

Desafios para a Criação de um Léxico baseado em *Frames* para o Português: um estudo dos *frames Judgment e Assessing*

Anderson Bertoldi, Rove Luiza de Oliveira Chishman

Programa de Pós-graduação em Linguística Aplicada
Universidade do Vale do Rio dos Sinos (UNISINOS) – São Leopoldo, RS – Brasil

andersonbertoldi@yahoo.com, rove@unisinós.br

Abstract. *This paper presents a comparative study of Judgment and Assessing frames in English and Portuguese. The aim is to verify the possibility of using the FrameNet frames to construct a lexical database for Brazilian Portuguese. The research corpus is composed by 50 legal documents, totalizing 1.055,535 tokens and 39,108 types. Through a contrastive method the Judgment and Assessing frames were selected and translation equivalents for the English lexical units were established. The points considered in this research were the polysemy and the semantic relations of words. The polysemy is the main difficulty in applying FrameNet frames for Portuguese description.*

Resumo. *Este artigo apresenta um estudo comparativo entre o inglês e o português dos frames Judgment e Assessing. O objetivo é analisar em que medida é possível transpor os dados lexicais do inglês para o português, utilizando o FrameNet como ponto de partida. O corpus da pesquisa é formado de 50 textos jurídicos, totalizando 1.055.535 tokens e 39.108 types. O método de pesquisa consiste em selecionar um frame e encontrar equivalentes em português. Os equivalentes de tradução são comparados com os itens lexicais do inglês. A polissemia pode ser considerada uma das principais dificuldades na transposição dos dados do FrameNet para o português ao partir-se de um frame e não de uma unidade lexical.*

1. Introdução

Um crescente interesse tem motivado a aplicação do paradigma FrameNet à construção de léxicos de diferentes línguas, tais como o German FrameNet (Boas, 2002; 2005), o Spanish FrameNet (Subirats e Petruck, 2003) e o Japanese FrameNet (Hasegawa et al., 2006). Estudos comparativos voltados à investigação da polissemia das línguas, dos padrões de lexicalização, dos padrões valenciais e das paráfrases e equivalentes de tradução, aplicando a Semântica de *Frames*, vêm sendo realizados em diferentes línguas.

O presente trabalho avalia o método de criação de bases de dados utilizando o paradigma FrameNet. Esse método pode ser resumido em: (i) estudo de um determinado *frame* semântico e dos itens lexicais da língua inglesa que já foram codificados pelo FrameNet, (ii) procura de equivalentes de tradução na língua-alvo e (iii) estudo comparativo entre o inglês e a língua-alvo propondo soluções para as dificuldades computacionais de transferências de dados lexicais de uma língua a outra. O motivo de se partir dos *frames* é a procura de uma coerência estrutural entre a base de dados americana e suas versões multilíngües. Porém, o que se discute neste trabalho é que, devido a uma baixa cobertura de unidades lexicais no FrameNet, tal método de trabalho

pode ocasionar uma descrição insuficiente das unidades lexicais da língua-alvo. O que se propõe é uma extensão manual do FrameNet a partir de unidades lexicais, não de *frames*.

Para tratar desse tema, o presente trabalho está estruturado em cinco seções. Na seção 1, apresentam-se as diretrizes gerais do método de extensão multilíngües de bases de dados segundo o paradigma FrameNet. Na seção 2, discutem-se os termos fundamentais para o trabalho com o FrameNet. A seção 3 traz um estudo sobre os verbos de julgamento, a partir do qual se realiza toda a discussão apresentada neste trabalho e se avalia o método corrente de criação de bases de dados lexicais seguindo o paradigma FrameNet. Na seção 4, dois *frames* semânticos são comparados, *Judgment* e *Assessing*. A seção 5 apresenta as conclusões deste trabalho, que servem como roteiro metodológico para criação de bases de dados lexicais utilizando o paradigma FrameNet.

2. O FrameNet e a Semântica de *Frames*

A base de dados lexicais FrameNet, iniciada em 1997, é estruturada segundo a Semântica de *Frames* (Fillmore, 1982). A Semântica de *Frames* é considerada um programa de pesquisa em semântica empírica que, além de enfatizar as continuidades entre a linguagem e a experiência, providencia uma estrutura para a apresentação dos resultados da pesquisa (Fillmore, 1982). Os *frames* são pequenas “cenas” ou “situações” abstratas, de forma que, para entender a estrutura semântica de um verbo, é necessário entender as propriedades de tais cenas esquematizadas. Assim, pode-se dizer que os *frames* são estruturas esquemáticas de situações e dos participantes das situações (Fillmore, 1982).

A base de dados lexicais FrameNet pode ser considerada como essa estrutura para apresentação dos resultados. A base de dados lexicais FrameNet, disponível on-line (www.framenet.icsi.berkeley.edu), contém mais de 10.000 unidades lexicais da língua inglesa, sendo mais de 6.100 totalmente anotadas, e um conjunto de mais de 940 *frames*. No FrameNet, as unidades de análise lexical fundamentais são o *frame* e a *unidade lexical*. “Quando o sentido de uma palavra está baseado em um *frame* particular, nós dizemos que a palavra *evoca o frame*” (Fillmore et al., 2003, p. 236). A *unidade lexical* é concebida como uma unidade composta por uma palavra e um sentido. Ou seja, no FrameNet, cada sentido diferente de uma palavra representa uma unidade lexical distinta.

A descrição do léxico da língua inglesa leva em consideração a valência sintática e semântica dos itens lexicais. A valência sintática especifica os tipos frasais e as funções gramaticais. A valência semântica é descrita em termos de entidades de um *frame* evocado por uma unidade lexical, tais entidades são chamadas de “elementos *frame*” (Fillmore et al., 2003). A unidade lexical *acusar* é a evocadora do *frame Judgment* (*Julgamento*). A descrição semântica do *frame Judgment* inclui os elementos *frame*: *Cognizer*, *Evaluee*, *Expressor* e *Reason*. A descrição da estrutura esquemática representada por um *frame* é feita através dos elementos *frame*.

3. Os Verbos de Julgamento

Em seu trabalho *Verbs of Judging: an exercise in semantic description*, Fillmore (1971) apresenta um estudo dos verbos *accuse*, *blame*, *criticize*, *credit*, *praise*, *scold*, *confess*, *apologize*, *forgive*, *justify* e *excuse*. O critério utilizado para classificar esses verbos foi a

estrutura de papéis. Ou seja, os verbos de julgamento se referem a uma determinada **situação** que afeta favoravelmente ou desfavoravelmente um indivíduo, o **afetado**. Há alguém a quem é dada a responsabilidade pela situação, o **acusado**. O indivíduo que faz o julgamento moral da situação é chamado de **juiz**. O juiz pode simplesmente fazer um julgamento e guardá-lo para si, ou pode fazer alguma afirmação sobre a situação. Nesse caso, há a pessoa a quem a afirmação é endereçada, que é chamada de **destinatário**. O emitente do julgamento é a **fonte elocutiva** e o destinatário é o **alvo elocutivo**.

A força do trabalho de Fillmore (1971) está na descrição do significado de cada verbo através dos seus participantes e dos pressupostos subjacentes a cada verbo. Considerando-se o significado dos verbos *accuse* (*acusar*) e *criticize* (*criticar*), conforme Fillmore (1971), *accuse* é uma **situação** que envolve um **juiz** e um **acusado**. O verbo *accuse* pressupõe uma situação ruim, mas não há a pressuposição de que o **acusado** seja de fato responsável pela situação. Já o verbo *criticize*, que envolve os mesmos participantes pressupõe a responsabilidade do **acusado** pela situação ruim.

Porém, há algumas informações importantes sobre essa classe de verbos que não são sistematizadas em Fillmore (1971). Verbos como *excuse* (*desculpar(se)*), *justify* (*justificar(se)*), *confess* (*confessar*), *apologize* (*desculpar-se*) e *forgive* (*desculpar*) não possuem o **juiz** como um de seus participantes. Ou seja, verbos como *excuse*, *justify*, *confess* e *apologize* levam em consideração a perspectiva do **acusado**. Por exemplo, é o **acusado** quem confessa ter cometido alguma injúria, não o **afetado**. Verbos como *forgive* levam em consideração o ponto de vista do **afetado**. Cabe ao **afetado** a decisão de esquecer ou não uma ofensa. No FrameNet, esses verbos não estão agrupados no *frame Judgment*, talvez por não levarem em consideração justamente o ponto de vista do **juiz**, aquele que faz o julgamento da **situação**.

Na verdade, os verbos que levam em conta a perspectiva do **juiz** são *accuse* (*acusar*), *blame* (*culpar*), *criticize* (*criticar*), *credit* (*acreditar*), *praise* (*elogiar*) e *scold* (*ralhar*). Desses verbos que demonstram a avaliação do **juiz**, *accuse*, *blame*, *criticize*, *credit* e *scold* demonstram uma avaliação negativa. Há também os verbos que expressam atos de fala, como *praise* (*elogiar*) e *scold* (*ralhar*). Ou seja, não há apenas a avaliação, mas a avaliação é comunicada.

O FrameNet não descreve nem codifica as pressuposições, porém organiza melhor os verbos de julgamento, agrupando-os em novos frames. O *frame* semântico mais representativo para os verbos de julgamento é o *frame Judgment* (*julgamento*). Os verbos que comunicam um julgamento evocam *frames* mais específicos que o *frame Judgment*, como os *frames Judgment_direct_address* (*endereço direto de julgamento*) e *Judgment_communication* (*comunicação de julgamento*). O quadro abaixo apresenta a localização dos verbos de julgamento no FrameNet e os *frames* semânticos que cada verbo evoca:

Como podemos ver no quadro, os verbos *excuse* (*desculpar(se)*), *justify* (*justificar(se)*), *confess* (*confessar*), *apologize* (*desculpar-se*) e *forgive* (*desculpar*), que não possuem o **juiz** na sua *estrutura de papéis*, não são, na verdade, verbos de julgamento, tanto que são agrupados sob outros *frames*. Os verbos *blame* e *accuse* podem evocar, cada um, dois *frames* distintos: *Judgment* e *Judgment_communication*. Isso significa que esses verbos são polissêmicos, possuindo dois significados, um de julgamento e outro de comunicação de julgamento.

Quadro 1. Verbos de julgamento no FrameNet

Unidade Lexical	Frame
Accuse	Judgment / Judgment_communication
Apologize	---
Blame	Judgment / Judgment_communication
Confess	Reveal_secret
Credit	---
Criticize	Judgment_communication
Excuse	Forgiveness
Forgive	Forgiveness
Justify	Justifying
Praise	Judgment_communication
Scold	Judgment_direct_address

Este estudo parte do *frame Judgment* e, através dos dados lexicográficos, compara-o com o *frame Assessing* (*Avaliação*), outro *frame* importante no estudo do *corpus* utilizado nesta pesquisa. Os *frames Judgment* e *Assessing* são semanticamente muito próximos, porém, nenhuma relação os liga no FrameNet. O FrameNet traz a seguinte definição para o *frame Assessing*: “Um **Avaliador** examina um **Fenômeno** para calcular seu **Valor** de acordo com algumas **Características** do **Fenômeno**. Esse **Valor** é um fator determinante para a aceitabilidade do **Fenômeno**”. O FrameNet define o *frame Judgment* da seguinte maneira: “Um **Conhecedor** faz um julgamento sobre um **Avaliado**. O julgamento pode ser positivo (por exemplo, *respeitar*) ou negativo (por exemplo, *condenar*), e essa informação é registrada nos tipos semânticos Positivo e Negativo nas Unidades Lexicais deste *frame*”. Em negrito estão marcados os elementos *frame*. A denominação inicial dada por Fillmore (1971) não corresponde à atual denominação do FrameNet para os verbos de julgamento.

O verbo *julgar*, a princípio, parece evocar o *frame Judgment*, mas, na verdade, evoca o *frame Assessing*. O *frame Judgment* é muito próximo ao *frame Assessing*, de forma que se torna muito difícil diferenciar um do outro. O problema do *frame Assessing* para descrever um verbo como *julgar* é que ele não inclui entre seus elementos *frame* o **Avaliado**, que está incluído entre os elementos *frame* do *frame Judgment*. O estudo lexicográfico aborda de forma sistemática as dificuldades que surgem ao se utilizar o paradigma FrameNet na criação de bases de dados.

4. Comparação entre os Frames Judgment e Assessing

O *corpus* desta pesquisa foi composto de 50 textos jurídicos, totalizando 1.055.535 *tokens*, estes apresentando 39.108 *types* (palavras distintas). A pesquisa ao *corpus* iniciou-se pelos verbos dos *frames Judgment* e *Judgment_communication* que foram listados em Fillmore (1971) como verbos de julgamento: *acusar*, *culpar* e *criticar*. O estudo do item lexical *acusar*, que ocorreu nove vezes no *corpus*, aponta para quatro significados diferentes, o que caracterizaria quatro unidades lexicais distintas. Das nove ocorrências, duas remetem ao *frame Judgment* (exemplo 1) e três ao *frame Judgment_communication* (exemplo 2).

- (1) (...) **Acusa-se** a Polícia [evaluatee] como sendo a responsável pelo aumento da criminalidade [reason] (...).
- (2) (...) chegou ao exagero de dizer que **acusar** de corporativismo [reason] os advogados [evaluatee] (...).

Pode-se ver que o *frame Judgment_communication* é uma especificação do *frame Judgment*, pois requer uma comunicação do julgamento. Na sentença (1), o verbo *acusar* pode ser considerado como sinônimo de *avaliar*, *considerar*, enquanto que na sentença (2) o verbo *acusar* pode ser considerado como sinônimo de *criticar*, um verbo evocador do *frame Judgment_communication*.

As outras quatro ocorrências estão assim divididas: uma ocorrência em que *acusar* é sinônima de *apontar*, *indicar* (exemplo 3) e três ocorrências em que *acusar* está ligada à linguagem especializada (exemplo 4), indicando o ato jurídico de indicar um suspeito como réu de um processo.

- (3) (...) uma tradição que já **acusava** desgaste (...).
- (4) (...) em virtude da separação das funções de **acusar** e julgar (...).

O FrameNet não descreve toda essa polissemia. O FrameNet não é uma base de dados lexicalmente orientada, portanto a polissemia recebe tratamento distinto de outras bases de dados. Isso não significa que tais itens lexicais não sejam altamente polissêmicos também em Inglês. A polissemia do item lexical *acusar* no FrameNet está restrita a apenas dois sentidos: o *frame Judgment* e o *frame Judgment_communication*. Assim, um trabalho orientado à linguagem de especialidade pode criar bases de dados lexicais ainda mais robustas na língua alvo, à medida que se percebe essa polissemia e que se criam novos *frames* e relações para os novos sentidos não formalizados no FrameNet.

Quadro 2. Comparação entre os verbos de Julgamento e de Avaliação

Frame Assessing		
Avaliar	44	Assessing
Julgar	271	Assessing
Total	315	
Frames Judgment e Judgment_communication		
Acusar	09	Judgment / Judgment_communication
Criticar	38	Judgment_communication
Culpar	01	Judgment / Judgment_communication
Total	48	

O item lexical *culpar* apareceu apenas uma vez no *corpus*, o que impede a análise das dificuldades de criação de uma base de dados, tais como a polissemia e os equivalentes de tradução. O item lexical *criticar*, segundo o estudo do *corpus* e aplicando-se o paradigma FrameNet, não apresenta polissemia. Todas as 38 ocorrências do item lexical *criticar* remetem ao *frame Judgment_communication*, conforme exemplo (5). Talvez esse comportamento semântico esteja diretamente ligado ao significado do verbo *criticar*, ou seja, as pressuposições intrínsecas a esse verbo: criticar

envolve um ato de fala que requer um comunicador, que profere uma avaliação moral sobre uma pessoa, objeto ou fato avaliado, tendo uma razão para realizar a crítica.

- (5) Ávila [communicator] **critica** ambas as distinções (fraca e forte) entre regras e princípios [valuee].

A unidade lexical *avaliar* teve 44 ocorrências no *corpus*. Todas as ocorrências de *avaliar* remeteram ao *frame Assessing* (exemplo 6). Três pontos chamaram a atenção na análise deste item lexical: (i) a sinonímia dos seus equivalentes de tradução em língua inglesa, (ii) dois sentidos distintos no *corpus*, embora ambos estejam contemplados no *frame Assessing* e (iii) a quantidade de modalização junto com o verbo *avaliar*.

- (6) Neste caso, o STJ [assessor] **avalia** a situação econômica das partes [phenomenon].

O item lexical *avaliar* tem três equivalentes de tradução possíveis no FrameNet: *evaluate*, *assess* e *value*. Um estudo de *corpus* poderia revelar traços interessantes à tradução envolvendo a sinonímia em língua inglesa dos equivalentes de tradução do verbo *avaliar*. Um mesmo item lexical, que é polissêmico em língua portuguesa, apresenta um padrão de lexicalização diferente na língua inglesa, com diferentes itens lexicais para os diferentes sentidos em português.

No *corpus* encontra-se *avaliar* como “calcular o valor de um bem material” (exemplo 7) e “avaliar algo imaterial, verificar, conhecer” (exemplo 8). Cabe aqui a pergunta para a qual não se possui ainda a resposta: Trata-se de um caso de polissemia no nível conceitual? A resposta afirmativa a essa pergunta implicaria a divisão do *frame Assessing* em dois *frames* mais específicos para representar os significados diferentes.

- (7) A pessoa jurídica [assessor] deve **avaliar** seu patrimônio [phenomenon] (...).
- (8) Os tribunais [assessor] devem **avaliar** primeiramente qual foi o direito lesado [phenomenon] (...).

Diferentemente dos demais itens lexicais analisadas até aqui, o verbo *avaliar* destaca-se no pela grande quantidade de sentenças em que ele aparece junto de verbos modalizadores. Exemplos incluem verbos como *dever*, *poder* e *conseguir* (exemplo 9).

- (9) (...) o MPE **deverá avaliar** a conduta do Edil (...).

O item lexical *julgar* também apresentou polissemia. Dois sentidos distintos foram identificados. O verbo *julgar* ocorreu 271 vezes no *corpus*: 244 vezes indicando o processo jurídico de avaliação de um fato ou uma pessoa, ou seja, indicando um uso especializado para o verbo *julgar* (exemplo 10) e 27 vezes indicando um processo de reflexão, ligado às crenças e costumes sociais, ou seja, um uso não especializado (exemplo 11). Apenas o significado mais geral é contemplado pelo FrameNet.

- (10)(...) a Justiça Militar Estadual [assessor] **julgaria** o PM [phenomenon] e a Justiça Comum [assessor], **julgaria** o Civil [phenomenon] (...).
- (11)(...) para fazer com que os homens [assessor] que ainda a cometem parem de se **julgar** superiores [feature] às mulheres [phenomenon] (...).

Neste ponto levantam-se algumas perguntas para as quais ainda não se possui resposta: se são sentidos diferentes não deveriam remeter a *frames* distintos? Deve-se considerar que *julgar* possui sentidos diferentes, ou apenas usos diferentes? O que

devemos fazer com a linguagem especializada no contexto de formalização da língua em uma base de dados baseada em *frames*? Outro ponto de discussão chama a atenção ao se analisar o verbo *julgar*. Das 245 ocorrências em que o verbo remete a um conhecimento especializado, em 12 ocorrências o verbo *julgar* é utilizado para proferir o resultado da avaliação, como no exemplo (12), podendo ser também performativo, conforme exemplo (13). Nesses casos, o verbo *julgar* tem o valor de *declarar*.

(12)(...) o MM. Juiz Federal Dr. José Godinho Filho [assessor] **julgou** improcedente [feature] o pedido [phenomenon] (...).

(13)**Julgo** improcedente [feature] a ação direta [phenomenon].

Os verbos *julgar* e *avaliar*, embora ambos remetam ao *frame* *Assessing*, possuem comportamento sintático-semântico distinto. O verbo *avaliar* tende a ter o avaliador (*assessor*) como argumento externo e o fenômeno em avaliação (*phenomenon*) como argumento interno, posto imediatamente após o verbo (exemplo 14). Já o verbo *julgar* repetirá o padrão comum ao verbo *avaliar* somente no sentido especializado, quando não indicar declaração de sentença (exemplo 15). No sentido especializado indicando declaração de sentença (exemplo 16), o padrão comum ao verbo *julgar* é o avaliador como argumento externo (*assessor*), o fenômeno em avaliação (*phenomenon*) como argumento interno e a característica avaliada (*feature*) posta imediatamente após o verbo, compondo uma predicação secundária. Já nos casos em que *julgar* indica um processo de reflexão, ligado às crenças e costumes sociais, a tipologia é mais livre, podendo o verbo vir acompanhado de adjetivo que qualifica o fenômeno em avaliação, compondo uma predicação secundária (exemplo 17), ou com adjetivo e oração reduzida de infinitivo (exemplo 18), ou somente com adjetivo (exemplo 19).

(14)Tanto a corrente marxista como a funcionalista [assessor] **avaliam** que as políticas sociais [phenomenon] são úteis e funcionais [feature] para o capitalismo [beneficiary] (...).

(15)O Tribunal da Relação de Lisboa [assessor], (...), **julgou** interessante caso em que cidadã alemã, (...), intentou no Tribunal do Trabalho de Lisboa ação declarativa de condenação na forma comum contra Embaixada da Áustria [phenomenon] (...).

(16)(...) o STF [assessor], posteriormente, **julgou** procedente [feature] a Ação Declaratória de Constitucionalidade (ADC) [phenomenon].

(17)(...) os idealizadores do código [assessor] **julgavam** a coação e o castigo [phenomenon] essenciais [feature] para se evitar o caos na sociedade [purpose] (...).

(18)(...) os sindicatos [assessor] (...) **julgaram** proveitoso [feature] empregá-lo em larga escala [phenomenon].

(19)(...) para que esse tome as providencias que **julgar** cabíveis [feature].

Os dados apresentados demonstram as dificuldades de se empregar o paradigma FrameNet para a criação de bases de dados lexicais. A polissemia pode estar ligada aos diferentes sentidos entre língua geral e linguagem especializada. Os padrões de lexicalização necessitam de um estudo de equivalências de tradução mais específico. Por fim, a relação sintático-semântica entre os diferentes sentidos de um mesmo item

lexical também devem ser analisados para a inclusão desses itens em uma base de dados lexicais.

5. Conclusão

O estudo apresentado demonstra que a polissemia é um dos grandes desafios para a construção de bases de dados lexicais segundo o paradigma FrameNet. Projetos de extensão bilíngüe do FrameNet, em geral, partem dos *frames*, procurando equivalentes de tradução na língua-alvo. O que defendemos aqui é o caminho inverso: o estudo do item lexical na língua-alvo e a identificação de um *frame* apropriado no FrameNet. Percebemos que em alguns casos, como em *julgar*, o FrameNet não descreve todos os significados possíveis de um item lexical. Tal estudo poderia ampliar o poder descritivo do FrameNet, à medida que seria possível criarem-se relações de *plug-in* entre unidades lexicais mais genéricas e unidades lexicais mais especializadas.

Podemos, então, apontar três desafios a serem enfrentados no uso do paradigma FrameNet: (i) a polissemia, no português e no inglês, (ii) os padrões de lexicalização e (iii) a proximidade conceitual de alguns *frames*. O estudo de cada item lexical individualmente, ao invés da identificação de equivalentes de tradução, é uma forma de se gerenciar a polissemia. Estudos futuros abordarão as questões relativas aos padrões de lexicalização e de proximidade conceitual de *frames*.

Agradecimento

Trabalho realizado com o apoio da CAPES através de bolsa de doutorado.

Referências

- Boas, H. (2002). Bilingual FrameNet dictionaries for machine translation. Third International Conference on Language Resources and Evaluation. Las Palmas, Spain. Vol. IV, p. 1364-1371.
- Boas, H. (2005). Semantic frames as interlingual representations for multilingual lexical databases. *International Journal of Lexicography*. Vol. 18, N.4, p. 445-478.
- Fillmore, C. (1971). Verbs of judging: an exercise in semantic description. In *Studies in linguistic semantics*. Edited by Fillmore and Langendoen. Holt, Rinehart and Winston: New York, p. 272-89.
- Fillmore, C. (1982). Frame semantics. In *Linguistics in the morning calm*, Edited by The linguistic society of Korea, Seoul: Hanshin Publishing Co., p. 111-137.
- Fillmore, C., Johnson, C., and Petruck, M. (2003). Background to FrameNet. In *International Journal of Lexicography*, Vol. 16, N. 3, p. 235-250.
- Hasegawa, Y. Ohara, K. Lee-Goldman, R., Fillmore, C. (2006). Frame integration, head switching, and translation: RISK in English and Japanese. Forth International Conference on Construction Grammar.
- Subirats, C. and Petruck, M. (2003). Surprise: Spanish FrameNet. International Congress of Linguistics. Workshop on Frame Semantics.