

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

ANDERSON BERTOLDI

**SEMÂNTICA DE *FRAMES* E RECURSOS LEXICAIS JURÍDICOS:
UM ESTUDO CONTRASTIVO**

SÃO LEOPOLDO

2011

Anderson Bertoldi

**SEMÂNTICA DE *FRAMES* E RECURSOS LEXICAIS JURÍDICOS:
Um Estudo Contrastivo**

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

Orientador: Profa. Dra. Rove Luiza de Oliveira Chishman

Co-orientador: Prof. Dr. Hans C. Boas

SÃO LEOPOLDO

2011

Anderson Bertoldi

**SEMÂNTICA DE *FRAMES* E RECURSOS LEXICAIS JURÍDICOS:
Um Estudo Contrastivo**

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

Aprovada em 05 de janeiro de 2011.

BANCA EXAMINADORA

Profa. Dra. Maria Margarida Martins Salomão – UFJF

Profa. Dra. Maria José Bocorny Finatto – UFRGS

Profa. Dra. Maria Eduarda Giering – UNISINOS

Profa. Dra. Rove Luiza de Oliveira Chishman – ORIENTADORA

*Dedico esta tese a meus pais,
por seu amor, apoio e compreensão.*

AGRADECIMENTOS

Gostaria de expressar meu agradecimento àquelas pessoas e entidades que me auxiliaram neste empreendimento. Em primeiro lugar, gostaria de agradecer o auxílio e apoio recebido de minha orientadora, Dra. Rove Chishman. Suas sugestões me ajudaram no desenho final desta tese. Em segundo lugar, gostaria de agradecer a meu co-orientador, Dr. Hans Boas, que me recebeu no ano de 2010 para um estágio doutoral no Departamento de Estudos Germânicos da Universidade do Texas em Austin.

Este trabalho só foi realizado devido ao apoio da CAPES. A pesquisa e as reflexões aqui apresentadas são resultado de uma bolsa de doutorado, modalidade 1, e de uma bolsa de estágio de doutorando no exterior. Sem o apoio da CAPES, este trabalho não existiria.

Agradeço o apoio do Programa de Pós-Graduação em Linguística Aplicada da Universidade do Vale do Rio dos Sinos. Foi muito importante para a minha formação profissional e acadêmica ter estudado neste PPG. Agradeço a todas as professoras que me acompanharam durante todos os anos em que fui aluno deste PPG e me ensinaram não somente conteúdos acadêmicos, mas também lições para a vida.

Por fim, agradeço aos meus pais e aos meus amigos que me acompanharam durante esse período de intenso trabalho. O apoio dos meus pais, em especial durante o ano em que estive nos EUA, foi fundamental para eu chegar ao final do doutorado. Meu mais profundo agradecimento a todos.

RESUMO

A presente tese teve como objetivo avaliar os desafios a serem enfrentados no desenvolvimento de recursos lexicais multilíngües segundo o paradigma FrameNet. Abordou-se aqui a Semântica de *Frames* com uma teoria da lingüística cognitiva e a forma como a FrameNet trata o conceito de *frame* semântico em sua base de dados. Para tratar dos desafios no uso de *frames* semânticos para a descrição de informação jurídica em diferentes línguas, optou-se por investigar o *frame* *Criminal_process*. Em um primeiro momento, identificaram-se, com o auxílio de um dicionário jurídico bilíngüe, os equivalentes em português das unidades lexicais do inglês relacionadas ao processo penal. Em um segundo momento, essas unidades lexicais foram contrastadas para verificar se o conhecimento jurídico evocado pela unidade lexical em inglês era o mesmo conhecimento jurídico evocado pelo seu equivalente em português. A abordagem contrastiva permitiu a criação de *frames* jurídicos descrevendo o processo penal brasileiro. Constatou-se que os *frames* semânticos apresentam diferentes níveis de equivalência. Este trabalho procurou explicar as divergências de *frames* semânticos por meio da diferenciação entre *frames* inatos e *frames* aprendidos. Uma vez que constatada a falta de equivalência perfeita de *frames* entre as línguas, é necessário se repensar o uso dos *frames* semânticos como interlíngua em recursos lexicais multilíngües, especialmente se esses recursos representarem conhecimento jurídico. Este trabalho apontou como uma possível solução para a falta de equivalência entre *frames* o uso de um recurso ainda pouco abordado pela FrameNet: os tipos semânticos. Os tipos semânticos poderiam ser utilizados em uma base de dados jurídica multilíngüe para marcar o papel social exercido por cada participante dos eventos jurídicos.

PALAVRAS-CHAVE: *Frames* Jurídicos, *Frames* Semânticos, Semântica de *Frames*, Lingüística Contrastiva.

ABSTRACT

This Ph.D. dissertation discusses the challenges to be faced in developing multilingual lexical resources according to the FrameNet paradigm. This work was based on Frame Semantics. The starting point was the `Criminal_process` frame. First, the equivalents in Portuguese for the lexical units were identified using a bilingual legal dictionary. Second, the lexical units were contrasted in order to verify the legal knowledge evoked by lexical units in English was the same legal knowledge evoked by their equivalents in Portuguese. Based on the contrastive study, legal frames representing the Brazilian legal system were created. This work could verify that semantic frames present different levels of equivalence. The differences in frame equivalence were analyzed according to the difference between innate and learned frames. Considering that the semantic frames do not present a perfect equivalence among languages, it is necessary to rethink the use of semantic frames as interlingual index to connect multilingual resources, especially legal lexical resources. This dissertation considers the possibility of using semantic types as a possible solution in a future.

KEY-WORDS: Legal Frames, Semantic Frames, Frame Semantics, Contrastive Linguistics.

LISTA DE FIGURAS

Figura 1: Entrada Lexical para <i>arrest.v</i>	27
Figura 2: Similaridade Organizacional.....	31
Figura 3: Inter-relação de Recursos Lexicais Multilíngües.....	32
Figura 4: Anotação de <i>Corpus</i> com a Ferramenta SALTO.....	51
Figura 5: Estrutura do Kicktionary	53
Figura 6: Cena <i>Pass</i> e seus <i>frames</i>	53
Figura 7: Cena e <i>frame</i>	54
Figura 8: Descrição das unidades lexicais do <i>frame</i> Intercept.....	55
Figura 9: Progressão do Processo Penal nos EUA	62
Figura 10: O <i>frame</i> <i>Criminal_process</i>	68
Figura 11: O <i>frame</i> <i>Processo_penal</i>	92
Figura 12: Relatório de <i>Frame</i>	105

LISTA DE TABELAS

Tabela 1: Unidades Lexicais Evocadoras do <i>Frame Arrest</i>	70
Tabela 2: Unidades Lexicais Evocadoras do <i>Frame Arraignment</i>	74
Tabela 3: Unidades Lexicais Evocadoras do <i>Frame Notification_of_Charges</i>	75
Tabela 4: Unidades Lexicais Evocadoras do <i>Frame Entering_a_plea</i>	77
Tabela 5: Unidades Lexicais Evocadoras do <i>Frame Bail_decision</i>	79
Tabela 6: Unidades Lexicais Evocadoras do <i>Frame Try_defendant</i>	80
Tabela 7: Unidades Lexicais Evocadoras do <i>Frame Trial</i>	82
Tabela 8: Unidades Lexicais Evocadoras do <i>Frame Court_examination</i>	84
Tabela 9: Unidades Lexicais Evocadoras do <i>Frame Jury_deliberation</i>	86
Tabela 10: Unidades Lexicais Evocadoras do <i>Frame Verdict</i>	87
Tabela 11: Unidades Lexicais Evocadoras do <i>Frame Sentencing</i>	89
Tabela 12: Níveis de Equivalência de <i>Frames</i>	119

SUMÁRIO

1 INTRODUÇÃO	12
2 A ABORDAGEM DO SIGNIFICADO NA SEMÂNTICA DE <i>FRAMES</i>	16
2.1 A Semântica de <i>Frames</i> e a Linguística Cognitiva	17
2.2 A Semântica de <i>Frames</i>	21
2.3 A FrameNet e o Tratamento do Significado	25
2.4 Os <i>Frames</i> Semânticos e a Multilingualidade	30
3 A FRAMENET E A ANÁLISE CONTRASTIVA	37
3.1 A Faceta Contrastiva	38
3.1.1 A Linguística Contrastiva.....	39
3.1.2 A Equivalência	41
3.2 A Faceta Jurídica	45
3.2.1 A Terminologia Jurídica.....	46
3.2.2 Os Sistemas Jurídicos e a Equivalência de Termos Jurídicos	47
3.3 A Faceta Computacional	49
3.3.1 Criação de FrameNets em Diversas Línguas.....	50
3.3.2 Anotação Semântica e Criação Automática de FrameNets	51
3.3.3 A FrameNet em Domínios Específicos	52
4 ANÁLISE CONTRASTIVA DO <i>FRAME CRIMINAL_PROCESS</i>	58
4.1 Entendendo o Processo Penal	59
4.1.1 Etapas do Processo Penal nos Estados Unidos.....	60
4.1.2 O Processo Penal no Brasil.....	63
4.2 Estudo Contrastivo do <i>Frame Criminal_process</i>	66
4.2.1 O <i>frame</i> Arrest.....	69
4.2.2 O <i>frame</i> Arraignment e seus <i>subframes</i>	72
4.2.3 O <i>frame</i> Try_defendant.....	80

4.2.4 O <i>frame</i> Trial e seus <i>subframes</i>	81
4.2.5 O <i>frame</i> Sentencing.....	88
4.3 O <i>frame</i> Processo_penal.....	91
4.3.1 O <i>frame</i> Prisão.....	93
4.3.2 O <i>frame</i> Denúncia.....	95
4.3.3 O <i>frame</i> Audiência_de_instrução.....	97
4.3.4 O <i>frame</i> Pronúncia.....	98
4.3.5 O <i>frame</i> Julgamento e seus <i>subframes</i>	98
4.3.6 O <i>frame</i> Julgar_acusado.....	103
5 DIVERGÊNCIA DE <i>FRAMES</i> E IMPLICAÇÕES COMPUTACIONAIS.....	108
5.1 <i>Frames</i> Semânticos: Universais ou Específicos?.....	110
5.2 A Multilinguagem em FrameNets Jurídicos.....	118
6 CONSIDERAÇÕES FINAIS.....	123
REFERÊNCIAS.....	127

1 INTRODUÇÃO

Esta tese discute o uso do paradigma FrameNet (Fillmore et al., 2003) para a criação de um recurso lexical computacional para a linguagem jurídica do português brasileiro. A opção por esse tópico foi resultado da dissertação de mestrado do autor (Bertoldi, 2007), pesquisando a representação da semântica dos adjetivos em ontologias jurídicas, e do seu trabalho como bolsista CNPq no projeto FrameCorp¹, trabalhando com o uso das etiquetas semânticas da FrameNet para anotação semântica (Chishman et al., 2008).

Ao utilizar as etiquetas semânticas da FrameNet para anotação de *corpora*, no âmbito do Projeto FrameCorp, o autor passou a investigar a possibilidade de uso da FrameNet para anotação de textos jurídicos. A primeira percepção foi a de que os *frames* semânticos descrevendo eventos jurídicos variavam em nível de equivalência entre o inglês e o português, alguns encontrando correspondência no sistema judicial brasileiro, outros não. Assim, este trabalho se interessa pela aplicação da Semântica de *Frames* para a criação de léxicos computacionais jurídicos.

Essa pesquisa se inspira em outras iniciativas de criação de léxicos computacionais jurídicos, como a JurWordNet (Sagri e Tiscornia, 2004) e a LOIS (Sagri e Tiscornia, 2004), ambos os projetos foram desenvolvidos pelo *Istituto di Teoria e Tecniche della Informazione Giuridica* (ITTIG-CNR)². A JurWordNet é um léxico terminológico do domínio jurídico italiano. Seu desenvolvimento é baseado na WordNet de Princeton (Miller, 1995; Fellbaum, 1999). Esse léxico é, portanto, estruturado em grupos de sinônimos, chamados de *synsets*. Já o projeto LOIS (*Lexical Ontologies for Legal Information Sharing*) tem por objetivo facilitar a

¹ O Projeto FrameCorp, coordenado pela prof^a. Dr^a. Rove Luiza de Oliveira Chishman, tem por objetivo a investigação semântico-computacional do léxico do Português do Brasil a partir da utilização de corpus eletrônico, verificando em que medida a Semântica de *Frames* (Fillmore, 1982; 1985) se presta à descrição semântica de verbos.

² www.ittig.cnr.it

extração de informação em bases de dados jurídicos através da construção de uma WordNet jurídica multilíngüe (Dini et al., 2005).

Léxicos baseados em redes de sinônimos possuem uma limitação quando aplicados a domínios especializados. Em geral, a sinonímia nos domínios especializados é reduzida. Assim, a função dessas redes de sinônimos é mais voltada para a identificação de variantes sócio-dialetais que de termos técnico-científicos. Essa função das redes semânticas baseadas em sinonímia, quando aplicada à recuperação de informação, pode diminuir as barreiras entre o especialista e o leigo na busca por informação jurídica. No entanto, algumas aplicações computacionais, como anotação semântica de *corpora* e sumarização, podem necessitar de um recurso lexical de maior potencial descritivo, contendo informação de papéis semânticos, não apenas de relações semânticas. Este trabalho pretende contribuir para suprir essa lacuna por léxicos computacionais jurídicos com informação de papéis semânticos.

Uma série de iniciativas têm aplicado o paradigma FrameNet a diferentes línguas, com finalidades que vão desde a criação de léxicos até a anotação de *corpora*. Iniciativas como a Spanish FrameNet (Subirats, 2009), a Japanese FrameNet (Ohara, 2009) e a FrameNet Brasil (Salomão, 2009) são exemplos de criação de léxicos baseados em *frames* em diferentes línguas. Já o Kicktionary (Schmidt, 2009) é um dicionário multilíngüe baseado em *frames* para a linguagem do futebol. O Projeto Salsa (Burchardt et al., 2009), por sua vez, utiliza as etiquetas semânticas da FrameNet para anotação de *corpora* em alemão e criação automática de *frames*. No domínio jurídico, Giulia Venturi (*Istituto di Linguistica Computazionale – CNR – Pisa*) desenvolve pesquisa de doutorado aplicando o paradigma FrameNet à criação de um léxico jurídico para o italiano (Venturi, 2009). Essas pesquisas servem de embasamento aplicado para a pesquisa que é apresentada aqui.

Assim, este trabalho tem por objetivo avaliar em que medida os *frames* da FrameNet se prestam à criação de uma FrameNet do domínio jurídico em português. Assim, parte-se das unidades lexicais da FrameNet e, através da busca de equivalentes em português para essas unidades lexicais, avalia-se se o evento jurídico evocado pelas unidades lexicais em inglês e português é correspondente. Alguns trabalhos, como Padó e Lapata (2005), Padó (2007) e Tonelli e Pianta (2008), defendem a possibilidade de criação automática de léxicos multilíngües através do uso de *corpora* paralelos. Com essa técnica, seria apenas necessário anotar um *corpus* em inglês com as etiquetas semânticas da FrameNet e transferir a anotação do *corpus* em inglês para um *corpus* em outra língua. Este trabalho sustenta que tal técnica

não seria eficiente para o domínio jurídico, dada a forte orientação social dessa área de conhecimento.

Para discutir essa questão, esta pesquisa está focada no *frame* *Criminal_process*. Trata-se de um *frame* não-lexical. Segundo Huppenhof et al. (2010, p.80), *tais frames não possuem unidades lexicais e são apresentados puramente para conectar dois (ou mais) frames semanticamente*. O *frame* *Criminal_process*, conforme a definição da FrameNet, indica um evento em que *um suspeito é preso por autoridades sob certas acusações e é processado como réu*.

O *frame* *Criminal_process* se divide em *subframes* que descrevem as diferentes etapas de um processo penal nos Estados Unidos, como *Arrest*, *Arraignment*, *Trial* e *Sentencing*. O *frame* *Arraignment* se divide em três partes: *Notification_of_charges*, *Entering_a_plea* e *Bail_decision*. O *frame* *Trial* tem como *subframes* *Court_examination*, *Jury_deliberation* e *Verdict*. Além desses *frames*, faz parte desta pesquisa o *frame* *Try_defendant*, que está em relação de perspectiva com o *frame* *Trial*. A partir da análise desses onze *frames*, busca-se a resposta para a seguinte pergunta: É possível a criação de uma FrameNet jurídica para o português a partir dos *frames* semânticos propostos para as unidades lexicais do inglês?

A análise contrastiva das sentenças anotadas, ou seja, a comparação dos exemplos em inglês anotados pela FrameNet com as sentenças em português que foram anotadas para esta pesquisa, levou em consideração os seguintes pontos: (i) a compatibilidade do *frame* com o significado da sentença em português, (ii) a compatibilidade dos elementos de *frame* com as sentenças em português e (iii) a necessidade de criação de novos elementos de *frame*. Por fim, os resultados foram avaliados e um *frame* *Processo_penal* foi proposto para o sistema judicial brasileiro. É a partir do estudo contrastivo que se defende a tese de níveis de equivalência de *frames*, como explicação para alguns problemas ligados à criação de léxicos computacionais jurídicos baseados em *frames*.

A estrutura desta tese organiza os conteúdos importantes para a discussão do tema proposto. O capítulo 2, organizado em três seções, apresenta a teoria lingüística central a esta pesquisa, a Semântica de *Frames*. Assim, a seção 2.1 apresenta a lingüística cognitiva, o modelo lingüístico que serve de base à teoria lingüística conhecida como Semântica de *Frames*. A seção 2.2 apresenta a Semântica de *Frames*. A seção 2.3 traz a FrameNet, léxico computacional que tem como base teórica a Semântica de *Frames* e que é o ponto de partida

da análise contrastiva apresentada nesta pesquisa. A seção 2.4 apresenta as questões multilíngües relacionadas à criação de FrameNets multilíngües.

Por se tratar de uma análise contrastiva em uma área especializada, faz-se necessário abordar questões como tradução e terminologia. O capítulo 3, portanto, foi organizado em três seções. A seção 3.1 apresenta a lingüística contrastiva e as questões de pesquisa atualmente exploradas por essa área da lingüística. A seção 3.2 aborda questões de tradução jurídica, da diferença entre sistemas jurídicos e da equivalência de termos jurídicos. Já a seção 3.3 apresenta os trabalhos que tem aplicado a Semântica de *Frames* em estudos contrastivos multilíngües e que servem de embasamento a esta pesquisa.

O capítulo 4 discute, através da análise contrastiva, a possibilidade de uso da FrameNet para a criação de um léxico jurídico computacional para o português. Esse capítulo se divide em três seções. Na seção 4.1, é apresentado um estudo da progressão do processo penal nos Estados Unidos e no Brasil. Na seção 4.2, é apresentada a análise contrastiva do *frame* `Criminal_process` em inglês e português. Na seção 4.3, é proposto o *frame* `Processo_penal` de acordo com a progressão do processo penal no Brasil.

O capítulo 5 desenvolve, de forma mais aprofundada, a constatação feita no capítulo anterior de que existem níveis de equivalência de *frames* entre as línguas. Assim, a seção 5.1 discute a universalidade dos *frames* semânticos em aplicações computacionais multilíngües e a seção 5.2 trata da multilingualidade em FrameNets jurídicas. O capítulo 6 aponta as direções futuras deste trabalho.

2 A ABORDAGEM DO SIGNIFICADO NA SEMÂNTICA DE *FRAMES*

O objetivo deste capítulo é apresentar a Semântica de *Frames* (Fillmore, 1982; 1985), desde seus princípios teóricos, que permitem considerá-la como uma das teorias lingüísticas da lingüística cognitiva, até a sua aplicação como base teórica do léxico computacional FrameNet (Fillmore et al., 2003). Embora a Semântica de *Frames* tenha sido proposta já na década de 70, é apenas com o desenvolvimento da FrameNet, no final da década de 90, que essa teoria encontrou amplo reconhecimento. A explicação para esse interesse na Semântica de *Frames* se deve a seu uso como base teórica para estudos contrastivos (Fontenelle, 2000; Boas, 2002, 2005) e para a criação de léxicos computacionais baseados em *frames* (Subirats, 2009; Ohara, 2009; Salomão, 2009).

Os diferentes manuais de lingüística cognitiva (Croft e Cruse, 2004; Evans e Green, 2006; Ungerer e Schmid, 2006; Geeraerts e Cuyckens, 2007) apresentam a Semântica de *Frames* como uma teoria integrante do empreendimento cognitivista. No entanto, essa teoria se desenvolve à parte das demais teorias que dão início à lingüística cognitiva, como a teoria das metáforas conceituais (Lakoff e Johnson, 1980) e a gramática cognitiva (Langacker, 1987; 1991). Ao defender os princípios de não distinção entre conhecimento lingüístico e conhecimento enciclopédico e de contextualização cultural da língua, a Semântica de *Frames* se torna parte integrante do empreendimento cognitivista.

A Semântica de *Frames* se desenvolve como abordagem alternativa às teorias do significado baseadas em listas de condições, muito comum nas abordagens estruturalistas. A Semântica de *Frames* tem como estrutura fundamental o *frame*. O conceito de *frame* é proveniente da computação (Minsky, 1974) e das ciências sociais (Goffman, 1974). Fillmore propõe a Semântica de *Frames* na década de 70 (Fillmore, 1975, 1976, 1977). No entanto, a sua teoria vai se tornar conhecida pelos dois trabalhos publicados na década de 80 (Fillmore, 1982; 1985).

O léxico computacional FrameNet nasce como um recurso lexical organizado com base na Semântica de *Frames*. Assim, a FrameNet tem sido utilizada como recurso lexical para aplicações em processamento de linguagem natural, como paradigma para a construção de léxicos baseados em *frames* para diferentes línguas e como repositório do léxico da língua inglesa, utilizado especialmente em análises contrastivas. Este trabalho parte da premissa de que a FrameNet é um paradigma que pode ser aplicado à construção de léxicos computacionais em diferentes línguas, além de ser repositório das unidades lexicais da língua inglesa. É por esse motivo que esta pesquisa utiliza os exemplos anotados da FrameNet na análise contrastiva com o português, buscando o desenvolvimento de recurso lexical computacional semelhante para o domínio jurídico.

Assim, para tratar desse tema, este capítulo foi estruturado em três seções. A seção 2.1 discute os princípios norteadores da lingüística cognitiva comuns às diferentes teorias lingüísticas que compõem esse amplo modelo lingüístico. A seção 2.2 apresenta a Semântica de *Frames* e a sua abordagem do significado. A seção 2.3 apresenta a FrameNet, e os termos chave para compreensão desse recurso lexical, como unidade lexical, evocador de *frame*, *frame* e elementos de *frame*. Já a seção 2.4 discute a criação de FrameNets multilíngües.

2.1 A Semântica de *Frames* e a Lingüística Cognitiva

A Semântica de *Frames* é uma teoria lingüística pertencente ao modelo lingüístico conhecido como lingüística cognitiva. A lingüística cognitiva tem início na década de 80 com o descontentamento de pesquisadores como Lakoff, Langacker e Talmy com a proposta gerativista de Noam Chomsky. As obras *Women, Fire and Dangerous Things* (Lakoff, 1987) e *Foundations of Cognitive Grammar* (Langacker, 1987) podem ser consideradas como marcos fundamentais de uma nova concepção de linguagem, abrindo caminho para o empreendimento que seria conhecido como lingüística cognitiva. Embora a gramática gerativa seja considerada uma abordagem cognitivista, não se pode confundir a gramática gerativista com a lingüística cognitiva, visto que o movimento hoje denominado de lingüística cognitiva surgiu em oposição ao empreendimento gerativista.

Uma das características mais marcantes da lingüística cognitiva é a multiplicidade de teorias. Diferentemente do empreendimento gerativista que possui o lingüista Noam Chomsky como seu patrono, o empreendimento cognitivista não foi motivado pelo trabalho de um único pesquisador. As teorias lingüísticas consideradas hoje como parte integrante do

empreendimento cognitivista se desenvolveram de forma autônoma. Tendo tido um desenvolvimento independente e tendo como objeto de interesse diferentes fenômenos lingüísticos, o que essas teorias têm de semelhante é a oposição às idéias gerativistas de modularidade da mente, separação entre conhecimento lingüístico e conhecimento enciclopédico, não-motivação semântica da sintaxe e geração da linguagem por regras lógicas.

Talmy (2000), na introdução de sua obra *Toward a Cognitive Semantics*, define a lingüística cognitiva como uma abordagem conceitual da linguagem. Segundo Talmy (2000), há três sistemas de abordagem para a análise da linguagem: o formal, o psicológico e o conceitual. Seguindo essa distinção, a abordagem formal se caracteriza pelo estudo dos padrões estruturais das formas lingüísticas, considerando a língua como um sistema autônomo. Como temas centrais da abordagem formal estão o estudo das estruturas lexical, sintática e morfológica.

A abordagem psicológica analisa a linguagem através do sistema cognitivo, ou seja, a linguagem é analisada sob a perspectiva da percepção, da memória e do raciocínio. As investigações nesse tipo de abordagem incluem a análise da memória semântica, o estudo da categorização, a inferência e o conhecimento contextual. Segundo Talmy (2000), a abordagem conceitual é influenciada pelo estudo da categorização realizado pelas abordagens psicológicas.

A abordagem conceitual, segundo Talmy (2000), se dedica ao estudo de como a linguagem estrutura o conteúdo conceitual. São temas de interesse dessa abordagem o mapeamento conceitual da metáfora, a estrutura conceitual dos *frames* semânticos, a relação entre o texto e o contexto, o agrupamento de estruturas conceituais em sistemas de estruturas mais amplos. A abordagem conceitual estuda a estrutura da língua a partir da perspectiva conceitual. Assim, a língua não é vista como um sistema autônomo e as estruturas gramaticais são vistas como representação da estrutura conceitual. De acordo com Talmy (2000), a abordagem conceitual objetiva a integração das perspectivas que a lingüística e a psicologia possuem sobre a organização da cognição em um entendimento unificado da estrutura conceitual humana.

Considerando as observações de Talmy (2000) sobre a centralidade da categorização para a lingüística cognitiva, é importante dizer que é a categorização já vem sendo um tema de interesse desde os gregos. Aristóteles, na obra *Metafísica* (Aristotle, 1933), já definia os traços “essenciais” e “acidentais” como forma de categorização. O estruturalismo vai tratar da

categorização através da atribuição de traços semânticos, tais como [+HUMANO], [-HUMANO]. Essa abordagem também ficou conhecida como análise componencial. O gerativismo, para tratar da descrição e homogeneidade das categorias linguísticas, também adota os traços. Assim, uma palavra com os traços [+N] e [-V] vai ser um substantivo e uma palavra com os traços [+V] e [-N] vai ser um verbo. No entanto, não é categorização nesse sentido que interessa à linguística cognitiva, mas sim a idéia de categorização por protótipos.

O significado das palavras não é algo fixo capaz de ser definido com traços semânticos. Em seu estudo *The Bondaries of Words and their Meanings*, Labov (1973), demonstra que não há uma fronteira fixa entre o que é uma xícara, uma caneca e uma tigela, sendo que uma série de características, como material, forma e função, vão atuar na diferenciação desses três objetos. A categorização por protótipos, hipótese explorada pela psicóloga Eleanor Rosch (Rosch, 1973, 1975, 1978; Rosch e Mervis, 1975; Rosch et al., 1976), vai ser fundamental para explicar os limites difusos entre as categorias. De acordo com essa teoria, a categorização se daria a partir de um exemplar prototípico que conteria a maioria das características comuns aos demais integrantes da mesma categoria. Isso explicaria por que, quando o ser humano pensa em pássaro, ele provavelmente pensará primeiramente em um pardal, um rouxinol ou um canário, e não em um avestruz, uma galinha ou um peru.

Segundo a visão da linguística cognitiva, a língua possui papel fundamental no processo de categorização. Assim sendo, o significado pode ser considerado o fenômeno linguístico primário. É por esse motivo que a semântica possui primazia na análise linguística em uma abordagem cognitivista. A análise linguística, por sua vez, diferencia-se da abordagem gerativa não apenas pela primazia da semântica, mas também pela análise empírica. Langacker (1988) denomina a metodologia empregada pela linguística cognitiva como *modelo baseado no uso (usage-based model)*.

A análise linguística baseada em *corpora* é significativa para a linguística cognitiva devido à importância dada ao fator cultural envolvido na linguagem. A linguística cognitiva parte do princípio de que a linguagem é fundamentada na cognição humana e a cognição, por sua vez, é culturalmente orientada. A teoria da metáfora conceitual (Lakoff e Johnson, 1980) explora a relação entre cognição, cultura e linguagem. A teoria de modelos culturais e teorias populares (Holland e Quinn, 1987) demonstra que o conhecimento cultural influencia tanto o léxico quanto a gramática da língua. Essa influência da cultura sobre a língua pode ser constatada em Hiraga et al. (1999) e Dirven et al. (2003). É devido a essa influência da cultura

na língua que Langacker (1990) afirma que não há divisão rígida entre o que é conhecimento lingüístico e conhecimento de mundo.

As diferentes teorias da lingüística cognitiva abordam tanto o significado lexical quanto a gramática da língua e o processamento cognitivo da linguagem. A teoria da metáfora e da metonímia conceitual (Lakoff e Johnson, 1980) defende que a metáfora e a metonímia não são recursos estilísticos, mas um mecanismo cognitivo para criação de significado. A teoria dos esquemas imagéticos (Johnson, 1987) defende que a experiência se manifesta no nível cognitivo em termos de esquemas imagéticos. A teoria dos espaços mentais (Fauconnier, 1985; 1997) é uma teoria de construção do significado, que defende que o significado é construído através de espaços mentais que se conectam durante o discurso, permitindo a compreensão entre os falantes. Outras teorias não menos importantes incluem a teoria dos modelos cognitivos idealizados (Lakoff, 1987), a Semântica de *Frames* (Fillmore, 1982; 1985), a teoria da mesclagem conceitual (Fauconnier e Turner, 1987), a gramática cognitiva (Langacker, 1987, 1991) e a gramática das construções (Fillmore et al., 1988; Goldberg, 1995, 2006).

A lingüística cognitiva representa uma nova forma de abordar a língua, integrando fatores culturais com a estrutura e a função das formas lingüísticas. Evans et al. (2007) apontam quatro conquistas da lingüística cognitiva. Em primeiro lugar, a lingüística cognitiva permitiu uma abordagem integrada da organização conceitual e lingüística. Em segundo lugar, essa abordagem conseguiu uma revalorização dos dados empíricos. Em terceiro lugar, as teorias da lingüística cognitiva têm mostrado que a linguagem revela processos cognitivos sistemáticos e que são centrais à organização do pensamento. Por fim, a lingüística cognitiva tem permitido a integração de questões formalistas e funcionalistas. Os lingüistas cognitivos se interessam tanto pela representação do conhecimento da linguagem na mente como pelas funções sociais e comunicativas no uso da linguagem.

Delineadas as características fundamentais da lingüística cognitiva, é o momento de tratar da teoria que serve de fundamentação teórica a este trabalho. Assim, a seção seguinte aborda a Semântica de *Frames*.

2.2 A Semântica de *Frames*

A semântica pode ser considerada uma área relativamente nova da lingüística. As abordagens para o significado até a década de 70 não avançaram muito na descrição do significado da língua. O estruturalismo decidiu tratar do significado através de lista de condição, a chamada análise componencial, através da qual se elencava uma lista de condições que deveriam ser satisfeitas para o reconhecimento do significado da palavra. Segundo as abordagens mais formalistas, o significado era um conhecimento intuitivo, o que constituía um impedimento para a sua formalização. Assim, o gerativismo optou em um primeiro momento por não tratar do significado.

Motivados pelo excesso de formalismo no tratamento da língua, diferentes propostas para a abordagem do significado começam a se desenvolver a partir da década de 60. Katz e Fodor (1963) são os primeiros a pensarem em como deveria ser uma teoria semântica compatível com o gerativismo. O movimento da Semântica Gerativa, iniciado por alunos de Chomsky, representa um movimento de oposição ao tratamento excessivamente sintático da língua. Na psicologia cognitiva estudos sobre categorização começam a discutir os limites difusos dos conceitos. É em meio a essa realidade cultural da década de 70 que Charles J. Fillmore começa a pensar em uma nova maneira de tratar o significado. Esse é o início da Semântica de *Frames*.

A Semântica de *Frames* nasce a partir da combinação de duas idéias muito discutidas na década de 70, **protótipo** (Rosch, 1973) e **frame** (Minsky, 1974; Goffman, 1974). Em *An Alternative to Checklist Theories of Meaning* (1975), Fillmore lança as bases da Semântica de *Frames* quando propõe a relevância dessas duas idéias para a teoria semântica. Segundo Fillmore (1975), o protótipo é uma forma de entender o significado de uma categoria através do reconhecimento de um exemplar ou um protótipo da categoria e os *frames* são uma estruturas de conceitos que se conectam entre si, dando coerência a alguma área do conhecimento humano.

Fillmore (1975) também apresenta os conceitos de **cena** e **frame**. A **cena** seria não apenas uma cena visual, mas todo um conjunto de tipos familiares de transações interpessoais, cenários padrões definidos culturalmente, estruturas institucionais, experiências inativas, imagem corporal, crenças humanas, ações experiências e imagens. O **frame** seria um sistema de escolhas lingüísticas, sejam palavras, regras ou categorias gramaticais, associadas à determinada instância prototípica de uma cena. Através dessa distinção, Fillmore (1975)

distingue entre uma estrutura “cognitiva” e uma estrutura “lingüística” que se ativam mutuamente na construção do significado. O conhecimento da cena permitira ao falante escolher as palavras apropriadas para descrever a cena e o conhecimento das palavras permitiria acenar a estrutura cognitiva e entender o significado das palavras. A partir dessa distinção, Fillmore (1975) lança o termo **cenas-e-frames** para se referir a essa interação entre a estrutura cognitiva e a estrutura lingüística.

Em *Frame Semantics and the Nature of Language* (1976) Fillmore adiciona a idéia de **contexto** às noções anteriormente propostas de protótipo e *frame*. Segundo Fillmore (1976), tanto o significado da palavra como o significado do enunciado dependem do contexto extralingüístico, por isso o contexto deve ser considerado pela teoria semântica. A partir dessa consideração, Fillmore divide os *frames* em cognitivos e interacionais. Os **frames interacionais** representam as escolhas lingüísticas adequadas para uma determinada situação. Por exemplo, um homem pode optar pelo verbo *mijar* em uma conversar entre amigos do mesmo sexo e, no entanto, utilizar o verbo *urinar* em uma consulta médica. Outros exemplos de *frames* interacionais incluem saber qual cumprimento utilizar em uma situação formal e em uma situação informal. Já os **frames cognitivos** não demonstram escolhas que o falante faz motivado pelo nível de formalidade de uma situação, mas o conhecimento da cena que envolve determinado evento, os seus participantes e o papel desempenhado por cada participante no evento.

Fillmore alia às noções de *frame* interacional e *frame* cognitivo a idéia de **estruturamento lexical** (*lexical framing*). Para Fillmore (1976), os usuários de uma língua interpretam seu ambiente, formulam suas próprias mensagens, entendem as mensagens dos outros usuários da língua e criam um modelo mental de seu mundo através de *frames* cognitivos e interacionais. Nesse contexto, estruturamento lexical é o processo de selecionar, dentre as palavras que compõem um *frame*, aquela que melhor expresse o significado que se deseja comunicar. Fillmore (1976) cita como exemplos *on land* e *on the ground*. Apesar de se referirem à mesma entidade, ou seja, *terra* (*chão*), quando um falante de inglês utiliza a expressão *on land*, pressupõe-se que ele estava no mar antes de estar em terra e quando usa *on the ground* pressupõe-se que o falante estava no ar. É por esse motivo que Fillmore afirma que entender uma palavra requer o conhecimento de todo o *frame* no qual essa palavra está inserida.

Em *Scenes-and-frames semantics* (1977), Fillmore apresenta um primeiro esboço do que virá a ser a Semântica de *Frames*, associando a gramática de casos (Fillmore, 1968) à

relação semântica de perspectiva. Em seu artigo *The case for case* (1968), Fillmore estuda as funções sintáticas que seriam universais. Para isso, ele adota a posição de Tesnière (1959), que afirma que a divisão sujeito/predicado não é apropriada para descrever a língua. É então que surge a estrutura de casos (*case frame*) com os seis casos, ou papéis semânticos, que dariam origem a todo o estudo de papéis temáticos: *agentivo*, *instrumental*, *dativo*, *factivo*, *locativo* e *objetivo*.

Fillmore (1977) demonstra através do evento de transação comercial que os verbos *comprar*, *vender* e *custar* representam diferentes perspectivas do mesmo evento. O vendedor cede a mercadoria em troca de dinheiro e o comprador cede o dinheiro em troca da mercadoria. Um evento como transação comercial marca a troca de posse de dois bens: o dinheiro passa da posse do comprador para o vendedor e a mercadoria passa da posse do vendedor para o comprador. A análise da relação de perspectiva no evento de transação comercial já apresenta um primeiro esboço do que será chamado de elementos de *frame* posteriormente. Esses elementos de *frame* vêm substituir a proposta de casos (Fillmore, 1968).

Os artigos *Frame Semantics* (1982) e *Frames and the Semantics of Understanding* (1985) apresentam a Semântica de *Frames* em sua versão mais apurada, desde as primeiras propostas de Fillmore (1975, 1976, 1977). Pela primeira vez, o termo **Semântica de Frames** é definido:

Com o termo ‘semântica de *frames*’ eu tenho em mente um programa de pesquisa em lingüística empírica e uma estrutura descritiva para apresentar os resultados de tal pesquisa. 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, para adicionar novos sentidos às palavras e para reunir os significados dos elementos em um texto no significado total do texto. Pelo termo ‘*frame*’ eu tenho em mente qualquer sistema de conceitos relacionados de tal forma que para entender qualquer um deles você tem que entender toda a estrutura na qual ele se encaixa; quando um dos conceitos em tal estrutura é introduzido em um texto ou uma conversa, todos os outros são automaticamente disponibilizados (Fillmore, 1982, p.111).

A Semântica de *Frames* adota explicitamente uma metodologia de pesquisa **empírica**. Fillmore (1985) afirma que uma abordagem semântica como a Semântica de *Frames* é ‘empírica’, ao invés de ‘formal’, à medida que procura encontrar as formas

detalhadas em que expressões específicas se encaixam em suas situações (Fillmore, 1985, p.231). Observa-se já o posicionamento da Semântica de *Frames* em direção à lingüística cognitiva, distanciando-se do empreendimento gerativista.

Fillmore (1982, 1985) abandona a separação entre cenas e *frames*. A distinção entre cena como estrutura cognitiva e frame como estrutura lingüística é abandonada. A separação apresentada em Fillmore (1977) entre cenas cognitivas, cenas interacionais e *frames* lingüísticos perde a razão de existir a partir da concepção de *frame* como uma estrutura de conceitos relacionados que se ativam mutuamente. Assim, a noção de *frame* engloba a dimensão cognitiva e lingüística ao mesmo tempo, uma vez que as palavras são consideradas como representantes de categorias cognitivas, quebrando com a tradicional distinção entre conhecimento lingüístico e conhecimento enciclopédico.

Na visão que eu estou apresentando, as palavras representam categorizações de experiências, e cada uma dessas categorias é sustentada por uma situação motivadora baseada em conhecimento e experiência anteriores. Com relação ao significado lexical, a pesquisa em semântica de *frames* pode ser vista como um esforço para entender que razão uma comunidade de falantes pode ter encontrado para criar a categoria representada pela palavra e explicar o significado da palavra apresentando e clarificando essa razão (Fillmore, 1982, p.112).

Novamente se observa a orientação da Semântica de *Frames* em direção à lingüística cognitiva ao tratar o fenômeno do significado como categorização, não como uma lista de condições, e ao levar em consideração o fator cultural envolvido na linguagem. Para a Semântica de *Frames*, as palavras têm a capacidade de **evocar** todo um conhecimento de mundo que é organizado através de uma estrutura cognitiva chamada de *frame*:

Frames interpretativos podem ser introduzidos no processo de entendimento de um texto através de duas formas; sendo invocados pelo interprete ou sendo evocados pelo texto. Um *frame* é invocado quando o interprete, tentando depreender o sentido de um segmento textual, é capaz de dar-lhe uma interpretação situando o conteúdo textual em um padrão que é conhecido independentemente do texto. Um *frame* é evocado pelo texto se alguma forma ou padrão lingüístico é convencionalmente associado com o *frame* em questão (Fillmore, 1985, p.232).

O artigo *Toward a Frame-Based Lexicon: The Semantics of Risk and its Neighbors* (Fillmore e Atkins, 1992) apresenta o primeiro exercício de análise semântica baseada em *frames* e aponta a futura criação de um dicionário *on line* baseado em *frames*. A partir do estudo de unidades lexicais que expressam risco, como *risk*, *danger* e *hazard*, Fillmore e Atkins propõem onze categorias para descrever os participantes do *frame Risk*. Essas categorias são: *chance*, *harm*, *victim*, *valued object*, *risky situation*, *deed*, *actor*, *intended gain*, *purpose*, *beneficiary* e *motivation*. No entanto, esses papéis semânticos ainda não são chamados de elementos de *frame*, nome que será dado, posteriormente, aos papéis semânticos desenvolvidos no contexto do Projeto FrameNet. O estudo do *frame Risk* (Fillmore e Atkins, 1992) dará origem a *frames* mais específicos na FrameNet: *Daring*, *Run_risk*, *Risky_situation* e *Being_at_risk*.

Fillmore (1985, p.232) considera (...) os *frames* como ferramentas para a descrição e explicação do significado lexical e gramatical. No entanto, como se verá na seção seguinte, essa teoria tem sido mais utilizada para a descrição do significado lexical. A Semântica de *Frames* não tem sido aplicada à descrição do significado gramatical. Na verdade, outras teorias da lingüística cognitiva, como a gramática das construções (Fillmore et al., 1988; Goldberg, 1995, 2006), têm sido aliadas à Semântica de *Frames* com o objetivo de explicar questões gramaticais (Fillmore, 2008). A seção seguinte aborda a forma como a FrameNet aplica a Semântica de *Frames* na descrição do significado das palavras.

2.3 A FrameNet e o Tratamento do Significado

A FrameNet é um projeto lexicográfico desenvolvido pelo *International Computer Science Institute*, Berkeley, desde 1998. O nome FrameNet é inspirado na WordNet (Miller, 1995; Fellbaum, 1998), um recurso lexical que organiza as palavras segundo relações semânticas. A FrameNet é um recurso lexical que descreve o significado segundo os princípios da Semântica de *Frames*. Conforme a Semântica de *Frames*, as formas lingüísticas evocam a informação contextual “armazenada” na estrutura cognitiva conhecida como *frame*.

Os itens lexicais na FrameNet são tratados como unidades lexicais. A unidade lexical é a soma de uma forma (palavra) com um significado. Cada novo significado de uma palavra representa uma nova unidade lexical. Assim, são as unidades lexicais que evocam os *frames*.

Embora a FrameNet não trate da polissemia, segundo a Semântica de *Frames*, a polissemia é pode ser concebida como uma palavra que apresenta diferentes unidades lexicais.

O método de análise lexical adotado pela FrameNet, conforme Fillmore e Baker (2010), segue cinco etapas:

1. **Caracterização do *frame***. Caracteriza-se a situação descrita pelas unidades lexicais, por exemplo, a prisão de um suspeito, como no caso do *frame* Arrest.

A acusa B de ter cometido um crime e A prende B.

2. **Descrição e nomeação dos elementos de *frame***. Após a caracterização de um *frame* específico, identificam-se todos os possíveis participantes da situação e criam-se nomes para cada participante.

No caso do *frame* Arrest, A é chamado de AUTORIDADES, B é chamado de SUSPEITO, o crime é chamado de OFENSA e a acusação contra o suspeito é chamada de ACUSAÇÕES.

3. **Seleção das unidades lexicais**. Após a descrição da situação e da identificação e nomeação dos elementos de *frame*, as unidades lexicais e expressões evocadoras do *frame* são identificadas. No caso do *frame* Arrest, as unidades lexicais elencadas pela FrameNet são:

apprehend.v, apprehension.n, arrest.n, arrest.v, book.v, bust.n, bust.v, collar.v, cop.v, nab.v, summons.v

4. **Anotação de sentenças**. Sentenças selecionadas para exemplificar os padrões sintáticos e semânticos de cada unidade lexical são anotadas com elementos de *frame*. As sentenças abaixo demonstram dois exemplos anotados pela FrameNet:

Are [you AUTORIDADES] **arresting** [me SUSPEITO] [for the murder of Topaz Brown? OFENSA]

When he gave them his name [they AUTORIDADES] **arrested** [him SUSPEITO] [on a charge of rape. ACUSAÇÕES]

Lexical Entry

arrest.v

Frame: Arrest

Definition:

COD: seize (someone) by legal authority and take them into custody.

Frame Elements and Their Syntactic Realizations

The Frame Elements for this word sense are (with realizations):

Frame Element	Number Annotated	Realization(s)
Authorities	(58)	CNI.-- (20) DNI.-- (1) NP.Ext (29) PP[by].Dep (8)
Charges	(39)	DNI.-- (30) INI.-- (2) PP[for].Dep (2) PP[on].Dep (4) Sforto.Dep (1)
Offense	(19)	2nd.-- (3) PP[for].Dep (2) PP[over].Dep (1) PPing[for].Dep (7) PP[in].Dep (2) PP[on].Dep (1)
Place	(15)	PP[at].Dep (2) PP[in].Dep (9) PP[on].Dep (1) VPing.Dep (1)
Source of legal authority	(3)	PP[under].Dep (3)
Suspect	(58)	NP.Ext (28) INI.-- (1) NP.Obj (29)
Time	(23)	PP[on].Dep (3) PP[after].Dep (3) PPing[while].Dep (1) PP[at].Dep (2) NP.Dep (6) Sub.Dep (4) PP[in].Dep (4) AVP.Dep (2) PPing[after].Dep (1)

Figura 1: Entrada Lexical para *arrest.v*

5. **Geração automática de entradas lexicais.** Os exemplos anotados para cada unidade lexical são transformados automaticamente em uma entrada lexical contendo a definição da unidade lexical, as realizações sintáticas de cada elemento de *frame* e os padrões valências. A figura 1 mostra a entrada lexical da unidade lexical *arrest.v*.

Na FrameNet, a informação sobre valência é especificada em dois níveis: o sintático e o semântico. A valência sintática especifica os tipos sintagmáticos (sintagma nominal, preposicional etc.) e as funções gramaticais (sujeito, objeto etc.). A valência semântica é descrita em termos de entidades que podem participar de um *frame* evocado por uma palavra, ou seja, os elementos de *frame* (Fillmore et al., 2003).

Conforme Fillmore e Baker (2010), os elementos de *frame* representam propriedades ou entidades que podem ou devem estar presentes em qualquer instância de um *frame*. A FrameNet diferencia os elementos de *frame* em **centrais**, **periféricos** e **extratemáticos**. Segundo Fillmore e Baker (2010), a distinção entre esses tipos nem sempre é clara.

De uma forma geral, elementos de *frame* que são obrigatoriamente expressos são centrais. No caso de verbos, elementos de *frame* que expressam funções sintáticas centrais como sujeito e objeto também devem ser centrais. Em alguns casos, determinados elementos de *frame* que são centrais não necessitam ser expressos. O *frame* *Arrest*, por exemplo, possui como elementos de *frame* centrais *AUTORIDADES*, *SUSPEITO*, *ACUSAÇÕES* e *OFENSA*. No entanto, quando o elemento *ACUSAÇÃO* é expresso, o elemento *OFENSA* é suprimido.

Os elementos de *frame* periféricos expressam em geral funções de adjuntos, expressando tempo, lugar ou modo. Se algumas unidades lexicais expressam lugar como um elemento periférico, outras vão ter o elemento de *frame* indicando lugar como elemento de *frame* central. A diferença entre elementos centrais e periféricos depende essencialmente da necessidade de complementação da unidade lexical. Segundo Fillmore (2007), a diferença entre central e periférico é análoga à distinção entre **actantes** e **circunstantes** apresentada por Tesnière (1959). Os elementos de *frame* extratemáticos introduzem informação referente a outro *frame*, tal como o propósito motivador de algum evento ou ação. Os elementos de *frame* periféricos e extratemáticos são agrupados na FrameNet sob a denominação de elementos não-centrais.

Às vezes, os elementos de *frame* de uma unidade lexical podem ser simplesmente omitidos. Segundo Fillmore e Baker (2010), há uma explicação gramatical e duas lexicais

para os elementos de *frame* não realizados. A explicação gramatical está relacionada com estruturas gramaticais que permitem a omissão de algum argumento. Esse tipo de omissão de argumento é chamado de **instanciação nula construcional** (INC). Exemplos de instanciação nula construcional são a omissão do sujeito em orações imperativas e a omissão do agente em oração em voz passiva.

Os outros dois casos são chamados de **instanciação nula indefinida** (INI) e **instanciação nula definida** (IND). Os casos de instanciação nula indefinida envolvem verbos que podem assumir tanto uma forma transitiva como intransitiva, como o verbo *to eat*, em inglês. Nesses casos, a FrameNet trata o verbo como transitivo e considera o objeto omitido como um caso de instanciação nula indefinida. Já os casos de instanciação nula definida envolvem os casos em que o objeto do verbo é essencial, porém ambas as partes envolvidas na conversa o omitem por compartilharem do mesmo conhecimento.

As relações *frame-a-frame* na FrameNet são classificadas por Fillmore e Baker (2010) em: relações de generalização, relações de estrutura de evento e relações sistemáticas.

1. Relações de generalização:

Herança: Todos os elementos de *frame* do *frame* pai são ligados aos elementos de *frame* do *frame* filho, porém os elementos de *frame* do *frame* filho não necessitam ter o mesmo nome dos elementos de *frame* do *frame* pai. O *frame* filho representa um subtipo do pai. Por exemplo, o *frame* `Judgment_direct_adress` herda os elementos de *frame* do *frame* `Judgment_communication`, na medida em que o *frame* pai descreve um ato de comunicação de uma avaliação e o *frame* filho especifica esse evento, descrevendo casos em que a avaliação é comunicada diretamente ao avaliado.

Perspectiva: Duas unidades lexicais, por exemplo, *sell* e *buy*, evocam *frames* com diferentes perspectivas de um evento de transação comercial. Assim, *sell* (*vender*) apresenta a perspectiva do vendedor e o verbo *buy* (*comprar*) apresenta a perspectiva do comprador.

Uso: Nesta relação, o *frame* filho depende do contexto geral provido pelo *frame* pai. Embora nem todos os elementos de *frame* estejam ligados aos elementos de *frames* do pai, alguns podem estar relacionados. O *frame* `Judgment_comunication` usa o *frame* `Judgment`. Assim, alguns elementos de *frame* são semelhantes a ambos os *frames*.

2. Relações de estrutura de evento:

Subframe: Eventos complexos apresentam *subframes* com ordenação temporal. O *frame* `Criminal_process` apresenta três *subframes*, ordenados temporalmente através da relação **precede**: `Arrest`, `Arraignment`, `Trial` e `Sentencing`.

Precede: Especifica a relação temporal entre *frames*. `Arrest` precede `Arraignment`, `Arraignment` precede `Trial` e `Trial` precede `Sentencing`.

3. Relações sistemáticas:

Causativo de: O *frame* pai representa o causativo do *frame* filho. O *frame* `Cause_change_of_position_on_a_scale` é o causativo de `Cause_position_on_a_scale`.

Incoativo de: O *frame* pai representa o incoativo e o *frame* filho representa o estativo. `Change_position_on_a_scale` é o incoativo de `Position_on_a_scale`.

A estrutura da FrameNet pode ser mais bem compreendida pela distinção entre relações sintagmáticas *versus* relações paradigmáticas (Cruse, 1986). A FrameNet descreve o significado lexical através da relação associativas das unidades lexicais com os demais componentes da sentença. Assim, os elementos de *frame* descrevem a natureza semântica dos participantes da predicação. Recursos léxico-conceituais como a WordNet descrevem o significado lexical das unidades lexicais por meio de relações paradigmáticas, ou seja unidades lexicais de significado semelhante que são agrupadas formando um grupo de sinônimos.

2.4 Os *Frames* Semânticos e a Multilinguagem

A presente seção se ocupa de questões relacionadas ao desenvolvimento de FrameNets multilíngües, tais como a estruturação desses recursos e a sua conexão por meio de relações. A criação de recursos lexicais multilíngües enfrenta o desafio de representar, computacionalmente, a riqueza semântica das línguas.

Altenberg e Granger (2002) defendem que os padrões de polissemia entre as línguas variam, podendo, inclusive, não haver a equivalência entre as línguas. Por esse motivo, os recursos lexicais multilíngües encontram dificuldades em conectar bases de dados em diferentes línguas, pois, mesmo havendo equivalência entre as línguas, a polissemia dos equivalentes varia entre as línguas. Uma solução para esse problema apresentada pela

EuroWordNet (Vossen, 1999) foi a criação de relações multilíngües de quase-equivalência, baseada na relação semântica de quase-sinonímia.

Para tratar da problemática da equivalência entre línguas e a construção de FrameNets, serão apresentadas a seguir metodologias de desenvolvimento de recursos multilíngües. Será discutido aqui o relatório de Lönneker-Rodman (2007) sobre a multilingualidade na FrameNet. Esse relatório faz uma revisão de diferentes modelos estruturais para construção de recursos multilíngües e apresenta detalhes sobre o desenvolvimento de *frames* semânticos no âmbito do projeto FrameNet e de suas extensões multilíngües.

Os dois primeiros pontos apontados por Lönneker-Rodman (2007) para a construção de recursos multilíngües são a **similaridade organizacional** e a **inter-relação** dos recursos lingüísticos. A **similaridade organizacional** ocorre quando diferentes bases de dados utilizam o mesmo paradigma para a construção de recursos em diferentes línguas. Como exemplo dessa estratégia, tem-se a EuroWordNet. No contexto do projeto EuroWordNet, WordNets foram criadas em diferentes línguas seguindo o mesmo princípio organizacional: os grupos de sinônimos (*synsets*). No caso da FrameNet, projetos como a Spanish FrameNet, a Japanese FrameNet e a FrameNet Brasil seguem a mesma organização estrutural do projeto pioneiro FrameNet, organizando a informação lexical em *frames*. A figura 2 apresenta a similaridade organizacional dos recursos lexicais multilíngües.

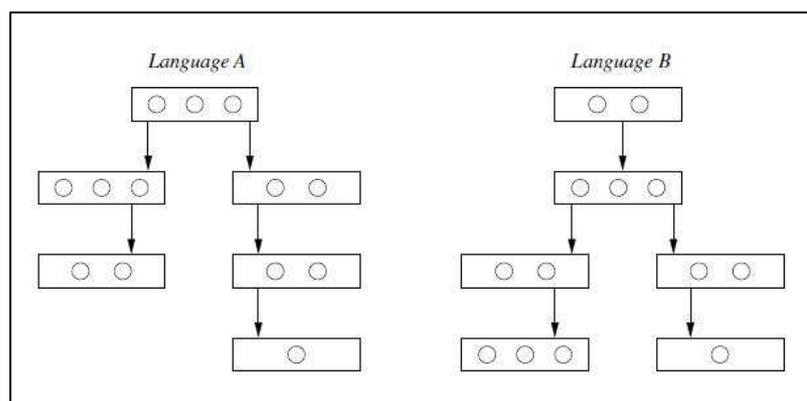


Figura 2: Similaridade Organizacional (Lönneker-Rodman, 2007, p.3)

A **inter-relação** está ligada à similaridade estrutural. Recursos lexicais construídos seguindo os mesmos princípios estruturais podem ser inter-relacionados através de diferentes relações multilíngües. Segundo Lönneker-Rodman (2007), a verdadeira inter-relação de recursos lexicais envolve a ligação entre componentes internos, ou seja, conexão de pares de

unidades lexicais em diferentes línguas. Isso equivaleria à perfeita sobreposição entre equivalentes em duas línguas. Como tal fenômeno é raro, ou mesmo inexistente, os recursos lexicais necessitam criar alternativas para inter-relacionarem as diferentes bases de dados, com relações multilíngües que vão além do mapeamento perfeito entre unidades lexicais. A figura 3 apresenta a inter-relação de recursos lexicais multilíngües.

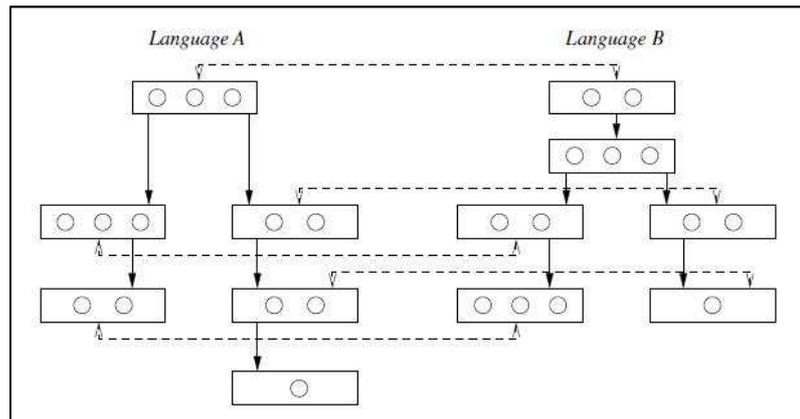


Figura 3: Inter-relação de Recursos Lexicais Multilíngües (Lönneker-Rodman, 2007, p.3)

Segundo Lönneker-Rodman (2007), há duas metodologias possíveis para a inter-relação de recursos lexicais multilíngües: a **fusão** e a **expansão**. Na metodologia de **fusão**, recursos para diferentes línguas são construídos de forma independente. Depois de concluída a construção dos diferentes recursos, esses são relacionados um ao outro por meio de relações multilíngües. Essa foi a metodologia adotada pela EuroWordNet (Vossen, 1999). Já na metodologia de **expansão** um recurso em uma língua é transferido para outra língua. Conforme Lönneker-Rodman (2007), essa metodologia assume que a estrutura conceitual permanece imutável, sendo apenas a informação específica de língua substituída. Essa tem sido a metodologia adotada por projetos como a Spanish FrameNet. No entanto, essa metodologia implica uma simplificação dos fenômenos lingüísticos, tais como a polissemia, uma vez que se parte de um recurso já existente em outra língua e se tenta manter a mesma organização estrutura do recurso lexical. Segundo Lönneker-Rodman (2007, p.5), *a abordagem de Expansão tende a produzir recursos estruturalmente similares com o risco de se negligenciar as diferenças específicas de língua na lexicalização e, portanto, na estrutura do léxico.*

Lönneker-Rodman (2007) apresenta duas formas pelas quais a representação da conexão entre as expressões lingüísticas nos recursos multilíngües podem ser realizadas: o

mapeamento direto e o **mapeamento indireto**. No caso do **mapeamento direto**, duas expressões lingüísticas correspondentes na língua A e na língua B são diretamente relacionadas. No **mapeamento indireto**, as expressões lingüísticas são conectadas por meio de um índice intermediário chamado de **índice de interlíngua** (Vossen, 1999). Esse índice é representado separadamente das demais bases específicas de língua. As relações para conectar as expressões lingüísticas, seja por meio de mapeamento direto ou de mapeamento indireto, podem ser de diferentes tipos.

De acordo com Lönneker-Rodman (2007), a **relação multilíngüe** mais utilizada pelos diferentes recursos multilíngües é a **equivalência**. No entanto, a equivalência de significado entre as línguas é difícil de ser definida, podendo haver expressões lingüísticas em uma língua que não apresentem equivalentes em outra língua, como o caso de alguns conceitos jurídicos. Assim, se nenhuma outra relação multilíngüe além da equivalência for estabelecida para conectar as expressões lingüísticas em recursos lexicais multilíngües, pode-se correr o risco de não haver a conexão entre as expressões lingüísticas. Como solução para essa problemática, o projeto EuroWordNet (Vossen, 1999) optou pela criação de diferentes relações multilíngües além da equivalência, como a quase-equivalência entre as línguas e a sinonímia entre diferentes classes gramáticas, como entre *branco* e *branquear*.

O desenvolvimento de FrameNets pelo método de expansão, como no caso da Spanish FrameNet e da Japanese FrameNet, tem resultado em recursos similares do ponto de vista organizacional. A FrameNet está dividido em três bases de dados: a base de dados contendo as unidades lexicais, a base de dados de sentenças e a base de dados de *frames*. Segundo Ruppenhoff et al. (2010), dessas três bases de dados, é a base de *frames* que apresenta um nível mais razoável de independência de língua. É devido a essa maior similaridade dos *frames* entre as línguas que Boas (2005) sugere o uso dos *frames* como índice de interlíngua na construção de recursos lexicais multilíngües.

No entanto, para Lönneker-Rodman (2007), os *frames* em diferentes línguas podem refletir aspectos particulares de cada língua, o que resultaria num desvio da equivalência de *frames* entre determinadas línguas. Assim, Lönneker-Rodman (2007) analisa as possibilidades de diferenças dos *frames* entre as línguas em quatro níveis: o *frame*, os elementos de *frame*, a centralidade dos elementos de *frame* e as relações entre *frames*.

Equivalência de frames: Segundo Lönneker-Rodman (2007), há pelo menos duas possibilidades para a criação de um novo *frame* em uma FrameNet de outra língua. Uma delas é a inadequação das descrições de *frame* da FrameNet quando comparadas com o *frame* de

outra língua. Um exemplo de discordância é o *frame* Statement, que na Japanese FrameNet é descrito por dois *frames* diferentes, Statement_verbal_act e Statement_verbal_transfer (Lönneker-Rodman, 2007). A outra possibilidade é a cobertura inadequada de um determinado campo semântico na FrameNet. Um exemplo de cobertura incompleta é o caso do *frame* Return na Spanish FrameNet, para o qual ainda não há cobertura na FrameNet (Lönneker-Rodman, 2007). Os *frames* jurídicos analisados neste trabalho evidenciam a inadequação de vários *frames* semânticos da FrameNet para a descrição dos eventos jurídicos brasileiros. Essa falta de correspondência entre os eventos jurídicos nos EUA e no Brasil levou ao desenvolvimento de novos *frames* para descrever determinados eventos jurídicos, como Denúncia e Pronúncia, que são propostos a partir do estudo das unidades lexicais do *frame* Notification_of_charges.

Equivalência de elementos de *frame*: A necessidade de criação de novos elementos de *frame* para comportar as diferenças entre as línguas também implica a criação de um novo *frame*. Esse posicionamento com relação à necessidade de imutabilidade dos *frames* para eles serem considerados correspondentes entre as línguas só enfatiza a diferença dos *frames* jurídicos. O *frame* Instrução, apesar de ser comparável ao *frame* Court_examination na etapa do processo penal que ambos os *frames* representam, apresentou em português a necessidade de divisão do elemento de *frame* WITNESS em dois elementos de *frame*: TESTEMUNHA e RÉU. Isso o tornaria um novo *frame*. Segundo Lönneker-Rodman (2007), a adição de informação específica de língua a um *frame* cria um *frame* mais específico. Esse *frame* mais específico poderia ser ligado ao *frame* semântico da FrameNet por meio de relação de herança multilíngüe. Assim, o *frame* mais específico de língua seria um *frame* filho do *frame* semânticos da FrameNet, que, por ser mais geral, seria o *frame* pai (Lönneker-Rodman, 2007). No entanto, os *frames* jurídicos são podem ser conectados por meio de relação de herança multilíngüe, uma vez que eles descrevem eventos socialmente criados. Assim, o *frame* Instrução não poderia ser considerado como *filho* do *frame* Court_examination, pois não se trata de um *frame* mais específico.

Equivalência de tipos semânticos e de centralidade de elementos de *frame*: Segundo Lönneker-Rodman (2007), diferenças entre a centralidade dos elementos de *frame* entre as línguas não podem ser representadas através da relação de herança multilíngüe. Assim, *frames* semânticos da FrameNet cujos elementos de *frame* centrais não possuam *status* de elementos de *frame* centrais em FrameNets de outras línguas não podem ser considerados equivalentes. Assim, diferenças de centralidade de *frames* e de tipos semânticos entre as

línguas devem ser consideradas como indicadores de diferenças de elementos de *frame*. Lönneker-Rodman (2007) afirma que essas diferenças produzem um novo *frame*, mesmo que as diferenças sejam sutis.

Equivalência de relações entre elementos *frames*: Toda vez que o número de elementos de *frame*, os tipos ou a centralidade semânticos dos elementos de *frame* são alterados, muda também a relação dos elementos de *frame* entre si (Lönneker-Rodman, 2007). Assim, qualquer mudança necessária no novo *frame* para acomodar as diferenças lingüísticas implicará a não equivalência com o *frame* semântico da FrameNet.

As observações sobre a correspondência multilíngüe dos *frames* semânticos apresentadas em Lönneker-Rodman (2007) nos levam à conclusão de que o uso dos *frames* como representação de interlíngua para a conexão de recursos lexicais multilíngües pode ser bastante restrito. Inclusões de elementos de *frame* em um *frame* criado para outra língua leva à criação de um *frame* mais específico, conectado ao *frame* semântico da FrameNet através de relação de herança multilíngüe. No entanto, mudanças na definição do *frame*, na centralidade dos elementos de *frame* e nos tipos semânticos requerem a criação de um novo *frame* na nova FrameNet. É importante lembrar que as observações de Lönneker-Rodman (2007) levam em conta a metodologia de expansão da FrameNet, em que as novas FrameNets utilizam toda a estrutura da FrameNet norte-americana, inclusive a definição e os elementos de *frames* em inglês.

O objetivo deste capítulo foi apresentar a Semântica de *Frames* e a FrameNet. Pode-se dizer que o sucesso dessa teoria semântica se deve em grande parte ao sucesso da FrameNet. De fato, a Semântica de *Frames* e a FrameNet andam lado a lado, a teoria alimentando o recurso lexical e o recurso lexical alimentando a teoria. Assim, esse capítulo apresentou a teoria que está em estudo neste trabalho e o recurso lexical que (i) foi utilizado para a análise contrastiva apresentada no capítulo 4 e (ii) serviu de paradigma para a proposta de *frame* para o português apresentado no capítulo 5.

O caráter contrastivo e especializado deste trabalho pede algumas considerações a esse respeito. Portanto, o capítulo 3 apresentará:

- A) A lingüística contrastiva em sua abordagem mais atualizada, não a tradicional lingüística contrastiva aplicada ao ensino de línguas, mas aplicada ao desenvolvimento de léxicos computacionais multilíngües.
- B) A problemática da equivalência de tradução em domínio jurídico.
- C) Os trabalhos realizados em lingüística contrastiva sob essa nova orientação metodológica, em língua geral e em linguagem de especialidade.

3 A FRAMENET E A ANÁLISE CONTRASTIVA

O capítulo anterior explorou os fundamentos teóricos e aplicados deste trabalho; de um lado a Semântica de *Frames* como paradigma teórico orientado pelos princípios da lingüística cognitiva e de outro a FrameNet como paradigma de aplicação da Semântica de *Frames* para a criação de recursos lexicais. Este capítulo, por sua vez, aborda a orientação metodológica desta pesquisa, que está baseada em três pontos interdependentes: (i) a lingüística contrastiva, (ii) a interface entre língua geral e linguagem de especialidade e (iii) a criação de recursos lexicais computacionais.

A lingüística contrastiva neste trabalho é inspirada em Altenberg e Granger (2002), partindo da visão de lingüística contrastiva aplicada à criação de recursos lexicais. Assim, o objetivo aqui é analisar em que medida os eventos jurídicos representados pelos *frames* semânticos da FrameNet são equivalentes aos eventos jurídicos brasileiros. Para tanto, a opção foi contrastar as unidades lexicais relacionadas ao processo penal americano e brasileiro. Nesse contexto, a abordagem contrastiva não deve ser confundida com a análise contrastiva tradicional³, cujo principal objetivo era o contraste de línguas para fins pedagógicos. Também não deve ser confundida com a tipologia lingüística, análise contrastiva que visa encontrar universais lingüísticos⁴.

Apesar de este trabalho não investigar as unidades lexicais sob o ponto de vista da terminologia, muitas das unidades lexicais analisadas aqui são termos jurídicos. No entanto, definir termos jurídicos não faz parte do escopo desta pesquisa. O interesse aqui é analisar em que medida os *frames* semânticos da FrameNet que representam um conhecimento jurídico são correspondentes em inglês e português.

³ São exemplos desse tipo de análise contrastiva tradicional as gramáticas contrastivas do inglês e espanhol (Stockwell et al., 1965) e do inglês e polonês (Fisiak et al., 1978).

⁴ São exemplos desse tipo de abordagem os trabalhos de Comrie (1989) e Croft (2002) sobre universais lingüísticos e tipologia lingüística.

A orientação computacional deste trabalho está relacionada ao desenvolvimento de recursos lexicais. Uma série de projetos tem se ocupado com descrição semântica do léxico para a criação de recursos computacionais (Altenberg e Granger, 2002; Boas, 2009a). Esta pesquisa se orienta por esses projetos anteriores e utiliza a Semântica de *Frames* e a FrameNet como paradigma para a análise contrastiva, conforme Boas (2009a).

Assim, para discutir os três eixos metodológicos que impulsionam esta pesquisa, o presente capítulo foi estruturado em três seções. A seção 3.1 apresenta a lingüística contrastiva e a equivalência de unidades lexicais entre línguas, parte fundamental da análise lingüística que se apresentará no capítulo seguinte. A seção 3.2 aborda as diferenças entre os sistemas jurídicos e os reflexos que essa falta de correspondência entre sistemas tem na equivalência de termos jurídicos entre línguas. A seção 3.3 apresenta trabalhos contrastivos que têm como finalidade a aplicação do paradigma FrameNet para a criação de recursos lexicais de língua geral e de linguagem especializada.

3.1 A Faceta Contrastiva

Esta seção trata da lingüística contrastiva, um dos eixos metodológicos desta pesquisa. Embora tanto lingüística contrastiva quanto estudos de tradução se interessem pela equivalência⁵, sem a qual não seria possível nenhum contraste entre línguas, é importante enfatizar que este trabalho não se trata de um estudo de tradução.

Embora lingüística contrastiva e estudos de tradução se encontrem em um momento de convergência (Granger, 2003), essas duas disciplinas se originaram de programas de pesquisa bem distintos. A tradução tem seu foco inicial na tradução de texto entre línguas. O início dos estudos de tradução remonta aos tempos de Cícero e São Jerônimo⁶, ainda na Roma Antiga (Granger, 2003). A tradução foi por muito tempo uma atividade desenvolvida dentro

⁵ Não se pode deixar de mencionar que o conceito de equivalência é compartilhado pelos estudos de tradução e pela lingüística contrastiva. Faz-se referência aqui aos estudos de tradução, mesmo não sendo o foco do trabalho, para chamar a atenção ao fato de o conceito de equivalência ser contestado também pelos estudos de tradução. Segundo Oliveira (2007), a equivalência nos estudos de tradução era, inicialmente, encarada como a busca por igualdade de valores entre as línguas. Segundo essa visão, o significado seria estável e seria partilhado entre as línguas. Com o desenvolvimento dos estudos lingüísticos na última metade do século 20, a equivalência cai em uma progressiva desvalorização e a ênfase passa a ser dada à recepção do texto traduzido. A tradução não é mais vista como uma procura por equivalentes perfeitos que possam transmitir na língua meta o mesmo significado da língua fonte. Essa crescente desvalorização da equivalência chega a seu ápice com a abordagem desconstrucionista. Para os desconstrucionistas, não há qualquer tipo de equivalência entre o texto original e a tradução.

⁶ São Jerônimo é conhecido por religiosos como o padroeiro dos estudos bíblicos pelo seu trabalho na tradução da Bíblia do grego antigo e do hebraico para o latim.

de mosteiros, sendo a tradução da Bíblia o trabalho de tradução por excelência. Assim, os estudos de tradução têm seu início na tradução de textos clássicos, tornando-se uma disciplina acadêmica apenas na década de 80 (Granger, 2003).

A lingüística contrastiva, inicialmente chamada de análise contrastiva, encontra seu auge após a Segunda Guerra Mundial, motivada pela possibilidade de descrição da estrutura das línguas e aplicação pedagógica. Segundo Granger (2003, p.17), *a análise contrastiva foi, inicialmente, um empreendimento puramente aplicado, objetivando produzir métodos e ferramentas de ensino de língua estrangeira mais eficientes*. Nesse período, há uma profusão de gramáticas contrastivas (Stockwell et al., 1965; Fisiak et al., 1978). Segundo a visão da época, se contrastadas as estruturas diferentes entre as línguas, seria mais fácil para o estudante de língua estrangeira aprender a nova estrutura.

Apesar de terem interesses de pesquisa diversos, tanto a análise contrastiva como os estudos de tradução compartilham o interesse pela equivalência entre línguas. A equivalência aqui será abordada em dois blocos. A subseção 3.1.1 apresenta a lingüística contrastiva. A subseção 3.1.2 discute a equivalência sob o ponto de vista da lingüística contrastiva.

3.1.1 A Lingüística Contrastiva

Os estudos contrastivos podem ser divididos em diferentes abordagens. A lingüística comparativa histórica, durante o século 19, levou muitos estudiosos, motivados pelas semelhanças entre as línguas, a compararem a estrutura das línguas em busca de uma origem comum. O século 20 viu o auge de duas outras abordagens contrastivas. A tipologia lingüística buscava, através da comparação entre as línguas, os universais lingüísticos, ou seja, uma série de características comuns às línguas. A análise contrastiva, por sua vez, tinha objetivos pedagógicos: o ensino de segunda língua⁷.

Segundo Altenberg e Granger (2002, p.5), *quando a análise contrastiva emerge como uma disciplina acadêmica nas décadas posteriores à Segunda Guerra Mundial, ela era tida basicamente como um ramo aplicado da lingüística, servindo para fins pedagógicos no ensino de língua estrangeira e de segunda língua*. Impulsionada pelas idéias de autonomia do sistema lingüístico e estrutura lingüística, a análise contrastiva tinha como princípio a idéia de que explicitar as diferenças estruturais das línguas ajudaria no ensino de línguas. São

⁷ Uma visão mais ampla da história das abordagens contrastivas é apresentada em Krzeszowski (1990), Altenberg e Granger (2002) e Granger (2003).

exemplos desses empreendimentos contrastivos as gramáticas contrastivas do inglês e espanhol (Stockwell et al., 1965) e do inglês e do polonês (Fisiak et al., 1978).

Conforme Altenberg e Granger (2002) apontam, a análise contrastiva teve seus dias de glória na década de 60, até ser associada com o estruturalismo e cair em desprestígio. Foi também nesse período que o gerativismo despontou como novo modelo lingüístico. Outro motivo também foram os avanços na compreensão do processo de aquisição de segunda língua (Granger, 2003). Assim, fatores estruturais, como conhecer em que nível duas línguas se diferenciavam, passaram a ser menos importantes que fatores cognitivos e emocionais envolvidos no processo de aprendizado, como a generalização de regras na língua-alvo e a motivação para o aprendizado. Esses fatores levaram ao um declínio acentuado da análise contrastiva.

Nas últimas décadas, a análise contrastiva vem ganhando novo impulso, mas dessa vez com um novo objetivo. A lingüística contrastiva, como a análise contrastiva tem sido chamada atualmente, deslocou o foco de pesquisa da gramática para o léxico. Diferentes projetos multilíngües cujo objetivo é a representação do significado lexical tem sido desenvolvidos desde a década de 90⁸, como o AQUILEX (Calzolari e Briscoe, 1995), o SIMPLE (Lenci et al., 2000) e a EuroWordnet (Vossen 1998)⁹.

A explicação para esse renascimento da lingüística contrastiva, segundo Altenberg e Granger (2002), se deve a três fatores. A crescente integração da Europa tem provocado uma maior demanda por tradução, interpretação e ensino de línguas. Ao mesmo tempo, uma crescente valorização do léxico, do discurso e da evidência empírica tem motivado o uso de *corpora* em pesquisa. Por fim, o advento da computação e a possibilidade de analisar a língua com base em grandes *corpora* têm provocado um grande impacto no processamento de linguagem natural.

Este trabalho se alinha a essa visão de lingüística contrastiva aplicada à criação de recursos lexicais computacionais (Altenberg e Granger, 2002; Boas, 2009a). Assim, este trabalho não vai tratar de tipologia lingüística, nem da comparação de estruturas lingüísticas para fins pedagógicos, mas comparação lingüística para fins de construção de léxicos computacionais. No caso desta pesquisa, interessa a comparação dos *frames* semânticos em

⁸ O grande impulso para o desenvolvimento de léxicos multilíngües tem sido a unificação econômica da Europa. Segundo as normas da União Européia, todos os documentos produzidos pelo Parlamento Europeu devem ser traduzidos para todas as línguas oficiais dos países membros. Isso provoca uma demanda muito grande por tradução, o que faz a União Européia investir em projetos para o desenvolvimento de recursos lexicais que possam auxiliar na tradução.

⁹ Uma lista extensa de projetos multilíngües pode ser encontrada em Boas (2009b).

inglês e português para verificar em que medida eles são correspondentes, pois a resposta para essa questão influencia a metodologia usada para a criação de léxicos baseados em *frames*.

3.1.2 A Equivalência

Para Krzeszowski (1990, p.35), um estudo contrastivo inclui três etapas metodológicas: (i) descrição, (ii) justaposição e (iii) comparação. A descrição é a etapa em que são definidos o que será comparado e que paradigma será usado para a comparação. A justaposição envolve a busca e a identificação dos equivalentes a serem comparados. A comparação envolve a análise dos equivalentes. Desde já se percebe que a equivalência é o ponto fundamental da lingüística contrastiva.

Para manter a coerência do fenômeno analisado, Krzeszowski (1990) sugere o *tertium comparationis* como medida de comparação. *Todas as comparações envolvem a suposição básica de que os objetos a serem comparados compartilham algo em comum, através do qual as diferenças podem ser elencadas. Esta plataforma de referência é chamada de tertium comparationis* (Krzeszowski, 1990, p.15). Assim, Krzeszowski (1990) estabelece sete tipos de equivalência que podem servir de *tertium comparationis*: equivalência estatística, equivalência de tradução, equivalência de sistema, equivalência semântico-sintática, equivalência de regra, equivalência substantiva e equivalência pragmática.

Segundo Krzeszowski (1990), dois itens lingüísticos são estatisticamente equivalentes se cada um é o equivalente mais freqüente do outro. A equivalência de tradução envolve dois textos que estão em relação de equivalência de tradução, um sendo a tradução do outro. A equivalência de sistema envolve a comparação do sistema lingüístico, tal como a ocorrência de determinadas categorias gramaticais nas línguas em contraste. A equivalência semântico-sintática é definida por Krzeszowski (1990) como estrutura profunda idêntica, em que a estrutura profunda é entendida como estrutura capaz de derivar as estruturas sintáticas. Quando Krzeszowski (1990) fala em equivalência de regras, ele está se referindo a regras para a geração de sentenças, de acordo com o paradigma gerativista. A equivalência substantiva envolve os níveis fonológico e lexical. Por fim, a equivalência pragmática envolve o estudo dos níveis estilístico e sócio-lingüístico da língua.

Krzeszowski (1990) aplica o modelo gerativista-transformacional ao estudo contrastivo. Sua concepção de equivalência semântico-sintática e de equivalência de regras

expressa bem a orientação gerativista. Altenberg e Granger (2002) afirmam que o *tertium comparationis* é fundamentado na equivalência em sentido estrito, o que impede o seu uso em alguns casos de estudos menos formalistas. James (1980, p.178) também critica as abordagens formalistas para o estudo da equivalência quando afirma que (...) *identidade de estrutura profunda não garante equivalência de tradução*.

Assim, o que Altenberg e Granger (2002) sugerem é iniciar por uma similaridade percebida ou assumida entre itens lingüísticos de diferentes línguas, ou seja, uma similaridade de significado. A similaridade de significado entre línguas é muito mais uma questão de julgamento que de correspondência formal. Chesterman (1998) defende que o nível de similaridade depende do julgamento do pesquisador. Portanto, identificar equivalentes em diferentes línguas demonstra a competência bilíngüe do pesquisador.

Se a intuição do analista tem papel fundamental na definição da equivalência, como garantir a confiabilidade da equivalência entre línguas? Spalatin (1967) propõe a reversibilidade da tradução (*back-translation*) para verificar a correspondência do significado dos itens lingüísticos entre as línguas. A reversibilidade de tradução propõe o isolamento de elementos lingüísticos formais do texto fonte, que são traduzidos para a língua-alvo e da língua-alvo novamente para a língua-fonte para se verificar a correspondência formal e a preservação do significado. O método de reversibilidade da tradução demonstra a importância da equivalência de tradução da língua-fonte para a língua-alvo e da língua-alvo para a língua-fonte.

Uma segunda alternativa, baseada na proposta de reversibilidade da tradução de Spalatin (1967), é proposta por Altenberg (1999): a correspondência mútua de dois itens em um corpus de tradução bidirecional. Se x na língua A é sempre traduzido por y na língua B e y na língua B é sempre traduzido por x na língua A, então x e y terão uma correspondência mútua de 100%. Segundo Altenberg e Granger (2002), a correspondência mútua de categorias em diferentes línguas dificilmente chega a 100%, sendo 80% uma correspondência já alta.

O que faz da correspondência entre as línguas uma meta tão difícil de ser atingida? A resposta para essa pergunta pode ser buscada na lingüística cognitiva. Diferentes culturas categorizam o mundo de formas variadas e as diferentes formas de se ver o mundo podem ser detectadas na língua. Assim, cada língua vai possuir um elenco diferente de itens lexicais com relações de sentido diferentes entre si em cada língua. O resultado dessa interação entre cognição, língua e cultura é a falta de correspondência dos itens lexicais entre as línguas.

Para Altenberg e Granger (2002), as diferentes extensões de significados nas línguas dão origem a diferenças nos processos de lexicalização, de gramaticalização e de extensão metafórica, o que resulta em padrões complexos de polissemia nas línguas. As diferenças nos padrões de polissemia entre as línguas provocam três tipos de relações (Altenberg e Granger, 2002, p. 22):

- (a) **Polissemia sobreposta** (*overlapping polysemy*): os itens lingüísticos em duas línguas possuem basicamente as mesmas extensões de significado. No entanto, essa completude de correspondência é muito difícil de existir.
- (b) **Polissemia divergente** (*diverging polysemy*): os itens lingüísticos em duas línguas têm diferentes extensões de significado. Segundo Altenberg e Granger (2002), os casos em que uma única palavra em uma língua corresponde a diversas palavras em outra língua é uma variante da polissemia divergente.
- (c) **Correspondência inexistente** (*no correspondence*): um item lingüístico em uma língua não encontra equivalente em outra língua.

A proposta de Altenberg e Granger (2002) é baseada nas relações semânticas que os itens lexicais apresentam entre si nas diferentes línguas e no nível de sobreposição dessas relações entre as línguas. Snell-Hornby (1983) propõe quatro níveis de equivalência, baseadas, não em relações semânticas entre as palavras, mas no núcleo semântico de cada item lexical:

- (a) **Equivalência total** (*total equivalence*): os itens lexicais em duas línguas são intensionalmente e extensionalmente correspondentes. Esse é o caso de termos técnicos, como *oxygen/oxigênio*, e termos de difusão internacional, como *typewriter/máquina de escrever*.
- (b) **Equivalência de trabalho** (*working equivalence*): um item lingüístico em uma língua possui diversos correspondentes em outra língua, no entanto, o equivalente necessitará ser encontrado com base no contexto. Esse é o caso do verbo *to know* em inglês, que em português possui dois correspondentes, *saber* e *conhecer*.
- (c) **Cobertura parcial** (*partial coverage*): um item lingüístico em uma língua encontra correspondência parcial em um item lingüístico em outra língua, sendo que o restante do significado do verbo necessita ser expresso por paráfrase. Esse é o caso do verbo inglês *tiptoe*; cujo significado de movimento encontra correspondente em um verbo e a maneira como o movimento foi realizado

necessita ser parafraseada: *caminhar na ponta dos pés* ou *entrar na ponta dos pés*.

- (d) **Cobertura-zero** (*nil-coverage*): quando um determinado item lingüístico expressa um conhecimento cultural muito específico, esse item tende a não encontrar correspondência. Esse é o caso de alguns termos jurídicos em inglês, como *arraignment*.

As relações de equivalência de Snell-Hornby (1983) foram propostas a partir de seu estudo sobre a correspondência dos verbos em inglês e alemão. Kade (1968, *apud* Pym, 2010, p.29) apresenta quatro tipos de relação de equivalência. As relações de equivalência de Kade (1968) possuem uma certa sobreposição com as duas outras taxonomias anteriores:

- (a) **Uma-para-um** (*Eins-zu-Eins*): um item lingüístico em uma língua encontra um correspondente exato em outra língua. Essa relação é semelhante à relação de equivalência total proposta por Snell-Hornby (1983).
- (b) **Um-para-muitos** ou **muitos-para-um** (*Viele-zu-Eins*): um item lingüístico em uma língua corresponde a vários em outra língua. Essa relação equivale à relação de polissemia divergente proposta por Altenberg e Granger (2002).
- (c) **Um-para-parte** (*Eins-zu-Teil*): um item lingüístico encontra equivalência apenas parcial em um item lingüístico em outra língua. Um exemplo trazido por Pym (2010) para ilustrar essa relação é a palavra *brother* em inglês, que em chinês e coreano não possui equivalente pleno, pois nessas duas línguas possuem palavras diferentes para indicar um irmão mais velho ou um irmão mais novo. Essa relação é parcialmente equivalente à relação de cobertura parcial de Snell-Hornby (1983). As relações de equivalência propostas por Snell-Hornby (1983) são pensadas para os verbos, pois partem de seu estudo sobre a equivalência dos verbos em inglês e alemão.
- (d) **Um-para-nenhum** (*Eins-zu-Null*): nenhum equivalente é encontrado na língua-alvo. Essa relação equivale à relação de correspondência inexistente (Altenberg e Granger, 2002) e à relação de cobertura-zero (Snell-Hornby, 1983).

Como se pode ver nesta seção, a lingüística contrastiva tem como peça fundamental o conceito de equivalência. A equivalência, por sua vez, não é um fenômeno trivial. Boa parte do sucesso do estudo contrastivo depende da competência bilíngüe do analista.

Isso se deve ao fato de (i) as palavras possuírem uma ampla rede de polissemia que nem sempre é equivalente entre as línguas (Kade, 1968; Altenberg e Granger, 2002), (ii) o núcleo de significado de um item lexical não corresponder totalmente ao equivalente de tradução (Snell-Hornby, 1983), (iii) a correspondência ser parcial por questão de diferenças culturais (Kade, 1968) ou (iv) simplesmente não haver correspondência (Kade, 1968; Snell-Hornby, 1983; Altenberg e Granger, 2002).

Para este trabalho interessa particularmente a falta de equivalência entre as unidades lexicais causada por diferenças de sistemas, como é o caso das diferenças entre os sistemas jurídicos que resultarão em termos jurídicos distintos em diferentes países. Assim sendo, a próxima seção discute os problemas de equivalência no domínio jurídico.

3.2 A Faceta Jurídica

Esta seção discute a problemática da equivalência no Direito. O trabalho de criação de léxicos computacionais jurídicos multilíngües¹⁰ necessita gerenciar a falta de equivalência entre os termos jurídicos em diferentes línguas. Enquanto as taxonomias das ciências naturais como a botânica ou a biologia independem de cultura, o Direito é uma construção social. Assim, cada sociedade terá seu grupo de leis próprias. À medida que as leis mudam, os termos jurídicos que acompanham essas leis mudam também. Isso provoca a falta de equivalência entre os termos jurídicos em diferentes línguas.

Não são apenas as leis que mudam de um país para outro, muda também o sistema jurídico. Dois sistemas jurídicos interessam a este trabalho, a *Common Law*, de origem inglesa e seguida em países como os Estados Unidos e a Inglaterra, e a *Civil Law*¹¹, de origem romana e seguida em países como Itália, Portugal e Brasil. Assim, a subseção 3.2.1 tem por objetivo traçar a relação entre os termos jurídicos e os sistemas jurídicos e a subseção 3.2.2 aborda os sistemas jurídicos e a equivalência de termos jurídicos.

¹⁰ Está-se utilizando a palavra multilíngüe em um sentido amplo, envolvendo qualquer tipo de contraste entre línguas. Esse contraste pode ser entre duas línguas, bilíngüe, ou em mais de duas línguas, multilíngües. Como a diferença entre bilíngüe e multilíngüe não interessa a este trabalho, utilizar-se-á a palavra *multilíngüe* de forma a abranger ambos os conceitos.

¹¹ A tradução de *civil law* é *direito civil*. No entanto, não se está utilizando a tradução para não confundir o sistema jurídico *civil law* com o direito civil, ramo do direito que trata dos direitos dos cidadãos. Outra possibilidade seria utilizar o termo equivalente em português, direito romano. Essa possibilidade, porém, não expressa o contraste que há entre *common law* e *civil law*.

3.2.1 A Terminologia Jurídica

Segundo Rey (1995, p.10), *a terminologia é fundamentalmente preocupada com os nomes e o processo de nomeação*. A terminologia, apesar de ser uma disciplina nova, tem uma longa tradição no mundo ocidental. Pode ser ver a preocupação com a terminologia desde a Grécia Antiga, com *Crátilo* de Platão, até o século dezoito, com a *Encyclopedie* de Diderot (Rey, 1995). Segundo Cabré (1993), a internacionalização das ciências no século dezenove provocou nos cientistas uma necessidade de dispor de regras para a formação de termos. O fato de ciências como a biologia e a botânica criarem seus termos a partir de palavras latinas demonstra uma necessidade por padronização e universalidade, pois era necessário garantir o reconhecimento das espécies animais e vegetais por cientistas de diferentes países. No entanto, será apenas no século vinte que serão lançadas as bases metodológicas da terminologia, com os trabalhos de pesquisadores como Wüster e Lotte (Cabré, 1993).

Segundo Cabré (1993), o termo se divide em três eixos: o lingüístico, o cognitivo e o comunicativo. O termo é uma unidade lingüística que nomeia um conceito de uma área especializada com o fim de comunicação entre especialistas dessa área. Assim, o objeto de estudo da terminologia são as palavras pertencentes a um domínio específico do conhecimento. No eixo cognitivo do termo está a categorização. Os termos nomeiam categorias especializadas, ou seja, conceitos específicos de uma área de conhecimento. A finalidade do termo, portanto, é nomear categorias criadas por profissionais, permitindo a comunicação. A comunicação especializada se diferencia da comunicação em geral justamente pelo uso de terminologia específica.

Essa tripartição do conceito entre a sua natureza lingüística, cognitiva e comunicativa explicará os desencontros lexicais que são abordados no capítulo 4. Apesar de alguns termos jurídicos em português e inglês apresentarem equivalência, ou seja, eles possuírem uma forma lingüística semelhante, o conceito que eles expressam é diferente. Isso se deve às diferenças entre os sistemas jurídicos adotados pelos Estados Unidos e pelo Brasil, o que provoca diferenças profundas no Direito de ambos os países, desde as leis até os procedimentos legais adotados por ambos.

Conforme Mattila (2006), enquanto *um conceito é a representação mental de um objeto (...)* *um termo é a designação técnica de um conceito* (p.108). A diferença entre termos jurídicos e outras palavras não é facilmente identificada, pois, (...) *mais do que outras*

linguagens para fins específicos, a linguagem jurídica contém palavras da linguagem comum cujo significado é definido com mais precisão ou mesmo divergente daquele da linguagem comum (p.113).

Os termos jurídicos podem ser criados ou excluídos pelos legisladores. Quando um termo é excluído e substituído por outro, esse novo termo é imediatamente incorporado à linguagem especializada. Na linguagem dos cidadãos, no entanto, a substituição ocorre de forma lenta. Mattila (2006) apresenta três procedimentos para a criação de termos jurídicos: (i) uma palavra da língua geral, ou de outra área especializada, passa por uma especialização, tornando-se um termo jurídico, (ii) um neologismo é criado para expressar um determinado conceito ou (iii) um empréstimo de uma palavra de uma língua estrangeira para suprir a falta de uma unidade lexical na própria língua. A tradução jurídica, por necessitar recriar na língua-alvo o sentido da língua-fonte, necessita criar constantemente novos termos jurídicos para suprir a falta de equivalência de termos jurídicos na língua-alvo¹².

3.2.2 Os Sistemas Jurídicos e a Equivalência de Termos Jurídicos

Como explicar tamanha variação nos conceitos de uma ciência como o Direito? Segundo Mattila (2006, p.105), *o direito não existe no mundo físico. Por ser inteiramente criado pelos humanos, o direito é sempre relacionado à cultura de uma determinada sociedade. O direito constitui, portanto, um fenômeno social*. Quanto mais diferentes os conceitos de dois sistemas jurídicos, menos interação histórica houve entre essas culturas. É por esse motivo que países que adotam a *civil law*, de origem romana, tendem a ter um sistema jurídico mais semelhante, enquanto que a diferença aumenta significativamente ao se comparar o sistema de um país que adota a *civil law*, como o Brasil, com um país que adota a *common law*, como os Estados Unidos.

Os sistemas jurídicos podem ser divididos em grupos, de acordo com sua semelhança conceitual. Dois grupos apresentam maior significância na atualidade: a *common law* e a *civil law*. A *common law* é aplicada em países de língua inglesa, com notoriedade na Inglaterra e nos Estados Unidos. Já a *civil law* é o sistema adotado por países da Europa continental e da América Latina. As diferenças conceituais entre esses dois sistemas se devem a fatores históricos. Enquanto a *civil law* se desenvolveu nas universidades medievais européias, tendo como base o Direito Romano, a *common law* se formou nos tribunais da Inglaterra durante a

¹² Sarcevic (1997) discute a criação de novos termos jurídicos através da tradução.

invasão normanda (Mattila, 2006). Os sistemas baseados na *civil law* têm nas leis e nos códigos a base legal para as decisões judiciais, já as decisões dos tribunais de países baseados na *common law* são sempre baseadas em decisões de casos anteriores semelhantes.

Essa relação estreita entre cultura, Direito e língua pode ser claramente percebida em organizações internacionais, como a União Européia, ou países multilíngües, como o Canadá. A União Européia é a maior criadora de termos jurídicos multilíngües. Devido ao grande número de línguas oficiais, os documentos necessitam ser traduzidos para todas as línguas oficiais da União. Conforme Sarcevic (1997), as línguas de trabalho da União Européia são, geralmente, o francês e o inglês. Os textos produzidos durante os trabalhos da União Européia necessitam ser traduzidos para todas as línguas oficiais da União. Assim, novos termos nascem constantemente a partir da tradução, para dar conta das diferenças entre os sistemas jurídicos dos diferentes Estados membros.

Segundo Sarcevic (1997), a tradução legal no Canadá é uma questão delicada, pois o país não é apenas bilíngüe, mas possui dois sistemas jurídicos distintos: cinco províncias falantes de inglês adotam o sistema da *common law*, uma província bilíngüe (inglês-francês) adota o sistema da *civil law* (Québec) e outras quatro províncias que seguem o sistema da *common law* apresentam algum nível de bilingüismo. A presença de duas línguas e dois sistemas jurídicos distintos leva a uma maior inconsistência na tradução legal. Por exemplo, o termo *civil law*, em um país que segue o sistema da *common law*, indica o sistema jurídico baseado no Direito Germano-Românico. Se o termo *civil law* for traduzido literalmente para o francês, *droit civil*, esse termo não indicará mais o sistema jurídico, mas o ramo do Direito que regulamenta a vida do cidadão.

As questões abordadas nesta seção não se restringem à tradução jurídica, pois a equivalência é peça fundamental também para a lingüística contrastiva. Esta seção teve como objetivo chamar a atenção para as diferenças entre os sistemas jurídicos, visto que essas diferenças serão demonstradas a partir do estudo contrastivo apresentado no capítulo 4. A seção seguinte aborda a faceta computacional desta pesquisa.

3.3 A Faceta Computacional

A FrameNet foi concebido inicialmente como um recurso lexical monolíngüe, porém o desenvolvimento de projetos semelhantes para o alemão¹³ (Boas, 2002; 2005), espanhol¹⁴ (Subirats e Petruck, 2003) e japonês¹⁵ (Ohara et al., 2003) tem estimulado as discussões a respeito da possibilidade de uso da FrameNet como interlíngua. A possibilidade em discussão é o uso da base de dados de língua inglesa como ponto de conexão entre as diferentes FrameNets, possibilitando, assim, um léxico multilíngüe (Boas, 2005).

Padó (2007) também defende o uso dos *frames* semânticos como interlíngua em recursos multilíngües. A suposição é que os *frames* estão no nível conceitual, o que faz com que eles sejam lingüisticamente neutros. Os *frames* poderiam, portanto, ser utilizados como ponto de conexão entre os itens lexicais das diferentes línguas, superando o problema da falta de paralelismo entre as línguas. No entanto, este trabalho considera essa posição uma simplificação de uma questão complexa.

Se a lingüística cognitiva defende um modelo enciclopédico para a semântica, como identificar o significado sem levar em conta as questões culturais que envolvem as línguas? Este trabalho, a partir dos dados analisados no capítulo 4, contesta essa visão simplificadora no capítulo 5. Para tanto, é necessário abordar aqui os projetos que estão sendo desenvolvidos em lingüística computacional e lexicografia computacional tendo como base a FrameNet. Isso inclui desde o uso das etiquetas semânticas da FrameNet à criação de FrameNet para diferentes línguas.

A opção por apresentar esses projetos aqui se deve ao fato de ele servirem de reflexão para o capítulo 5, que trata das divergências entre *frames*, contestando algumas conclusões simplificadoras que se tem apresentado a respeito do uso dos *frames* semânticos como interlíngua¹⁶. O foco desta seção está no uso do paradigma FrameNet em outras línguas além do inglês. Assim, esta seção divide os diferentes projetos em três grupos. A subseção 3.3.1 aborda a criação de FrameNets em diferentes línguas. A subseção 3.3.2 apresenta o uso das etiquetas da FrameNet para anotação semântica de textos. Por fim, a subseção 3.3.3 traz os projetos voltados para a aplicação da Semântica de *Frames* em domínios especializados.

¹³ <http://www.laits.utexas.edu/gframenet/>

¹⁴ <http://gemini.uab.es:9080/SFNsite>

¹⁵ <http://jfn.st.hc.keio.ac.jp/index.html>

¹⁶ Interlíngua, no contexto deste trabalho, deve ser entendida como um recurso lexical projetado com a finalidade de interligar recursos lexicais em diferentes línguas, criando um recurso lexical multilíngüe.

3.3.1 Criação de FrameNets em Diversas Línguas

A Spanish FrameNet (Subirats, 2009), em desenvolvimento conjunto na Universidade Autônoma de Barcelona e no *International Computer Science Institute*, tem por objetivo desenvolver um recurso lexical com anotação semântica e sintática que possa servir como *corpus* de treino para tarefas de anotação automática de papéis semânticos. Algumas tarefas, como a extração de construções sintáticas, são realizadas automaticamente, outras, como a anotação sintática e semântica de textos, é realizada de forma semi-automática ou manual (Subirats, 2009).

A partir de uma análise da base de dados da Spanish FrameNet, pode-se ver que a opção metodológica da Spanish FrameNet foi iniciar o recurso lexical espanhol a partir dos *frames* já descritos pela FrameNet, metodologia chamada de *expansão*. Assim, a Spanish FrameNet utiliza todos os relatórios de *frame* da própria FrameNet, tais como *frames*, elementos de *frame* e descrição dos *frames*. As sentenças anotadas do inglês são substituídas por sentenças anotadas do espanhol, e uma lista de unidades lexicais do espanhol é fornecida para cada *frame*.

A Japanese FrameNet está em desenvolvimento na Universidade de Keio, no Japão, desde 2002. Conforme o *site* do projeto Japanese FrameNet, o objetivo atual é a construção de uma base de dados no estilo da FrameNet contendo as unidades lexicais do japonês. A Japanese FrameNet está trabalhando em conjunto com a FrameNet para construírem uma base de dados bilíngüe através da conexão das bases de dados americana e japonesa (Ohara, 2009).

Também estão em construção FrameNets para o alemão (Boas, 2002; 2005), o hebraico (Petrucci, 2009), o português brasileiro (Salomão, 2009), o italiano (Venturi et al., 2009), o francês (Pitel, 2009) e outros projetos estão iniciando ainda. As técnicas para criação de FrameNets variam. Alguns projetos se baseiam exclusivamente na análise manual, como a FrameNet brasileira. Em outros casos, utilizam-se técnicas automáticas, baseadas na anotação semântica e geração automática de *frames* (Padó e Lapata, 2005; Tonelli e Pianta, 2008). A subseção seguinte aborda a anotação semântica a partir das etiquetas semânticas da FrameNet e o uso da anotação semântica para geração de FrameNets.

3.3.2 Anotação Semântica e Criação Automática de FrameNets

O projeto SALSA (*Saarbrücken Lexical Semantics Annotation and Analysis*) vem realizando a anotação de *corpora* em língua alemã a partir das etiquetas semânticas da FrameNet (Burchardt et al., 2009). O projeto SALSA parte da suposição de que é possível reutilizar as etiquetas semânticas da FrameNet para a análise semântica do alemão. Assim, esse projeto inclui (i) a anotação de um *corpus* de grande porte em alemão e a geração de léxico baseado em *frames* a partir da anotação do *corpus* e (ii) a indução de modelos baseados em dados para análise semântica automática e aplicações em processamento de linguagem natural (Burchardt et al., 2009).

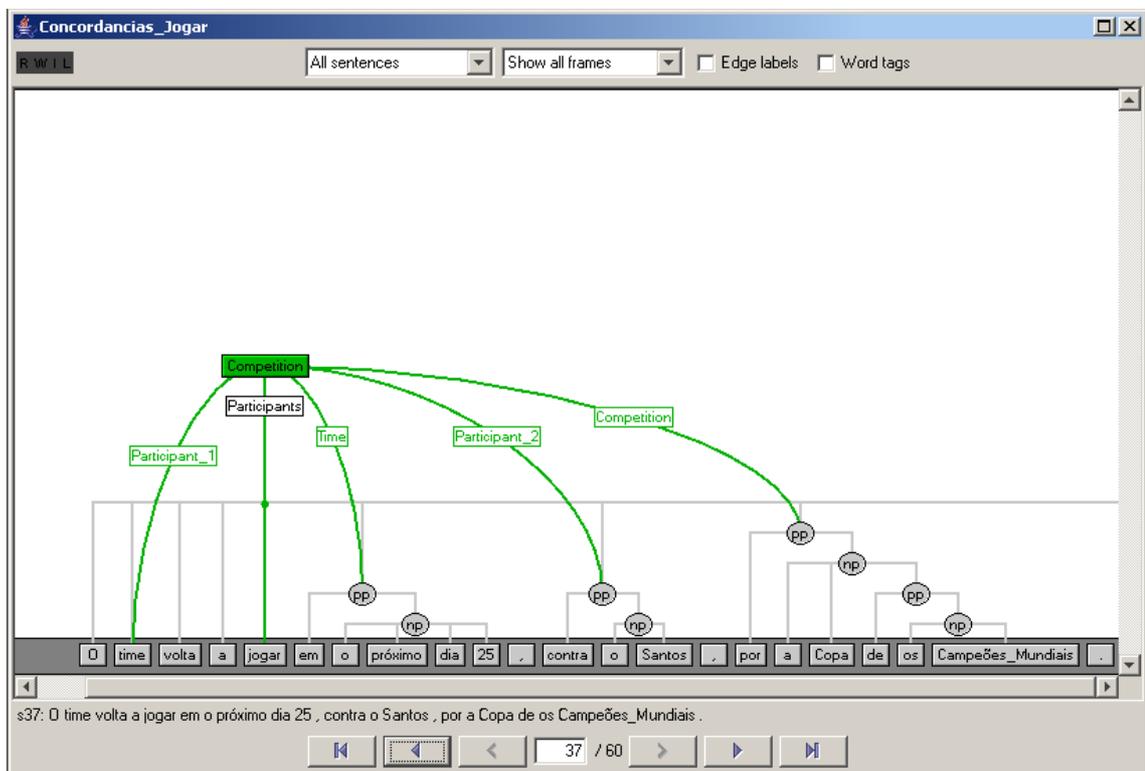


Figura 4: Anotação de *Corpus* com a Ferramenta SALTO

Diferentemente da FrameNet, o projeto SALSA volta sua atenção para a anotação extensiva de *corpus*. A primeira versão do *corpus*, finalizada em 2007, contém 20.000 instâncias anotadas (Burchardt et al., 2009). A anotação do *corpus* foi realizada através de ferramenta SALTO (Burchardt et al., 2006), desenvolvida pelo próprio projeto. A ferramenta SALTO pode ser utilizada para anotação semântica em diferentes línguas, basta dispor da base de dados de *frame* em *xml* da FrameNet, que é fornecida pelo próprio grupo de pesquisa

da FrameNet, e fazer a anotação sintática do *corpus* a ser utilizado com o auxílio de um *parser* sintático, no caso do português, o PALAVRAS (Bick, 2000). A figura 4 mostra um exemplo de sentença anotada para o português extraído de Chishman et al. (2007).

A ferramenta SALTO também permite a criação de *frames*. Esse recurso é importante por dois motivos. Em primeiro lugar, a FrameNet ainda está em construção e não há todos os *frames* necessário para a anotação extensiva de *corpora*. Em segundo lugar, nos casos em que há confronto entre os *frames* da FrameNet e a língua em análise é possível criar novos *frames*. O projeto FrameCorp (Chishman et al., 2008) utilizou a ferramenta SALTO para a anotação de *corpora* em português, sendo a figura 4 um exemplo da anotação realizada por esse projeto.

O trabalho do projeto SALSa com a anotação de *corpora* de grande extensão, geração automática de entradas lexicais a partir de *corpora* anotados e análise semântica automática influenciou vários trabalhos de criação automática de FrameNet. A proposta de Padó e Lapata (2005) sugere o uso de *corpora* paralelos para a criação automática de entradas lexicais baseadas em *frames*. A partir da anotação de um *corpus* em inglês com as etiquetas da FrameNet, seria possível transferir a anotação do *corpus* em inglês para um *corpus* de outra língua. Essa técnica vai inspirar vários trabalhos de transferência automática de anotação semântica, como Padó e Pitel (2007), para o francês, Tonelli e Pianta (2008), Dini e Bosca (2009) e Venturi et al. (2009), para o italiano.

3.3.3 A FrameNet em Domínios Específicos

Esta seção tem por objetivo abordar a construção de léxicos computacionais multilíngües baseados em *frames* para domínios especializados. A aplicação da Semântica de *Frames* e da FrameNet a domínios especializados é ainda tímida se comparada a todos os esforços que vêm sendo realizados para a criação de recursos lexicais de língua geral. Dolbey et al. (2006) propõem a criação de uma BioFrameNet, uma FrameNet especializado para o domínio da biomedicina. Uma proposta de FrameNet jurídica para o italiano está em construção por Venturi (2009). No entanto, a proposta multilíngüe que está mais bem desenvolvida neste momento é o Kicktionary (Schmidt, 2009).

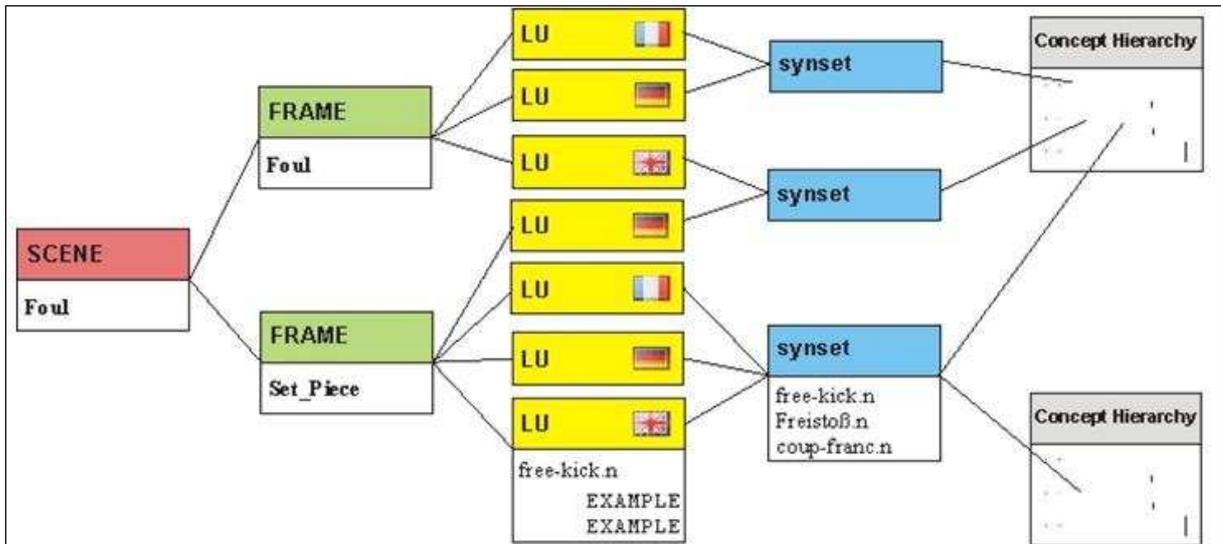


Figura 5: Estrutura do Kicktionary

O Kicktionary¹⁷ é um léxico computacional multilíngüe da linguagem do futebol (Schmidt, 2009). Diferentemente da FrameNet, que tem como unidades mínimas de descrição o *frame* e a unidade lexical, no Kicktionary as unidade mínimas de descrição são a cena, o *frame* e a unidade lexical. A distinção entre cena e *frame*, apresentada em Fillmore (1977), havia sido abandonada em trabalhos posteriores (Fillmore, 1982; 1985). A figura 5 demonstra a organização dos dados no Kicktionary.



Figura 6: Cena *Pass* e seus *frames*

O Kicktionary reativa a distinção entre cenas e *frames*. As cenas são entidades conceituais e os *frames* são entidades lingüísticas. Assim, os *frames* congregam as unidades lexicais, enquanto agrupam os *frames* que estão relacionados por cenários típicos. A figura 6

¹⁷ <http://www.kicktionary.de/index.html>

apresenta os diferentes *frames* associados à cena *Pass* (*Passe*). Pode-se ver que a cena *Pass* possui 11 *frames*, que estão divididos, segundo a sua natureza, em quatro diferentes categorias: *A pass* (*Pass*, *Pass_back*), *The path of the pass at the recipient's end* (*Mark*, *Being_free*, *Control*, *Connect*, *Flick_on*), *Intercepting a pass* (*Intercept*, *Bad_pass*) e *Others* (*Pass_combination*, *Supply_pass*). O Kicktionary também apresenta os *frames* relacionados à cena *Pass* (*Set_piece*, *Goal_kickoff*).

Segundo Schmidt (2009), o *frame* é a entidade estrutural usada para agrupar expressões lingüísticas que compartilham uma perspectiva em comum sobre uma certa cena conceitual. Enquanto a cena é definida em termos de porções de conhecimento abstrato (e possivelmente não-lingüístico), a noção de *frame* diz respeito às propriedades de meios lingüísticos concretos de expressar esse tipo de conhecimento. A figura 7 apresenta uma estrutura esquemática para a cena *Pass* e um dos *frames* relacionados a essa cena, o *frame Intercept*.

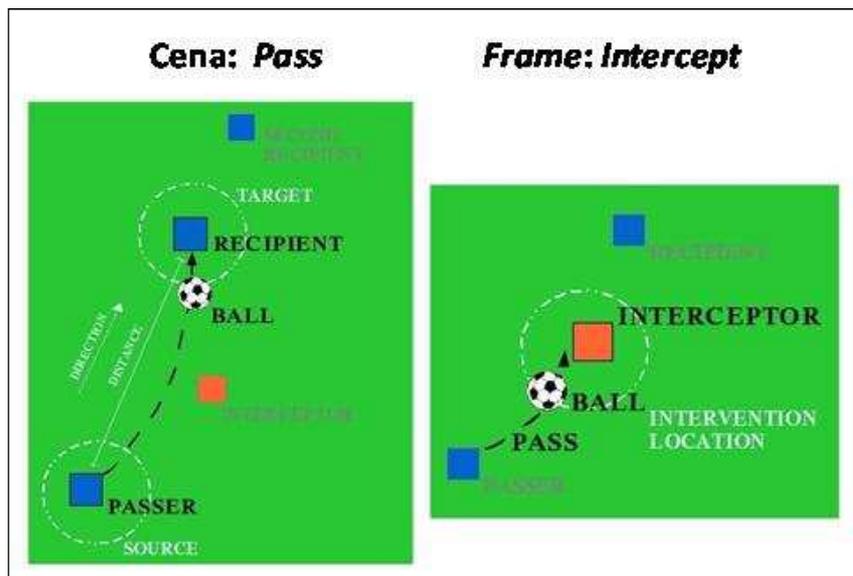


Figura 7: Cena e *frame*

A cena “*Pass*” é uma descrição mais ampla. Essa cena envolve diferentes eventos, como o passe (*frame Pass*), a interceptação do passe do adversário (*frame Intercept*), o controle da bola (*frames Mark* e *Control*). Os *frames* agrupam unidades lexicais de significado semelhante. A figura 6 apresenta a descrição das unidades lexicais do *frame Intercept*.

Intercept [Scene: Pass]

Lexical Units / Lexikalische Einheiten / Unités lexicales

■ abfangen aufnehmen sich_verschätzen unterlaufen
■ cut_out intercept interception misjudge
■ interceptor interception juger_mal

Frame-elements / Frame-Elemente / Éléments de frame

1. INTERCEPTOR
2. PASS
3. INTERVENTION_LOCATION
4. BALL
5. TARGET

*	1	2	3	4	5
abfangen.v					
aufnehmen.v					
sich_verschätzen.v					
unterlaufen.v					
cut_out.v					
intercept.v					
interception.n					
misjudge.v					
interceptor.v					
interception.n					
juger_mal.v					

Figura 8: Descrição das unidades lexicais do *frame* Intercept

Como se pode ver na figura 8, o nível lexical do Kicktionary traz as unidades lexicais equivalentes e a informação sintática para as três línguas. A tabela de realizações sintáticas marca a possibilidade de ocorrências de elementos de *frame* em cada língua. Podemos ver diferenças sintáticas significativas entre os itens lexicais do alemão, em preto, e os itens lexicais do francês, em azul, e do inglês, em vermelho. A tabela mostra que os elementos de *frame* BALL e TARGET só ocorrem em construções de itens lexicais do alemão. Esse exemplo enfatiza a importância dos trabalhos multilíngües para o processamento de linguagem natural. Atividades multilíngües envolvendo sumarização e geração de textos, além da polêmica tradução automática, têm muito a se beneficiar com trabalhos contrastivos.

O presente capítulo teve como objetivo apresentar os pilares metodológicos deste trabalho: (i) a lingüística contrastiva, (ii) a interface entre língua geral e língua de especialidade e (iii) a criação de recursos lexicais computacionais. Aqui, resume-se em que medida cada um desses três pontos se fazem presentes no próximo capítulo.

Com relação à faceta contrastiva:

- A) A lingüística contrastiva neste trabalho tem finalidade aplicada à criação de recursos lexicais multilíngües. Desde já, estão excluídas, dentre as finalidades deste trabalho, a análise contrastiva com finalidade pedagógica ou a tipologia lingüística.
- B) A equivalência das unidades lexicais é o ponto inicial e fundamental desta pesquisa. A análise dos *frames* semânticos que é apresentada no capítulo seguinte parte da identificação de equivalentes em português para as unidades lexicais em inglês.

Com relação à faceta jurídica:

- A) Grande parte das unidades lexicais (talvez todas) que são analisadas neste trabalho são termos jurídicos. No entanto, a visão adotada aqui não é a terminológica. Embora já haja trabalhos utilizando a Semântica de *Frames* para o estudo da terminologia (Verhagen e Spranger, 2006), essa não é a opção deste trabalho. Isso significa que não se está analisando termos, e sim unidades lexicais. Ou seja, parte-se da Semântica de *Frames* para realizar uma análise contrastiva de *frames* e suas unidades lexicais.
- B) Diferentes sistemas jurídicos, como a *common law* e a *civil Law*, resultam em diferentes terminologias. Esse fato leva à reflexão de quão equivalente é um equivalente de tradução no Direito.

Com relação à faceta computacional:

- A) Este trabalho se alinha a uma série de esforços na criação de recursos lexicais baseados em *frames*. A opção por um recurso lexical computacional se deve ao interesse futuro de utilização dos *frames* semânticos em processamento de linguagem natural, tal como é feito pelo projeto FrameNet ao disponibilizar ao público a sua base de dados de *frames* em formato *XML*.
- B) Apesar de vários projetos partirem de anotação *corpora* e criação automática de recursos lexicais baseados em *frames*, está-se optando pela análise manual. Com isso, objetiva-se demonstrar como os domínios especializados socialmente construídos, como o Direito, diferem de cultura para cultura. Nesses casos, os *frames* semânticos apresentam mais desencontros conceituais quando comparados em diferentes línguas.

- C) Este trabalho deve ser visto como o início de um esforço mais amplo na construção de um recurso lexical computacional para a linguagem jurídica do Brasil. O que se está propondo, a partir deste trabalho, é a criação de um recurso lexical especializado que possa ser empregado em recuperação de informação jurídica.

4 ANÁLISE CONTRASTIVA DO FRAME CRIMINAL_PROCESS

Diferentes projetos têm utilizado com sucesso métodos automatizados para criação de recursos lexicais baseados em *frames* (Padó e Lapata, 2005; Padó, 2007; Tonelli e Pianta, 2008). Esses trabalhos, no entanto, não estão focados em um domínio especializado. No caso do Direito, que é uma criação social que não existe no mundo concreto (Mattila, 2006), os *frames* jurídicos variam de uma língua para outra, não por questões lingüísticas, mas por questões culturais. Assim, a aplicação de métodos automáticos para a criação de recursos lexicais especializados, em domínios socialmente motivados como o Direito, pode não apresentar o mesmo sucesso que os resultados obtidos com a língua geral.

O objetivo deste capítulo, então, é demonstrar, a partir do estudo contrastivo do *frame Criminal_process*, em que medida os eventos jurídicos nos EUA e no Brasil são contrastantes e como essas diferenças jurídicas são representadas nos *frames* semânticos. Assim, este trabalho trata das repercussões que as diferenças entre sistemas jurídicos apresentam no desenvolvimento de léxicos computacionais jurídicos segundo o paradigma FrameNet. Para tratar desse tema, é necessário se levar em consideração os aspectos jurídico, contrastivo e computacional.

Em primeiro lugar, o conhecimento do processo penal nos EUA e no Brasil é fundamental para se entender a equivalência entre as unidades lexicais do inglês e do português. Como se pode ver no estudo contrastivo, algumas unidades lexicais do inglês possuem equivalente em português. No entanto, o conhecimento jurídico expresso por uma unidade lexical em inglês é diferente do conhecimento jurídico expresso pelo seu equivalente em português. Isso ocorre devido às diferenças entre o processo penal nos EUA e no Brasil. Assim, o primeiro passo para se realizar um estudo contrastivo com unidades lexicais do domínio jurídico é entender as diferenças nos dois sistemas que estão em contraste.

Em segundo lugar, o estudo contrastivo aqui tem o objetivo de verificar a correspondência entre os eventos jurídicos relativos ao processo penal nos EUA e no Brasil.

Esse estudo contrastivo parte da busca por equivalentes no português para as unidades lexicais do inglês. Não se trata de uma análise jurídica, e sim lingüística, tendo a Semântica de *Frames* como paradigma teórico. A análise contrastiva com finalidade pedagógica e a análise tipológica das línguas estão excluídas dos objetivos desta pesquisa.

Em terceiro lugar, este trabalho se alinha a uma série de projetos cujo objetivo é aplicar a Semântica de *Frames* e o paradigma FrameNet à criação de recursos lexicais especializados (Dolbey et al., 2006; Schmidt, 2009; Venturi, 2009). O aspecto computacional deste trabalho não está relacionado a algum método computacional utilizado nesta pesquisa, mas aos seus objetivos futuros: a criação de um recurso lexical especializado para a linguagem jurídica brasileira que possa futuramente ser utilizado para a anotação semântica de textos jurídicos.

O presente capítulo está estruturado em três seções, cada uma tratando de um desses aspectos. A seção 4.1 apresenta o aspecto jurídico deste estudo contrastivo. Assim, a subseção 4.1.1 aborda a progressão de um processo penal nos EUA e a subseção 4.1.2 a progressão de um processo penal no Brasil. Entender como se desenvolve um processo penal em ambos os países é fundamental para o entender o porquê da falta de correspondência entre os *frames* do inglês e do português. A seção 4.2.1 apresenta o estudo contrastivo do *frame* *Criminal_process* em inglês e português. Por fim, a seção 4.3 apresenta o *frame* *Processo_penal*, uma proposta de *frame* jurídico para o português do Brasil. Essa seção ilustra o tipo de informação que um futuro léxico computacional jurídico baseado em *frames* pode apresentar.

4.1 Entendendo o Processo Penal

O objetivo desta seção é esclarecer as questões jurídicas relacionadas ao processo penal que serão discutidas, diretamente e indiretamente, no estudo contrastivo do *frame* *Criminal_process*. Por serem fundamentados em sistemas jurídicos diferentes, o sistema americano na *common law* e o sistema brasileiro na *civil law*, é necessário revisar os passos de um processo penal nos EUA e no Brasil. As diferenças nos sistemas jurídicos e nos procedimentos relacionados ao processo penal resultam em falta de correspondência nos *frames* em inglês e português.

Por um motivo de organização didática, optou-se por apresentar a progressão dos processos penais americano e brasileiro em seções separadas. Assim, a subseção 4.1.1

apresenta o processo penal americano e a subseção 4.1.2 apresenta o processo penal brasileiro.

4.1.1 Etapas do Processo Penal nos Estados Unidos

A progressão de um processo penal nos Estados Unidos pode variar de um estado para outro. As etapas apresentadas aqui dizem respeito a um procedimento padrão. Segundo o *Oxford Companion to American Law*, o processo penal se inicia com a investigação do crime (2002, p.651). A partir do momento em que uma investigação policial consegue obter provas que façam acreditar que determinada pessoa cometeu um crime, a polícia tem o poder de prender essa pessoa.

A prisão ocorre quando as autoridades mantêm uma pessoa sob custódia para acusá-la por um crime. Conforme Lafave et. al. (2009), a polícia necessita obter um mandato de prisão para prender uma pessoa, mas a maioria das prisões em casos de crimes graves é realizada sem mandato. Isso se deve ao fato de que, em casos de crimes graves, a polícia não necessita de um mandato de prisão para prender um suspeito. Casos de ofensas menores em que o suspeito é pego em flagrante também não necessitam de um mandato de prisão. Conforme será apresentado na seção seguinte, na FrameNet, o *frame Arrest (prisão)* é o primeiro *subframe* do *frame Criminal_process*.

Após a prisão, é feito o *fichamento* do suspeito em uma delegacia. Esse processo envolve o registro de ocorrência, a coleta de impressões digitais e a fotografia do suspeito. Nesse momento um policial revisa as acusações contra o suspeito e decide pela continuidade da acusação. Optando por dar continuidade à acusação, o suspeito permanece preso até ser levado ao tribunal, dentro de um período de 24 a 48 horas, para a primeira aparição, quando será preenchido um documento oficial de acusação, chamado de *complaint*. Embora a polícia seja responsável pelo registro da acusação que pesa contra o suspeito, a decisão de levar adiante a acusação é do promotor público.

Em muitas jurisdições dos Estados Unidos, o promotor público revisa a acusação apresentada pela polícia e decide sobre a consistência da acusação e das provas ainda antes de o suspeito ser levado ao tribunal para a sua primeira aparição. Em outras, a polícia é responsável por preencher uma petição inicial no momento da primeira aparição, cabendo ao

promotor público revisar essa petição em um momento entre a primeira aparição e a audiência preliminar.

A primeira aparição, como é comumente chamada a apresentação do acusado diante do magistrado para a sua primeira audiência, é um procedimento breve em que o acusado é informado das acusações contra ele, os seus direitos, em especial o direito de permanecer em silêncio para não produzir provas contra si próprio, e a data da audiência preliminar. Outra função da primeira aparição é estabelecer a fiança para que o acusado possa responder ao processo em liberdade. A fiança é uma caução, geralmente em dinheiro, exigido pelo tribunal para a soltura do acusado, sob a condição de comparecer ao tribunal sempre que solicitado. A decisão sobre a fiança é representada, no *frame* `Criminal_process`, pelo *subframe* `Bail_decision`.

A etapa seguinte a primeira aparição é a audiência preliminar. A audiência preliminar é um procedimento conduzido em um tribunal de hierarquia mais baixa para verificar se há provas que permitam levar o acusado para julgamento, etapa essa que ocorre em um tribunal de hierarquia mais elevada. Segundo Renstrom (1991, p.162), *o propósito da audiência preliminar é determinar se o promotor público pode estabelecer a causa provável de se acreditar que um acusado cometeu um determinado crime*. Esta etapa do processo só ocorre se assim o acusado desejar, e tanto a acusação quanto a defesa podem chamar testemunhas para depor.

Em dezoito estados americanos, além do Distrito de Columbia e do sistema judicial federal o acusado de crimes graves tem o direito de ser denunciado somente após seu processo ser submetido à apreciação de um grande júri. A função do grande júri é verificar se as provas apresentadas pela acusação são suficientes para levar o acusado a julgamento. No caso de haver provas consideradas suficientes, o grande júri pronuncia o acusado através do preenchimento de um documento chamado *indictment*, um documento equivalente à *pronúncia* para o sistema judicial brasileiro. O *indictment* substitui o *complaint* como peça acusatória no processo. Em casos em que não há necessidade da revisão do processo pelo grande júri, um documento chamado *information* é preenchido. Tanto o *indictment* quanto o *information* substituem o *complaint* como peça acusatória, com a exceção de que o *indictment* é preenchido pelo grande júri, enquanto o *information* é preenchido pelo procurador.

Após a pronúncia do acusado, que é feita através do preenchimento do *indictment* ou do *information*, o acusado deve comparecer a uma audiência chamada de *arraignment*. A

audiência chamada de *arraignment* é representada, na FrameNet, pelo *frame* Arraignment. Nessa audiência, o acusado é informado das acusações que pesam contra ele e dos seus direitos. Essa etapa do *arraignment* é representada pelo *frame* Notification_of_charges, um *subframe* do *frame* Arraignment.

Após a notificação das acusações, o acusado é solicitado a apresentar a sua declaração de inocência ou culpa. Essa declaração é chamada de *plea*, sendo essa uma das características mais distintivas de um processo penal nos Estados Unidos. Essa declaração de inocência ou de culpa é representada, na FrameNet, pelo *frame* Entering_a_plea, um *subframe* do *frame* Arraignment.

Considerando que o acusado assuma a sua culpa, a sua declaração de culpa deverá ainda ser analisada por um tribunal, para a certificação de que o acusado está consciente das conseqüências de assumir a autoria dos crimes de que é acusado. Aceitando a declaração de culpa, o tribunal estabelece a data em que o acusado receberá a sentença. Em caso de uma declaração de inocência, o acusado é levado a julgamento. O julgamento é representado pelo *frame* Trial, que está dividido em três *subframes*.

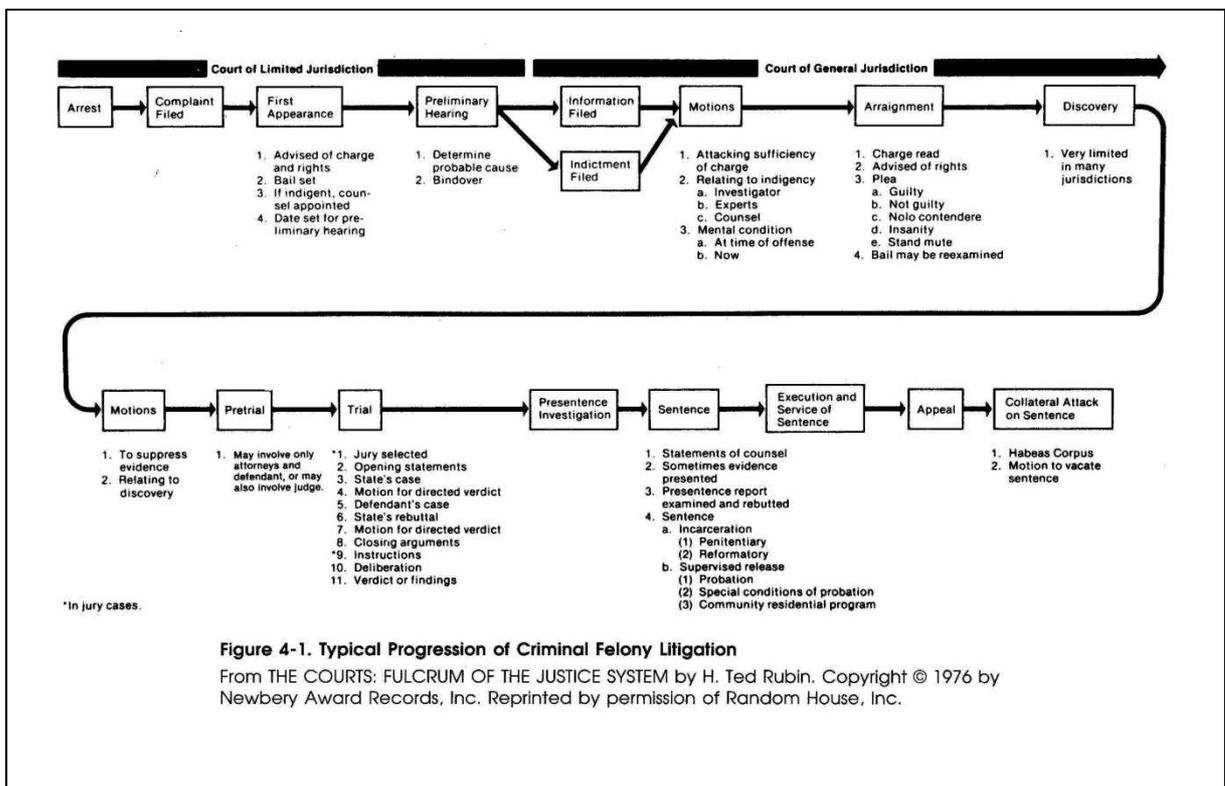


Figura 9: Progressão do Processo Penal nos EUA (Renstrom, 1991, p.121)

O artigo III, Seção 2, da Constituição dos Estados Unidos diz que o julgamento de todos os crimes, exceto em casos de impedimento, será por júri; e tal julgamento será realizado no estado onde o dito crime tenha sido cometido; (...). O julgamento inicia-se com a seleção dos jurados. Em geral, o júri é composto de 12 cidadãos, cuja função é declarar o réu culpado ou inocente. O julgamento se inicia com as alegações da defesa e da acusação. Essa primeira etapa do julgamento é descrita pelo *subframe* `Court_examination`.

No passo seguinte do julgamento, o júri recebe do juiz as instruções de como proceder para analisar o caso e os jurados se retiram para uma sala onde passam a deliberar até chegarem a um veredito. Essa etapa do processo penal americano é representada pelos *subframes* `Jury_deliberation` e `Verdict`. Sendo o réu considerado culpado, o juiz o sentencia. A etapa do sentenciamento do réu é descrita pelo *frame* `Sentencing`. O *frame* `Sentencing`, na FrameNet, encerra o processo penal, sendo esse o último *subframe* do *frame* `Criminal_process`.

Segundo o Black's Law Dictionary (1999), a sentença é *o julgamento que um tribunal pronuncia formalmente após condenar um réu; a pena imposta a um criminoso*. Cabe ao juiz impor a sentença ao réu, que em casos de crimes graves podem envolver desde prisão até pena de morte em alguns estados. No entanto, a ação do juiz é limitada pelas leis de cada jurisdição, não podendo ele agir de forma arbitrária nem abusiva na imposição de uma pena. Após a condenação, o réu ainda pode recorrer a uma corte de apelação para tentar reverter a sentença. A apelação é descrita pelo *frame* `Appeal`.

4.1.2 O Processo Penal no Brasil

O julgamento de um crime no Brasil pode seguir diferentes procedimentos. Segundo o *Código de Processo Penal*, título I, capítulo I, o procedimento pode ser comum ou especial. O procedimento comum será ordinário, sumário ou sumaríssimo. O procedimento ordinário trata de casos cuja pena máxima é igual ou superior a quatro anos. O processo sumário trata de casos em que a pena máxima seja inferior a quatro anos. Já o processo sumaríssimo trata de infrações penais de menor potencial ofensivo. O procedimento especial trata de crime de falência, crimes de responsabilidade de funcionários públicos e crimes contra propriedade imaterial. Nesta seção, são apresentadas as etapas de um processo comum de competência do Tribunal do júri.

Segundo *Constituição da República Federativa do Brasil de 1988*, artigo 5º, inciso XXXVIII, alínea d, a instituição do júri tem *a competência para o julgamento dos crimes dolosos contra a vida*. Diferentemente do processo penal americano, que parte do princípio de que todo cidadão tem o direito de ser julgado por seus pares, o júri no Brasil é adotado somente em casos de crimes dolosos contra a vida, ou seja, em que há a intenção de matar. O júri, também chamado de Conselho de Sentença, é composto de sete jurados, além do juiz, que será o presidente do Conselho de Sentença.

No momento em que a polícia toma conhecimento de uma infração penal, deverá dirigir-se ao local e assegurar-se de que a cena do crime não seja alterada. Após a perícia, é de responsabilidade da polícia apreender os objetos que tenham relação com o crime. A polícia deve também ouvir o ofendido e o suspeito, que será indiciado no inquérito policial, identificar o indiciado e verificar a sua vida pregressa. O inquérito policial deve ser remetido ao juiz competente. Se o juiz identificar a existência de um crime de ação pública, ou seja, um crime cuja competência de processar seja do Estado, o magistrado remete para o Ministério Público os documentos necessários para a apresentação da denúncia.

Após receber os autos do inquérito policial, o Ministério Público verifica a necessidade de maiores esclarecimentos sobre o inquérito e oferece denúncia dos indiciados. A denúncia, ou peça acusatória, é um documento que contém a exposição do fato criminoso, a qualificação do acusado, a classificação do crime e o rol das testemunhas. Uma vez que o Ministério Público oferece denúncia contra um acusado, ele não pode mais voltar atrás e a ação necessita correr todas as etapas de um processo penal. O oferecimento da denúncia, pelo Ministério Público no Brasil, é parcialmente equivalente ao preenchimento do *complaint*, na justiça americana.

Com exceção da prisão em flagrante, a prisão do acusado só poderá ser efetuada por ordem escrita da autoridade competente. Em alguns casos, o preso tem direito a liberdade provisória até que sua sentença seja transitada em julgado. A liberdade provisória pode ser concedida com ou sem fiança. Em caso de liberdade provisória com fiança, a fiança servirá de caução para pagar as custas processuais caso o acusado seja condenado. Sendo o acusado absolvido, o dinheiro lhe é devolvido. O sistema de concessão de liberdade provisória no Brasil é parcialmente equivalente ao americano, com a diferença de que nos EUA o acusado é apresentado a um juiz que lhe dará a fiança, enquanto no Brasil a fiança não exige uma audiência especial, sendo que, em alguns casos, o próprio delegado pode definir o valor da fiança para conceder liberdade provisória ao acusado.

Tendo o Ministério Público denunciado o acusado, o juiz competente faz a citação do réu para defesa preliminar, por escrito, em um prazo de dez dias. Apresentada a defesa, o juiz ouve a contra-resposta do Ministério Público, em um prazo de cinco dias. O juiz ainda determina a inquirição das testemunhas, a realização das diligências eventualmente requeridas pelas partes e marca a data da audiência de instrução. Essa audiência de instrução tem finalidade semelhante à audiência preliminar no processo penal americano.

O objetivo da audiência de instrução é coletar provas para a solução do caso. A audiência de instrução segue a seguinte ordem: depoimento do ofendido, depoimento das testemunhas de acusação, depoimentos das testemunhas de defesa, esclarecimentos de peritos, acareações, reconhecimentos de coisas e pessoas, qualificação e interrogatório do réu, debates e, por fim, a decisão do juiz. Caso o juiz considere a ação procedente, ele pronuncia o réu. Segundo Magalhães e Malta (1984, p.719), a pronúncia é *apreciação preliminar da prova feita pelo juiz presidente do Tribunal do Júri, em sentença, para submeter o réu posteriormente a julgamento definitivo. A pronúncia se baseia na materialidade do crime e nos indícios de autoria (...)*. A pronúncia é um procedimento que, devido à sua finalidade de verificar a procedência da acusação, é parcialmente correspondente ao *indictment* no procedimento americano.

Uma vez que o réu é pronunciado, considerando-se apenas os crimes dolosos contra a vida, ele será julgado pelo Tribunal do Júri. Talvez o julgamento seja a etapa do processo penal no Brasil em que haja maior correspondência com o processo americano, pois o Tribunal do Júri no Brasil foi inspirado pelo sistema americano. Segundo o artigo 447, do *Código de Processo Penal*, o Tribunal do Júri é composto por 1 (um) juiz togado, seu presidente e por 25 (vinte e cinco) jurados que serão sorteados dentre os alistados, 7 (sete) dos quais constituirão o Conselho de Sentença em cada sessão de julgamento. O primeiro procedimento do Tribunal do Júri é a formação do Conselho de Sentença. Uma vez escolhidos os sete jurados, eles fazem o juramento e, em seguida, recebem uma cópia da pronúncia.

O passo seguinte do Tribunal do Júri é a instrução plenária, seguindo a mesma ordem da Audiência de instrução: depoimento do ofendido, se possível, depoimento das testemunhas de acusação, depoimentos das testemunhas de defesa, esclarecimentos de peritos, acareações, reconhecimentos de coisas e pessoas, qualificação e interrogatório do réu e, por fim, os debates. Encerrada a instrução em plenário, o Conselho de Sentença se reúne para a votação. Segundo o artigo 482, do *Código de Processo Penal*, o Conselho de Sentença será *questionado sobre matéria de fato e se o acusado deve ser absolvido*. As decisões do Tribunal do Júri são tomadas por maioria de votos. Encerrada a votação e registrado o resultado do

juízo pelo escrivão em um termo, esse será assinado pelo presidente, pelos jurados e pelas partes. Em seguida, o juiz profere a sentença.

A exposição das etapas do processo penal brasileiro aqui já permite traçar algumas diferenças e semelhança entre os dois sistemas. Essa correspondência parcial entre os sistemas americano e brasileiro vai ocasionar um desencontro conceitual entre os equivalentes em inglês e português. Uma pesquisa em um dicionário jurídico bilíngüe vai indicar um equivalente para a maioria das unidades lexicais em estudo. Esse equivalente, no entanto, não significa que os *frames* evocados por essas unidades lexicais, em inglês e em português, são os mesmos. Diferenças entre os dois sistemas, como a não correspondência das etapas judiciais entre os dois procedimentos em estudo, vão ocasionar algum nível de diferença entre os *frames* jurídicos. Essa questão é abordada na próxima seção.

4.2 Estudo Contrastivo do Frame *Criminal_process*

A presente seção apresenta o estudo contrastivo das unidades lexicais relacionadas ao *frame Criminal_process*. O objetivo aqui é avaliar se as unidades lexicais do inglês e do português evocam o mesmo conhecimento dos eventos jurídicos, ou seja, se as unidades lexicais do português e do inglês evocam o mesmo *frame*. A resposta para essa questão é importante para os projetos multilíngües que utilizam as etiquetas semânticas da FrameNet e para os projetos de criação de recursos lexicais que partem dos *frames* da FrameNet para a criação automática de léxicos monolíngües baseados em *frames*. No caso deste trabalho, interessa saber o nível de compatibilidade entre os *frames* da FrameNet com os eventos jurídicos brasileiros porque desta resposta vai depender a metodologia futura empregada na criação de um recurso lexical baseado em *frames* para a linguagem jurídica brasileira.

O método de pesquisa adotado aqui pode ser dividido em três etapas, conforme Krzeszowski (1990): (i) descrição, (ii) justaposição e (iii) comparação. A etapa da descrição diz respeito à definição de o que está em comparação. Assim, estão em comparação, neste estudo, as unidades lexicais evocadoras dos *subframes* que compõem o *frame Criminal_process*. Portanto, o estudo comparativo se inicia pelas unidades lexicais para verificar se os *frames* evocados pelas unidades lexicais em português correspondem aos *frames* evocados pelas unidades lexicais em inglês.

A justaposição é a etapa em que se encontram equivalentes para os itens lexicais em comparação. Como visto no capítulo anterior, a equivalência depende, em grande parte, da

competência bilíngüe do pesquisador (Chesterman, 1998). Para tentar minimizar essa influência da competência lingüística do pesquisador, optou-se por utilizar um dicionário jurídico bilíngüe para definir quais são os equivalentes em estudo. Isso não significa que todas as opções apresentadas pelo dicionário foram seguidas indiscriminadamente. Em alguns casos, a palavra em inglês apresentava equivalentes em português que remetiam a ramos do Direito que não interessavam a este trabalho. Neste caso, restringiu-se o estudo àquelas unidades lexicais que estavam relacionadas ao contexto de um processo penal.

A etapa da comparação é o estudo contrastivo em si. Considerando que estou partindo dos *frames* semânticos da FrameNet, as unidades lexicais do inglês já estão descritas, com entradas lexicais informando os padrões valências de cada unidade lexical e exemplos anotados. Assim, o estudo contrastivo necessita encontrar exemplos de sentenças em português contendo os equivalentes das unidades lexicais em inglês e português, anotá-las e comparar com as sentenças anotadas para o inglês. Este trabalho segue a metodologia da FrameNet para anotação de sentenças (Huppenhof et al., 2010). Isso significa que serão escolhidas sentenças para exemplificar os padrões valenciais da língua, ou seja, a anotação extensiva não faz parte dos objetivos deste trabalho.

Optou-se por utilizar nesta pesquisa o *corpus* NILC¹⁸, desenvolvido pelo Núcleo Interinstitucional de Lingüística Computacional, da Universidade de São Paulo em São Carlos, contendo 32.300 milhões de palavras. Essa opção se deve ao fato de esse *corpus* ser, no momento, o recurso em português brasileiro contendo exemplos de sentenças com as unidades lexicais em estudo.

O *frame* *Criminal_process* é, segundo a terminologia da FrameNet, um *frame* não-lexical (Huppenhof et al., 2010). A função dos *frames* não-lexicais é ligar *frames* semanticamente relacionados, sem possuir unidades evocadoras de *frame*. Esse *frame* descreve os diferentes estágios de um processo penal nos EUA. Assim, o *frame* *Criminal_process* é dividido em *subframes* que se ligam através de relações temporais.

¹⁸ O *corpus* NILC pode ser acessado pelo site www.linguateca.pt.

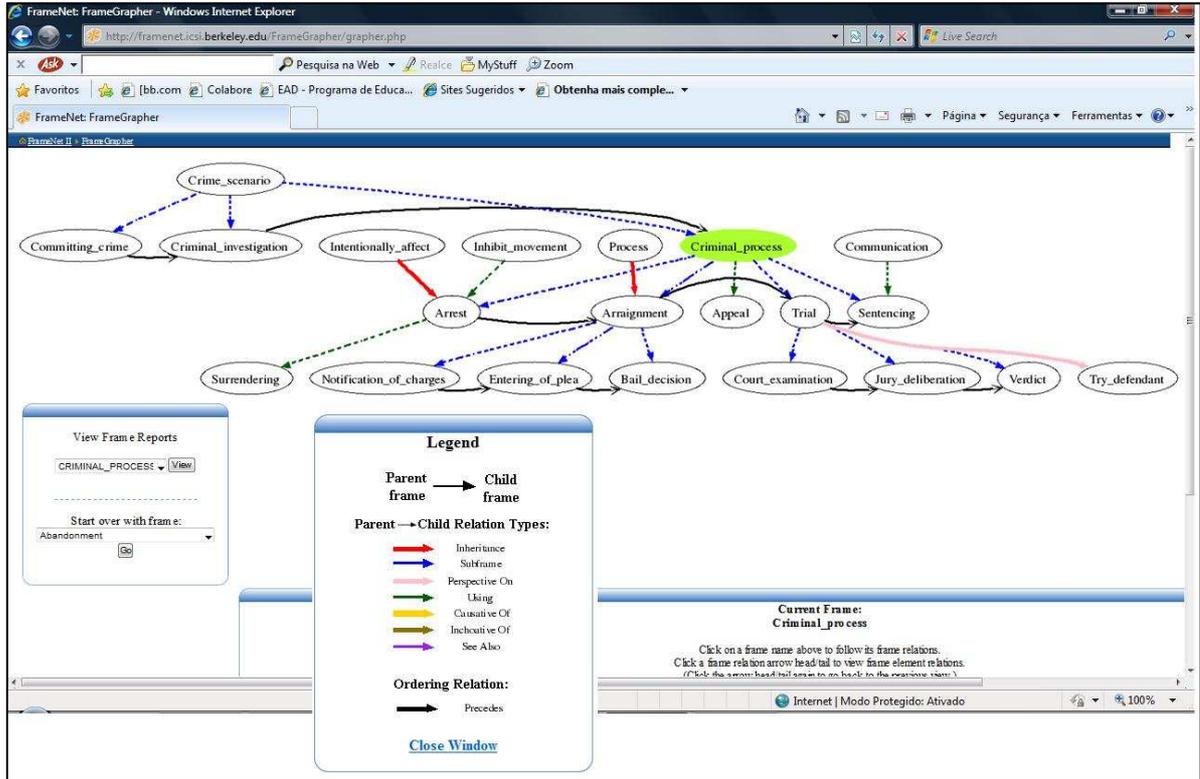


Figura 10: O *frame* Criminal_process

A figura 10 apresenta o *frame* Criminal_process e seus *subframes* relacionados através das relações *subframe* (em azul) e *precedes* (em preto). Em casos de *frames* complexos que representam uma seqüência de eventos ou de estados, cada etapa é representada como um *frame*. Esses *frames*, chamados de *subframes*, são ligados ao *frame* complexo através da relação *subframe*. A relação *precedes* ocorre entre dois *subframes* de um *frame* complexo, definindo a seqüência de determinados eventos ou estados (Huppenhof et al., 2010). Além dessas relações, a relação *perspectives_on* demonstra que os *frames* Trial e Try_defendant apresentam perspectivas diferentes de um mesmo evento e a relação *using* indica que o *frame* Criminal_process apresenta informação importante para se compreender o *frame* Appeal.

Assim, o *frame* Criminal_process está dividido em quatro *subframes* que ocorrem em ordem temporal: Arrest, Arraignment, Trial e Sentencing. Os *frames* Trial e Try_defendant, segundo a FrameNet, apresentam diferentes perspectivas de um mesmo evento. O *frame* Arraignment se divide em três *subframes*: Notification_of_charges, Entering_a_plea e Bail_decision. O *frame* Trial também apresenta três *subframes*: Court_examination, Jury_deliberation e Verdict. O estudo contrastivo se delimitará aos *frames*

relacionados pelas relações *subframe*, *precedes* e *perspective_on*. Isso dá um total de onze *frames* e suas respectivas unidades lexicais. Portanto, fica excluído deste trabalho o *frame Appeal*.

O estudo contrastivo foi dividido em cinco subseções para tratar de cada *subframe* separadamente. A subseção 4.2.1 apresenta o *frame Arrest*. A subseção 4.2.2 traz o *frame Arraignment* e seus *subframes*. A subseção 4.2.3 analisa o *frame Try_defendant*. A subseção 4.2.4 discute o *frame Trial* e seus *subframes*. Por fim, a subseção 4.2.5 aborda o *frame Sentencing*.

4.2.1 O *frame Arrest*

O *frame Arrest* descreve o evento da prisão, início de um processo penal nos Estados Unidos. A FrameNet define o *frame Arrest* como: “AUTORIDADES acusam um SUSPEITO, que está sob suspeita de ter cometido um crime e levam-no sob custódia”. Este *frame* traz como elementos de *frame* centrais AUTORIDADES, ACUSAÇÕES, OFENSA e SUSPEITO. Na FrameNet, o crime pode receber o nome de OFENSA ou ACUSAÇÕES. OFENSA é o elemento de *frame* geral que descreve um crime. ACUSAÇÕES representa o nome que o crime recebe judicialmente, no processo de investigação, denúncia e julgamento do suspeito. O exemplo (1), coletado da FrameNet, exemplifica o *frame Arrest*:

(1) [The police AUTORIDADES] **arrested** [Harry SUSPEITO] [on charges of manslaughter ACUSAÇÕES].

A FrameNet apresenta onze unidades lexicais evocadoras do *frame Arrest*: *apprehend.v*, *apprehension.n*, *arrest.n*, *arrest.v*, *book.v*, *bust.n*, *bust.v*, *collar.v*, *cop.v*, *nab.v*, *summons.v*. A tabela 1 apresenta os equivalentes que o *Noronha Dicionário Jurídico* (Goyos Júnior, 1992) traz para essas onze unidades lexicais. As entradas do dicionário apresentam os equivalentes em grupos de sinônimos. O que as entradas das unidades lexicais mostram é que os padrões de polissemia entre o inglês e o português, para essas unidades lexicais, são divergentes (Altenberg e Granger, 2002). Ou seja, os itens lexicais em inglês e português apresentam diferentes extensões de significado nessas duas línguas (Ver discussão na seção 3.1.2). Para comparar as unidades lexicais do português com as unidades lexicais do inglês,

foram escolhidos a unidade lexical *arrest.v* e os três equivalentes em português apresentados pelo *Noronha Dicionário Jurídico*: *prender*, *deter* e *capturar*.

Inglês	Português	Inglês
<i>apprehend.v</i>	<i>apreender</i>	<i>to arrest; seize; apprehend; learn.</i>
	<i>prender</i>	<i>to arrest; imprison; incarcerate.</i>
	<i>deter</i>	<i>to detain; intercept; arrest.</i>
<i>apprehension.n</i>	<i>apreensão</i>	<i>seizure; apprehension; arrest; capture.</i>
	<i>prisão</i>	<i>prison; arrest; imprisonment;</i>
	<i>detenção</i>	<i>detention; withholding; confinement.</i>
<i>arrest.n</i>	--	--
<i>arrest.v</i>	<i>prender</i>	
	<i>deter</i>	
	<i>capturar</i>	<i>to capture; seize; arrest; conquer; occupy.</i>
<i>book.v</i>	<i>reservar lugar</i>	--
	<i>tomar nota do nome</i>	--
<i>bust.n</i>	--	--
<i>bust.v</i>	--	--
<i>collar.v</i>	--	--
<i>cop.v</i>	--	--
<i>nab.v</i>	--	--
<i>summons.v</i>	<i>citar</i>	<i>to quote; name; mention; cite; appear in court; summon; subpoena.</i>
	<i>convocar</i>	<i>to call; to summon.</i>
	<i>intimar</i>	<i>to notify.</i>
	<i>chamar</i>	<i>to call; to summon.</i>

Tabela 1: Unidades Lexicais Evocadoras do *Frame Arrest*

Os exemplos a seguir demonstram anotações realizadas pela FrameNet tendo o verbo *to arrest* como evocador do *frame Arrest*. O exemplo (2) apresenta o elemento de *frame* OFENSA, enquanto o exemplo (3) apresenta as ACUSAÇÕES:

- (2) Prior to this, when [police AUTORIDADES] **arrested** [a man SUSPEITO] [for assaulting his partner, OFENSA], the onus was on the woman to bring about a prosecution.

- (3) When he gave them his name [they AUTORIDADES] **arrested** [him SUSPEITO] [on a charge of rape ACUSAÇÃO].

Em português, tendo o verbo *prender* como evocador, apresentam-se exemplos como os seguintes:

- (4) [México AUTORIDADES] **prende** [segundo suspeito SUSPEITO] [pela morte do candidato Colosio OFENSA].
- (5) [Polícia AUTORIDADES] **prende** [libanês suspeito SUSPEITO] [de atentado contra jovens judeus OFENSA].

Os exemplos com o verbo *deter* como evocador são mais diversificados com relação aos elementos de *frame* que apareceram nos exemplos anotados:

- (6) [A polícia AUTORIDADES] **deteve** [ontem TEMPO] [em Gravatá (83 km de Recife) LUGAR] [três pessoas SUSPEITO] [acusadas de aliciar eleitores com idades entre os 40 e 45 anos ACUSAÇÕES].
- (7) [A Polícia Militar de São Paulo AUTORIDADES] **deteve** [ontem TEMPO] [32 pessoas SUSPEITOS] [durante blitz MODOS] [no morro do bairro cota 95, no km 52 da rodovia Anchieta, em Cubatão (62 km a sudeste de SP) LUGAR].

O último equivalente escolhido para esta análise contrastiva é *capturar*.

- (8) Até ontem, [a polícia AUTORIDADES] ainda não havia **capturado** [os assaltantes SUSPEITO].
- (9) Em 13 de outubro, [ele SUSPEITO] foi **capturado** [por policiais AUTORIDADES] e trazido a Brasília algemado.

As unidades lexicais *prender*, *deter* e *capturar* se comportam de forma semelhante ao seu equivalente em inglês, *arrest.v*, sendo a posição de sujeito preenchida preferencialmente por um sintagma nominal apresentando como elemento de *frame*

AUTORIDADES, o objeto verbal sendo preenchido por um sintagma nominal apresentando como elemento de *frame* SUSPEITO e os elementos de *frame* OFENSA ou ACUSAÇÕES ocorrendo como um sintagma preposicional. Em casos de uso da voz passiva, o elemento de *frame* SUSPEITO passa para a posição de sujeito, sendo o elemento de *frame* AUTORIDADES expresso através de um sintagma preposicionado.

O *frame* Arrest descreve, a partir das unidades lexicais da língua inglesa, um evento em que autoridades prendem um suspeito de ter cometido um crime. As unidades lexicais em português possuem um alto nível de similaridade de significado com as unidades lexicais em inglês, tanto que os elementos de *frame* utilizados para anotar as sentenças em português são os mesmo utilizados para anotar as sentenças em inglês.

Se considerarmos o *frame* semântico como uma estrutura cognitiva que organiza um determinado conhecimento, pode-se dizer, então, que o conhecimento no contexto norte-americano é similar ao conhecimento que se tem no Brasil sobre prisão. No Brasil, assim como nos Estados Unidos, a polícia é responsável pela investigação dos crimes e pela apresentação dos suspeitos à justiça, para que sejam julgados pela autoridade competente. Nesse sentido, o conhecimento organizado pelo *frame* Arrest em inglês é correspondente ao português, tanto que é possível utilizar os mesmo elementos de *frame* para anotar sentenças nas duas línguas.

Ao se anotarem as sentenças coletadas do *corpus*, percebe-se que a similaridade de significado das unidades lexicais em inglês e português permite o uso do mesmo *frame* para a descrição semântica. Isso permitiria concluir que o *frame* Arrest em inglês poderia ser transposto para o português sem maiores dificuldades, seja para a criação de recursos lexicais multilíngües ou para a anotação de *corpora*. Tamanha compatibilidade não é encontrada no próximo *frame* apresentado.

4.2.2 O *frame* Arraignment e seus *subframes*

O *frame* Arraignment sucede o *frame* Arrest. O *arraignment* é uma etapa do processo penal americano que não encontra correspondente no processo penal brasileiro. Como foi abordado na subseção 4.1.1, nessa audiência, o réu ouve a leitura das acusações pelas quais ele irá responder em juízo e declara a sua culpa ou a sua inocência. A audiência preliminar (*preliminary hearing*), etapa que precede o *arraignment* (cfe. figura 9), não é descrita pelo *frame* Criminal_process. O *frame* Arraignment é uma combinação de

duas etapas diversas do processo penal: primeira aparição (*first appearance*) e *arraignment*. O *subframe* *Entering_a_plea* remete à etapa judicial do *arraignment*, o *subframe* *Bail_decision* remete à etapa da primeira aparição e o *subframe* *Notification_of_charges* remete tanto à primeira aparição quanto ao *arraignment*.

Uma explicação para essa quebra entre as etapas de um processo judicial penal e as etapas desse processo descritas pela FrameNet é a orientação semasiológica da FrameNet. A FrameNet descreve os conceitos que são evocados por unidades lexicais. Além disso, a FrameNet não é um recurso lexical especializado. Assim sendo, unidades lexicais que evocam *frames* especificamente jurídicos e que não fazem parte do cotidiano do cidadão podem acabar não sendo descritas. Outra justificativa para essa diferença é o fato de a progressão de um processo penal nos Estados Unidos poder variar de estado para estado.

A FrameNet define o *frame* *Arraignment* como: “Em uma audiência no tribunal, um RÉU é informado das ACUSAÇÕES contra ele, (geralmente) dá sua RESPOSTA de inocência ou culpa e o JUIZ decide o montante da fiança, caso haja fiança”. O *frame* *Arraignment* possui como elementos de *frame* centrais ACUSAÇÕES, RÉU e JUIZ (cfe. exemplo 10). Não há nenhum evento semelhante no processo penal brasileiro com o *arraignment*.

- (10) [He_{RÉU}] was **arraigned** [on eight counts of treason_{ACUSAÇÕES}] [before Sir Travers Humphreys_{JUIZ}] [at the Old Bailey_{LUGAR}] [on 28 November 1945_{TEMPO}].

As unidades lexicais evocadoras do *frame* *Arraignment* são *arraign.v* e *arraignment.n*. A inexistência do evento jurídico faz com que as unidades lexicais relacionadas a tal evento não possuam equivalentes em português. A tabela 2 mostra como o *Noronha Dicionário Jurídico* aborda a questão da falta de equivalência para as unidades lexicais *arraign* e *arraignment*.

Como se pode ver na tabela 2, o *Noronha Dicionário Jurídico* não apresenta equivalentes para a unidade lexical *arraignment*. No lugar de equivalentes, o dicionário explica o que é um *arraignment*. A unidade lexical *arraign.v* apresenta três equivalentes, *acusar*, *denunciar* e *processar*. No entanto, quando se procura por essas palavras no dicionário, *denunciar* e *processar* não apresentam como equivalente o verbo *to arraign*. O verbo *acusar*, por sua vez, é um verbo geral podendo ser usado em diferentes etapas do processo penal, sem poder ser conectado a apenas uma etapa do processo penal.

Inglês	Português	Inglês
<i>arraign.v</i>	<i>acusar</i>	<i>to accuse; prosecute; charge; indict; incriminate; blame; arraign; renounce; acknowledge.</i>
	<i>denunciar</i>	<i>to denounce; accuse; inform against; report; proclaim.</i>
	<i>processar</i>	<i>to sue; proceed.</i>
<i>arraignment.n</i>	<i>leitura do libelo na presença do acusado.</i>	--

Tabela 2: Unidades Lexicais Evocadoras do *Frame Arraignment*

O verbo *acusar* pode até ser considerado como um equivalente parcial do verbo *to arraign*, visto que a etapa do processo penal americano chamada de *arraignment* não apresenta correspondência com nenhuma etapa do processo penal brasileiro. O fato de o verbo *acusar* ser apresentado como equivalente pode estar relacionado a uma das etapas do *arraignment*, que é a leitura das acusações pelas quais o réu responde. No entanto, a etapa do processo penal brasileiro que corresponde à acusação do suspeito, ou seja, quando o Ministério Público denuncia o acusado, não é equivalente ao *arraignment*. Esse fato leva à análise dos *subframes* que compõem o *frame Arraignment*. O *frame Arraignment* se divide em três *subframes*.

Frame Notification_of_charges: *Notification_of_charges* descreve a etapa inicial da primeira aparição e do *arraignment*, em que o juiz informa o acusado das acusações que pesam contra ele. Conforme a FrameNet define o *frame Notification_of_charges*, “a AUTORIDADE, o juiz ou outro oficial do tribunal, informa o ACUSADO das ACUSAÇÕES contra ele”. Os elementos de *frame* centrais deste *frame* são: ACUSADO, AUTORIDADE e ACUSAÇÃO. O exemplo (11) apresenta a anotação de uma sentença coletada na FrameNet.

- (11) A witness at the trial of [two men _{ACUSADO}] **accused** [of the murder of a police informer _{ACUSAÇÕES}] has claimed that one defendent drove the victim to a meeting with the other, who them killed him. [DNI _{AUTORIDADE}]

Esse *frame* descreve a parte inicial de um *arraignment*, quando um juiz lê para o réu as acusações que pesam contra ele. Segundo a FrameNet, as unidades lexicais que evocam esse *frame* são: *accuse.v*, *charge.n*, *charge.v*, *indict.v*, *indictment.n*. Embora o significado dessas cinco unidades lexicais seja diferente do significado de *arraign.v* e *arraignment.n*, pode-se ver, conforme a tabela 3, que o *Noronha Dicionário Jurídico* aponta vários equivalentes semelhantes ou iguais para as unidades lexicais evocadoras dos *frames* *Arraignment* e *Notification_of_charges*.

Inglês	Português	Inglês
<i>accuse.v</i>	<i>acusar</i>	<i>to accuse; prosecute; charge; indict; incriminate; blame; arraign; renounce; acknowledge.</i>
	<i>Denunciar</i>	<i>to denounce; accuse; inform against; report; proclaim.</i>
	<i>Despesa</i>	
	<i>Encargo</i>	
<i>charge.n</i>	<i>Gravame</i>	
	<i>Responsável</i>	
	<i>acusação</i>	<i>accusation; charge; incrimination; denunciation; prosecution; indictment.</i>
<i>charge.v</i>	<i>Debitar</i>	
	<i>Onerar</i>	
	<i>Carregar</i>	
	<i>acusar</i>	
<i>indict.v</i>	<i>pronunciar</i>	<i>to indict; arraign.</i>
	<i>acusar</i>	
	<i>Denunciar</i>	
<i>indictment.n</i>	<i>pronúncia</i>	<i>indictment; arraignment.</i>

Tabela 3: Unidades Lexicais Evocadoras do *Frame* *Notification_of_Charges*

Essa imprecisão em encontrar um equivalente apropriado se deve à não correspondência entre as etapas do processo penal nos Estados Unidos e no Brasil. Tanto o verbo *to arraign* (cfe. tabela 2) quanto o verbo *to indict* (cfe. tabela 3) apontam como equivalente o verbo *acusar*. Quando se procura o equivalente de *acusar* encontra-se, entre outros verbos, *to indict* e *to arraign*. No entanto, o *indictment* é uma fase anterior ao *arraignment*.

O *Noronha Dicionário Jurídico* apresenta como equivalentes da unidade lexical *to indict* três unidades lexicais: *pronunciar*, *acusar* e *denunciar*. No entanto, como se viu na subseção 4.1.2, essas unidades lexicais têm sentidos diferentes no contexto legal brasileiro. A acusação é algo mais genérico, qualquer pessoa pode acusar alguém. A denúncia é o documento pelo qual o Ministério Público dá início à ação penal, somente o Ministério Público pode denunciar. Já a pronúncia é a sentença preliminar apresentada por um juiz presidente do Tribunal do Júri para submeter um réu posteriormente a julgamento definitivo, ou seja, só um juiz pode pronunciar um acusado. Como se pode ver, esses três verbos são usados em etapas diferentes do processo penal no Brasil.

Apesar de as unidades lexicais do *frame* *Notification_of_charges* possuírem equivalentes em português, o evento jurídico descrito pelo *frame* *Notification_of_charges* não possui correspondência completa com o processo penal brasileiro. Isso se deve, em primeiro lugar, ao fato de o *frame* *Notification_of_Charges* ser *subframe* de um *frame* que não encontra correspondência no processo penal brasileiro. Em segundo lugar, as etapas do processo penal brasileiro que sucedem a prisão do suspeito até o seu julgamento não são compatíveis com o processo penal americano.

Em geral, no Brasil, é apresentada denúncia contra o acusado e após a denúncia é realizada a audiência de instrução para coletar provas. A partir das provas coletadas na audiência de instrução, o juiz faz a pronúncia do acusado e, como réu, ele vai para o julgamento. Nos EUA o acusado é levado ao juiz pela primeira vez para receber a fiança, comparecendo posteriormente para audiência preliminar. Após a audiência preliminar, o processo é revisado por um para um grande júri, que faz o pronúncia o réu. Após o réu ser pronunciado pelo grande júri, ele deve comparecer ao *arraignment*, que é uma audiência em que lhe são informadas as acusações contra ele. Somente após o *arraignment* o réu é levado a julgamento.

O *frame* *Notification_of_charges* necessita ser desdobrado em dois *frames* para representar corretamente o processo penal brasileiro: Denúncia e Pronúncia. No *frame* *Denúncia*, as unidades lexicais evocadoras de *frame* seriam *acusar.v*, *acusação.n*, *denunciar.v*, e *denúncia.n*. Já no *frame* *Pronúncia*, as unidades lexicais evocadoras de *frame* seriam *pronúncia.n* e *pronunciar.v*. Assim, de nada adiantaria tentar anotar sentenças em português com os elementos de *frame* apresentados pela FrameNet para o *frame*

Notification_of_charges, pois seria necessário a criação de dois grupos diferentes de etiquetas para os *frames* Acusação e Pronúncia¹⁹.

Frame Entering_a_plea: Segundo a FrameNet, no *frame* Entering_a_plea, “o ACUSADO dá sua RESPOSTA de inocência ou culpa em reação às ACUSAÇÕES. Em muitas jurisdições, as afirmações possíveis são inocente, culpado ou *nolo contendere*. Pessoas acusadas podem permanecer em silêncio, nesse caso, o tribunal geralmente faz a afirmação da culpa em nome do acusado”. Os elementos de *frame* centrais são: ACUSADO, ACUSAÇÕES, TRIBUNAL, JUIZ e RESPOSTA. O exemplo (12) demonstra uma anotação da FrameNet para o *frame* Entering_a_plea.

(12) [Simpson ACUSADO] **pleaded** [not guilty RESPOSTA] [to the murder of his wife ACUSAÇÕES].

Inglês	Português	Inglês
<i>plea.n</i>	<i>contestação</i>	<i>plea; answer; appearance by an answer; rebutting; defendant's initial reply; defendant's pleading.</i>
	<i>resposta</i>	<i>answer; reply; response.</i>
	<i>argüição</i>	<i>challenge; arguing; argumentation; censure; pleading.</i>
	<i>defesa</i>	<i>defense; answer; plea; protection; justification; defender.</i>
<i>plead.v</i>	<i>postular</i>	<i>to plead; request; demand; postulate.</i>
	<i>contestar</i>	<i>to contest; dispute; controvert; deny; question; challenge.</i>
	<i>defender</i>	<i>to defend; protect; prohibit.</i>
	<i>Responder</i>	--
	<i>pleitear</i>	<i>to litigate; sue; plead.</i>
	<i>demandar</i>	<i>to demand; claim; file suit; implead; sue.</i>
	<i>argüir</i>	<i>to allege; challenge; object.</i>
	<i>pedir</i>	<i>to ask; demand; claim; request; order.</i>
	<i>requer</i>	<i>to formulate a petition; apply; petition.</i>

Tabela 4: Unidades Lexicais Evocadoras do *Frame* Entering_a_plea

¹⁹ A seção 4.3 apresenta um esboço de como seria um *frame* representativo do processo penal brasileiro.

Esse *frame* descreve a segunda etapa do *frame* Arraignment, após o acusado ser informado das acusações. Nesta etapa, o acusado é solicitado a declarar-se inocente ou culpado. Não há no processo penal brasileiro nenhuma etapa correspondente a esta. Segundo a FrameNet, as unidades lexicais que evocam esse *frame* são: *plea.n*, *plead.v*. As entradas do *Noronha Dicionário Jurídico* para *plea* e *to plead* demonstram a falta de equivalência. Como não há no processo penal brasileiro evento similar, é impossível traduzir o termo do inglês. Assim, a opção adotada pelo dicionário é elencar uma série de palavras de natureza semântica similar. Como *plea* é a confissão formal, diante do juiz, de culpa ou afirmação da inocência, o dicionário busca como equivalentes palavras relacionadas à comunicação verbal, como *resposta*, *arguição*, *contestar*, *pedir*.

Enquanto o *frame* Notification_of_charges possui uma determinada similaridade em inglês e português, visto que tanto no processo penal americano como no brasileiro há o evento jurídico da acusação de um suspeito, o *frame* Entering_a_plea não apresenta nenhuma correspondência. As suas unidades lexicais, *plea* e *to plead*, não encontram equivalentes em português, necessitando serem parafraseadas com outras palavras relacionadas à comunicação. Assim, considerando-se a criação de um *frame* do processo penal brasileiro, o *frame* Entering_a_plea teria de ser totalmente desconsiderado.

Frame Bail_decision: Segundo a FrameNet, “este *frame* cobre uma etapa do processo penal em que o JUIZ decide se o ACUSADO deve ser libertado sob fiança e, em caso afirmativo, qual o montante da fiança”. O *frame* Bail_decision apresenta como elementos de *frame* centrais o ACUSADO, o JUIZ e o STATUS. O elemento de *frame* STATUS indica os termos de concessão da fiança, sendo essa com depósito de dinheiro, sem depósito de dinheiro e o valor do depósito. O exemplo (13) traz uma das sentenças anotadas pela FrameNet para o *frame* Bail_decision.

(13) [He_{ACUSADO}] was *granted* [conditional_{STATUS}] **bail** [by Teesside Magistrates_{JUIZ}].

As duas unidades lexicais evocadoras do *frame* Bail_decision são *bail.n* e *bond.n*. Segundo o *Noronha Dicionário Jurídico* (cfe. tabela 5), a unidade lexical *fiança* está entre os equivalentes possíveis. Embora exista a equivalência lexical, o procedimento pelo qual o acusado de um crime recebe fiança nos Estados Unidos é diferente do Brasil. Enquanto nos Estados Unidos o acusado recebe a decisão sobre a fiança na audiência chamada de

primeira aparição, no Brasil a mesma autoridade que autoriza a prisão, seja o delegado ou o juiz, já deve informar o acusado das condições de responder ao processo penal em liberdade provisória. Essa liberdade pode ser concedida com ou sem fiança, dependendo da natureza do crime.

Inglês	Português	Inglês
	<i>fiança</i>	<i>surety; bond; guaranty; security.</i>
<i>bail.n</i>	<i>caução</i>	<i>pledge; collateral; bond; guarantee; guaranty; securitization; bail.</i>
	<i>título (bônus) de dívida geralmente com garantia</i>	--
	<i>apólice</i>	<i>bond; bill; policy.</i>
<i>bond.n</i>	<i>obrigação</i>	<i>obligation; duty; task; bond; bill; liability; debenture; charge.</i>
	<i>Fiança</i>	
	<i>garantia</i>	<i>guarantee; warrant; plevin; guaranty; surety; warranty.</i>
	<i>Caução</i>	
	<i>debênture</i>	<i>Debênture</i>
<i>fix.v</i>		
<i>order.v</i>		
<i>set.v</i>		

Tabela 5: Unidades Lexicais Evocadoras do *Frame Bail_decision*

Assim, apesar de haver equivalência entre as unidades lexicais, o *frame Bail_decision* corresponde apenas parcialmente ao procedimento de fiança no Brasil. Dessa forma, considerando que a audiência de *arraignment* não existe no Brasil, e que o *frame Arraignment* não faz parte do *frame Processo_penal* brasileiro, o *frame Bail_decision* necessita de algumas alterações referentes aos seus elementos de *frame* e ser deslocado do lugar onde está no *frame Criminal_process*, para um novo lugar no *frame Processo_penal*.

Seguindo as mesmas etapas do processo judicial penal (cfe. figura 9), o *Arraignment* precede o *frame Trial*. Os *frames Trial* e *Try_defendant* se encontram em relação de perspectiva. Enquanto o *frame Trial* é um *frame* geral que traz as etapas de um julgamento, o *frame Try_defendant* apresenta o evento do julgamento.

4.2.3 O *frame Try_defendant*

O *frame Try_defendant* é definido pela FrameNet como “um RÉU é julgado por um JÚRI ou JUIZ em um TRIBUNAL por ACUSAÇÕES. Os elementos de *frame* centrais são: ACUSAÇÕES, RÉU, AUTORIDADE_GOVERNAMENTAL, JUIZ e JÚRI. O elemento de *frame* AUTORIDADE_GOVERNAMENTAL representa a autoridade, em geral o país, que promulgou as leis que serão aplicadas à punição do réu. O exemplo (14) mostra uma sentença anotada pela FrameNet para o *frame Try_defendant*.

- (14) [The court_{JUIZ}] would **try** [people_{RÉU}] [for war crimes, crimes against humanity and other serious human-rights violations_{ACUSAÇÕES}].

Segundo a definição da FrameNet, o *frame Try_defendant* descreve uma situação em que um RÉU é julgado por um JÚRI ou um JUIZ, em um TRIBUNAL, por ACUSAÇÕES contra ele. O evento jurídico descrito por esse *frame* corresponde ao julgamento do réu no processo penal brasileiro. Assim como nos Estados Unidos, no Brasil as acusações contra alguém são decididas em um tribunal por um juiz e, em casos de crimes dolosos contra a vida, por um júri.

Essa correspondência de sistema jurídico é acompanhada de equivalência entre as unidades lexicais. Como se pode ver na tabela 6, o *Noronha Dicionário Jurídico* apresenta apenas um equivalente para o único evocador de *frame* apresentado pela FrameNet para o *frame Try_defendant*.

Inglês	Português	Inglês
<i>try.v</i>	<i>julgar</i>	<i>to judge; try a case; adjudge; decide; decree.</i>

Tabela 6: Unidades Lexicais Evocadoras do *Frame Try_defendant*

A correspondência entre os eventos jurídicos é expressa através da compatibilidade dos elementos de *frame* em inglês e português. Ao se anotarem sentenças em português, percebe-se que os elementos de *frame* se ajustam perfeitamente à descrição do evento de um julgamento no Brasil. Os exemplos da FrameNet incluem:

(15) [“He RÉU] will be **tried** [in court TRIBUNAL] [for his actions,” ACUSAÇÕES] he added.
[CNI JUIZ]

(16) [The soldiers RÉU] were expected to be **tried** [by a military court JUIZ] [next
month. TEMPO] [DNI ACUSAÇÕES]

Em português, percebe-se a semelhança não só semântica, mas também sintática das sentenças com o verbo *julgar*:

(17) [Amin Elou RÉU] foi **julgado** [por júri popular JUIZ].

(18) [Paiakan e Irekran RÉU] não serão **julgados** [por um júri popular JUIZ] porque as acusações são de crimes contra a honra.

4.2.4 O *frame Trial* e seus *subframes*

O *frame Trial* representa a etapa do julgamento no processo judicial penal. Este *frame* se divide em três *subframes* relacionados através da relação *precedes*: *Court_examination*, *Jury_deliberation* e *Verdict*. Segundo a definição da FrameNet, “um julgamento é um procedimento conduzido em um TRIBUNAL diante de um JUIZ e, geralmente, um JÚRI, quando os participantes se encontram para decidir um CASO. O julgamento é uma disputa entre duas partes em que “a PROMOTORIA tenta provar que o RÉU é culpado das ACUSAÇÕES e a DEFESA tenta provar que o RÉU é inocente”. Os elementos de *frame* centrais de: CASO, ACUSAÇÕES, TRIBUNAL, RÉU, DEFESA, JUIZ, JÚRI e PROMOTORIA. O exemplo (19) apresenta uma sentença anotada pela FrameNet.

(19) Indeed, the [**case** CASO] [against the soldier RÉU] involved rests primarily on his evidence, not hers.

Assim como o *frame Try_defendant*, o evento jurídico representado pelo *frame Trial* também apresenta correspondência com o evento jurídico brasileiro. Segundo a definição da FrameNet, o julgamento é um procedimento conduzido em um tribunal, diante de um JUIZ e (geralmente) um JÚRI, em que os participantes se encontram para decidir um CASO.

O julgamento nos EUA e no Brasil encontra correspondência por dois motivos. Em primeiro lugar, pela função semelhante, tanto nos Estados Unidos quanto no Brasil, de arbitrar disputas comerciais e pessoais e de aplicar a lei em casos de crimes. Em segundo lugar, o processo penal brasileiro se inspirou no processo penal americano ao utilizar o júri para julgar casos de crimes dolosos contra a vida. Considerando que o Direito é uma criação social e que cada sociedade tem suas próprias leis e procedimentos jurídicos, as semelhanças entre os procedimentos judiciais, ou seja, a forma como os tribunais procedem, indicam que, em algum ponto da história, esses sistemas tiveram alguma interação.

Inglês	Português	Inglês
<i>case.n</i>	<i>caso</i>	<i>case; event; occurrence; matter; file.</i>
	<i>ação judicial</i>	<i>case; legal action; action at law; suit; legal suit; cause; legal cause; judicial proceeding.</i>
	<i>processo judicial</i>	<i>action at law; judicial proceeding; suit.</i>
	<i>lide</i>	<i>suit; action; judicial proceedings; cause.</i>
	<i>litígio</i>	<i>litigation; lawsuit; dispute.</i>
<i>trial.n</i>	<i>juízo</i>	<i>judgment; court session; trial; verdict; decision; sentence.</i>
	<i>processo</i>	<i>process; proceeding; procedure; cause; suit.</i>
	<i>juízo</i>	<i>court; curia; trial; lawsuit; action; process; discernment; reason; opinion.</i>
	<i>Prova</i>	
	<i>Teste</i>	

Tabela 7: Unidades Lexicais Evocadoras do *Frame Trial*

Essa correspondência entre os sistemas jurídicos se expressa também na equivalência das unidades lexicais. Segundo a FrameNet, as unidades lexicais evocadoras do *frame Trial* são *case.n* e *trial.n*. Como se pode ver na tabela 7, as unidades lexicais evocadoras do *frame Trial* possuem equivalentes em português. Este é mais um caso de polissemia divergente (Altenberg e Granger, 2002), em que um item lexical em uma língua encontra diversos itens lexicais como equivalentes na outra língua. Exemplos de sentenças coletadas em *corpus* com esses equivalentes e anotados com os elementos de *frame* do *frame Trial* demonstram a equivalência semântica dessas unidades lexicais nas duas línguas. Os exemplos (20) e (21) apresentam sentenças anotadas pela FrameNet.

(20) One of the marks of a competent lawyer is his ability to know what gaps there are in the facts of [his DEFESA] [**case** CASO].

(21) He had been awaiting [**trial** CASO] [at the High Court TRIBUNAL] [on assault, robbery and firearms charges. ACUSAÇÕES]

Os exemplos (22), (23), (24) e (25) trazem sentenças anotadas para o português.

(22) Juntos, os soldados Valter da Silva Leme e Joel Celestino são réus em 15 **processos** [por lesão corporal, tentativa de homicídio e homicídio ACUSAÇÕES].

(23) Em levantamento feito na 1ª Vara de Homocídios de Recife, o Centro Luiz Freire identificou 50 **processos** [contra supostos matadores RÉU], abertos entre 1991 e 1993.

(24) O recurso pode provocar, em 95, um novo **juízo** [dos acusados RÉU] [pelos desembargadores do Tribunal de Justiça do Estado. PROMOTORIA]

(25) O réu que, por não argüir na sua resposta fato impeditivo, modificativo ou extintivo do direito do autor, dilatar o **juízo** [da lide CASO], será condenado nas custas [1] a partir do saneamento do processo e perderá, ainda que vencedor na causa, o direito a haver do vencido honorários advocatícios.

As etapas de um julgamento nos EUA e no Brasil também são correspondentes, porém com algumas diferenças com relação aos procedimentos adotados pelos tribunais. O *frame* Trial está dividido em três *subframes*: Court_examination, Jury_deliberation e Veredict. Esses *frames* serão analisados a seguir.

Frame Court_examination: O *frame* Court_examination corresponde à parte de depoimentos de um julgamento. Segundo a FrameNet, “um INQUIRIDOR pergunta a uma TESTEMUNHA (ou a um réu), perguntas no contexto de um procedimento de tribunal”. A FrameNet não representa com seus *frames* semânticos o processo de escolha das testemunhas,

que, na verdade, é o procedimento inicial de um julgamento. Os elementos de *frame* centrais são INQUIRIDOR e TESTEMUNHA, conforme o exemplo apresentado pela FrameNet (26).

(26) [Kempler_{TESTEMUNHA}] was also **cross-examined** [by two new defense attorneys, Gabi Shaha and Shmuel Flishman._{INQUIRIDOR}]

O *frame* Court_examination corresponde à fase do julgamento em que defesa e acusação apresentam as suas argüições e as testemunhas depõem. Essa fase, no processo penal brasileiro, é conhecida como instrução em plenário. Embora o evento jurídico representado pelo *frame* Court_examination tenha um correspondente no processo penal brasileiro, algumas unidades lexicais em inglês não possuem equivalentes em português. A tabela 8 apresenta os equivalentes listados pelo *Noronha Dicionário Jurídico*.

Inglês	Português	Inglês
<i>cross-examination.n</i>	<i>interrogatório da testemunha em julgamento pela parte contrária.</i>	
<i>cross-examine.v</i>	--	
<i>cross.n</i>	--	
<i>examination.n</i>	<i>exame</i>	<i>examination; test; analysis; inspection.</i>
	<i>inspeção</i>	<i>inspection.</i>
	<i>perícia</i>	<i>examination; audit; check.</i>
	<i>inquérito</i>	<i>Inquiry</i>
<i>examine.v</i>	<i>examinar</i>	<i>to examine; search; investigate; discuss.</i>
	<i>reconhecer</i>	<i>to recognize; acknowledge; admit; avow; affirm; accept.</i>
	<i>Registrar</i>	--
	<i>interrogar</i>	<i>to interrogate; examine; question; inquiry; cross-examine.</i>
	<i>fazer perícia</i>	

Tabela 8: Unidades Lexicais Evocadoras do *Frame* Court_examination

As unidades lexicais *cross.n*, *cross-examination.n* e *cross-examine.v* não encontram em português uma unidade lexical capaz de traduzir o seu significado. Ou seja, o conceito em inglês é representado através de uma única unidade lexical, mas não em português. Apenas *examine.v* e *examination.n* encontram equivalente em português.

No entanto, ao anotarem-se sentenças em português, percebe-se a necessidade de se ampliar os elementos de *frame* apresentados pela FrameNet. O *frame* *Court_examination* apresenta apenas dois elementos de *frame*, INQUIRIDOR e TESTEMUNHA, sendo o elemento de *frame* TESTEMUNHA utilizado tanto para a testemunha como para o réu. Em português, segundo os exemplos do *corpus*, o verbo *interrogar*, considerado equivalente de *to examine*, é utilizado preferencialmente para o réu, enquanto o verbo *depor* é utilizado preferencialmente para a testemunha. Assim, faz-se necessário criar também um elemento de *frame* ACUSADO. O exemplo (27) demonstra um sentença anotada pela FrameNet.

(27) Should [the parties _{INQUIRIDOR}] be allowed to **examine** and cross-examine [witnesses? _{TESTEMUNHA}]

O exemplo (28) demonstra um exemplo de sentença contendo o elemento de *frame* ACUSADO e o exemplo (29) traz o elemento de *frame* TESTEMUNHA, conforme a lista original de elementos de *frame* da FrameNet.

(28) [Sete traficantes -- entre eles Romildo de Souza Costa, o Miltinho do Dendê; Ernaldo Pinto de Medeiros, o Uê; e Odir dos Santos _{ACUSADO}] -- foram **interrogados** [ontem _{TEMPO}] [no 3º Tribunal do Júri, _{LUGAR}] [sob acusação de terem matado oito traficantes, na Vila do João, em Bonsucesso, em março de 1995. _{ACUSAÇÕES}]

(29) [Principal testemunha da chacina _{TESTEMUNHA}] **depõe** [no II Tribunal do Júri _{LUGAR}] reafirma denúncias e diz que Emanuel mentiu ao inocentar Côrtes.

O que se pode depreender da análise do *frame* *Court_examination* é a necessidade de se revisarem os elementos de *frame* propostos pela FrameNet. Por questões de semântica da língua, em português as unidades lexicais *interrogar* e *depor* evocam o mesmo *frame*, porém em perspectivas diferentes. Enquanto o verbo *interrogar* tende a ser utilizado para o RÉU, o verbo *depor* tende a ser utilizado para a TESTEMUNHA.

Frame Jury_deliberation: O *frame* *Jury_deliberation* descreve a etapa em que os jurados se reúnem para discutir o veredito. Segundo a definição apresentada

pela FrameNet, “o JÚRI discute o CASO e as ACUSAÇÕES para avaliar a POSSÍVEL_SENTENÇA do acusado. Os elementos de *frame* centrais são: JÚRI, CASO, ACUSAÇÕES e POSSÍVEL_SENTENÇA, como se pode ver na sentença anotada pela FrameNet (30):

(30) He plans to send [jurors JÚRI] out to **deliberate** [their verdicts on Monday POSSÍVEL_SENTENÇA].

O *frame* *Jury_deliberation* apresenta uma situação oposta ao *frame* *Court_examination*. As unidades lexicais apresentam equivalência, conforme se pode ver na tabela 9. No entanto, a decisão do júri no Brasil não é feita através de deliberação dos jurados, mas de votação. Assim, as unidades lexicais são equivalentes, pois possuem significado similar em ambas as línguas. No entanto, essa etapa do evento jurídico americano não encontra correspondência no evento jurídico brasileiro, visto que o evento representado por este *frame* não existe no processo penal brasileiro.

Inglês	Português	Inglês
<i>deliberate.v</i>	<i>deliberar</i>	<i>to deliberate; consider; consult; decide; resolve; determine.</i>
<i>deliberation.n</i>	<i>deliberação</i>	<i>deliberation; determination; resolution; decision.</i>
	<i>consideração</i>	<i>esteem; respect.</i>
	<i>reflexão</i>	--

Tabela 9: Unidades Lexicais Evocadoras do *Frame* *Jury_deliberation*

Frame Verdict: Assim como o veredito é a etapa final de um julgamento nos Estados Unidos, o *frame* *Verdict* é o último *subframe* do *frame* *Trial*. Conforme a FrameNet, “um JUIZ dá um VEREDITO sobre uma ACUSAÇÃO contra um RÉU em um CASO”. O JUIZ, neste contexto, pode ser tanto o juiz quanto o júri. Os elementos de *frame* centrais deste *frame* são: CASO, ACUSAÇÕES, RÉU, VEREDITO e JUIZ. O exemplo (31) traz uma sentença anotada pela FrameNet.

(31) [An Old Bailey jury that included nine women JÚRI] **found** [Sheila Beeson, 29, RÉU] [not guilty of child cruelty VEREDITO].

Inglês	Português	Inglês
<i>acquit.v</i>	<i>absolver</i>	<i>to acquit; dismiss; discharge; absolve.</i>
	<i>remir</i>	
	<i>quitar</i>	
	<i>saldar</i>	
	<i>desobrigar</i>	
<i>acquittal.n</i>	<i>absolvição</i>	<i>absolution; acquittal; discharge.</i>
	<i>perdão</i>	<i>acquittal; condonation; pardon.</i>
	<i>remissão</i>	<i>remission; pardon; forgiveness.</i>
<i>clear.v</i>	<i>compensar (cheque)</i>	
	<i>apurar (lucro líquido)</i>	
<i>convict.v</i>	<i>condenar</i>	<i>to condemn; declare guilty; convict; sentence.</i>
	<i>sentenciar</i>	<i>to decide; pass judgement; sentence.</i>
<i>conviction.n</i>	<i>condenação</i>	<i>conviction; judgement against; sentence.</i>
<i>find.v</i>	<i>encontrar</i>	
	<i>achar</i>	
	<i>descobrir</i>	
	<i>determinar</i>	
<i>finding.n</i>	<i>veredito</i>	<i>decision; resolution; verdict; judgement.</i>
	<i>decisão</i>	
<i>guilty.a</i>	<i>culpado</i>	<i>Guilty</i>
<i>not_guilty.a</i>	<i>inocente</i>	<i>innocent; guiltless.</i>
<i>pronounce.v</i>	<i>pronunciar</i>	<i>to indict; arraign.</i>
	<i>declarar</i>	<i>to declare; announce; proclaim; manifest.</i>
<i>ruling.n</i>	<i>decisão em vigor</i>	
	<i>vigorante</i>	
	<i>vigente</i>	
<i>verdict.n</i>	<i>veredito</i>	<i>verdict; sentence.</i>
	<i>decisão do júri</i>	

Tabela 10: Unidades Lexicais Evocadoras do *Frame Verdict*

O *frame Verdict* apresenta uma longa lista de unidades lexicais indicando a decisão dos jurados sobre um caso. Todas essas unidades lexicais, como se pode ver na tabela 10, apresentam um equivalente em português. O evento jurídico representado pelo *frame*

Verdict apresenta correspondência entre os sistemas norte-americano e brasileiro e as unidades lexicais desse *frame* apresentam equivalentes em português.

Conforme a definição da FrameNet, no *frame* Verdict “um JUÍZ dá um VEREDITO a respeito das ACUSAÇÕES contra um RÉU em um CASO particular”. Esse evento é o mesmo no processo penal brasileiro. Dessa forma, todos os elementos de *frame* do *frame* em inglês podem ser utilizados em português.

Os exemplos (32), (33), (34) e (35) ilustram o uso dos elementos de *frame* para anotação de sentenças em português.

- (32) [O réu Élvio Olivar do Prado Miranda, 21, RÉU] [acusado de tentativa de homicídio ACUSAÇÕES], foi [absolvido VEREDITO] [por 4 votos a 3 CIRCUNSTÂNCIA].
- (33) [O júri JUÍZ] [a RÉU] **considerou** [não culpada VEREDITO] [por motivo de insanidade temporária. RAZÃO]
- (34) [O júri JUÍZ] **decidiu** [por quatro votos a três CIRCUNSTÂNCIA] [culpar VEREDITO] [Divino RÉU] [por homicídio qualificado ACUSAÇÕES].
- (35) [Sem provas RAZÃO], [os seis acusados RÉU] foram [inocentados VEREDITO].

4.2.5 O *frame* Sentencing

A sentença é o último passo do processo penal, cabendo apenas a apelação a um tribunal superior para tentar reverter a decisão. Conforme a FrameNet define o *frame* Sentencing, “uma SENTENÇA, normalmente uma punição, é imposta a um CONDENADO por um TRIBUNAL, geralmente representado por um juiz”. Os elementos de *frame* centrais, neste caso, são: CONDENADO, TRIBUNAL, OFENSA, SENTENÇA e TERMO_DA_SENTENÇA. O elemento de *frame* TERMO_DA_SENTENÇA denota a duração da pena. O exemplo (36) traz a anotação apresentada pela FrameNet:

- (36) [A court in Bootle TRIBUNAL] **sentenced** [Fields CONDENADO] [on July 11 TEMPO] [to 60 days' imprisonment SENTENÇA] [for refusing to pay the community charge or "poll tax" OFENSA].

Assim como o *Verdict*, o evento jurídico representado pelo *frame Sentencing* em inglês é correspondente ao evento jurídico em português. Como se pode ver na tabela 11, a FrameNet apresenta cinco unidades lexicais evocadoras do *frame Sentencing*.

Inglês	Português	Inglês
<i>condemn.v</i>	<i>condenar</i>	<i>to condemn; declare guilty; convict; sentence.</i>
	<i>desaprovar</i>	--
<i>order.v</i>	<i>Ordenar</i>	--
<i>send up.v</i>	--	
<i>sentence.n</i>	<i>sentença</i>	<i>sentence; judgement; decision; award; finding; finding of fact; adjudication; administrative adjudication; administrative determination.</i>
	<i>veredicto</i>	<i>verdict; sentence.</i>
	<i>decisão</i>	<i>decision; resolution; verdict; judgement.</i>
<i>sentence.v</i>	--	

Tabela 11: Unidades Lexicais Evocadoras do *Frame Sentencing*

Os exemplos (37) e (38) trazem sentenças anotadas com o verbo *condenar* como evocador de *frame*:

(37) [Ele _{CONDENADO}] também foi **condenado** [por tentativa de homicídio contra Orlando Canuto _{OFENSA}], que sobreviveu à chacina.

(38) [Ubirajara _{CONDENADO}] foi **condenado** [a 19 anos _{SENTENÇA}] [para cada homicídio _{OFENSA}] e [a 12 anos _{SENTENÇA}] [pela tentativa de homicídio de Orlando _{OFENSA}].

Esta seção analisou a possibilidade de se empregarem os *frames* semânticos da FrameNet para a descrição da linguagem jurídica brasileira. A comparação entre as unidades lexicais em inglês e os equivalentes apresentados por um dicionário jurídico bilíngüe teve como objetivo verificar a correspondência dos eventos jurídicos norte-americanos com os eventos jurídicos brasileiros.

A pressuposição que impulsionou esta comparação foi a de que se as unidades lexicais evocam *frames* e se as unidades lexicais em inglês possuem equivalentes em

português, então aquelas unidades lexicais que possuem equivalentes devem apresentar correspondência de *frames* semânticos também. No entanto, os resultados não foram tão homogêneos como se previa inicialmente.

O *frame* Arrest apresentou correspondência em inglês e português. As unidades lexicais do *frame* Arrest apresentam equivalentes em português, o evento descrito pelo *frame* é semelhante nos EUA e no Brasil e os elementos de *frame* podem ser utilizados sem problemas para a anotação das sentenças em português. Isso leva à conclusão de que se esse *frame* fosse transposto do inglês para o português, através de um método automático, não apresentaria problemas de correspondência.

O *frame* Arraignment, no entanto, não apresenta correspondência. As unidades lexicais que evocam este *frame* não possuem equivalentes e o evento jurídico americano não encontra correspondente no sistema brasileiro. Considerando que o *frame* Arraignment é um *frame* complexo, os seus *subframes* apresentam diferentes graus de correspondência com o português.

As unidades lexicais do *frame* Notification_of_charges apresentam equivalentes em português, no entanto, o evento jurídico representado pelo *frame* não corresponde ao português. Como o *frame* Notification_of_charges é uma etapa do *frame* Arraignment, o evento de acusação em português necessita ser deslocado para uma outra posição no *frame* Processo_penal brasileiro. O *frame* Entering_a_plea, que é o evento jurídico central do *frame* Arraignment, não possui correspondência em português. As unidades lexicais evocadoras do *frame* Entering_a_plea não possuem equivalentes em português, nem o evento jurídico possui correspondente no processo penal brasileiro. As unidades lexicais do *frame* Bail_decision, assim como as do *frame* Notification_of_charges, apresentam equivalentes em português, porém o evento jurídico não é correspondente com o processo penal brasileiro.

O *frame* Try_defendant apresenta um alto nível de correspondência em português. O evento jurídico é correspondente e a unidade lexical que evoca esse *frame* possui apenas um único equivalente, segundo o dicionário bilíngüe consultado. Esse fato sugere que, quanto maior o grau de correspondência entre os sistemas, maior a equivalência e a semelhança entre os *frames* das línguas em comparação. Correspondência semelhante é observada nos *frames* seguintes.

As unidades lexicais do *frame* Trial apresentam equivalentes em português, o evento jurídico representado pelo *frame* em inglês é semelhante ao português e os elementos de *frame* podem ser utilizados para a anotação de sentenças em português. Assim como o *frame* Try_defendant, o *frame* Trial apresenta um alto grau de correspondência em português. É preciso lembrar que ambos os *frames* estão, segundo a FrameNet, em relação de perspectiva, ou seja, representam o mesmo evento, mas de pontos de vista diferentes.

O *frame* Trial se divide em três *subframes*: Court_examination, Jury_deliberation e verdict. A etapa do julgamento representada pelo *frame* Court_examination é equivalente ao julgamento no Brasil. No entanto, por questões de morfologia do inglês, algumas unidades lexicais em inglês não possuem equivalentes no português. As unidades lexicais do *frame* Jury_deliberation possuem equivalentes em português, porém essas unidades lexicais não evocam, em português, um conhecimento jurídico como no inglês. Essa falta de correspondência se deve ao fato de o evento jurídico representado pelo *frame* Jury_deliberation não ter correspondente no processo penal brasileiro. Por fim, o *frame* verdict apresenta correspondência entre unidades lexicais e eventos jurídicos.

O *frame* Sentencing é o fim do processo penal. As unidades lexicais desse *frame* apresentam equivalente em português. Também o evento jurídico representando por este *frame* é correspondente com o processo penal brasileiro. Essas observações levam a dois pontos a serem ainda desenvolvidos neste trabalho: (i) o *frame* Processo_penal em português e (ii) a correspondência entre *frames*. A seção seguinte apresentará um esboço de como seria um *frame* para o processo penal brasileiro levando em consideração o estudo contrastivo. Já a correspondência entre *frames* será o tema do próximo capítulo.

4.3 O *frame* Processo_penal

Esta seção apresenta um estudo preliminar, com base no estudo contrastivo, de como seria o *frame* Processo_penal brasileiro em um léxico computacional jurídico baseado em *frames*. Assim, apresenta-se aqui a organização dos *frames* e das unidades lexicais evocadoras de *frames*. O foco da atenção nesta seção será o agrupamento das unidades lexicais segundo o *frame* por elas evocado, a definição dos *frames* jurídicos, o reconhecimento dos elementos de *frame* centrais de cada *frame* jurídico e anotação de

exemplos sentenças. O estudo dos padrões valências das unidades lexicais não será apresentado neste trabalho por exceder os seus objetivos.

A anotação de sentenças apresentada aqui não é extensiva, e sim seletiva. Foram selecionadas apenas aquelas sentenças que melhor exemplificam a anotação de sentenças com os elementos de *frame*. Como a anotação não é extensiva, não foram criadas entradas lexicais para as unidades lexicais elencadas aqui. A criação dessas entradas lexicais sem um estudo mais aprofundado das ocorrências sintáticas dessas unidades poderia levar à criação de entradas lexicais distorcidas. Assim, desde já aponta-se a ampliação da anotação de sentenças apresentada aqui e a criação de entradas lexicais como um trabalho futuro.

Baseado no estudo das etapas do processo penal brasileiro, conforme subseção 4.1.2, e no estudo contrastivo das unidades lexicais relacionadas ao processo penal, conforme seção 4.2, propõe-se, nesta seção, um *frame* *Processo_penal* dividido em cinco *subframes*: *Prisão*, *Denúncia*, *Audiência_de_instrução*, *Pronúncia* e *Julgamento*. O *Julgamento* está dividido em três etapas: *Instrução*, *Veredito* e *Sentença*. Assim como na FrameNet, o *frame* *Julgamento* está em relação de perspectiva com o *frame* *Julgar_acusado*.

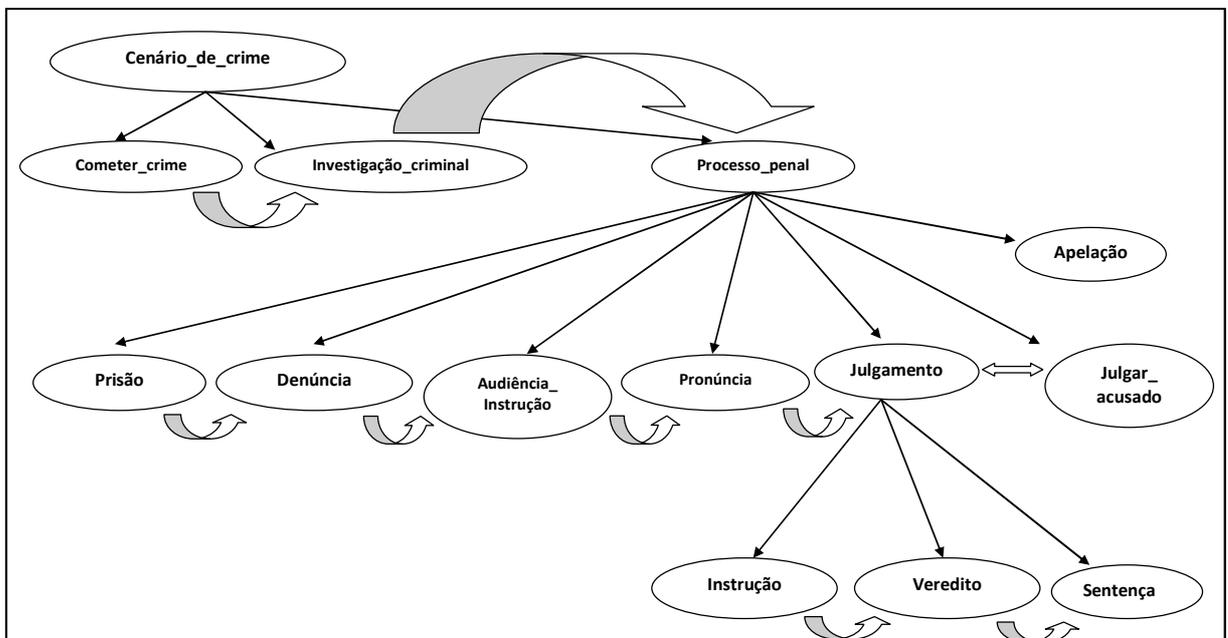


Figura 11: O *frame* *Processo_penal*

Apesar de os *frames* apresentados aqui serem uma proposta preliminar, necessitando ainda uma ampliação da anotação e uma revisão das definições dos *frames* e dos elementos de

frames, essa seção já ilustra o desencontro conceitual entre os processos penais norte-americano e brasileiro.

O *frame* proposto é baseado no *frame* *Criminal_process*. Como no *frame* *Criminal_process*, o *frame* *Processo_penal* é um *frame* não lexical, seu objetivo é organizar os diferentes eventos de um processo penal. Assim, as cinco diferentes etapas listadas para o processo penal formam *frames* independentes relacionados através da relação *precedes*. O *frame* *Julgamento*, por sua vez, é um *frame* complexo, que está dividido em três *frames* também relacionados através da relação *precedes*. As seções seguintes apresentam o estudo preliminar desses *frames*.

4.3.1 O *frame* **Prisão**

O *frame* *Prisão* descreve um ato em que *AUTORIDADES* privam um *SUSPEITO* da liberdade por *ACUSAÇÕES* contra ele. Os elementos de *frame* centrais a este *frame* são: *AUTORIDADES*, *SUSPEITO*, *OFENSA*, *ACUSAÇÕES*. As unidades lexicais evocadoras deste *frame* são: *prender*, *prisão*, *fichar*, *deter*, *capturar*, *em cana*. As sentenças anotadas a seguir exemplificam as ocorrências de cada unidade lexical e os elementos de *frame* que ocorrem com cada um delas.

Os exemplos (39) e (40) trazem ocorrências com o verbo *prender*:

(39) [França *AUTORIDADES*] **prende** [95 suspeitos *SUSPEITO*] [de colaboração com terror argelino. *OFENSA*]

(40) [México *AUTORIDADES*] **prende** [principal suspeito *SUSPEITO*] [de planejar assassinato de Ruiz. *OFENSA*]

Os exemplos (41) e (42) exemplificam o verbo *deter* como evocador do *frame* *Prisão*:

(41) [A PF *AUTORIDADES*] também **deteve** [para interrogatório *PROPÓSITO*] [o major reformado Luiz Montenaro, da PM (Polícia Militar) do Rio. *SUSPEITO*]

(42) [A polícia AUTORIDADES] **deteve** [mais de 40 pessoas SUSPEITO] [por badernas, embriaguez e uso de drogas OFENSA] [na pedreira Paulo Leminski, em Curitiba. LUGAR]

Os exemplos (43) e (44) trazem as sentenças com o verbo *capturar*:

(43) Até ontem, [a polícia AUTORIDADES] ainda não havia **capturado** [os assaltantes. SUSPEITO]

(44) Em 13 de outubro, [ele SUSPEITO] foi **capturado** [por policiais AUTORIDADES] e trazido a Brasília algemado.

Os exemplos (45) e (46) apresentam sentenças com o verbo *fichar*:

(45) [O economista Pécio Arida, que deve assumir a presidência do Banco Central, SUSPEITO] foi **fichado** [pelo Deops AUTORIDADES] [com os codinomes de Renato, Abel ou Daniel. MODO]

(46) Nomes como Bebezão, Pão com Ovo e Uê designam [a nova geração da criminalidade SUSPEITO] **fichada** [na polícia do Rio. AUTORIDADES]

Os exemplos (47) e (48) trazem exemplos de sentenças anotadas com o substantivo *prisão*:

(47) Mas a **prisão** [por consumo OFENSA] é um contrasenso, uma contradição que dificulta qualquer política, disse.

(48) A **prisão** [de Maria Alice Queiroz SUSPEITO] aconteceu na manhã de ontem, em Cabo Frio.

Por fim, os exemplos (49) e (50) apresentam exemplos com a construção *em cana*:

(49) [Em país com Proálcool, LUGAR] [usineiro SUSPEITO] não vai **em cana**.

(50) Vai ter ensaio, sim, nem que [eu SUSPEITO] vá **em cana**, disse Ivo

O *frame* *Prisão* é constituído por um número expressivo de evocadores de *frames*, indo desde verbos até construções como *em cana*. Em um procedimento típico, após a prisão do suspeito, é realizada a denúncia do acusado pelo Ministério Público. Assim, o *frame* seguinte é a Denúncia.

4.3.2 O *frame* Denúncia

O *frame* Denúncia representa o ato pelo qual a AUTORIDADE_DE_ACUSAÇÃO, representada pelo Ministério Público, responsabiliza o ACUSADO por ACUSAÇÕES. Os elementos de *frame* centrais desse *frame* são ACUSADO, AUTORIDADE_DE_ACUSAÇÃO e ACUSAÇÕES. No caso deste *frame*, também são importantes os elementos de *frame* periféricos: AUTORIDADE_JULGADORA, DENÚNCIA, EVIDÊNCIAS, BASE_LEGAL e PROCESSO. As unidades lexicais deste *frame* são *acusar*, *acusação*, *denunciar* e *denúncia*.

Os exemplos (51) e (52) apresentam a anotação de sentenças com o verbo *acusar*:

(51) [Estes suspeitos ACUSADO] podem ser **acusados** [pelos promotores AUTORIDADE_DE_ACUSAÇÃO] [em denúncia a ser apresentada, DENÚNCIA] [na próxima semana, TEMPO] [ao juiz José Geraldo Antonio AUTORIDADE_JULGADORA]

(52) Chefe do Estado Maior do serviço secreto, [Jairo ACUSADO] é **acusado** [pela promotora Léa Pelegrino AUTORIDADE_DE_ACUSAÇÃO] [de planejar o assassinato do ex-capitão e bicheiro José de Souza Oliveira Filho, que escreveu um diário denunciando o envolvimento da Polícia Militar, políticos e magistrados com a máfia do jogo do bicho no estado ACUSAÇÕES]

Já as sentenças (53) e (54) trazem exemplos com o substantivo *acusação*:

(53) Argumentou o juiz que leu ainda as **acusações** [que pesam contra o cabo ACUSADO] [de ser integrante da segurança de traficantes e ter participado de assaltos a três farmácias. ACUSAÇÕES]

- (54) [Os responsáveis AUTORIDADE_DE_ACUSAÇÃO] pela **acusação** [no processo que investiga a chacina da Candelária PROCESSO] estudam a possibilidade de pedir o adiamento do julgamento dos PMs Marcelo Côrtes e Cláudio Luiz dos Santos e do serralheiro Jurandir Gomes, marcado para o próximo dia 27 .

Os exemplos (55) e (56) trazem sentenças anotadas para o verbo *denunciar*:

- (55) [A partir desses documentos, EVIDÊNCIAS] [o Ministério Público AUTORIDADE_DE_ACUSAÇÃO] **denunciou** [os bicheiros ACUSADO] novamente e ficou comprovado que eles mantinham suas atividades mesmo de trás das grades
- (56) [O empresário ACUSADO] foi **denunciado** [em março TEMPO] [pelo procurador da República Artur Gueiros, AUTORIDADE_DE_ACUSAÇÃO] [com base na Lei 1.533/51, que cataloga crimes contra a economia popular. BASE_LEGAL]

Os exemplos (57) e (58) exemplificam sentenças com o substantivo *denúncia*:

- (57) Dentro de cinco dias, baseado em novas investigações, [o Ministério Público AUTORIDADE_DE_ACUSAÇÃO] apresentará **denúncia** [contra novos acusados. ACUSADO]
- (58) Novos indiciados -- Piñeiro confirmou ontem que outros nomes de policiais podem ser incluídos na nova **denúncia** [que o Ministério Público fará AUTORIDADE_DE_ACUSAÇÃO] [contra Cunha, Alcântara, Carlos Jorge Liaffa Coelho e Nilton de Oliveira, que foram presos dias antes do julgamento do PM Marcos Vinícius Emanuel. ACUSADO]

Neste *frame* o que chama a atenção é a alta incidência de elementos de *frame* periféricos, ao contrário do *frame* *Prisão*. Após a denúncia ser apresentada pelo Ministério Público, o acusado é chamado para a audiência de instrução. Nessa audiência, ele apresenta o seu depoimento sobre o caso. O *frame* *Audiência_de_instrução* é o próximo *frame* a ser apresentado.

4.3.3 O *frame* Audiência_de_instrução

O *frame* *Audiência_de_instrução* representa a audiência preliminar em que um JUIZ ouve o ACUSADO de um crime com o objetivo de decidir sobre o prosseguimento do processo. São elementos de *frame* centrais JUIZ, ACUSADO, TESTEMUNHAS e ACUSAÇÕES. As unidades lexicais evocadora deste *frame* são *interrogar* e *depor*.

Os exemplos (59), (60), (61) e (62) trazem exemplos de sentenças anotadas para os verbos *interrogar* e *depor*:

- (59) [A juíza Giselle Lima e Silva Rocha, da 36^a Vara Criminal, JUIZ] **interrogou** [ontem TEMPO] [nove dos 14 bicheiros ACUSADO] [condenados por formação de quadrilha no ano passado. ACUSAÇÕES]
- (60) **Interrogado** [o réu, ACUSADO] ser-lhe-á concedido, se o requerer, o prazo de três dias para apresentar defesa, arrolar testemunhas até o máximo de três e requerer diligências.
- (61) [Eles ACUSADO] foram denunciados pelo Ministério Público e deverão ser **interrogados** no dia 30 de maio.
- (62) [As testemunhas TESTEMUNHAS] **depõem**, na audiência de instrução, perante o juiz da causa, exceto:

Aqui é importante uma observação: as unidades lexicais evocadoras do *frame* *Audiência_preliminar* são as mesmas evocadoras do *frame* *Instrução*. O que diferencia uma situação da outra é o contexto. No exemplo (61) é possível reconhecer que se trata de um evocador do *frame* *Audiência_preliminar* devido à sucessão de eventos expressa pela sentença. Já no exemplo (62) o termo *audiência de instrução* aparece na própria sentença. Depois de realizada a audiência de instrução, o juiz emite o seu parecer pronunciando ou não o acusado. Assim sendo, o próximo *frame* é a *Pronúncia*.

4.3.4 O *frame* Pronúncia

O *frame* Pronúncia descreve o evento jurídico em que o JUIZ presidente do Tribunal do Júri faz a apreciação preliminar das provas, em sentença, para submeter o RÉU posteriormente a julgamento. Os elementos de *frame* centrais são JUIZ e RÉU. As unidades lexicais evocadoras do *frame* Pronúncia são *pronúncia* e *pronunciar*.

Os exemplos (63) e (64) apresentam sentenças anotadas com o verbo *pronunciar*:

(63) Nos termos do que decidiu o STF, não é necessária a prova incontroversa da existência do crime para que [o réu RÉU] seja **pronunciado**.

(64) [O juiz JUIZ] deve **pronunciar** [o réu RÉU] (TJSP, RCrim 71.325, RT 648 / 275).

Já os exemplos (65) e (66) apresentam exemplos de sentenças anotadas para o substantivo *pronúncia*:

(65) Suponha-se que morra a vítima da tentativa de homicídio após a **pronúncia** [do réu. RÉU]

(66) Defensor dativo que se manifesta pela **pronúncia** [do réu. RÉU]

Sendo o réu pronunciado, ele será julgado. O *frame* Julgamento é o próximo *frame* a ser descrito.

4.3.5 O *frame* Julgamento e seus *subframes*

O *frame* Julgamento descreve o evento jurídico em que um JUIZ, presidente do Tribunal do Júri, e um corpo de jurados, o JÚRI, devem decidir sobre a culpa ou inocência de um RÉU. A PROCURADORIA tenta provar a culpa e a DEFESA tenta provar a inocência do RÉU. Os elementos de *frame* centrais a este *frame* são JUIZ, JÚRI, PROCURADOR, RÉU, DEFESA, ACUSAÇÕES, TRIBUNAL e AÇÃO. As unidades lexicais evocadoras de *frame* são *juulgamento*, *processo* e *ação penal*.

Os exemplos (67) e (68) trazem a anotação de sentenças com a unidade lexical *juízo*:

(67) O réu que, por não argüir na sua resposta fato impeditivo, modificativo ou extintivo do direito do autor, dilatar o **juízo** [da lide, ACÇÃO] será condenado nas custas [1] a partir do saneamento do processo e perderá, ainda que vencedor na causa, o direito a haver do vencido honorários advocatícios.

(68) O recurso pode provocar, em 95, um novo **juízo** [dos acusados RÉU] [pelos desembargadores do Tribunal de Justiça do Estado. JUIZ]

Os exemplos (69) e (70) apresentam as sentenças anotadas com a unidade lexical *processo*:

(69) Em levantamento feito na 1ª Vara de Homocídios de Recife, o Centro Luiz Freire identificou 50 **processos** [contra supostos matadores, RÉUS] abertos entre 1991 e 1993.

(70) Celestino é acusado em quatro processos por lesões corporais e Leme é réu em 11 **processos** [por homicídio, tentativa de homicídio e lesão corporal. ACUSAÇÕES]

Os exemplos (71) e (72) trazem exemplos de sentenças com *ação penal* como evocador de *frame*:

(71) Ex.: **ação penal** [por crimes de homicídio, aborto, infanticídio, lesão corporal, furto, estelionato, peculato etc. ACUSAÇÕES]

(72) Oferecimento de denúncia, em **ação penal** [por crime de homicídio, sem o laudo necroscópico. ACUSAÇÕES]

O *frame* *Juízo*, diferentemente do *frame* *Processo_penal*, não é um *frame* do tipo não lexical, pois ele tem unidades lexicais evocadoras de *frame*. Esse *frame* é também um *frame* complexo, ou seja, divide-se em diferentes *subframes* que caracterizam

diferentes etapas do procedimento de julgamento. Os três *subframes* estão relacionados pela relação *Precedes*: Instrução, Veredito e Sentença.

O *frame* Instrução: O *frame* Instrução descreve a fase de instrução em plenário, em que o JUIZ interroga o RÉU e as TESTEMUNHAS da defesa e da acusação depõem. Os elementos de *frame* são RÉU, TESTEMUNHA e JUIZ. Os elementos de *frame* periféricos neste *frame* se encontram presentes na maioria das sentenças anotadas: ACUSAÇÕES, TRIBUNAL, AÇÃO e TEMPO. As unidades evocadoras de *frame* são *depor*, *interrogar* e *testemunhar*.

Os exemplos (73) e (74) trazem sentenças anotadas com *interrogar* como evocador do *frame* Instrução:

(73) [Sete traficantes -- entre eles Romildo de Souza Costa, o Miltinho do Dendê; Ernaldo Pinto de Medeiros, o Uê; e Odir dos Santos RÉU] -- foram **interrogados** [ontem TEMPO] [no 3º Tribunal do Júri, TRIBUNAL] [sob acusação de terem matado oito traficantes, na Vila do João, em Bonsucesso, em março de 1995. ACUSAÇÕES]

(74) Mas, ao ser **interrogado** [no 1º Tribunal do Júri, TRIBUNAL] negou a confissão.

Os exemplos (75) e (76) trazem sentenças anotadas contendo a unidade lexical *depor*:

(75) [Principal testemunha da chacina TESTEMUNHA] **depõe** [no II Tribunal do Júri TRIBUNAL] reafirma denúncias e diz que Emanuel mentiu ao inocentar Côrtes.

(76) O Ministério da Justiça montou um esquema especial de segurança para receber [Wagner, TESTEMUNHA] que hoje vive na Suíça e será trazido ao Brasil para **depor** [no julgamento dos acusados da morte de oito meninos de rua. AÇÃO]

Por fim, o exemplo (77) apresenta uma sentença contendo a unidade lexical *testemunhar* como evocador do *frame* Instrução:

(77) [Orlando TESTEMUNHA] **testemunhou** [ontem TEMPO] [no tribunal TRIBUNAL] e afirmou que Ubirajara era um dos assassinos.

Uma diferença importante de ser enfatizada neste *frame* é a semântica dos verbos *testemunhar* e *depor*. Pode-se perceber por meio das sentenças aqui analisadas que o verbo *depor* tende a ser empregado para expressar o depoimento do RÉU, enquanto o verbo *testemunhar* é reservado para a TESTEMUNHA. No entanto, o testemunho de uma TESTEMUNHA não deixa de ser um depoimento.

O *frame* Veredito: O *frame* Veredito descreve a fase de votação em que o JÚRI decide sobre a culpa ou inocência do RÉU. Os elementos de *frame* são JUIZ, DECISÃO, ACUSAÇÕES. Os elementos de *frame* periféricos neste *frame* se encontram presentes na maioria das sentenças anotadas: CIRCUNSTÂNCIAS, MOTIVO, AÇÃO e TEMPO. As unidades evocadoras de *frame* são *decidir*, *considerar*, *absolver*, *inocentar*, *condenar*, *condenação* e *veredito*.

Os exemplos (78) a (88) trazem sentenças anotadas com os evocadores de *frame* do *frame* Veredito:

- (78) Quanto a essa acusação, o [júri JUIZ] **decidiu** [absolver DECISÃO] [o réu Alexandre Cardoso, o Topeira, RÉU] e [condenar DECISÃO] [Sandro Baggi e André Rodrigues da Silva, o Gargamel. RÉU]
- (79) [O júri JUIZ] **decidiu** [por quatro votos a três CIRCUNSTÂNCIAS] [culpar DECISÃO] [Divino RÉU] [por homicídio qualificado. ACUSAÇÕES]
- (80) [O júri JUIZ] [a RÉU] **considerou** [não culpada DECISÃO] [por motivo de insanidade temporária. MOTIVO]
- (81) [O réu RÉU] foi [**absolvido** DECISÃO] [pela maioria de dois votos CIRCUNSTÂNCIAS], e por conseqüência o empenho de dois compadres e de três amigos e a minha miopia moral decidiram da sentença.
- (82) [Três integrantes de uma gangue headbangers RÉU] foram [**absolvidos** DECISÃO] [anteontem TEMPO] [no 1º Tribunal do Júri de São Paulo TRIBUNAL] [da acusação de homicídio de dois jovens de uma gangue rival em junho de 93 em um bar na Saúde (zona sul de SP). ACUSAÇÕES]

- (83) [O comerciário Sérgio Aroldo Moraes, 22, de Pindamonhangaba, RÉU] foi [absolvido DECISÃO] [anteontem TEMPO] [por unanimidade CIRCUNSTÂNCIAS] [da acusação de homicídio. ACUSAÇÕES]
- (84) [O réu Élvio Olivar do Prado Miranda, 21, RÉU] [acusado de tentativa de homicídio ACUSAÇÕES], foi [absolvido DECISÃO] [por 4 votos a 3. CIRCUNSTÂNCIAS]
- (85) [Sem provas, MOTIVO] [os seis acusados RÉU] foram [inocentados DECISÃO]
- (86) [A condenação DECISÃO] [do PM Marcus Vinícius Emanuel RÉU] e as recentes confissões dos soldados Nelson Cunha e Marco Aurélio Alcântara não encerram o caso Candelária.
- (87) [O caso AÇÃO] deverá ser analisado até o final do ano [por um júri popular JUIZ], que dará o **veredito**.
- (88) O protesto por novo júri não impedirá a interposição da apelação, quando, pela mesma sentença, [o réu RÉU] tiver sido [condenado DECISÃO] [por outro crime, ACUSAÇÕES] em que não caiba aquele protesto.

Neste *frame* é importante observar que, no caso das unidades lexicais *absolver*, *inocentar*, *condenar* e *condenação*, o elemento de *frame* DECISÃO vem incorporado à unidade lexical. O próximo *frame* representa a etapa final do *frame* julgamento: o *frame* Sentença.

O *frame* Sentença: O *frame* Sentença descreve a fase do *frame* julgamento em que o JUIZ profere a sentença ao RÉU. Os elementos de *frame* centrais são CONDENADO, TRIBUNAL, OFENSA e PENA. A unidade lexical evocadora deste *frame* é *condenar*.

Os exemplos (89) a (91) apresentam sentenças com a unidade lexical *condenar* anotadas com os elementos de *frame*:

- (89) [Ubirajara CONDENADO] foi **condenado** [a 19 anos SENTENÇA] [para cada homicídio OFENSA] e [a 12 anos PENA] [pela tentativa de homicídio de Orlando OFENSA]

(90) [Três policiais e um serralheiro CONDENADO] podem ser **condenados** [a até 400 anos de prisão PENA].

(91) Mais complicado do que encontrar todos os autores da chacina da Candelária é esclarecer as dúvidas surgidas depois que [o juiz José Geraldo Antonio TRIBUNAL] **condenou** [o assassino confesso Marcus Vinicius Borges Emanuel CONDENADO] a [309 anos de cadeia, PENA] [anteontem. TEMPO]

Apesar de a unidade lexical *condenar* ser também evocadora do *frame* Veredito, o que caracteriza o *frame* Sentença é a presença do elemento de *frame* PENA. Esse elemento de *frame* indica a pena que o CONDENADO recebeu pela OFENSA que foi condenado. Este *subframe* indica o fim do processo penal e a condenação do réu.

Como se pode ver nos *subframes* do *frame* julgamento, os *subframes* herdam os elementos de *frame* do *frame* “pai”. É por esse motivo que elementos de *frames* como TRIBUNAL e AÇÃO, que não são utilizados na anotação das sentenças com os evocadores do *frame* Julgamento, fazem parte do *frame* Julgamento, pois eles serão herdados pelos *frames* “filhos” do *frame* Julgamento. A seção seguinte apresenta o *frame* Julgar_acusado, que está em relação de perspectiva com o *frame* Julgamento.

4.3.6 O *frame* Julgar_acusado

O *frame* Julgar_acusado descreve o julgamento de um RÉU, que é acusado de um crime, um corpo de jurados, o JUIZ, é responsável por avaliar as ACUSAÇÕES e decidir se o RÉU é culpado pelo crime, a OFENSA. Os elementos de *frame* centrais são JUIZ, RÉU, OFENSA e ACUSAÇÕES. A unidade lexical evocadora deste *frame* é *condenar*.

Os exemplos (92) a (99) trazem sentenças anotadas para a unidade lexical *julgar*:

(92) Para o governador, o fato de [os acusados RÉU] serem **juulgados** [por um júri popular JUIZ] é muito positivo.

(93) O corregedor ligou para a Polínter e descobriu que os policiais daquela divisão estavam em missão oficial, transportando [o preso Jorge Luís Ribeiro, RÉU] que seria **juulgado** [por homicídio naquela cidade. ACUSAÇÕES]

- (94) [O Júri _{JUIZ}] [da Comarca em que foi praticado o homicídio _{LUGAR}] **julgará** também [o estupro. _{OFENSA}]
- (95) Anulado parcialmente o julgamento do Júri, [o acusado _{RÉU}] voltará a ser **juogado** [pelo Tribunal Popular _{JUIZ}] [somente em relação ao homicídio. _{OFENSA}]
- (96) [Amin Elou _{RÉU}] foi **juogado** [por júri popular. _{JUIZ}]
- (97) [Ele é o último policial civil _{RÉU}] que será **juogado** [pelo júri _{JUIZ}] [por causa da chacina de 18 presos no 42º DP, no Parque São Lucas (zona leste). _{OFENSA}]
- (98) O juiz José Antônio de Paula Santos Neto, do 2º Tribunal do Júri, decidiu que [o falso médico Bruno César Pascini _{RÉU}] será **juogado** [sob a acusação de ser o mandante da morte da mulher, a decoradora Gisele Balesteiros Ribeiro Pascini. _{ACUSAÇÃO}]
- (99) [Paiakan e Irekran _{RÉU}] não serão **juogados** [por um júri popular _{JUIZ}] [porque as acusações são de crimes contra a honra. _{MOTIVO}]

A título de exemplo de como os *frames* propostos seriam exibidos em um recurso lexical jurídico, a figura 12 apresenta um modelo de relatório de *frame* para o *frame* Julgar_acusado.

Julgar_acusado	
Definição:	
Este <i>frame</i> descreve o julgamento de um Réu , que é acusado de um crime, um corpo de jurados, o Juiz , é responsável por avaliar as Acusações e decidir se o Réu é culpado pelo crime, a Ofensa .	
Elementos de <i>Frame</i> (EF):	
Centrais:	
Juiz	Este EF identifica a entidade que decide o caso. No caso do processo penal brasileiro, o papel de Juiz é exercido pelo júri em decisões de primeira instância. O Júri da Comarca em que foi praticado o homicídio julgará também o estupro.
Réu	Este EF identifica o Réu que está sendo julgado por ter sido acusado de um crime, a Ofensa . Amin Elou foi julgado por júri popular.
Ofensa	Este EF identifica o crime de que o Réu é acusado. Ele é o último policial civil que será julgado pelo júri por causa da chacina de 18 presos no 42º DP, no Parque São Lucas (zona leste) .
Acusações	Este EF identifica as Acusações pelas quais o Réu é julgado. O corregedor ligou para a Polínter e descobriu que os policiais daquela divisão estavam em missão oficial, transportando o preso Jorge Luís Ribeiro, que seria julgado por homicídio naquela cidade .
Periféricos:	
Lugar	Este EF indica o Lugar onde o caso é julgado. O Júri da Comarca em que foi praticado o homicídio julgará também o estupro.

Figura 12: Relatório de *Frame*

É importante enfatizar que o elemento de *frame* JUIZ tem um caráter de “modo como o julgamento se realizou”. Isso ocorre no *frame* julgar_acusado porque o Tribunal do Júri é um dos procedimentos de julgamento de crimes no Brasil. Há outros procedimentos, sendo que o Tribunal do Júri é reservado apenas para casos de crimes dolosos contra a vida. Assim, o elemento de *frame* JUIZ, quando realizado por expressões como *júri*, *júri popular* ou *tribunal popular*, tem um caráter duplo de JUIZ e de MODO.

Este capítulo teve por objetivo contrastar as unidades lexicais relacionadas ao processo penal em inglês e português e avaliar, através da comparação entre as línguas, a possibilidade de se utilizarem os *frames* semânticos da FrameNet para a construção de um recurso lexical da linguagem jurídica brasileira. A opção foi analisar o *frame* *Criminal_process*, um dos poucos *frames* relacionados ao Direito.

Assim, este capítulo apresentou na seção 4.1 a progressão do processo penal nos EUA e no Brasil. Essa seção teve como objetivo trazer referencial teórico para o estudo contrastivo apresentado na seção 4.2. A seção 4.2, por sua vez, apresentou o estudo contrastivo das unidades lexicais evocadoras dos *subframes* do *frame* *Criminal_process*. A partir dos resultados do estudo contrastivo, a seção 4.3 apresentou um esboço de como seria o *frame* *Processo_penal* para o Direito brasileiro. Como explicitado no início deste capítulo, as três seções apresentadas aqui, foram organizadas segundo três princípios: jurídico, contrastivo e computacional. No entanto, o estudo contrastivo levanta algumas questões que ainda não foram abordadas neste capítulo:

- A) Qual a relação entre *frames* semânticos, cultura e correspondência de *frames* entre línguas? Enquanto trabalhos como Padó e Lapata (2005), Padó (2007) e Tonelli e Pianta (2008) afirmam que é possível a transposição de *frames* entre diferentes línguas, pois os *frames* representam um nível conceitual, este capítulo mostrou que essa transposição não é assim tão simples. A explicação de que *frames* representam um nível conceitual relativamente independente de língua é frágil. Pode-se constatar, por meio da análise contrastiva, a falta de correspondência entre *frames* e de elementos de *frame* em vários casos.
- B) O crescente uso da Semântica de *Frames* em aplicações computacionais multilíngües automatizadas leva a outra questão. Como ficaria a relação cognição, cultura e linguagem, postulada pela lingüística cognitiva, quando as aplicações computacionais que utilizam a Semântica de *Frames* para fins multilíngües tendem a passar por cima das diferenças culturais presentes na língua, através da visão simplificada de que os *frames* são um nível cognitivo e, por isso, não possuem tanta variação entre as línguas como o léxico?

Tendo em vista essas questões, resta, para o próximo capítulo, tratar dos seguintes pontos:

- A) Como os princípios norteadores da lingüística cognitiva podem contestar a visão generalista de correspondência de *frames*?
- B) Como explicar a correspondência de *frames* em alguns casos e a não correspondência em outros?
- C) Em que medida os *frames* semânticos estudados são correspondentes entre o inglês e o português?
- D) Quais as implicações dessas diferenças de *frames* entre as duas línguas para as aplicações multilíngües?

5 DIVERGÊNCIA DE *FRAMES* E IMPLICAÇÕES COMPUTACIONAIS

O presente capítulo discute as implicações computacionais deste estudo para a criação de recursos lexicais multilíngües segundo o paradigma FrameNet. O estudo contrastivo demonstrou que os *frames* semânticos em domínios socialmente construídos, como o Direito, apresentam um alto grau de divergência entre as línguas. Essa divergência entre os sistemas jurídicos norte-americano e brasileiro demonstrou a necessidade de criação de *frames* específicos para o processo penal brasileiro. Essas divergências de *frames* trazem à tona questões sobre a falta de correspondência de *frames* e suas implicações para recursos lexicais multilíngües.

Uma das propostas de uso multilíngüe da FrameNet é como índice de interlíngua para a conexão de bases de dados em diferentes línguas, sugerido por Boas (2005). A partir da comparação dos dados da FrameNet e da Spanish FrameNet para o domínio *comunicação*, Boas (2005) sugere o uso dos *frames* semânticos como interlíngua, unindo léxicos computacionais baseados em *frames*. Essa interlíngua, por ser baseada em *frames*, evitaria problemas que uma interlíngua baseada no léxico apresentaria, como polissemia, padrões de valência sintática e semântica, diferenças em padrões de lexicalização e relações de paráfrase e de equivalência de tradução entre as línguas. Assim, os *frames* são considerados como um nível *conceitual* relativamente isento das dificuldades oferecidas pelo léxico. Boas (2005) afirma que para o sucesso de tal empreendimento é necessário se distinguir os *frames universais* e os *frames específicos de língua*²⁰.

Padó (2007) defende a idéia de que os *frames* semânticos apresentam um alto nível de paralelismo entre as línguas, pois os *frames* representariam um nível conceitual da língua. Isso o leva a sugerir a criação automática de léxicos baseados em *frames*. O método consiste, basicamente, em um processo de análise e anotação semântica automática de *corpora* em

²⁰ A questão da universalidade dos *frames* será retomada na seção 5.1.

inglês, utilizando as etiquetas semânticas (elementos de *frame*) da FrameNet. O segundo passo é a transferência das anotações do *corpus* em inglês para um *corpus* em outra língua. Esse método exige, em primeiro lugar, a existência de *corpora* paralelos, sem os quais não é possível realizar o mapeamento das estruturas lingüísticas semelhantes. Esse *corpus* necessita, portanto, ser composto de textos traduzidos²¹.

Outra aplicação da Semântica de *Frames* que merece destaque é a criação de léxicos multilíngües para domínios especializados. Nessa área, o Kicktionary (Schmidt, 2009) merece destaque por ser o léxico computacional multilíngüe baseado em *frames* mais bem desenvolvido até o momento. Uma vez que a FrameNet não apresenta *frames* capazes de descrever os eventos de uma partida de futebol, Schmidt (2009) opta por construir *frames* específicos para a linguagem do futebol. Apesar de o esporte poder ser considerado uma expressão cultural, os *frames* semânticos para o esporte são diferentes dos *frames* semânticos para o Direito. Uma vez que o Direito é uma ciência socialmente construída, seus *frames* são diferentes em cada língua e cultural, enquanto os *frames* do futebol são basicamente os mesmo para todas as culturas, pois as regras desse esporte são as mesmas, mudando apenas os itens lexicais nas diferentes línguas.

É levando em consideração a problemática da divergência dos *frames* jurídicos entre as línguas que este trabalho discute as implicações que uma visão de *frame* mais culturalmente orientada tem para as aplicações computacionais. Para tratar das questões referentes à multilingüidade e criação de recursos lexicais jurídicos, o presente capítulo foi dividido em três seções. A seção 5.1 discute a similaridade dos *frames* semânticos entre as línguas. A seção 5.2 discute a multilingüidade em recursos lexicais jurídicos baseados em *frames*.

²¹ Uma diferenciação entre *corpora* paralelos e *corpora* comparáveis necessita ser delineada aqui. Os *corpora* comparáveis são compostos por textos comparáveis em línguas diferentes. Ou seja, a natureza dos textos que compõem o *corpus* na língua A é semelhante à natureza dos textos que compõem o *corpus* na língua B. Por esse motivo eles são comparáveis. Já *corpora* paralelos requerem uma similaridade maior entre os textos do *corpus* na língua A e na língua B. Tamanha similaridade só é alcançada através da tradução. Por isso os *corpora* paralelos são *corpora* de textos traduzidos. Uma discussão sobre a natureza dos *corpora* paralelos e comparáveis é apresentada em Teubert (1996). Um dos *corpora* paralelos mais utilizados na atualidade é o Europarl, contendo os textos do Parlamento Europeu. O *corpus* Europarl é disponibilizado gratuitamente no site <http://www.statmt.org/europarl/>.

5.1 *Frames* Semânticos: Universais ou Específicos?

Esta seção tem por objetivo discutir o que se vai chamar de *universalidade de frames*. O trabalho com os *frames* jurídicos leva, inevitavelmente, para a discussão sobre a similaridade dos *frames* semânticos em uma abordagem multilíngüe. Boas (2005) sugere o uso dos *frames* semânticos como interlíngua em bases de dados multilíngües. Boas (2005) reconhecendo que os *frames* semânticos não são os mesmos em todas as línguas, afirma que o sucesso no uso de *frames* semânticos como interlíngua depende de se distinguirem os *frames universais* e dos *frames específicos de língua*. Os *frames universais* seriam comuns à maioria das línguas, servindo como ponto de conexão entre as bases de dados de cada língua.

Projetos em lingüística computacional que utilizam as etiquetas semânticas da FrameNet para anotação de *corpus* também partem do princípio da *universalidade de frames*. Projetos de anotação extensiva de *corpora* como o SALSA (Burchardt et al., 2009) e de desenvolvimento automático de léxicos computacionais multilíngües a partir de anotação de *corpora* paralelos (Padó e Lapata, 2005; Johansson e Nugues, 2005, 2006; Padó e Pitel, 2007; Tonelli e Pianta, 2008) partem do princípio de que há paralelismo entre as línguas. O paralelismo seria, de forma resumida, a correspondência lexical e semântica entre línguas quando comparadas. Assim, a forma das unidades lexicais variaria entre as línguas, mas o significado, representado pelo *frame* semântico, seria equivalente. Essa tese justificaria os altos índices de correspondência entre as línguas ao se transferir a anotação de *frames* semânticos de um *corpus* em inglês para um *corpus* em outra língua²². No entanto, como Padó (2007) ressalta, o sucesso de tal técnica depende do grau de paralelismo entre as línguas.

A proposta de indução automática de léxicos apresentada por Padó (2007) requer o paralelismo entre as línguas para o sucesso da anotação semântica, conforme o próprio autor reconhece. Outro fator de dificuldade para a metodologia proposta é a polissemia das línguas. Padó (2007) aponta, entre os casos difíceis, a distinção entre os *frames* Judgment_communication e Notification_of_charges. Ambos os *frames* possuem a unidade lexical *accuse* como evocadora de *frame*. A diferença entre esses dois

²² Essa técnica de transferência de anotação de um *corpus* em inglês para um *corpus* em outra língua, utilizando um *corpus* paralelo, foi discutida na seção 3.3.2. Padó e Lapata (2005) sugerem que se utilize as etiquetas semânticas da FrameNet para anotar um *corpus* em inglês e que, através de algoritmos para alinhamento das unidades lexicais entre o *corpus* na língua inglesa e o *corpus* na língua-alvo, se transfira a anotação semântica do inglês para a língua-alvo. Essa metodologia requer um *corpus* paralelo, ou seja, um *corpus* de textos traduzidos. O *corpus* que mais tem sido utilizado para aplicação desta metodologia é o *Corpus Paralelo Europarl*, um *corpus* dos anais do Parlamento Europeu em onze línguas da União Européia.

frames está no fato de um estar relacionado a um domínio especializado (*Notification_of_charges*) enquanto o outro é um *frame* de significado geral (*Judgment_communication*). A proposta de Padó (2007), apesar de representar uma alternativa rápida para a criação de recursos lexicais, esbarra em questões centrais à natureza das línguas, como a polissemia e a falta de paralelismo.

A metodologia de indução automática de FrameNets tem sido utilizada para diversas línguas, tais como o alemão (Padó e Lapata, 2005), o espanhol (Johansson e Nugues, 2005), o sueco (Johansson e Nugues, 2006), o francês (Padó e Pitel, 2007) e o italiano (Tonelli e Pianta, 2008). Esses trabalhos, apesar de relevantes para a lingüística computacional, suscitam perguntas se analisados sob a perspectiva da lingüística cognitiva, visto que todos partilham como princípio teórico a Semântica de *Frames*. Esses trabalhos partem de uma visão simplificada de que os *frames* semânticos são estruturas conceituais relativamente independentes de língua. Em primeiro lugar, é necessário questionar o que é um conceito e se os conceitos são estruturas cognitivas realmente independentes de língua. Segundo a visão de Croft (2006, p.271), (...) *um conceito é uma estrutura semântica simbolizada por uma palavra*. Devido a essa ligação estrita entre significado e forma, Croft afirma que (...) *a noção de conceito é suficientemente difícil de ser identificada independentemente de língua* (...) (2006, p.271).

Depois de ter explicitar o uso da FrameNet em lingüística computacional, há dois pontos que devem ser discutidos sobre a *universalidade de frames*, levando-se em consideração essas aplicações computacionais: (i) a relação *frames* semânticos e conhecimento enciclopédico e (ii) a relação *frame* de língua geral e *frame* de linguagem especializada. Está-se chamando de *universalidade de frames* a essa visão linguisticamente descompromissada de se defender uma correspondência de *frames* semânticos entre línguas sem uma análise lingüística do *corpus* em estudo. Não se trata de uma crítica aos trabalhos automatizados e estatísticos, mas um olhar lingüístico sobre os fenômenos que não interessam a esses trabalhos: os casos de falta de correspondência entre *frames*. A partir da discussão apresentada nesta seção, duas idéias são defendidas. Em primeiro lugar, a *universalidade de frames* é uma simplificação da natureza cognitiva e cultural dos *frames* semânticos. Em segundo lugar, a tese da *universalidade de frames* não resiste a um estudo contrastivo em domínios especializados socialmente construídos, como o Direito²³.

²³ Está-se chamando de domínios especializados socialmente construídos as áreas de conhecimento inteiramente criadas pela sociedade, como o Direito. Segundo Mattila (2006, p.105), *o direito não existe no mundo físico. Por*

Os trabalhos multilíngües com a FrameNet levam a uma reflexão lingüística sobre a relação entre o nível conceitual e o nível lexical. Segundo a lingüística cognitiva, a língua não é um sistema autônomo. Assim, a língua possui uma estreita ligação com a cultura e a cognição. Holland e Quinn (1987) sustentam que os modelos culturais estruturam a experiência e o comportamento humano. Sweetser (1987) discute, por meio da análise da palavra *lie* (*mentira*), como a estrutura semântica é dependente dos modelos culturais e como essa estrutura reflete esses modelos. Isso significa dizer que a cultura interfere na língua. A língua, por sua vez, estrutura o conteúdo conceitual (Talmy, 2000). Assim, a cultura tem influência sobre a categorização, visto que, para a lingüística cognitiva, a linguagem tem papel fundamental no processo de categorização. Isso explica por que a lingüística cognitiva não faz distinção entre conhecimento lingüístico e conhecimento de mundo.

A abordagem do significado pela lingüística cognitiva é enciclopédica. Segundo Fillmore (1982, p. 112), *as palavras representam categorizações de experiências, e cada uma dessas categorias é sustentada por uma situação motivadora baseada em conhecimento e experiência anteriores*. A visão de Fillmore (1982) é a de que o significado das palavras é entendido a partir de conhecimento e experiência anteriormente adquiridos pelo falante de uma determinada comunidade. Apesar de a categorização ser central para o entendimento da linguagem, na visão da lingüística cognitiva, essa categorização é orientada pelo contexto cultural. Assim, a teoria da *universalidade de frames* é posta à prova, pois, se os *frames* são estruturas conceituais que representam o contexto necessário para se entender o significado de uma unidade lexical, logo o *frame* semântico será culturalmente orientado também, não apenas a unidade lexical. É essa sutil relação entre cultura, língua e categorização que leva a lingüística cognitiva a rejeitar a teoria da autonomia da língua e da distinção entre conhecimento lingüístico e conhecimento de mundo.

A Semântica de *Frames* é uma teoria calcada na visão enciclopédica do significado que, em sua origem (Fillmore 1982, 1985), alerta, justamente, para a necessidade de o significado das palavras necessitarem de conhecimento de mundo. É no sentido de afirmar a necessidade do conhecimento cultural para a compreensão do significado das palavras que

ser inteiramente criado pelos humanos, o direito é sempre relacionado à cultura de uma determinada sociedade. Talvez, dessas áreas de conhecimento criadas pela sociedade, o Direito seja a mais particular de todas, por apresentar um conjunto de normas válidas apenas para a sociedade que a criou. Assim, as leis criadas pela sociedade brasileira valem para o Brasil, enquanto as leis criadas pela sociedade americana valem apenas para a sociedade americana. Outro exemplo de domínio especializado socialmente construído seriam as unidades de medidas, como *polegada*, *pé*, *centímetro* ou *metro*. As unidades de medidas variam, em alguns casos, de país para país. Assim, os EUA adotam unidades de medidas como *polegada*, *pé* e *milha*. O Brasil adota unidades diferentes: *centímetro*, *metro*, *quilômetro*.

Fillmore (1982) traz o exemplo da palavra café-da-manhã (*breakfast*). O café-da-manhã pode ser a primeira refeição do dia. No entanto, se um americano ler um anúncio de um restaurante dizendo: *servimos café-da-manhã o dia todo*, ele vai entender que não se trata da primeira refeição do dia, mas do tipo de alimento tipicamente servido no café-da-manhã. A busca de *frames universais* e *frames dependentes de língua* a partir de uma teoria enciclopédica que não faz distinção entre conhecimento contextual, conhecimento lingüístico e conhecimento de mundo é uma simplificação da proposta teórica original. Apesar de a simplificação dos princípios cognitivos da Semântica de *Frames*, não se pode deixar de reconhecer a importância de todos esses empreendimentos na construção de recursos lexicais para o processamento de linguagem natural.

A Semântica de *Frames* é, em sua essência, uma teoria que investiga a forma como a língua reflete a cultura. Segundo Fillmore (1982, p. 112), (...) *a semântica de frames pode ser vista como um esforço para entender que razão uma comunidade de falantes pode ter encontrado para criar a categoria representada pela palavra e explicar o significado da palavra, apresentando e clarificando essa razão*. Os *frames* semânticos não se prestam a uma distinção clara entre *frames dependentes de língua* (*frames* próprios a uma determinada língua e cultura) e *frames universais* (*frames* independentes de língua e de cultura), visto que a sua agenda de pesquisa está comprometida com a relação entre cultura e significado lexical. No entanto, o trabalho com os *frames* jurídicos leva à verificação de que há uma similaridade maior entre alguns *frames* e menor entre outros *frames*.

Na verdade, explicar a similaridade de *frames* entre as línguas pelo viés da *universalidade de frames* é uma simplificação do fenômeno lingüístico. A busca de respostas para a similaridade de *frames* deveria pôr o foco da pesquisa sobre outra questão: a distinção entre *frames inatos*²⁴ e *frames aprendidos*:

Alguns *frames* são indubitavelmente inatos, no sentido de que eles aparecem naturalmente e inevitavelmente no desenvolvimento cognitivo de cada humano (o conhecimento das características do rosto humano pode ser um exemplo). Outros são aprendidos através da experiência ou do treino (por exemplo, conhecimento de artefatos e instituições sociais), o caso extremo seriam aqueles *frames* cuja existência dependeria inteiramente dos usos lingüísticos associados (tais como as unidades de medidas – *polegadas, pés, jardas* etc. – e unidades de tempo como *semana* e os nomes dos meses). (Fillmore, 1985, p.232-3)

²⁴ Usa-se o termo *inato* na falta de um termo que expresse melhor para expressar o que, talvez, poderia se chamar *frames* perceptuais, corporais, físicos. No entanto, reconhece-se que o termo *inato* não é isento de controvérsias.

Os *frames* inatos implicam a *corporificação* da estrutura conceitual (Lakoff e Johnson, 1999), enquanto os *frames* aprendidos implicam o conhecimento enciclopédico. O entendimento que o ser humano tem do mundo a sua volta é mediado, em grande parte, pelo entendimento do seu corpo e pelas capacidades sensoriais e motoras de seu corpo. Conceitos como *esquerda* e *direita* ou *acima* e *abaixo* são compreendidos através do conhecimento que o ser humano tem de seu corpo. Assim, determinados conceitos são, necessariamente, relativos à realidade física do ser humano. Só é possível entender o conceito de *esquerda* ou *direita* tendo como referência um corpo ou objeto. Outro ponto importante com relação à corporificação da estrutura conceitual diz respeito aos órgãos sensoriais e motores do ser humano.

A capacidade de categorização do ser humano é relativa à capacidade física de seu corpo. As cores exemplificam a relação existente entre os órgãos sensoriais e a capacidade de categorização do ser humano. Mesmo que a categorização sobre as cores seja determinada, em grande parte, por fatores culturais (ver discussão de Berlin e Kay, 1969), a percepção que os seres humanos têm sobre as cores está limitada a fatores físicos. Segundo Varela et al. (1991), os seres humanos possuem três tipos de fotorreceptores, esquilos e coelhos possuem dois tipos de fotorreceptores, outros possuem quatro tipos de fotorreceptores, como alguns tipos de pássaros e peixes. Isso significa que só se pode categorizar segundo as suas capacidades. Se faltam meios físicos para se reconhecer determinadas cores, não se estará apto a criar categorias que representem as cores que não são vistas.

A similaridade de *frames* entre línguas não deve ser equacionada como uma característica universal às línguas, mas sim como um resultado da estrutura conceitual corporificada (Lakoff e Johnson, 1999). O significado representado pelos *frames* inatos depende mais da noção de corpo, da motricidade, das percepções sensoriais e da cognição que do conhecimento cultural. Como os seres humanos são todos dotados do mesmo aparato físico, ou seja, o mesmo corpo biológico e o mesmo sistema cognitivo, salvo em situações de problemas físicos, a categorização dos movimentos corporais, das percepções sensoriais e dos processos cognitivos pode apresentar semelhanças consideráveis entre as línguas.

Como exemplo, pode-se citar o *frame* Perception_experience. Segundo a FrameNet, o *frame* Perception_experience contém unidades lexicais que indicam percepção. Essa percepção é pretendida pelo *perceptor*, acontecendo por acaso. Por essa razão, o *perceptor* é chamado de PERCEPTOR_PASSIVO. Esse *frame* é evocado por unidades

lexicais como *feel.v* (*sentir*), *see.v* (*enxergar*), *smell.v* (*cheirar*). O *frame* *Perception_experience* poder ser considerado um *frame* inato à medida que descreve um conhecimento relacionado à experiência corporal e essa experiência corporal tende a ser a mesma em diferentes culturais.

Essa similaridade dos *frames* inatos entre as línguas é o que faz os criadores de recursos lexicais multilíngües baseados em *frames* defenderem a idéia da *universalidade* dos *frames* semânticos. Segundo esses projetos, os *frames* semânticos, por serem estruturas conceituais, são isentos da influência da língua, podendo ser aplicados sem grandes dificuldades em trabalhos multilíngües. Na verdade, essa visão requer uma atenção maior. Em primeiro lugar, os *frames* inatos, ou seja, esses *frames* que representam informações referentes à realidade corporal do ser humano, são apenas uma parte dos *frames* possíveis em uma língua. As línguas ainda apresentam os *frames* aprendidos, que representam informações mais culturais. Em segundo lugar, mesmo os *frames* inatos podem apresentar diferenças culturais.

Um exemplo de *frame* inato que apresenta diferenças de categorização entre línguas é o *frame* *Self_motion*. Segundo a FrameNet, o *frame* *Self_motion* descreve um evento em que um ser vivo se movimenta por suas próprias forças através de um caminho. As unidades lexicais evocadoras desse *frame* incluem *fly.v* (*voar*), *tiptoe.v* (*caminhar na ponta dos pés*), *walk.v* (*caminhar*). Se compararmos os pares *tiptoe* e *walk*, percebe-se que se trata de duas categorizações diferentes para o evento de movimento por conta própria, se tomarmos por base a definição de Croft (2006), de que um conceito é uma estrutura semântica simbolizada por uma palavra. As diferenças entre o inglês e o português para os verbos *walk/caminhar* e *tiptoe/caminhar na ponta dos pés* não podem ser consideradas apenas como uma diferença lexical entre as línguas. Há uma diferença na forma como cada língua categoriza as diferentes formas de movimento, tanto que no português não há um verbo equivalente ao inglês *tiptoe*²⁵.

Baseado na distinção entre *frames* inatos e *frames* aprendidos apresentada por Fillmore (1985), passa-se agora à discussão dos *frames* aprendidos. Se, por um lado, os *frames* inatos representam o conhecimento adquirido pelos humanos através das suas

²⁵ Essas diferenças lexicais são discutidas em Talmy (2000). No capítulo *Lexicalization Patterns*, Talmy trata das relações sistemáticas na língua entre significado e expressão de superfície. Em primeiro lugar, ele assume que é possível isolar elementos separadamente no domínio do significado e no domínio da expressão de superfície. Os elementos semânticos são *Motion*, *Path*, *Figure*, *Ground*, *Manner* e *Cause*. Já os elementos de superfície são verbo, oração subordinada etc. Em segundo lugar, Talmy examina quais elementos semânticos são expressos por cada elemento de superfície. Essas relações não se dão uma-a-uma. Uma combinação de elementos semânticos pode ser expressa por um único elemento de superfície, ou um único elemento semântico para uma combinação de elementos de superfície.

percepções sensoriais e da sua cognição, os *frames* aprendidos representam o conhecimento de mundo (conhecimento enciclopédico). Como já foi comentado anteriormente, a Semântica de *Frames*, assim como as demais abordagens da lingüística cognitiva para o significado, é uma abordagem enciclopédica. Assim, não é feita distinção entre conhecimento lingüístico e conhecimento de mundo, ou seja, o conhecimento do significado lingüístico implica o conhecimento do contexto no qual determinada expressão lingüística é usada e da cultura na qual os falantes estão inseridos. Voltando ao exemplo do café-da-manhã (*breakfast*), é o conhecimento cultural que permite aos falantes reconhecerem que, no caso dos EUA, ter café-da-manhã para o jantar significa ter no jantar uma refeição tipicamente consumida no café-da-manhã: ovos, panquecas, *waffles*, *maple syrup* etc. Nesse caso, não se trata da ordem das refeições, *café-da-manhã*, *almoço* e *jantar*, mas do protótipo de café-da-manhã. Refeição é um *frame* culturalmente aprendido, pois em cada cultura as pessoas terão momentos diferentes para se alimentarem, tendo seus protótipos de como deve ser uma refeição apropriada para determinado momento do dia.

Outro caso de *frame* aprendido são os *frames* jurídicos. Mattila (2006, p.105) sustenta que *por ser inteiramente criado pelos humanos, o direito é sempre relacionado à cultura de uma determinada sociedade*. Devido a essa relação estreita com a cultura de cada país, os *frames* jurídicos podem ser considerados como *frames* aprendidos. Esses *frames* não são adquiridos através da percepção, como acontece com a audição, a visão ou o tato. Não são também compreendidos pela relação do corpo humano com o mundo que o rodeia, como no caso dos conceitos de *esquerda* e *direita*. Como esses *frames* são construções sociais, eles apresentam um alto nível de divergência entre sistemas jurídicos.

O estudo contrastivo apresentado no capítulo 4 já apontava para a necessidade de se criarem *frames* jurídicos específicos para o sistema jurídico brasileiro. Alguns *frames* semânticos da FrameNet descreviam eventos jurídicos compatíveis com o sistema judicial brasileiro, como o caso do *frame* Trial. Outros *frames*, como o *Notification_of_charges*, representavam eventos que correspondiam ao sistema judicial brasileiro somente parcialmente. Outros ainda, como o *Arraignment*, não apresentavam correspondência com o sistema judicial brasileiro. Deve se levar em conta que o Direito americano é baseado na *common law*, enquanto o Direito brasileiro é fundamentado na *civil law*.

Essa diferença nos sistemas é a raiz das divergências entre os *frames*, como no caso do *frame* *Arraignment*, visto que o *arraignment* é uma etapa típica de um processo penal

segundo a *common law*. No entanto, a mesma diferença entre os sistemas é responsável por algumas compatibilidades entre os *frames*, como no caso do *frame Trial*. O julgamento de crimes dolosos contra a vida no Brasil é baseado no julgamento norte-americano, ou seja, o sistema judicial brasileiro adota o Tribunal do Júri do sistema norte-americano. Outros países que seguem a *civil law*, como a Alemanha, não fazem uso do Tribunal do Júri. No caso dos *frames* jurídicos, a interação cultural permitira a adoção de práticas sociais de outra cultura, fazendo com que os *frames* jurídicos tornem-se, ao longo do tempo, semelhantes. Mattila (2006) aponta para uma crescente integração entre a *common law* e a *civil law* com a unificação econômica da Europa. Uma vez que as regulamentações do Parlamento Europeu são criadas por todos os membros da comunidade, esses dois sistemas estão em processo de conversão devido à intensa interação por meio dos tribunais que interpretam e aplicam as leis criadas pela União Européia.

Os *frames* jurídicos colocam à prova a separação entre conhecimento lingüístico e conhecimento de mundo. Como as leis são construções sociais, criadas por um grupo social muito restrito de especialistas, o significado de muitos conceitos jurídicos não chega a ser conhecido por toda a população. Esse fato só reforça a tese de que os *frames* jurídicos são aprendidos. O estudo contrastivo dos *frames* semânticos leva à refutação da *universalidade de frames*. Em primeiro lugar, os *frames* semânticos não são universais. Em segundo lugar, alguns *frames* semânticos apresentam uma natureza inata por se tratarem de *frames* que representam conhecimentos adquiridos na relação do ser humano com o ambiente. Essas relações são mediadas pelo sistema perceptual, como visão, audição, localização espacial etc., o que Lakoff e Johnson (1999) vão chamar de realismo corporificado.

A visão de correspondência de *frames* entre línguas que se apresenta neste trabalho traz algumas conseqüências não só para o desenvolvimento de léxicos computacionais jurídicos baseados em *frames*, mas também para o desenvolvimento de recursos lexicais multilíngües baseados no paradigma FrameNet. A proposta de Boas (2005) de uso dos *frames* semânticos como representação de interlíngua para conectar léxicos multilíngües baseados em *frames* necessita ser revista. Considerando que nem todos os *frames* apresentam o mesmo grau de similaridade, sendo que em níveis especializados como o Direito pode haver até mesmo a falta total de correspondência, é preciso pensar em novas rotas para a multilingualidade na FrameNet. Considerando essa necessidade, a seção 5.2 traz uma reflexão sobre a equivalência de *frames* entre línguas e possíveis soluções para os léxicos computacionais multilíngües baseados em *frames*.

5.2 A Multilinguagem em FrameNets Jurídicos

A presente seção tem por objetivo discutir a multilinguagem em léxicos jurídicos baseados em *frames*. Segundo Lönneker-Rodman (2007), a metodologia de extensão de FrameNets exige equivalência completa de *frames*, elementos de *frame*, relações entre *frames* e tipos semânticos dos elementos de *frame*. A necessidade de equivalência completa entre as línguas torna a equivalência de *frames* quase impossível. Os *frames* jurídicos são evidências da falta de equivalência de *frames* entre línguas, uma vez que esses *frames* representam um conhecimento socialmente construído, sendo, portanto, específicos de cada cultura e de cada país.

O tratamento da equivalência de *frames* é relativamente recente. Em geral, os estudos de equivalência se detêm no estudo da equivalência de unidades lexicais. A idéia de equivalência de *frames* (Lönneker-Rodman, 2007) tem por objetivo a construção de recursos lexicais baseados no paradigma FrameNet. Assim, se um *frame* semântico em espanhol possui similaridade no nível da definição do *frame*, no número de elementos de *frame* e no tipo semântico desses elementos, esse *frame* manterá na FrameNet do espanhol o mesmo nome que possui na FrameNet do inglês. Se alguma falta de similaridade for constatada, como o acréscimo de elementos de *frame*, o *frame* em espanhol receberá um novo nome.

Percebe-se a partir do estudo contrastivo que os *frames* jurídicos apresentam diferentes níveis de correspondência, variando desde a quantidade de elementos de *frame* até a natureza do evento jurídico descrito pelo *frame*. Neste trabalho, alia-se essa idéia de níveis de correspondência à idéia de equivalência de *frames* de Lönneker-Rodman (2007), apresentando-se a idéia de **níveis de equivalência de *frames***. A questão dos níveis de equivalência de *frames* não é central a este trabalho. Ela surge a partir do estudo contrastivo e do desenvolvimento de *frames* jurídicos para o português. Assim, essa noção não será plenamente desenvolvida aqui. No entanto, o quão equivalente um *frame* jurídico é na comparação entre sistemas jurídicos distintos será fundamental para a conexão dos *frames* jurídicos em um recurso lexical multilíngüe.

Na seção 5.1, foram apresentadas as idéias de frames inatos e frames aprendidos (Fillmore, 1985) como uma possível resposta para uma maior ou menor correspondência entre frames, sugerindo que os frames aprendidos apresentavam menos correspondência entre línguas e culturas por apresentarem informação socialmente compartilhada por um

determinado grupo, como os frames jurídicos. No entanto, isso não significa que esses frames não possam apresentar equivalência entre línguas.

	Try_defendat	Court_examination	Notification_of_charges	Arraignment
	Julgar_acusado	Instrução	Denúncia	--
Equivalência:				
Unidades lexicais	Sim	+/-	Sim	--
Evento jurídico	Sim	Sim	Não	--
Elementos de <i>frame</i>	Sim	Não	Não	--

Tabela 12: Níveis de Equivalência de *Frames*

Como se pode ver na tabela 12, o par de *frames* Try_defendant e Julgar_acusado apresenta um nível maior de equivalência do que os outros pares. Essa equivalência de *frames* é expressa pela equivalência das unidades lexicais evocadoras de ambos os *frames*, pela equivalência do evento jurídico descrito por ambos os *frames* e pela equivalência de elementos de *frames*, conforme estudo apresentado na subseção 4.2.3.

No caso do par de *frames* Court_examination e Instrução, apesar de o evento jurídico ser equivalente, algumas unidades lexicais do inglês não apresentam equivalência. Além disso, o acréscimo de uma nova unidade lexical no *frame* Instrução mostrou a necessidade de se dividir o elemento de *frame* WITNESS em TESTEMUNHA e RÉU, conforme estudo apresentado na subseção 4.2.4.

No par de *frames* Notification_of_charges e Denúncia, conforme visto na subseção 4.2.2, as unidades lexicais do *frame* Notification_of_charges apresentam equivalentes em português. No entanto, como o evento jurídico descrito pelo *frame* Notification_of_charges não apresenta um evento correspondente em português, os equivalentes em português apresentam unidades lexicais que evocam dois *frames* distintos: Denúncia e Pronúncia. Logo, pode-se concluir que o significado descrito pelo *frame* semântico é preponderante para definir o nível de equivalência dos *frames* entre línguas. Tal constatação não pode ser tomada como novidade. Altenberg e Granger (2002) defendem que o significado é o fator preponderante para a definição da equivalência de itens lingüísticos entre línguas diversas. Assim também ocorre com o *frame*. O significado

do *frame*, ou seja, o evento que o *frame* representa, deve ser o fator primordial para se definir o nível de equivalência de *frames*.

O *frame* *Arraignment* não apresenta qualquer forma de equivalência. Um dos motivos de as unidades lexicais não apresentarem equivalentes em português pode ser justamente a ausência de um evento jurídico equivalente para o sistema brasileiro. No entanto, essa conclusão não pode ser categórica. Algumas unidades lexicais polissêmicas podem evocar *frames* diversos e apresentar uma unidade lexical equivalente para o *frame* X e não apresentar uma unidade lexical equivalente para o *frame* Y.

A partir dessas observações, como solucionar a questão da conexão entre *frames* jurídicos em um contexto multilíngüe? O estudo dos *frames* jurídicos mostra que os *frames* não são uma categoria conceitual universal. No caso dos *frames* jurídicos, eles não são universais, mas altamente variáveis, não somente entre as línguas, mas também entre os países de mesma língua. Sugere-se aqui duas direções que podem ser mais bem exploradas pelos léxicos computacionais multilíngües, em especial os jurídicos: os **tipos semânticos** e as **informações ontológicas**.

Alguns eventos jurídicos, como o processo penal e seus diversos sub-eventos estudados neste trabalho, podem não apresentar correspondência entre um sistema e outro, porém, percebe-se uma correspondência na ordem dos eventos. Por exemplo, o evento jurídico descrito pelo *frame* *Criminal_process* se inicia com o *frame* *Arrest* e termina com o *frame* *Sentencing*. O *frame* *Processo_penal* descreve um evento jurídico que se inicia com o *frame* *Prisão* e termina com o *frame* *Sentença*. Assim, apesar de as partes não serem equivalentes, o todo é correspondente. Talvez, uma possível solução para tais dificuldades esteja em apostar em uma categoria de descrição semântica, já constante na FrameNet, mas pouco explorada: os tipos semânticos.

Os tipos semânticos na FrameNet têm a função de adicionar informação que não é representável através da hierarquia de *frames* e elementos de *frame* (Ruppenhoff et al., 2010). Esses tipos semânticos adicionam três tipos de informação à base de dados: tipos de elementos de *frame*, tipos de *frames* e tipos de unidades lexicais (Ruppenhoff et al., 2010). Os tipos de elementos de *frame*, também chamados de tipos ontológicos, indicam os tipos básicos dos elementos de *frame*. O elemento de *frame* *JUDGE* dos *frames* *Arraignment* e *Trial* apresenta o tipo semântico *Sentiant*, porém o elemento de *frame* *DEFENDANT* desses mesmos *frames* não apresenta nenhum tipo semântico. Os tipos de *frame* indicam a natureza do *frame*,

se é um *frame* não lexical ou um *frame* não perspectivado. Os tipos lexicais incluem informação sobre a natureza semântica da unidade lexical, como os substantivos transparentes.

Na verdade, os tipos semânticos na FrameNet, em especial aqueles que acrescentam informação semântica aos elementos de *frame*, poderiam ser mais bem explorados. Há pouca informação sobre os tipos semânticos em Ruppenhoff et al. (2010) e os elementos de *frames* não apresentam tipos semânticos de forma consistente; alguns elementos de *frame* são anotados com tipos semânticos, outros não. O que os recursos lexicais jurídicos multilíngües poderiam explorar de forma mais consistente é o tipo semântico de cada elemento de *frame*. Os tipos semânticos poderiam trazer informação sobre a função de cada participante²⁶ nos eventos jurídicos, permitindo que a correspondência de *frames* entre os diferentes sistemas jurídicos fosse identificada através da função de seus participantes, e não a partir da definição dos *frames*.

Outro fator importante para a busca da correspondência seria a ordem em que os eventos jurídicos ocorrem em cada sistema. Esse tipo de informação poderia ser codificado através de uma ontologia de domínio jurídico. Assim, uma ontologia poderia indicar que, para o processo penal brasileiro, o *frame* *Prisão* representa a primeira etapa do processo e que o *frame* *Denúncia* representa a segunda etapa. Já para o processo penal americano, o *frame* *Arrest* representa a primeira etapa, enquanto o *frame* *Arraignment* representa a segunda etapa. A partir do cruzamento das informações sobre a função dos elementos de *frame*, ou seja, dos participantes do evento jurídico, e as etapas de cada sistema, seria possível buscar a similaridade de sistema. Encontrar etapas equivalentes entre os sistemas pode ser uma alternativa para os casos em que os *frames* jurídicos não encontram nenhum nível de equivalência entre sistemas jurídicos. No entanto, essas possibilidades necessitam de maior investigação.

O presente capítulo teve por objetivo discutir as implicações computacionais que um estudo contrastivo sobre *frames* jurídicos apresenta para o desenvolvimento de recursos lexicais multilíngües. Para tanto, a seção 5.1 discutiu a visão de *universalidade de frames*

²⁶ O desenvolvimento de ontologias para aplicações computacionais tem tratado desse tema como *papéis sociais* (ver Masolo et al., 2004).

apresentada por muitos trabalhos com o paradigma FrameNet. De acordo com essa posição, os *frames* semânticos seriam independentes de língua por representarem um nível conceitual. Por meio da distinção entre *frames* inatos e aprendidos (Fillmore, 1985), procurou-se uma explicação para a similaridade de alguns *frames* e a diferenças de outros quando contrastados.

A seção 5.2 apresentou os níveis de equivalência de *frames*. A proposta de níveis de equivalência de *frames* surge a partir do trabalho com os *frames* jurídicos em português. Ao se comparar os *frames* jurídicos criados para o português com os *frames* da FrameNet, percebe-se que a natureza da não correspondência de *frames* variava desde a não equivalência de unidades lexicais até a não correspondência dos eventos jurídicos descritos pela FrameNet com os eventos jurídicos brasileiros.

Os *frames* cujo nível de equivalência é mais elevado são os menos problemáticos para os recursos lexicais multilíngües. No entanto, há *frames* que descrevem eventos jurídicos muito particulares de uma determinada sociedade, não apresentando evento correspondente em outro sistema. Esses *frames* são problemáticos para os recursos lexicais multilíngües, uma vez que não há como conectá-los com as outras bases de dados de outras línguas. A proposta apresentada na seção 5.2 foi o uso de tipos semântico e informações ontológicas na tentativa de se criar uma correspondência entre sistemas, mesmo que não haja uma correspondência entre *frames*. O próximo capítulo trata das conclusões a que se chegou e dos desdobramentos futuros deste trabalho.

6 CONSIDERAÇÕES FINAIS

A presente tese teve como objetivo avaliar os desafios a serem enfrentados no desenvolvimento de recursos lexicais multilíngües segundo o paradigma FrameNet. Para atingir esse objetivo, abordou-se, no capítulo 2, a Semântica de *Frames* como uma teoria da lingüística cognitiva e a Semântica de *Frames* como base teórica da FrameNet. Esse capítulo teve, em primeiro lugar, o propósito de elucidar os princípios norteadores da Semântica de *Frames* como uma teoria calcada na visão enciclopédica do significado, em segundo lugar, de apresentar a forma como a FrameNet trata o conceito de *frame* semântico em sua base de dados.

O capítulo 3, por sua vez, tratou das três facetas que dão sustentação metodológica a esta pesquisa. A faceta contrastiva indica que a metodologia aplicada a essa pesquisa foi baseada em estudo contrastivo de unidade lexicais. A faceta jurídica diz respeito à natureza do estudo, ou seja, um estudo de unidades lexicais relacionadas a eventos jurídicos. A faceta computacional aqui diz respeito ao interesse desta pesquisa pelo desenvolvimento de recursos lexicais computacionais, em especial os recursos lexicais jurídicos. Levando em consideração esses três pontos, o capítulo foi estruturado em três seções para discutir em que medida cada uma dessas três facetas interferem no desenvolvimento de recursos lexicais computacionais multilíngües.

Esses três pontos são importantes de serem revistos devido à importância de cada um para o desenvolvimento de recursos lexicais multilíngües. Com relação ao contrastivo, o ponto fundamental a ser retomado aqui é o conceito de equivalência. Embora este trabalho tenha se detido à visão da lingüística contrastiva sobre a equivalência, pode-se dizer que tanto para a lingüística contrastiva, quanto para a tradução, a equivalência é tema de difícil tratamento. As unidades lexicais e os significados mudam entre as línguas, existindo teorias para a tradução que simplesmente negam a existência da equivalência. Dado que os padrões de polissemia entre as línguas variam, é difícil de se estabelecerem equivalentes que

apresentem a mesma rede de polissemia em diferentes línguas, o que faz com que a equivalência perfeita seja difícil de ser atingida.

O fator jurídico é outro ponto de divergência a ser enfrentado no desenvolvimento de léxicos especializados. Devido ao fator social envolvido no Direito e seu caráter normativo aplicado a uma determinada sociedade, é difícil de se encontrar a correspondência entre os eventos jurídicos de países diferentes. Esse fator dificulta ainda mais a busca por equivalentes em uma língua para as unidades lexicais especializadas de outra língua.

Por fim, os léxicos computacionais jurídicos necessitam dar conta dessa falta de equivalência lexical e jurídica. Considerando-se a impossibilidade de se conectar através de relações de equivalência entidades que não são equivalentes, os léxicos computacionais jurídicos representam um desafio para os pesquisadores que trabalham com o desenvolvimento de recursos especializados. Assim, este trabalho, ao propor a investigação de *frames* jurídicos, buscou avaliar em que medida as questões multilíngües e jurídicas são refletidas pelos *frames* semânticos.

Para tratar dos desafios no uso de *frames* semânticos para a descrição de informação jurídica em diferentes línguas, optou-se por investigar o *frame* *Criminal_process*. Em um primeiro momento, identificaram-se, com o auxílio de um dicionário jurídico bilíngüe, os equivalentes em português das unidades lexicais do inglês relacionadas ao processo penal. Em um segundo momento, essas unidades lexicais foram contrastadas para verificar se o conhecimento jurídico evocado pela unidade lexical em inglês era o mesmo conhecimento jurídico evocado pelo seu equivalente em português. A abordagem contrastiva permitiu a criação de *frames* jurídicos descrevendo o processo penal brasileiro.

O alto grau de divergência dos *frames* jurídicos evidencia que os *frames* não são estruturas conceituais independentes de língua. Aqui é importante retomar a orientação cognitiva da Semântica de *Frames* e dos *frames* semânticos. De acordo com a visão da lingüística cognitiva, a língua estrutura o conteúdo conceitual (Talmy, 2000). O conhecimento cultural, por sua vez, também influencia a língua (Holland e Quinn, 1987). Considerando-se que os *frames* são um sistema de conceitos relacionados no qual o conhecimento de um conceito requer o conhecimento de todos os conceitos que formam o sistema (Fillmore, 1985), a concepção de que os *frames* possam ser independentes de língua é uma interpretação simplificada da natureza cognitiva dos *frames* semânticos. Os *frames* são

estruturas de conhecimento organizadas pela língua a partir das experiências sociais, culturais e corporais de um determinado grupo social.

Os diferentes níveis de equivalência de *frames* evidenciam que não há um limite claro entre ***frames* universais** e ***frames* dependentes de língua**. A própria dicotomia universal *versus* lingüístico reacende a dicotomia defendida pela abordagem gerativista entre gramática *versus* léxico. A gramática como capacidade universal e inata seria o *lugar* das regularidades, enquanto o léxico seria o *lugar* da irregularidade. Os pares de *frames* jurídicos analisados apresentam diferentes níveis de equivalência. Enquanto, os pares com maior equivalência apresentam correspondência de evento jurídico, de elementos de *frame* e de unidades lexicais, os pares com menor equivalência apresentam apenas correspondência de unidades lexicais, ou seja, as unidades lexicais apresentam equivalentes em português, mas o evento jurídico não é equivalente. Assim, a diferenciação entre *frames* universais e *frames* dependentes de língua não apresenta uma explicação consistente sobre a equivalência de *frames* entre línguas.

Este trabalho procurou explicar as divergências de *frames* semânticos por meio da diferenciação entre ***frames* inatos** e ***frames* aprendidos** (Fillmore, 1985). Os *frames* inatos são estruturados a partir da relação do ser humano com seu próprio corpo. Assim, o sistema perceptual, motor e cognitivo são fundamentais para o desenvolvimento dos *frames* inatos. Já os *frames* aprendidos são estruturados através da relação do ser humano com a sociedade. É por meio do conhecimento do mundo que o ser humano desenvolve esses *frames*. Nesse sentido, os *frames* jurídicos são evidências de *frames* aprendidos, uma vez que o conhecimento descrito por esses *frames* é tão especializado que apenas um determinado grupo da sociedade tem *domínio* sobre esses *frames*. No entanto, essa explicação é uma divisão categórica entre *frames*. Os *frames* inatos não estão livres da influência cultural. Os *frames* aprendidos, por sua vez, não são sempre divergentes.

Uma vez que se abre mão da busca por *frames* com equivalência perfeita entre as línguas, é necessário se repensar o uso dos *frames* semânticos como interlíngua em recursos lexicais multilíngües, especialmente se esses recursos representarem conhecimento jurídico. Como os *frames* jurídicos descrevem um conhecimento socialmente construído, há *frames* que simplesmente não existem em determinadas sociedades.

Na introdução deste trabalho, apresentou-se a seguinte pergunta de pesquisa: É possível a criação de uma FrameNet jurídica para o português a partir dos *frames* semânticos propostos para as unidades lexicais do inglês? Pode-se concluir, com base na análise

apresentada aqui, que os *frames* semânticos podem ser aplicados com sucesso à construção de recursos lexicais computacionais jurídicos. No entanto, em caso de recursos multilíngües, a metodologia adotada para a construção de recursos jurídicos deve ser diferente da metodologia aplicada à criação de recursos de língua geral. O estudo contrastivo dos *frames* apresentado aqui demonstrou a falta de equivalência de muitos *frames* jurídicos. Essa falta de equivalência de *frames* dificulta a expansão dos *frames* jurídicos já desenvolvidos pela FrameNet para outras línguas.

A metodologia de expansão dos *frames* em inglês para o português poderia gerar resultados distorcidos, dadas as diferenças entre os sistemas jurídicos norte-americano e brasileiro. Este trabalho defende a idéia de que a criação de recursos lexicais jurídicos a partir da anotação manual pode resultar em uma melhor descrição do sistema jurídico de um país. Além disso, este trabalho também defende a idéia de a criação de recursos lexicais jurídicos ser monolíngüe e a ligação entre as bases de dados de diferentes línguas se dar de forma posterior a sua criação.

Apesar de concluir esta pesquisa aqui, o assunto não está esgotado. Este trabalho apontou como uma possível solução para a falta de equivalência entre *frames* o uso de um recurso ainda pouco abordado pela FrameNet: os **tipos semânticos**. Os tipos semânticos poderiam ser utilizados em uma base de dados jurídica multilíngüe para marcar o papel social exercido por cada participante dos eventos jurídicos. Uma ontologia jurídica poderia ajudar a organizar o conhecimento referente à função de cada participante. Este trabalho representa uma orientação metodológica para o desenvolvimento de um recurso lexical baseado em *frames* para a linguagem jurídica brasileira.

Levando-se em consideração as conclusões a que se chegou com este trabalho, pretende-se iniciar a criação de um recurso lexical jurídico monolíngüe. Isso não exclui a possibilidade futura de uma conexão a outra base de dados lexicais jurídica para a criação de um recurso multilíngüe. No entanto, dadas as especificidades do sistema judiciário de cada país, acredita-se que a criação de bases de dados independentes é mais produtiva, visto que, no caso do domínio jurídico, os *frames* apresentam muita divergência. Para a criação de um recurso lexical baseado em *frames* para a linguagem jurídica brasileira seria necessária, primeiramente, a criação de um *corpus* jurídico representativo contendo leis, doutrinas e jurisprudências, pois o Brasil ainda não dispõe de um *corpus* especializado. Assim, como desdobramentos futuros desta tese, apontam-se a criação de um *corpus* especializado e a criação de um recurso lexical baseado em *frames* para a linguagem jurídica brasileira.

REFERÊNCIAS

ALTENBERG, B. Adverbial connectors in English and Swedish: Semantic and lexical correspondences. In: HASSELGARD, H.; OKSEFJELL, S. (Eds.). *Out of corpora. Studies in honour of Stig Johansson*. Amsterdam/Atlanta: Rodolpi, 1999.

ALTERNBERG, B.; GRANGER, S. (Eds.). *Lexis in Contrast: Corpus-based Approaches*. Amsterdam: John Benjamins, 2002.

ARISTOTLE. *The Metaphysics*. With translation by H. Tredennick. Cambridge: Harvard, 1933.

BERLIN, B.; KAY, P. *Basic color terms: their universality and evolution*. Berkeley / Los Angeles: University of California Press, 1969.

BERTOLDI, A. *A semântica dos adjetivos: como e por que incluí-la em uma ontologia de domínio jurídico*. Dissertação (Mestrado em Linguística Aplicada). UNISINOS, São Leopoldo, 2007.

BICK, E. The Parsing System "Palavras": Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework. Dr.phil. thesis. Aarhus University. Aarhus, Denmark: Aarhus University Press. November 2000.

BRASIL. LEIS, decretos, etc. *Código de processo penal*. 10. ed. São Paulo: Revista dos Tribunais, 2005. 1046 p.

BRASIL. *Constituição da República Federativa do Brasil*. 26. ed. São Paulo: Saraiva, 2000. 311 p.

BLACK'S LAW DICTIONARY. Edited by Bryan A. Garner. 7th ed. St. Paul: West Group, 1999, 1738p.

BOAS, H. C. Bilingual FrameNet Dictionaries for Machine Translation. THIRD INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION. In Proceedings of the Third International Conference on Language Resources and Evaluation. Las Palmas: ELRA. Vol. IV, 2002, p.1364-1371.

_____. Semantic Frames as Interlingual Representations for Multilingual Lexical Databases. *International Journal of Lexicography*. 2005, Vol. 18, No.4, p. 445-478.

_____. (Ed.) *Multilingual FrameNets in computational lexicography – methods and applications*. Berlin/New York: Mouton de Gruyter, 2009a.

_____. Recent trends in multilingual computational lexicography. In: BOAS, H. C. (Ed.) *Multilingual FrameNets in computational lexicography – methods and applications*. Berlin/New York: Mouton de Gruyter, 2009b, p.1-34.

BURCHARDT, A.; ERK, K.; FRANK, A.; KOWALSKI, A.; PADÓ, S. SALTO – a versatile multi-level annotation tool. In: Proceedings of the 5th International Conference on Language Resources and Evaluation, 2006.

BURCHARDT, A.; ERK, K.; FRANK, A.; KOWALSKI, A.; PADÓ, S.; PINKAL, M. Using FrameNet for the semantic analysis of German: annotation, representation, and automation. In: BOAS, H. C. (Ed.) *Multilingual FrameNets in computational lexicography: Methods and applications*. Berlin/New York: Mouton de Gruyter, 2009, p.209-244.

CABRÉ, M. T. *La terminología. Teoría, metodología, aplicaciones*. Barcelona: Editorial Antártida/Empúries, 1993.

CALZOLARI, N.; BRISCOE, T. ACQUILEX-I and -II: Acquisition of lexical knowledge from machine readable dictionaries and text corpora. *Cahiers Lexicologique*, Vol. 67, n.2, 1995, 95–114.

CHESTERMAN, A. *Contrastive functional analysis*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 1998.

CHISHMAN, R.; TEIXEIRA, L.; BERTOLDI, A. Anotação semântica: um experimento a partir da semântica de frames. In: VI Encontro de Lingüística de Corpus, 2007, São Paulo. VI Encontro de Lingüística de Corpus. São Paulo: USP-UFSCAR, 2007. v. 1. p. 25-26

CHISHMAN, R. L. O.; BERTOLDI, A.; PADILHA, J. G.; LERMEN, L. Corpus e Anotação Semântica: um Experimento para a Língua Portuguesa a partir da Semântica de Frames. In: VI WORKSHOP EM TECNOLOGIA DA INFORMAÇÃO E DA LINGUAGEM HUMANA. XIV Simpósio Brasileiro de Sistemas Multimídia e Web. Vila Velha: SBSC, 2008, p.321-325.

COMRIE, B. *Language universals and linguistic typology: Syntax and morphology*. 2nd ed. Oxford: Blackwell, 1989.

CROFT, W. *Typology and universals*. 2nd ed. Cambridge: Cambridge University Press, 2002.

CROFT, W. The role of domains in the interpretation of metaphors and metonymies. In: GEERAERTS, D. *Cognitive linguistics: basic readings*. Berlin: Walter de Gruyter, 2006.

CROFT, W.; CRUSE, D. A. *Cognitive Linguistics*. Cambridge: Cambridge University Press, 2004.

CRUSE, A. *Lexical Semantics*. Cambridge: Cambridge University Press, 1986.

DINI L.; BOSCA, A. *Dependency Based Valence Induction for an Italian FrameNet*, In: Proceedings of GL2009, p. 27-35, 2009.

DINI, L.; PETERS, W.; LIEBWALD, D.; SCHWEIGHOFER, E.; MOMMERS, L.; VOERMANS, W. Cross-lingual legal information retrieval using a WordNet architecture. In: INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND LAW. Proceedings of the 10th International Conference on Artificial Intelligence and Law, Bologna. ACM Press: New York, 2005, p.163-167.

DIRVEN, R.; FRANK, R.; PÜTZ, M. (Eds.) *Cognitive Models in Language and Thought*. Ideology, Metaphors, and Meanings. Berlin: Mouton de Gruyter, 2003

DOLBEY, A.; ELLSWORTH, M.; SCHEFFCZYK, J. BioFrameNet: A Domain-specific FrameNet Extension with Links to Biomedical Ontologies. In: Proceedings of the Biomedical Ontology in Action” Workshop at KR-MED. 2006, p.87-94.

EVANS, V.; BERGEN, B. K.; ZINKEN, J. The Cognitive Linguistics Enterprise: An Overview. In: EVANS, V.; BERGEN, B. K.; ZINKEN, J. (Eds). *The Cognitive Linguistics Reader*. London: Equinox Publishing Co. 2007

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

FAUCONNIER, G. *Mental Spaces*. Cambridge, MA: MIT Press, 1985.

_____. *Mappings in Thought and Language*. Cambridge: Cambridge University Press, 1997.

FAUCONNIER, G.; TURNER, M. Blending as a central process of grammar. In: GOLDBERG, A. (Ed.). *Conceptual Structure, Discourse and Language*. Stanford: CSLI Publications, 1996, p113-130.

_____. Conceptual Integration Networks. *Cognitive Science*. Vol. 22, N°2, 1998, p. 133-187.

FELLBAUM, C. A semantic network of English: the mother of all wordnets. *Computers and the Humanities*. v.32, n.2-3, 1998, p.209-220.

FILLMORE, C. J. The case for case. In: BACH, E.; HARMS, R. T. (Eds.) *Universals in Linguistic Theory*. Volume 67. New York: Holt, Rinehart and Winston, 1968, p.1-88.

_____. 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.

_____. 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.

_____. 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.

_____. Frame semantics. In: The Linguistic Society of Korea (Eds.). *Linguistics in the Morning Calm*. Seoul: Hanshin, 1982, p.111-37.

_____. Frames and the semantics of understanding. *Quaderni di Semantica*. Vol.6, N.2, 1985, p.222-254.

_____. Valency Issues in FrameNet. In HERBST, T.; GÖTZ-VOTTELER, K. (Eds.). *Valency: Theoretical, Descriptive and Cognitive Issues*. Berlin, New York: Mouton de Gruyter, 2007, p.129-160.

_____. Border Conflicts: FrameNet Meets Construction Grammar. THE XIII EURALEX INTERNATIONAL CONGRESS. In: BERNAL, E.; DECEASARIS, J. (Eds.). *Proceedings of the XIII EURALEX International Congress*. Barcelona: Edicions a Petició, 2008, p. 49-68.

FILLMORE, C. J.; BAKER, C. A frames approach to semantic analysis. In: HEINE, B.; NARROG, H. (Eds.). *The Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press, 2010, p.313-339.

FILLMORE, C. J.; JOHNSON, C. R.; PETRUCK, M. R. L. Background to FrameNet. *International Journal of Lexicography*. Vol.16, N°3, 2003, p.235-250.

- FILLMORE, C. J.; KAY, P.; O'CONNOR, M. C. Regularity and idiomaticity in grammatical constructions. *Language*, Vol.64, No.3, 1988, p. 501-538.
- FISIAK, J.; LIPÍŃSKA-GRZEGOREK, M.; ZABROCKI, T. *An introductory English-Polish contrastive grammar*. Warszawa: Państwowe Wydawnictwo Naukowe, 1978.
- FONTENELLE, T. A Bilingual Lexical Database for Frame Semantics. *International Journal of Lexicography*, Vol.13, N°. 4, 2000, p.232 - 248.
- GEERAERTS, D.; CUYCKENS, H. *The Oxford Handbook of Cognitive Linguistics*. New York: Oxford University Press, 2007.
- GOFFMAN, E. *Frame Analysis*. New York: Harper, 1974
- GOLDBERG, A. *Constructions*. A Construction Grammar Approach to Argument Structure. Chicago: The University of Chicago Press, 1995.
- GOLDBERG, A. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press, 2006.
- GOYOS JÚNIOR, D. N. *Noronha's Legal Dictionary – Noronha Dicionário Jurídico: English-Portuguese, Portuguese-English – Inglês-Português, Português-Inglês*. 1.ed. São Paulo: Observador Legal, 1992.
- GRANGER, S. The corpus approach: a common way forward for contrastive linguistics and translation studies? In: GRANGER, S.; LEROT, J.; PETCH-TYSON, S. (Eds.) *Corpus-based approaches to contrastive linguistics and translation studies*. Amsterdam/New York: Rodolpi, 2003.
- HIRAGA, M.; SINHA, C.; WILCOX, S. (Eds.) *Cultural, Psychological and Typological Issues in Cognitive Linguistics*. Amsterdam: John Benjamins, 1999.
- HOLLAND, D.; QUINN, N. (Eds.). *Cultural Models in Language and Thought*. Cambridge: Cambridge University Press, 1987.
- RUPPENHOFER, J.; ELLSWORTH, M.; PETRUCK, M. R. L.; JOHNSON, C. R.; SCHEFFCZYK, J. *FrameNet II: Extended Theory and Practice*. Berkeley: International Computer Science Institute, 2010. Available at: <http://framenet.icsi.berkeley.edu/>
- JAMES, C. *Contrastive analysis*. Essex: Longman, 1980.
- JOHANSSON, R.; NUGUES, P. Using parallel corpora for automatic transfer of FrameNet annotation. In: Proceedings of the 1st ROMANCE FrameNet Workshop, Cluj-Napoca, Romania. 2005.

_____. A FrameNet-Based Semantic Role Labeler for Swedish. In: Proceedings of the joint ACL and COLING, 436–443, Sydney, Australia. 2006.

JOHNSON, M. *The Body in the Mind. The Bodily Basis of Meaning, Imagination, and Reason*. Chicago: The University of Chicago Press, 1987.

LAKOFF, G.; MARK JOHNSON, M. *Philosophy in the flesh: the embodied mind and its challenge to Western thought*. New York: Basic Books, 1999.

KADE, O. *Zufall und Gesetzmässigkeit in der Übersetzung*. Leipzig: VEB Verlag Enzyklopädie, 1968.

KATZ, J. J.; FODOR, J. A. The structure of a semantic theory. *Language*. Vol.39, N°.2, 1963, p.170-210.

KRZESZOWSKI, T. P. *Contrasting languages: The scope of contrastive linguistics*. Berlin/New York: Mouton de Gruyter, 1990.

LABOV, W. The Boundaries of Words and their Meanings. In: BAILEY, C. N.; SHUY, R. *New Ways of Analyzing Variation in English*. Washington D.C.: Georgetown University Press, 1973, p. 340-373.

LAFAVE, W. R.; ISRAEL, J. H.; KING, N. J.; KERR, O. S. *Principles of Criminal Procedure: post-investigation*. 2ed. St. Paul: West, 2009, 821p.

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, R. *Foundations of Cognitive Grammar, Volume I, Theoretical Prerequisites*. Stanford: Stanford University Press, 1987.

_____. A Usage-based Model. In: RUDZKA-OSTYN, B. (Ed.) *Topics in Cognitive Linguistics*. Amsterdam: John Benjamins, 1988, p.127-161.

_____. *Concept, Image, and Symbol: The Cognitive Basis of Grammar*. Berlin: Mouton de Guyter, 1990.

_____. *Foundations of Cognitive Grammar, Volume II, Descriptive Application*. Stanford: Stanford University Press, 1991.

LENCI, A.; BEL, N.; BUSA, F.; CALZOLARI, N.; GOLA, E.; MONACHINI, M.; OGOROWSKI, A.; PETERS, I.; PETERS, W.; RUIMY, N.; VILLEGAS, M.; ZAMPOLLI, A. SIMPLE: a general framework for the development of multilingual lexicons. *International Journal of Lexicography*. v.13, n.4, p.249-263, 2000

LÖNNEKER-RODMAN, B. *Multilinguality and FramNet*. Technical Report. TR-07-001. Berkeley: ICSI, 2007.

MASOLO, C., VIEU, L., BOTTAZZI, E., CATENACCI, C., FERRARIO, R., GANGEMI, A., GUARINO, N. Social Roles and their Descriptions. In: DUBOIS, D; WELTY, C.; WILLIAMS, M. A. (eds.), *Proceedings of the Ninth International Conference on the Principles of Knowledge Representation and Reasoning (KR2004)*, Whistler, Canada, June 2-5, 2004, p. 267-277.

MAGALHÃES, H. P.; MALTA, C. P. T. *Dicionário jurídico*. 4a ed. rev., atualizada e ampliada. 2 Vol. Rio de Janeiro: Edições Trabalhistas, 1984, 946p.

MATTILA, H. *Comparative Legal Linguistics. Target*. Vol. 21. Hampshire: ASHGATE, 2006.

MILLER, G. A. WordNet: a lexical database for English. *Communications of the ACM*. New York: ACM Press. v.38, n.11, 1995, p.39-41.

MINSKY, M. *A framework for representing knowledge*. Artificial Intelligence Memo N°. 306. Cambridge, MA: Massachusetts Institute of Technology, 1974.

OLIVEIRA, A. R. Equivalência: sinônimo de divergência. *Cadernos de Tradução*, Vol. 1, No 19, 2007, p.97-114.

THE NEW OXFORD AMERICAN DICTIONARY. 2nd ed. New York: Oxford University Press, 2005.

OHARA, K. H. Frame-based contrastive lexical semantics in Japanese FrameNet: The case of risk and kakeru. In: BOAS, H. C. (Ed.) *Multilingual FrameNets in computational lexicography: Methods and applications*. Berlin/New York: Mouton de Gruyter, 2009, p.163-182.

OHARA, K. H.; FUJII, S.; SAITO, H.; ISHIZAKI, S.; OHORI, T.; SUZUKI, R. The Japanese FrameNet Project: A Preliminary Report. In *PROCEEDINGS OF PACIFIC ASSOCIATION FOR COMPUTATIONAL LINGUISTICS (PACLING'03)*, 249-254. Halifax, Canada. August, 2003.

THE OXFORD COMPANION TO AMERICAN LAW. Edited by Kermit L. Hall, David S. Clark, James W. Ely, Jr., Joel Grossman, N. E. H. Hull. New York: Oxford, 2002, 912p.

PADÓ, S. *Cross-lingual Annotation Projection Models for Role-Semantic Information*. PhD Thesis. Saarbrücken: Universität des Saarlandes, 2007.

PADÓ, S.; LAPATA, M. Cross-lingual projection of role-semantic information. HUMAN LANGUAGE TECHNOLOGY CONFERENCE AND CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE. In: Proceedings of HLT/EMNLP-05, Vancouver: Association for Computational Linguistics, 2005, p.859-866.

PADÓ, S.; PITEL, G. Annotation précise Du français en sémantique de rôles par projection cross-linguistique. Proceedings of TALN 2007, Toulouse, France, 2007.

PITEL, G. Cross-lingual labeling of semantic predicates and roles: A low-resource method based on bilingual L(atent) S(ematic) A(nalysis). In: BOAS, H. C. (Ed.) *Multilingual FrameNets in computational lexicography: Methods and applications*. Berlin/New York: Mouton de Gruyter, 2009, p.245-284.

PYM, A. *Exploring translation theories*. London/New York: Routledge, 2010.

RENSTROM, P. *The American Law Dictionary*. Santa Barbara/Oxford: ABC-CLIO, 1991, 308p.

REY, A. *Essays on terminology*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 1995.

ROSCH, E. Natural Categories. *Cognitive Psychology*. Vol.4, N°.3, 1973, p.328-350.

ROSCH, E. Cognitive representations of semantic categories. *Journal of Experimental Psychology: General* Vol. 104, N°.3, 1975, p.192-233.

ROSCH, E. Principles of Categorization. In: ROSCH, E.; LLOYD, B. B. (Eds.). *Cognition and Categorization*. Hillsdale: Lawrence Erlbaum Associates, 1978, p. 27-48.

ROSCH, E.; MERVIS, C. B. Family resemblances: Studies in the internal structure of categories. *Cognitive psychology*. Vol.7, N°.4, 1975, p.573-605.

ROSCH, E.; SIMPSON, C.; MILLER, R. S. Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance*. Vol.2, N°.4, 1976, p.491-502.

SAGRI, M. T.; TISCORNIA, D. Semantic lexicons for accessing legal information. In: TRAUNMÜLLER, R. (ED.) *Electronic Government*. LNCS3183. Berlin/Heidelberg: Springer-Verlag, 2004, p. 72-81.

SALOMÃO, M. M.M. FrameNet Brasil: um trabalho em progresso. *Calidoscópico*. Vol.7, No.3, 2009, p.171-182.

SARCEVIC, S. *New Approach to Legal Translation*. The Hague/London/Boston: Kluwer Law International, 1997.

SCHMIDT, T. The Kicktionary – A multilingual lexical resource of football language. In: BOAS, H. C. (Ed.) *Multilingual FrameNets in computational lexicography: Methods and applications*. Berlin/New York: Mouton de Gruyter, 2009, p.102-132.

SNELL-HORNBY, M. *Verb-descriptivity in German and English: A contrastive study in semantic fields*. Heidelberg: Carl Winter - Universitätsverlag, 1983.

SPALATIN, L. Contrastive methods. *Studia Romanica et Anglica Zagrabiensia*. Vol.23, 1967, p.29-45.

STOCKWELL, R. P.; BOWEN, J. D.; MARTIN, J. W. *The grammatical structures of English and Spanish*. Chicago/London: The University of Chicago Press, 1965.

SUBIRATS, C.; PETRUCK, M. Surprise: Spanish FrameNet. INTERNATIONAL CONGRESS OF LINGUISTS. WORKSHOP ON FRAME SEMANTICS, Prague (Czech Republic), July 2003.

SUBIRATS, C. Spanish FrameNet: A frame-semantic analysis of the Spanish lexicon. In: BOAS, H. C. (Ed.) *Multilingual FrameNets in computational lexicography: Methods and applications*. Berlin/New York: Mouton de Gruyter, 2009, p.136-162.

SWEETSER, E. E. The definition of lie: an examination of the folk models underlying a semantic prototype. In: HOLLAND, D.; NAOMI QUINN, N. (eds.). *Cultural models in language and thought*, 43-66. Cambridge: Cambridge University Press, 1987.

TALMY, L. *Toward a Cognitive Semantics*. Volume I: Concept Structuring Systems. MIT: Cambridge (MA)/London, 2000.

TALMY, L. *Toward a Cognitive Semantics*. Volume II: Typology and Process in Concept Structuring. MIT: Cambridge (MA)/London, 2000.

TESNIÈRE, L. *Éléments de syntaxe structurale*. Paris: Klincksieck, 1959.

TEUBERT, W. Comparable or parallel corpora? In: SINCLAIR, J., PAYNE, J.; PÉREZ-HERNADÉZ, C. (eds.). *International Journal of Lexicography*. Corpus to corpus: a study of translation equivalence. Vol. 9, N. 3, 1996, p. 238-264.

TONELLI, S.; PIANTA, E. Frame Information Transfer from English to Italian. THE SIXTH INTERNATIONAL LANGUAGE RESOURCES AND EVALUATION (LREC'08). In: Proceedings of the Sixth International Language Resources and Evaluation (LREC'08). Marrakech: European Language Resources Association (ELRA), p.28-30, 2008.

UNGERER, F.; SCHMID, H. *An Introduction to Cognitive Linguistics*. 2nd ed. London: Pearson Longman, 2006.

UNITED STATES. *The Constitution of the United States*. For the use of schools and academies. By Geo. S. Williams, A.M. 3rd ed. Cambridge: Welch, Bigelow, and Company: 1862, 199p.

VARELA, F. J.; THOMPSON, E.; ROSCH, E. *The embodied mind: cognitive science and human experience*. Cambridge (MA): MIT, 1991.

VENTURI, G. Towards an Italian Legal FrameNet: Exploring evidence for extending and specializing FrameNet to the legal domain. FRAMENET MASTERCLASS AND WORKSHOP. In: The Eighth International Workshop on Treebanks and Linguistic Theories. Università Cattolica del Sacro Cuore: Milano, 2009.

VENTURI, G.; LENCI, A.; MONTEMAGNI, S.; VECCHI, E.; SAGRI, M.; TISCORNIA, D.; AGNOLONI, T. Towards a FrameNet Resource for the Legal Domain. In *Proceedings of the Third Workshop on Legal Ontologies and Artificial Intelligence Techniques, Workshop on Semantic Processing of Legal Texts (LOAIT '09)*. Barcelona, Spain. 8 June, 2009.

VERHAGEN, M.; SPRANGER, K. Extracting term candidates from text corpora using frame semantics. In: HACKEN, P. T. (ed.). *Terminology, computing and translation*. Tübingen: Gunter Narr Verlag, 2006.

VOSSSEN, P. Introduction to EuroWordNet. *Computers and the Humanities*. v.32, n.2-3, p.73-89, 1998.