]	
Results for P02: Creating synonyms as classes.	4 cases Minor 🟡
Results for P04: Creating unconnected ontology elements.	3 cases Minor 🟡
Results for P08: Missing annotations.	89 cases Minor 🟡
Results for P13: Missing inverse relationships.	39 cases Minor 🟡
Results for P22: Using different naming criteria in the ontology.	ontology* Minor 🟡
Results for P36: URI contains file extension.	ontology* Minor 🟡
SUGGESTION: symmetric or transitive object properties.	3 cases

Figura 37: Resultado avaliação automática – OOPS!

Fonte: elaborado pela autora

As duas questões mais relevantes apontadas pela ferramenta são: a falta de anotação para todas as classes. Realmente, anotamos somente as classes que tinham relação com os *frames* para reforçar esta interface. Temos ciência de que a anotação, tanto de classes quanto de propriedades de objeto, é muito importante para a documentação da ontologia. Contudo, é necessário destacar que o registro de informações sobre as classes e propriedades encontra-se descrito na seção de modelagem conceitual deste trabalho. E o outro aspecto relevante apontado pela ferramenta foi o fato de não serem definidas relações inversas do tipo *temReferencia* e *eReferidoPor*.

A segunda forma de avaliação realizada configurou-se efetivamente manual. Inicialmente consideramos separar aleatoriamente quatro documentos do *corpus*. No entanto, levando em conta o fato de que os documentos, com excessão da ata de audiência, referem-se à 2ª instância, conforme já explicado na seção sobre o *corpus* de pesquisa, entendemos a necessidade de modificar esta escolha; optamos por usar somente cinco termos de audiência para a avaliação manual da ontologia. Esses termos foram encaminhados ao Juiz de Direito da 2ª Vara Cível de Campo Bom a fim de que ele, como especialista da área, marcasse as informações relevantes.

De posse dos documentos marcados pelo especialista, verificamos a correspondência com a ontologia, ou seja, cada informação marcada foi analisada no sentido de verificar se havia uma classe a qual pudesse ser relacionada.

Na figura 38 é possível visualizar a análise realizada. As informações estão marcadas seguindo a seguinte notação: com tarja amarela, as informações relevantes marcadas pelo juiz; com letras vermelhas, as informações que identificamos como possíveis de representar na ontologia; e, nas caixas, as classes e/ou propriedades que permitiram a associação.

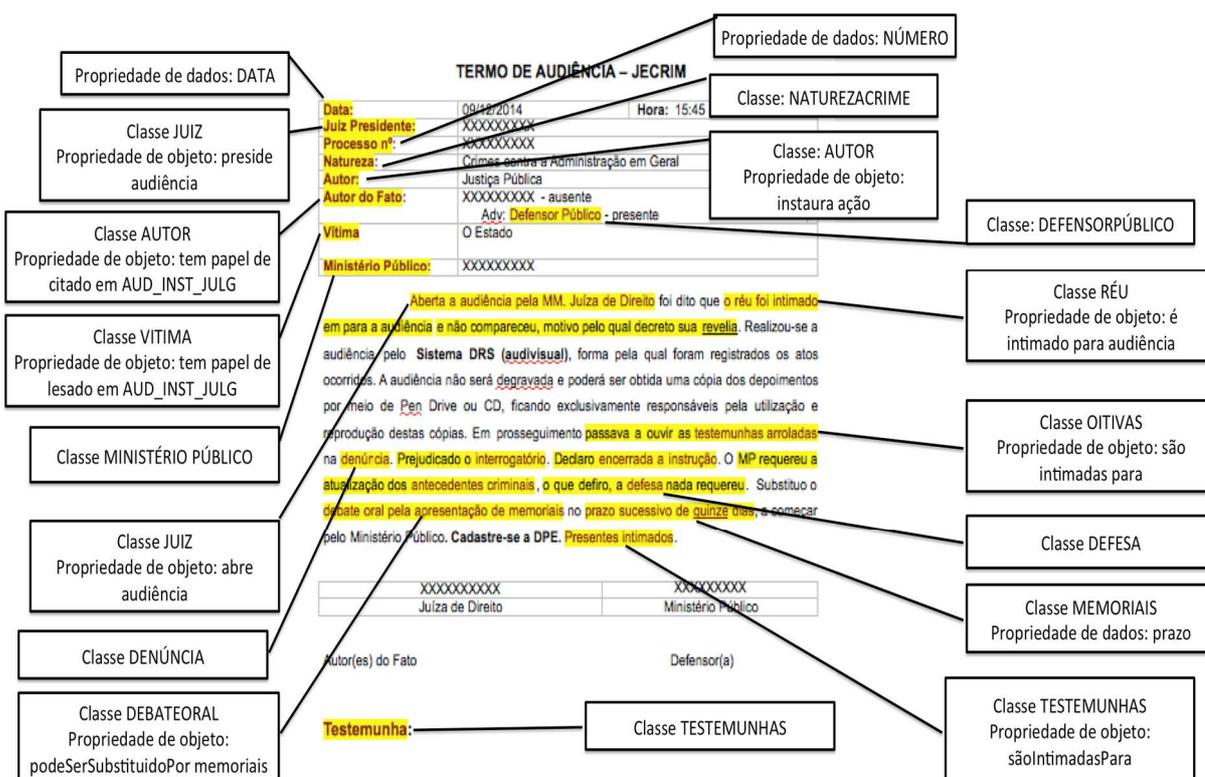


Figura 38: Avaliação manual - documentos

Fonte: elaborado pela autora

Organizamos a informação advinda da avaliação no quadro 30.

Documentos	Total de informações relevantes (C)	Total de informações mapeadas (AC)
AtaJECrim1 ANEXO B	16	10
AtaJECrim2 ANEXO C	20	17
AtaJECrim3 ANEXO D	21	17
AtaJECrim4 ANEXO E	15	12
AtaJECrim5 (figura 38)	21	21
TOTAL	77	65
RESULTADO (A=AC/C)		84.4%

Quadro 30: Resultado da avaliação manual

Fonte: Elaborado pela autora

Obtivemos como resultado 84.4% de informações relevantes presentes nos documentos que puderam ser mapeadas na ontologia. Com este resultado foi possível fazer duas constatações:

1. A ontologia não prevê a possibilidade de representar informações que não sejam casos prototípicos, como por exemplo: “réu não compareceu à audiência... juiz decreta sua revelia” ou “defesa nada requereu”. Isto se deve ao fato de optarmos por apresentar somente o caso típico do procedimento sumaríssimo, uma vez que este envolve muitas especificidades. Porém, entendemos que a ontologia, da forma como está estruturada, permite facilmente a inclusão destas novas informações.
2. A escolha pela avaliação manual tomando como base informações anotadas pelo especialista foi bastante produtiva, pois permitiu identificar exatamente o tipo de informação que pode ou não ser mapeado na ontologia.

Finalizando esta etapa de avaliação, cabe uma última observação. O ideal seria que a avaliação pudesse ser feita toda de forma automática, porém não encontramos nenhum sistema capaz de avaliar seguindo os mesmos critérios que usamos. Temos ciência de que a ontologia construída pode ser aperfeiçoada e ampliada com novas classes e propriedades.

Em relação à documentação da ontologia já incluímos alguns aspectos na própria estrutura OWL, tais como a glosa dos *frames*, a explicação sobre a inclusão de algumas classes (*FramesProcSum*, *Papel*, *UnidadeLexical*, entre outras) que foram necessárias para a implementação e a explicação sobre as propriedades de objeto. Além destas

informações, consideramos que os capítulos 4 e 5 deste trabalho correspondem ao guia relacionado à nossa metodologia, ou seja, são a documentação propriamente dita.

5.6 Síntese do capítulo: a OntoFrameJur

A figura 39 ilustra parcialmente a OntoFrameJur, mostrando seu tamanho e complexidade.

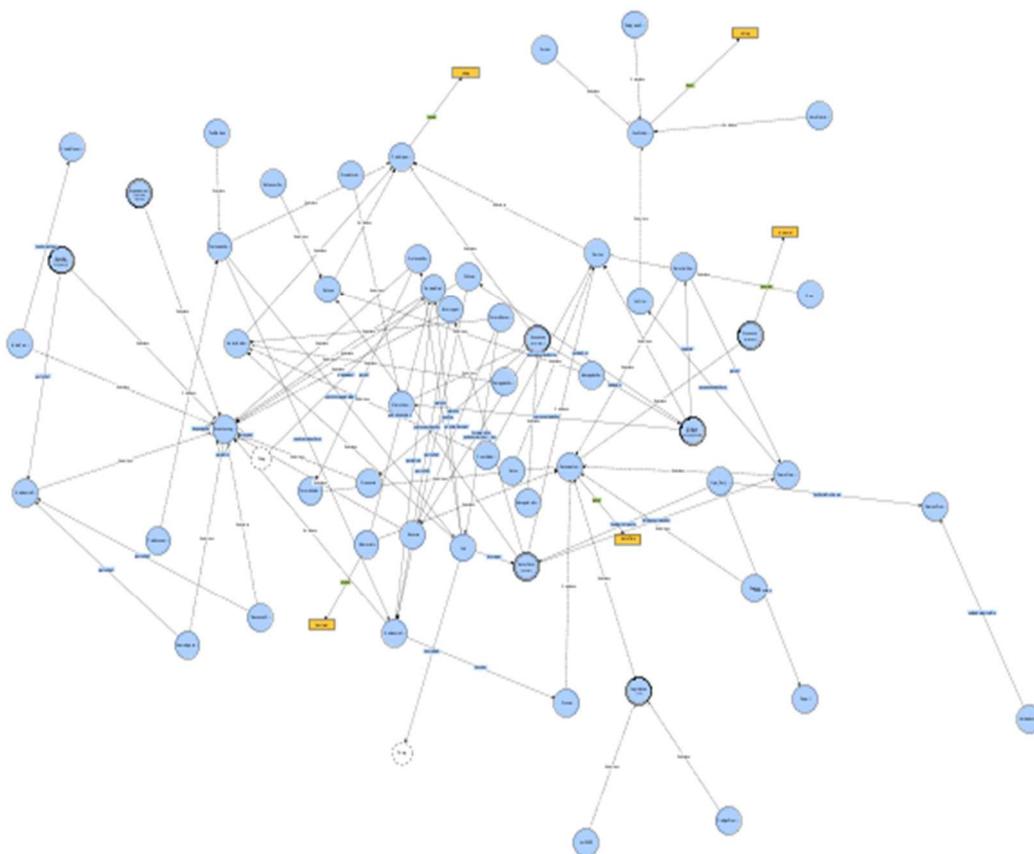


Figura 39: OntoFrameJur
Fonte: elaborado pela autora

Concluído este capítulo de modelagem e avaliação da ontologia, passamos às considerações finais sobre esta pesquisa.

6 CONSIDERAÇÕES FINAIS

O domínio jurídico se caracteriza pela complexidade de sua linguagem, adotando uma terminologia especializada. Além disso, constitui-se por uma grande quantidade de documentos gerados por inúmeros processos nas diferentes instâncias que precisam ser facilmente recuperados e localizados, a fim de sustentar novas decisões ou embasar novas solicitações. Tais motivos levam profissionais da área da Computação e do Direito a buscarem soluções no sentido de desenvolver sistemas semanticamente fortes capazes de contribuir com a recuperação da informação jurídica. A Linguística, através da Semântica Computacional, vem buscando seu espaço a fim de contribuir no desenvolvimento de soluções semânticas para a recuperação da informação.

Este trabalho insere-se nesta seara e visa à estruturação do Procedimento Sumaríssimo no Juizados Especiais Criminais através de um estudo conceitual do domínio, objetivando o desenvolvimento de uma ontologia baseada em *frames* capaz de cobrir os conceitos da área.

Tendo em mente o cenário descrito, a investigação realizada nesta tese procurou responder à questão: *Considerando o propósito de prover maior significado à descrição na ontologia do domínio jurídico tomando como base a Semântica de Frames, quais são os passos metodológicos a serem adotados?* e atender ao objetivo geral de *refletir sobre as convergências entre ontologias e frames, de modo a acreditar-se que a combinação destas duas formas de organizar o conhecimento garantem a representação mais precisa de conceitos e relações seguindo etapas metodológicas reinterpretadas à luz da Semântica de Frames.*

Com o intuito de prover a resposta ao questionamento inicial, a pesquisa possibilitou percorrer caminhos teóricos fundamentados nas áreas da Filosofia, da Ciência da Computação e da Ciência da Informação, bem como visualizar a intersecção entre o foco de estudo e a Linguística, sob apresentação dos diferentes aspectos das ontologias.

Seguindo o elemento motivador de prover a inclusão de maior significado às ontologias, aliamos a estrutura lógica normalmente associada às ontologias a uma descrição linguística baseada em *frames*, opção teórico-metodológica que, a nosso ver, resultou em uma resposta favorável ao nosso questionamento. Passemos às evidências.

Consideramos que esta tese contribui para a área das ontologias sob três aspectos. Primeiro por demonstrar a proficuidade da SF para a descrição dos conceitos em uma ontologia, provendo detalhamento semântico à estrutura descritiva do domínio através da inserção e distinção dos papéis dos participantes, por exemplo. Com esta abordagem pretende-se minimizar os problemas relacionados ao significado e à ambiguidade apontados pelos estudiosos da área da CI. Segundo, por apresentar uma metodologia reinterpretada a partir das vigentes, capaz de cobrir o domínio de forma diferenciada com base em *frames*. Além disso, a descrição das etapas propostas e os exemplos demonstrados na parte aplicada apresentam os detalhes necessários para o entendimento e possível reuso. E, terceiro, por apresentar uma maneira de uso do editor de ontologias com sugestões para a inclusão da informação conceitual.

A contribuição da tese na área da Linguística reside no fato de demonstrar que se pode ir além da Semântica Lógica ou da Semântica Lexical para descrever o domínio em ontologias. A pesquisa demonstrou que a descrição considerando a SF é viável e, principalmente, traz uma possibilidade de solução para o problema relacionado aos diferentes papéis que os participantes podem exercer, uma vez que a estrutura de *frames* permite explicitá-los na estrutura ontológica.

O estudo teórico acerca das ontologias e da SF amparou nossas escolhas e decisões na parte aplicada do trabalho, e fundamentou uma metodologia diferenciada para a construção de ontologias de domínio. Entendemos que a metodologia proposta corresponde a uma inovação na área das ontologias por apresentar a descrição do domínio a partir dos *frames*, diferentemente das propostas que pretendem transformar a FN em uma ontologia. Além disso, consideramos que esta metodologia pode ser facilmente utilizada para construção de ontologias de outros domínios. No entanto, temos ciência de que se trata de uma abordagem muito mais manual que automática. Esta constatação vai na direção contrária da maior parte de estudos recentes, os quais visam uma automatização cada vez maior para o processo de desenvolvimento de ontologias.

Defendendo nossa ideia de uma metodologia mais manual, consideramos que o tratamento automatizado para todo o processo de construção da ontologia é, justamente, o obstáculo para prover-lhe mais significado. Isto posto, entendemos a que nossa proposta metodológica possa ser utilizada para qualquer domínio que, assim como o jurídico, se preste à uma descrição através de *frames* justamente pela profusão de cenas e por sua dinâmica. Além disso, para facilitar o processo de conceitualização, a etapa de construção de *frames* pode ser

substituída pelo uso de *frames* existentes na FN, caso haja esta possibilidade (caso hajam *frames* relacionados ao domínio a ser representado).

A ontologia piloto que desenvolvemos para testar nossa metodologia pode ser usada em: (a) aplicações que visem à recuperação da informação jurídica nos portais dos Tribunais de Justiça; (b) sistemas de busca “inteligentes” associados aos *sites* de pesquisa; (c) recursos lexicais cuja finalidade seja disponibilizar informação jurídica para leigos, entre tantas outras aplicações. No endereço <https://www.dropbox.com/s/hdw0x1hxqtazca2/ontoframe.owl?dl=1> é possível baixar o arquivo OWL com a ontologia completa.

Concluimos a pesquisa com resultado satisfatório; no entanto, isto não significa que o assunto esteja esgotado. Conforme vimos nas etapas de avaliação, a ontologia ainda pode ser aprimorada e ampliada a fim de obter melhores resultados. Como desdobramentos futuros vislumbramos a aplicação da metodologia baseada em *frames* para a construção de outras ontologias para a área jurídica, compondo-se uma rede de ontologias integradas que visem a recuperação da informação.

REFERÊNCIAS

AGNOLONI, T.; BARRERA, M. F., SAGRI, M.T.; TISCORNIA, D.; VENTURI, G. Proceeding AICOL-I/IVR-XXIV'09 **Proceedings of the 2009 international conference on AI approaches to the complexity of legal systems: complex systems, the semantic web ontologies, argumentation, and dialogue**. p. 93-112. 2009.

AJANI, Gianmaria et al. **Legal Taxonomy Syllabus version 2.0**. IDT, p. 9, 2009.

ALMEIDA, Gladis Maria de Barcellos; OLIVEIRA, Leandro Henrique Mendonça de; ALUÍSIO, Sandra Maria. A terminologia na era da informática. **Ciência e Cultura**, v. 58, n. 2, p. 42-45, 2006.

ARAÚJO, Isabel et al. ONTOJURIS E ONTOTRIB: ontologias para a modelagem do conhecimento jurídico. In: **Anais Seminário de Pesquisa em Ontologia no Brasil**. Rio de Janeiro, 2008.

BAKER, Collin F. FrameNet: A Knowledge Base for Natural Language Processing. **ACL 2014**, v. 1929, p. 1-5, 2014.)

BAKER, Collin F.; FILLMORE, Charles J.; LOWE, John B. The Berkeley Framenet Project. In: **Proceedings of the 17th international conference on Computational linguistics-Volume 1**. Association for Computational Linguistics, 1998. p. 86-90.

BĀRZDIŅŠ, Guntis et al. Multidimensional ontologies: integration of Frame semantics and ontological semantics. In: **Proceedings of the 13th EURALEX International Congress**. 2008. p. 277-283.

BERNERS-LEE, Tim et al. The semantic web. **Scientific american**, v. 284, n. 5, p. 28-37, 2001.

BERTOLDI, Anderson. **Semântica de frames e recursos lexicais jurídicos um estudo contrastivo**. Tese de doutorado. São Leopoldo: UNISINOS, 2011.

BEVILACQUA, Cleci Regina. Por que e para que a Linguística de Corpus na Terminologia. In: TAGNIN, Stella; BEVILACQUA, Cleci Regina (org). **Corpora na Terminologia**. Hub Editorial. São Paulo, 2013.

BOELLA, Guido et al. Eunomos, a legal document and knowledge management system for regulatory compliance. In: **Information Systems: Crossroads for Organization, Management, Accounting and Engineering**. Physica-Verlag HD, 2012. p. 571-578.

BORCHERT, D. M. **Encyclopedia of Philosophy**. 2. ed. Detroit: Macmillan Reference. 2006.

BORST, Wilen Nico. **Construction of Engineering Ontologies for Knowledge Sharing and Reuse**. 1997. Tese de doutorado. Disponível em: <http://www.ub.utwente.nl/webdocs/inf/1/t0000004.pdf>. Acesso em: 01 jun. 2013.

BRASIL. Constituição (1988). **Constituição da República Federativa do Brasil**. Brasília, DF: Senado, 1988.

BRASIL. **Lei nº 11280/06**. Publicado em 16 de fevereiro de 2006. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2004-2006/2006/Lei/L11280.htm Acesso em: 29 agosto 2013.

BRASIL. **Lei nº 3689/41**. Publicado em 03 de outubro de 1941. Disponível em: http://www.planalto.gov.br/ccivil_03/decreto-lei/Del3689.htm Acesso em: 29 agosto 2013.

BRASIL. **Lei nº 12403/11**. Publicado em 04 de maio de 2011. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/112403.htm Acesso em: 29 agosto 2013.

BRASIL. **Lei nº 9099/95**. Publicado em 26 de setembro de 1995. Disponível em: http://www.planalto.gov.br/ccivil_03/Leis/L9099.htm Acesso em: 29 agosto 2013.

BREUKER, J. A. P. J.; HOEKSTRA, R. J. **Epistemology and ontology in core ontologies: FOLaw and LRI-Core, two core ontologies for law**. 2004.

BREUKER, Joost; VALENTE, Andre; WINKELS, Radboud. Legal ontologies: a functional view. In: **Procs. of 1st LegOut Workshop on Legal Ontologies**. 1997. p. 23-36.

BREWSTER, C., ALANI, H., DASMAHAPATRA, S. e WILKS, Y. Data Driven Ontology Evaluation. In: **Proceedings of International Conference on Language Resources and Evaluation (LREC 2004)**, Lisboa, Portugal, 2004.

BURCHARDT, Aljoscha. **Modeling Textual Entailment with Role-Semantic Information**. Ph.D. thesis, Universitat des Saarlandes. 2008

CASTRO, S. **Ontologia**. Rio de Janeiro: Jorge Zahar, 2008.

CALZOLARI, Nicoletta; ZAMPOLLI, Antonio; LENCI, Alessandro. Towards a standard for a multilingual lexical entry: The EAGLES/ISLE initiative. In: **Computational Linguistics and Intelligent Text Processing**. Springer Berlin Heidelberg, 2002. p. 264-279.

CHISHMAN, Rove. **Projeto Tecnologias Semânticas e Sistemas de Recuperação de Informação Jurídica**. Projeto aprovado no Edital CAPES/CNJ Acadêmico 2010.

CORCHO O.; GÓMEZ-PÉREZ, A.; FERNÁNDEZ-LÓPEZ, M. **Ontological Engineering**. London: Springer Press. 2004.

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

CROFT, W.; CRUSE, D. A. **Cognitive Linguistics**. Cambridge Textbooks in Linguistics. 2004

DAHLBERG, I. **Fundamentos teórico-conceituais da classificação**. *Revista de Biblioteconomia de Brasília*, v. 6, n. 1, p. 9-21, jan./jun. 1978a.

DAS, Dipanjan et al. **Frame-Semantic parsing**. *Computational Linguistics*. 2013

DA SILVA, Daniela Lucas; SOUZA, Renato Rocha; ALMEIDA, Maurício Barcellos. **Ontologias e vocabulários controlados: comparação de metodologias para construção.** Ci. Inf, v. 37, n. 3, p. 60-75, 2008.

DE ALMEIDA CAMPOS, Maria Luiza. O papel das definições na pesquisa em ontologia. **Perspectivas em Ciência da Informação**, v. 15, n. 1, p. 220-238, 2010.

DE OLIVEIRA, Leandro Henrique Mendonça; ALUISIO, Sandra Maria; DE BARCELOS ALMEIDA, Gladis Maria. **e-Termos: um ambiente colaborativo web para criação de produtos terminológicos.** 2006.

DIMOULIS, Dimitri. **Manual de Introdução ao estudo do Direito: definição e conceitos básicos, norma jurídica.** 4. Ed. Rev. Atual. e ampl. – São Paulo: Editora Revista dos Tribunais, 2011. Págs. 177 a 180.

DRUMOND, Lucas; GIRARDI, Rosário. **A multi-agent legal recommender system. Artificial Intelligence and Law.** Springer, 2008.

EVENS, M. **Relational models of the lexicon.** Cambridge: Cambridge University Press, 1988.

EVANS, Vyvyan, GREEN, Melanie. **Cognitive Linguistics: An Introduction**, (Edinburgh University Press.) 2006

ERK, Katrin; PADO, Sebastian. Shalmaneser—a toolchain for shallow semantic parsing. In: **Proceedings of LREC.** 2006.

FARIA, Domingo. **Encontro de Amor: Ensaios de Filosofia.** Lisboa: Sítio do Livro. [Depósito legal: 303288/09. ISBN: 978-989-20-1835-5. Data de Publicação: Dezembro de 2009].

FARRAR, S.; BATEMAN, J. **Linguistic ontology baseline.** OntoSpace Internal Report. Bremen: Collaborative Research Center for Spatial Cognition, 2005. Disponível em: <<http://www.ontospace.uni-bremen.de/pub/FarrarBateman05-i1-d3.pdf>> Acesso em: 16/06/2012.

FENSEL, D. **Ontologies: silver bullet for knowledge management and electronic commerce.** Berlin: Springer-Verlag, 2000.

FERNÁNDEZ-LÓPEZ, Mariano; GÓMEZ-PÉREZ, Asunción; JURISTO, Natalia. **Methontology: from ontological art towards ontological engineering.** 1997.

FIGUEIREDO, Daniel R. Introdução a redes complexas. **Atualizações em Informática**, p. 303-358, 2011.

FILLMORE, C. An alternative to checklist theories of meaning. FIRST ANNUAL MEETING OF THE BERKELEY LINGUISTICS SOCIETY. In: **Proceedings of the first annual meeting of the Berkeley Linguistics Society.** Berkeley: Berkeley Linguistics Society, 1975, p.123-131.

FILLMORE, C. Frame semantics and the nature of language. In **Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and**

Speech, Vol. 280, 1976, p. 20-32.

FILLMORE, C. Scenes-and-frames semantics. In ZAMPOLLI, A. (Ed.): **Linguistic Structures Processing: Fundamental Studies in Computer Science**, No. 59, North Holland Publishing, 1977, p.55-88.

FILLMORE, C. Frame Semantics. In: **The Linguistic Society of Korea. Linguistic in the Morning Calm**, Seoul, Hansinh Publishing Co, 1982.

FILLMORE, C. **Frames and the semantics of understanding**. Quaderni di Semantica. Vol.6, N.2, 1985, p.222-254.

FILLMORE, C. Frame semantics. In: GEERAERTS, D. (Ed.). **Cognitive linguistics: basic readings**. Berlin: Mouton de Gruyter, 2006. p.373-400.

FILLMORE, Charles J.; JOHNSON, Christopher R.; PETRUCK, Miriam RL. Background to framenet. **International journal of lexicography**, v. 16, n. 3, p. 235-250, 2003.

FULLY, Roberta Pimentel; BERNARDES, Ricardo Carellos da Silva. Processo Judicial Eletrônico e seu reflexo no Direito Processual. In: **Âmbito Jurídico**, Rio Grande, XIV, n. 90, jul 2011. Disponível em: <http://www.ambitojuridico.com.br/site/index.php?n_link=revista_artigos_leitura&artigo_id=9812>. Acesso em agosto de 2013.

GANGEMI, A.; GUARINO, N.; MASOLO, C.; OLTRAMARI, A.; OLTRAMARI, R.; SCHNEIDER, L. (2002). **Sweetening Ontologies with DOLCE**. (2002). Disponível em: <<http://www.loa.istc.cnr.it/Papers/DOLCE-EKAW.pdf>>. Acesso em 10 jan. 2012.

GARCÍA MARCO, F. J. **Ontologías y organización del conocimiento: retos y oportunidades para el profesional de la información**. El profesional de la información, v. 16, n. 6, p. 541- 550, 2007. Disponível em: <<http://www.elprofesionaldelainformacion.com>>. Acesso em: 10 jan 2013.

GEERAERTS, D. Cultural models of linguistic standardization. In: Dirven, R., Frank, R. e Pütz, M. (eds.), **Cognitive Models in Language and Thought. Ideology, Metaphors and Meanings**. Berlin/New York: Mouton de Gruyter, 25-68. 2003

GEERAERTS, Dirk. 1995. Cognitive Linguistics. In J. Verschueren, J.-O. Östman and J. Blommaert, eds., **Handbook of Pragmatics**. Amsterdam: John Benjamins, 111-116.

GEERAERTS, Dirk. **Cognitive Linguistics: Basic Readings**. Berlin: Mouton.2006

GEERAERTS, Dirk; CUYCKENS, Hubert. **Handbook of Cognitive Linguistics**. Oxford University Press. 2010.

GILDEA, Daniel; JURAFSKY, Daniel. Automatic labeling of semantic roles. **Computational linguistics**, v. 28, n. 3, p. 245-288, 2002.

GILCHRIST, A. **Thesauri, taxonomies and ontologies: an etymological note**. Journal of Documentation, v. 59, n. 1, p. 07-18, 2003.

GÓMEZ-PÉREZ, Asunción. **Evaluation of taxonomic knowledge in ontologies and knowledge bases**. 1999.

GOODRICH, Michael; TAMASSIA, Roberto; MOUNT, David. **Data structures and algorithms in C++**. John Wiley & Sons, 2011.

GREEN, Rebecca. **Inducing Semantic Frames from Lexical Resources**. Ph.D. thesis, University of Maryland, College Park. 2004.

GRUBER, Thomas R. **Ontolingua: A mechanism to support portable ontologies**. Stanford University, Knowledge Systems Laboratory, 1992.

GRUBER, T. R. **A Translation Approach to Portable Ontology Specifications**. Knowledge Acquisition, 5, 199-220. 1993.

GRUBER, T. R. **What is an ontology?** 1996. Disponível em: <<http://www.ksl.stanford.edu/kst/what-is-an-ontology.html>>. Acesso em: 12 jun. 2013.

GRUNINGER, Michael; FOX, Mark S. **Methodology for the Design and Evaluation of Ontologies**. 1995.

GUARINO, Nicola. Formal Ontology in Information Systems. In: **Proceedings of the First International Conference (FIOS'98)**, June 6-8, Trento, Italy. IOS press, 1998.

GUARINO, Nicola. Some ontological principles for designing upper level lexical resources. In: **International Conference on Language Resources and Evaluation, 1., 1998, Granada. Proceedings**. 1998b. Disponível em: <http://www.loa.istc.cnr.it/old/Papers/LREC98.pdf>. Acesso em: 13 abr. 2015.

HJØRLAND, Birger. Epistemology and the socio-cognitive perspective in information science. In: **Journal of the American Society for Information Science and Technology**, v. 53, n. 4, p. 257-270, 2002.

HJØRLAND, Birger. Semantics and knowledge organization. In: **Annual Review of Information Science and Technology**, v. 41, n. 1, p. 367-405, 2007a.

HOEKSTRA, Rinke et al. The LKIF Core Ontology of Basic Legal Concepts. **LOAIT**, v. 321, p. 43-63, 2007.

HOVY, E. Comparing Sets of Semantic Relations in Ontologies. In: Green, R., Bean, C. e Myaeng, S. editors, **The Semantics of Relationships: An Interdisciplinary perspective**, Kluwer, p. 91-110, 2002.

HUSSERL, E. **Logical Investigations**. Volume I. London: Routledge & Keagan Paul. 1970.

HUTCHINS, J. Retrospect and prospect in computer-based translation. In: **Machine Translation Summit**, Kent Ridge Labs, Singapore. Tokyo, 1999, p. 30-44. Disponível em: <[http://www.mind-consciousness-language.com/semantics outside language.pdf](http://www.mind-consciousness-language.com/semantics%20outside%20language.pdf)> Acesso em: 05/12/2014.

ICSI. **The FrameNet Project**. 2012. Disponível em: <https://framenet.icsi.berkeley.edu/fndrupal/> Acesso em 22/03/15.

ISO 25964-1:2011. **Thesauri and interoperability with other vocabularies**. Part 1: Thesauri for information retrieval. Geneva: International Standard Organization, 2011.

JOHANSSON, Richard; NUGUES, Pierre. Using WordNet to extend FrameNet coverage. In: **Building Frame Semantics Resources for Scandinavian and Baltic Languages**. Department of Computer Science, Lund University, 2007. p. 27-30.

LAKOFF, G. **Women, Fire, and Dangerous Things: What categories reveal about the mind**. Chicago: University of Chicago Press, 1987.

LAKOFF, G.; JOHNSON, M. **Metaphors We Live By**. Chicago: University of Chicago Press, 1980.

LANGACKER, Ronald. **Foundations of Cognitive Grammar: Theoretical prerequisites**. Stanford: Stanford University Press, 1987.

LANGACKER, Ronald. Subjectification. **Cognitive Linguistics**, 1990, 1, pp. 5-38.

LANGACKER, Ronald. **Foundations of Cognitive Grammar: Descriptive applications**. Stanford: Stanford University Press, 1991.

LIM, Soon Chong Johnson; LIU, Ying; LEE, Wing Bun. A methodology for building a semantically annotated multi-faceted ontology for product family modelling. **Advanced Engineering Informatics**, v. 25, n. 2, p. 147-161, 2011.

LUPO, Caterina et al. **General xml format (s) for legal sources-Estrella European Project IST-2004-027655**. 2007.

LYONS, J. **Semântica**. Lisboa: Editorial Presença, 1987.

MAEDCHE, A. e STAAB, S. Measuring Similarity between Ontologies. In: **Proc. Of the European Conference on Knowledge Acquisition and Management - EKAW-2002**. Madrid, Spain, LNCS/LNAI 2473, Springer, pp. 251-263, 2002.

MATTAR, João. **Introdução à Filosofia**. São Paulo: Pearson Prentice Hall, 2010.

MEALY, G. H. Another Look at Data. **AFIPS: American Federation of Information Processing Societies**, v. 31, Washington, DC: Thompson Books, London: Academic Press, p. 525-534, 1967. Disponível em: <<http://portal.acm.org/citation.cfm?id=1465682>>. Acesso em: 04/01/2015.

MEDIN, Douglas L.; SMITH, Edward E. Concepts and concept formation. **Annual review of psychology**, v. 35, n. 1, p. 113-138, 1984.

MENDONÇA, Fabrício Martins; ALMEIDA, Maurício Barcellos. Princípios metodológicos para desenvolvimento de ontologias: análise das práticas correntes e proposição de melhorias. In: **Encontro Nacional de Pesquisa em Ciência da Informação**, 15., 2014, Belo Horizonte. Anais... Belo Horizonte: UFMG, 2014.

MILLER, George A. et al. Introduction to wordnet: An on-line lexical database. **International journal of lexicography**, v. 3, n. 4, p. 235-244, 1990.

MINGHELLI, T. D. **A relação de meronímia em uma ontologia jurídica**. Dissertação de Metrado. São Leopoldo. UNISINOS. 2011.

MINGHELLI, T.D. **A representação semântica dos eventos e participantes do processo judicial com base na teoria da semântica de frames**. São Leopoldo. UNISINOS. 2015 (no prelo)

MINSKY, M. A. **A Framework for Representing Knowledge**. Artificial Intelligence Memo 306, MIT AI Lab, 1974.

MOHIT, Behrang; NARAYANAN, Srin. Semantic extraction with wide-coverage lexical resources. In: **Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: companion volume of the Proceedings of HLT-NAACL 2003--short papers-Volume 2**. Association for Computational Linguistics, 2003. p. 64-66.

MOMMERS, Laurens. Ontologies in the Legal Domain. In: R. Poli, J. Seibt (eds.), **Theory and Applications of Ontology: Philosophical 265 Perspectives**, Springer Science+Business Media B.V. 2010.

MORA, J. F. **Dicionário de Filosofia**, Lisboa: Dom Quixote. 1978.

MOREIRA, A.; ALVARENGA, L.; OLIVEIRA, A.P. **O nível do conhecimento e os instrumentos de representação: tesouros e ontologias**. Datagramazero, v.5, n.6, dez. 2004.

MOREIRA, Alexandra; SALOMÃO, Maria Margarida Martins. Análise ontológica aplicada ao desenvolvimento de frames. **ALFA: Revista de Linguística**, v. 56, n. 2, 2012.

NATIONAL INFORMATION STANDARDS ORGANIZATION (2005). **ANSI/NISO Z39.19-2003: guidelines for the construction, format, and management of monolingual thesauri**. 2005.

NEISSER, Ulric et al. Intelligence: knowns and unknowns. **American psychologist**, v. 51, n. 2, p. 77, 1996.

NILES, I., PEASE, A., (2001), Toward a Standard Upper Ontology, in **Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001)**.

NORVIG, Peter; RUSSELL, Stuart. **Inteligência Artificial, 3ª Edição**. Elsevier Brasil, 2014.

NOVAK, J.; GOWIN, B. **Aprender a aprender**. Lisboa: Plátano Editora. 1996

NOY, F. N.; McGUINNESS, D.L. **Ontology development 101: a guide to create your first ontology**. 2001.

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

OTLET, Paul. **Documentos e Documentação**. Tradução de Hagar Espanha. Paris, 1937. (Introdução aos trabalhos do Congresso Mundial da Documentação Universal, realizado em Paris, em 1937). Disponível em: <<http://www.conexaorio.com/bit/otlet/index.htm>>. Acesso em: 11 ago. 2012.

PAIVA, Mário. **Informática: o futuro da Justiça**. Revista Jurídica Consulex, ano XI, n. 244, p. 15, 2011.

PETRUCK, Miriam RL; DE MELO, Gerard. Precedes: A Semantic Relation in FrameNet. **Language Resources for Public Security Applications**, p. 45. 2012.

PETRUCK, Miriam RL. Frame semantics. **Handbook of pragmatics**, p. 1-13, 1996.

PIMENTEL SOBRINHO, A. C. **Modelo Conceitual da Ciência da Informação baseado nas visões Americana e Europeia**. Datagramazero (Rio de Janeiro), v. v.15, p. 35, 2014

POLI, R. Descriptive, Formal, and Formalized Ontologies. In D. Fisette, ed., **Husserl's Logical Investigations Reconsidered**, Kluwer, Dordrecht. 2003. Disponível em: <<http://robertopoli.co.cc>> Acesso em: 04/01/2015.

POVEDA-VILLALÓN, María; SUÁREZ-FIGUEROA, Mari Carmen; GÓMEZ-PÉREZ, Asunción. Validating ontologies with OOPS!. In: **Knowledge Engineering and Knowledge Management**. Springer Berlin Heidelberg, 2012. p. 267-281.

PREVOT, L. HUANG, C. CALZOLARI, N. GANGEMI, A. LENCI, A. OLTRAMARI, A. Ontology and the lexicon: a multidisciplinary perspective. In: HUANG, C. CALZOLARI, N. GANGEMI, A. LENCI, A. OLTRAMARI, A. PREVOT, L. **Ontology and the lexicon: a natural language processing perspective**. New York: Cambridge University Press, 2010.

RANGANATHAN, S. R. **Prolegomena to Library classification**. Bombay: Asia Publishing House, 1967. 640p

REALE, Giovanni. **Aristóteles-Metafísica**. São Paulo: Loyola, 2002. 3v. ISBN 8834305418.

REZENDE, Solange Oliveira. **Sistemas inteligentes: fundamentos e aplicações**. Editora Manole Ltda, 2003.

ROSCH, Eleanor; MERVIS, Carolyn B. Family resemblances: Studies in the internal structure of categories. **Cognitive psychology**, v. 7, n. 4, p. 573-605, 1975.

ROSCH, Eleanor. Principles of categorization. **Concepts: core readings**, p. 189-206, 1999.

ROVER, Aires José. Sistemas especialistas legais: pensando o sistema jurídico. **Seqüência: Estudos Jurídicos e Políticos**, v. 15, n. 29, p. 65-79, 2010.

RUSCHEL, A.J; ROVER, A.J.; O Uso de Ontologia no Sistema do Judiciário. In Nicolás Cabezudo Rodríguez (ed.) **Inclusión Digital: perspectivas y experiencias**, LEFIS Series 12, Zaragoza: Prensas Universitaria de Zaragoza, 2011, 414 páginas, (pp. 131-150), ISBN 978-84-15031-76-5

RUPPENHOFER, Josef et al. **FrameNet II: Extended theory and practice**. 2010.

SÁ, A. F. **Da Destruição Fenomenológica à Confrontação: Heidegger e a Incompletude da Ontologia Fundamental**. Lisboa: Lusosofia. 2008. Disponível em: <<http://www.lusosofia.net>> Acesso em: 15/12/2014.

SAGRI, Maria Teresa; TISCORNIA, Daniela; BERTAGNA, Francesca. Jur-WordNet. In: **Proceedings of the 2nd International Global Wordnet Conference**. 2004. p. 305-310.

SAGRI, M.T.; TISCORNIA, D. Semantic lexicons for accessing legal information. In: TRAUNMÜLLER, R. (ED) **Electronic Government**. LNCS 3183. Berlin/Heidelberg: Springer-Verlag, 2004, p. 72-81.

SAINT-DIZIER, P. & VIEGAS, E. **Computational Lexical Semantics**. Cambridge: Cambridge University Press, 1995.

SALES, L.F.; CAMPOS, M. L. A.; GOMES, H. E. Ontologia de domínio: um estudo das relações conceituais. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 13, n. 2, p. 62-76, maio/ago. 2008. Disponível em: <<http://www.eci.ufmg.br/pcionline/index.php/pci/article/viewFile/219>>. Acesso em: 20/10/2011

SALOMÃO, M. M. M. (2009): **FrameNet Brasil: Um trabalho em progresso**. In: *Calidoscópio*, Vol. 7.2.

SARMENTO, Luís; MAIA, Belinda; SANTOS, Diana. The Corpógrafo-a Web-based Environment for Corpora Research. In: **LREC**. 2004.

SAYAO, L. F. **Modelos teóricos em Ciência da Informação: abstração e método científico**. Ciência da Informação, Brasília: UnB, v. 30, n. 1, p. 82-91, jan./abr. 2001

SCHEWEIGHOFER, E.; LACHMAYER, Friedrich. Ideas, Visualisations and Ontologies. In: **International Workshop on Legal Ontologies**. Melbourne, Victoria, Australia. 1997. p. 7-13.

SCHMIDT, Thomas. **The Kicktionary: Combining corpus linguistics and lexical semantics for a multilingual football dictionary**. 2008.

SILVA, Augusto Soares. A Linguística Cognitiva uma Breve Introdução a Um Novo Paradigma em Linguística. **Revista portuguesa de humanidades**, v. 1, n. 1, p. 59-101, 1997.

SILVA, De Plácido e. *Vocabulário jurídico conciso*. Rio de Janeiro: Forense, 2008.

SINHA, Steve. **Answering Questions about Complex Events**. Ph.D. thesis, EECS Department, University of California, Berkeley, Dec. 2008.

SMITH, M. K. et al. OWL Web ontology language guide. In: **W3C recommendation**, 10 feb. 2004. Disponível em: <<http://www.w3.org/TR/2004/REC-owl-guide-20040210>>. Acesso em: 16 set. 2013.

SMITH, B.; WELTY, C. Ontology: towards a new synthesis. In: **International Conference on Formal Ontology in Information Systems**, Ogunquit, Maine, 2001. Disponível em: <<http://www.cs.vassar.edu/~weltyc/papers/fois-intro.pdf>> Acesso em 16/10/2014.

SOERGEL, D. **Functions of a Thesaurus / Classification / Ontological Knowledge Base**. 1997. Disponível em: <<http://ontolog.cim3.net>> Acesso em: 10/07/2013.

SOWA, John. **Roles and Relations**. 2012. Disponível em

<http://www.jfsowa.com/ontology/roles.htm> Acesso em julho de 2015.

STAAB, Steffen; STUDER, Rudi (Ed.). **Handbook on ontologies**. Springer Science & Business Media, 2013.

SUAREZ-FIGUEROA, Mari Carmen; GOMEZ-PEREZ, Asuncion; FERNANDEZ-LOPEZ, Mariano. The NeOn methodology for ontology engineering. In: **Ontology engineering in a networked world**. Springer Berlin Heidelberg, 2012. p. 9-34.

SUMO (2002), **Suggested Upper Merged Ontology**, web site, Disponível em <<http://ontology.teknowledge.com>> Acesso em 12/04/2012.

TALMY, Leonard. **Force dynamics in language and cognition**. Cognitive science, v. 12, n. 1, p. 49-100, 1988.

TISCORNIA, Daniela; SAGRI, Maria Tereza. **Legal Concepts and Multilingual Contexts in Digital Information**. Beijing Law Review, v. 3, n. 03, p. 73, 2012.

USCHOLD, Mike; GRUNINGER, Michael. Ontologies: Principles, methods and applications. **The knowledge engineering review**, v. 11, n. 02, p. 93-136, 1996.

USCHOLD, Michael; KING, Martin. **Towards a methodology for building ontologies**. Edinburgh: Artificial Intelligence Applications Institute, University of Edinburgh, 1995.

VALENTE, Andre; BREUKER, Joost. Towards principled core ontologies. In: **Proceedings of the Tenth Workshop on Knowledge Acquisition for Knowledge-Based Systems**. 1996.

VENTURI, Giulia et al. Towards a FrameNet resource for the legal domain. **LOAIT**, p. 67-76, 2009.

VENTURI, Giulia. Semantic annotation of Italian legal texts: a FrameNet-based approach. **Constructions and Frames**, v. 3, n. 1, p. 46-79, 2011.

VICKERY, B. C. Ontologies. **Journal of Information Science**. v. 23, n. 4, p. 272-286, 1997. Disponível em: <http://www.let.uu.nl/~paola.monachesi/personal/papers/vick97-onto.pdf> Acesso em: 10/07/2012.

VOSSSEN, P. EuroWordNet: Linguistic Ontologies in a Multilingual Database. **Communication and Cognition for Artificial Intelligence** - Special Issue, v. 15, n. 1-2, p. 37-80, 1998. Disponível em: <<http://vossen.info/docs/1998/VossenCCAI.pdf>> Acesso em: 20/05/2012.

WEAVER, W. Translation. In: **Machine translation of languages: fourteen essays**. (ed.) LOKED, W. N.; BOOTH, D. New York: MIT Press. 1955

WEINSTEIN, Peter C. **Ontology-Based Metadata: transforming the MARC Legacy**. Digital Libraries, Pittsburg, p. 254-263, 1998.

ANEXO A – Catálogo de armadilhas comuns – OOPS!

Catalogue of common pitfalls

Here you can find a catalogue of pitfalls that usually appear when developing ontologies. Some of them are very common and have been identified by several works about ontology evaluation (see [References](#)).

We would like to help you to find as many pitfalls as possible in your ontology developments. However, some of them depend on the domain being modelled or the requirements specified for each particular ontology. Up to now, OOPS! can identify semi-automatically those pitfalls in the catalogue with the title in **bold**. We encourage you to keep an eye of those pitfalls that OOPS! is not able to detect yet. It is a good idea to revise the ontology manually looking for them.

- **P01. Creating polysemous elements:** an ontology element whose name has different meanings is included in the ontology to represent more than one conceptual idea. For example, the class `theatre` is used to represent both the artistic discipline and the place in which a play is performed.
- **P02. Creating synonyms as classes:** several classes whose identifiers are synonyms are created and defined as equivalent. As an example we could define `car`, `Motorcar` and `Automobile` as equivalent classes. Another example is to define the classes `Waterfall` and `Cascade` as equivalents. This pitfall is related to the guidelines presented in [2] which explain that synonyms for the same concept do not represent different classes.
- **P03. Creating the relationship `isA` instead of using "rdfs:subClassOf", "rdf:type" or "owl:sameAs":** the `isA` relationship is created in the ontology instead of using OWL primitives for representing the subclass relationship (`rdfs:subClassOf`), the membership to a class (`rdf:type`), or the equality between instances (`owl:sameAs`). An example of this type of pitfall is to define the class `Actor` in the following way `Actor ≡ Person ∩ ∃interprets.Actuation ∩ ∃is.Man`. This pitfall is related to the guidelines for understanding the `isA` relation provided in [2].
- **P04. Creating unconnected ontology elements:** ontology elements (classes, relationships or attributes) are created with no relation to the rest of the ontology. An example of this type of pitfall is to create the relationship `memberOfTeam` and to miss the class representing teams; thus, the relationship created is isolated in the ontology.
- **P05. Defining wrong inverse relationships:** two relationships are defined as inverse relations when they are not necessarily inverse. For example, something is sold or something is bought; in this case, the relationships `isSoldIn` and `isBoughtIn` are not inverse.
- **P06. Including cycles in the hierarchy [1, 2]:** a cycle between two classes in the hierarchy is included in the ontology, although it is not intended to have such classes as equivalent. That is, some class A has a subclass B and at the same time B is a superclass of A. An example of this type of pitfall is represented by the class `Professor` as subclass of `Person`, and the class `Person` as subclass of `Professor`.
- **P07. Merging different concepts in the same class:** a class is created whose identifier is referring to two or more different concepts. An example of this type of pitfall is to create the class `StyleAndPeriod`, or `ProductOrService`.
- **P08. Missing annotations:** ontology terms lack annotations properties. This kind of properties improves the ontology understanding and usability from a user point of view.
- **P09. Missing basic information:** needed information is not included in the ontology. Sometimes this pitfall is related with the requirements in the ORSD that are not covered by the ontology. Other times it is related with knowledge that could be added to the ontology in order to make it more complete. An example of this type of pitfall is to create the relationship `startsIn` to represent that the routes have a starting point in a particular location; and to miss the relationship `endsIn` to show that a route has an end point. Another example is to create the relationship `follows` when modelling order relations; and do not create its inverse relationship `precedes`.
- **P10. Missing disjointness [1, 2, 3]:** the ontology lacks disjoint axioms between classes or between properties that should be defined as disjoint. For example, we can create the classes `Odd` and `Even` (or the classes `Prime` and `Composite`) without being disjoint; such representation is not correct based on the definition of these types of numbers.
- **P11. Missing domain or range in properties:** relationships and/or attributes without domain or range (or none of them) are included in the ontology. There are situations in which the relation is very general and the range should be the most general concept `Thing`. However, in other cases, the relations are more specific and it could be a good practice to specify its domain and/or range. An example of this type of pitfall is to create the relationship `hasWritten` in an ontology about art in which the relationship domain should be `Writer` and the relationship range should be `LiteraryWork`. This pitfall is related to the common error when defining ranges and domains described in [3].
- **P12. Missing equivalent properties:** when an ontology is imported into another, classes that are duplicated in both ontologies are normally defined as equivalent classes. However, the ontology developer misses the definition of equivalent properties in those cases of duplicated relationships and attributes. For example, the classes `CITY` and `City` in two different ontologies are defined as equivalent classes; however, relationships `hasMember` and `has-Member` in two different ontologies are not defined as equivalent relations.

- **P13. Missing inverse relationships:** this pitfall appears when a relationship (except for the symmetric ones) has not an inverse relationship defined within the ontology. For example, the case in which the ontology developer omits the inverse definition between the relations `hasLanguageCode` and `isCodeOf`, or between `hasReferee` and `isRefereeOf`.
- **P14. Misusing "owl:allValuesFrom" [3]:** this pitfall can appear in two different ways. In the first, the anomaly is to use the universal restriction (`allValuesFrom`) as the default qualifier instead of using the existential restriction (`someValuesFrom`). This means that the developer thinks that `allValuesFrom` implies `someValuesFrom`. In the second, the mistake is to include `allValuesFrom` to close off the possibility of further additions for a given property. An example of this type of pitfall is to define the class `Book` in the following way `Book ≡ ∃producedBy.Writer ∩ ∀uses.Paper` and closing the possibility of adding `Ink` as an element used in the writing.
- **P15. Misusing "not some" and "some not" [3]:** to mistake the representation of `some not` for `not some`, or the other way round. An example of this type of pitfall is to define a vegetarian pizza as any pizza which both has some topping which is not meat and also has some topping which is not fish. This example is explained in more detail in [3].
- **P16. Misusing primitive and defined classes [3]:** to fail to make the definition `complete` rather than `partial` (or `necessary and sufficient` rather than just `necessary`). It is critical to understand that, in general, nothing will be inferred to be subsumed under a primitive class by the classifier. This pitfall implies that the developer does not understand the open world assumption. A more detailed explanation and examples can be found in [3].
- **P17. Specializing a hierarchy exceedingly:** the hierarchy in the ontology is specialized in such a way that the final leaves cannot have instances, because they are actually instances and should have been created in this way instead of being created as classes. Authors in [2] provide guidelines for distinguishing between a class and an instance when modelling hierarchies. An example of this type of pitfall is to create the class `RatingOfRestaurants` and the classes `1fork`, `2forks`, and so on, as subclasses instead of as instances. Another example is to create the classes `Madrid`, `Barcelona`, `Sevilla`, and so on as subclasses of `Place`. This pitfall could be also named `Individuals are not Classes`.
- **P18. Specifying the domain or range exceedingly [2, 3]:** not to find a domain or a range that is general enough. An example of this type of pitfall is to restrict the domain of the relationship `isOfficialLanguage` to the class `City`, instead of allowing also the class `Country` to have official language or a more general concept such as `GeopoliticalObject`.
- **P19. Swapping intersection and union:** the ranges and/or domains of the properties (relationships and attributes) are defined by intersecting several classes in cases in which the ranges and/or domains should be the union of such classes. An example of this type of pitfall is to create the relationship `takesPlaceIn` with domain `OlympicGames` and with range the intersection of the classes `City` and `Nation`. Another example can be to create the attribute `Name` for the classes `City` and `Drink` and to define its domain as the intersection of both classes. This pitfall is related to the common error that appears when defining ranges and domains described in [3] and also related to the guidelines for defining these elements provided in [2].
- **P20. Misusing ontology annotations:** The contents of some annotation properties are swapped or misused. An example of this type of pitfall is to include in the Label annotation of the class `Crossroads` the following sentence `"the place of intersection of two or more roads"`; and to include in the Comment annotation the word `"Crossroads"`.
- **P21. Using a miscellaneous class:** to create in a hierarchy a class that contains the instances that do not belong to the sibling classes instead of classifying such instances as instances of the class in the upper level of the hierarchy. This class is normally named `Other` or `Miscellaneous`. An example of this type of pitfall is to create the class `HydrographicalResource`, and the subclasses `Stream`, `Waterfall`, etc., and also the subclass `OtherRiverElement`.
- **P22. Using different naming criteria in the ontology:** Ontology elements are not named using the same convention within the whole ontology. It is considered a good practice that the rules and style of lexical encoding for naming the different ontology elements is homogeneous within the ontology. One possibility for rules is that concept names start with capital letters and property names start with non-capital letters. In the case of style, there are different options such as camel case, hyp

Disponível em <http://oops.linkeddata.es/catalogue.jsp> Acesso em julho 2015.

ANEXO B – Termos de Audiência JECrim – etapa de avaliação

Termo de Audiência JECrim 1

TERMO DE AUDIÊNCIA – JECRIM

Data:	28/03/2013	Hora: 13:50
Juiz Presidente:	XXXXXXXXXX	
Processo nº:	XXXXXXXXXX	
Natureza:	Velocidade Incompatível - Lei 9503/97	
Autor:	Justiça Pública	
Autor do Fato:	XXXXXXXXXX	
Ministério Público:	XXXXXXXXXX	

Aberta a audiência pelo(a) MM. Juiz(a) de Direito foi dito que, presente o autor do fato, afirmou que cumpriu a condição de pagamento, conforme faz prova neste momento, juntando-se por cópia. Afirma que não havia entendido que necessitava se apresentar. Em razão disso, é concedida uma última oportunidade de cumprimento, passando, a contar desta data, a vigor a suspensão por mais dois anos, tendo o réu que se apresentar trimestralmente em juízo. Ao final do período, vista ao Ministério Público. Nada mais.

Decisão

XXXXXXXXXX	
Juiz de Direito	Ministério Público

Termo de Audiência JECrim 2

TERMO DE AUDIÊNCIA – JECRIM

Data:	01/10/2014	Hora: 14:00
Juiz Presidente:	XXXXXXXXXXXXX	
Processo nº:	XXXXXXXXXXXXX	
Natureza:	Crimes contra a Administração em Geral	
Autor:	Justiça Pública	
Autor do Fato:	XXXXXXXXXXXXX - presente Adv: Defensor Público - presente	
Vítima	O Estado	
Ministério Público:	XXXXXXXXXXXXX	

Aberta a audiência pela MM. Juíza de Direito foi dito que o autor do fato compareceu e esteve acompanhado do defensor público que apresentou defesa preliminar, requerendo para que a denúncia não fosse recebida por ausente a materialidade e indícios suficientes da autoria. Pela Juíza de Direito foi dito que recebia a denúncia. Designo audiência audiência de instrução para o dia 09/12/14 às 15h45min. Requisitem-se os policiais. Presentes intimados. Cadastre-se a DPE. Presentes intimados.

XXXXXXXXXXXXX	XXXXXXXXXXXXX
Juíza de Direito	Ministério Público

Termo de Audiência JECrim 3

TERMO DE AUDIÊNCIA – JECRIM

Data:	29/10/2013	Hora: 16:30
Juiz Presidente:	XXXXXXXXXXXXX	
Processo nº:	XXXXXXXXXXXXX	
Natureza:	Crimes contra a Administração em Geral	
Autor:	Justiça Pública	
Autor do Fato:	XXXXXXXXXXXXX Adv: XXXXXXXXXX – Defensor Público	
Ministério Público:	XXXXXXXXXXXXX	

Aberta a audiência pelo(a) MM. Juiz(a) de Direito foi dito que , presente o autor do fato e seu procurador, a parte requerida afirma que os fatos não ocorreram conforme narrado na denúncia, o que será provado durante a instrução, postulando a improcedência. Pelo MM Juiz foi dito que recebia a denúncia. Foram colhidos os depoimentos pelo Sistema de Gravação Audiovisual, os quais não serão degravados, ficando o CD encartado nos autos, à disposição das partes, para cópia ou consulta, e outro arquivado em cartório. A seguir, pelo MM Juiz foi declarada encerrada a instrução, determinando a atualização dos antecedentes do acusado. As partes não têm diligências a postular. Após, sejam os autos feitos com vista ao Ministério Público e Defensoria Pública para alegações finais Presentes intimados. Nada mais.

XXXXXXXXXXXXX	XXXXXXXXXXXXX
Juiz de Direito	Ministério Público