

Kicktionary_Br: um relato sobre a anotação semântica de um *corpus* voltado ao domínio do futebol

Rove Chishman (PPGLA – UNISINOS / CNPq)
Diego Spader de Souza (PPGLA – UNISINOS / CAPES)
João Gabriel Padilha (PPGLA – UNISINOS)

RESUMO: O presente trabalho tem como objetivo discutir os resultados relevantes do projeto Kicktionary_Br, que tem como proposta a construção de um *corpus* temático do futebol, em português brasileiro, semanticamente anotado com base na Semântica de Frames (FILLMORE, 1982; 1985). Esse projeto foi inspirado no recurso lexical The Kicktionary, criado por Thomas Schmidt, a partir do projeto FrameNet. Os principais resultados dessa investigação incluem o surgimento de casos que apresentam desafios para a metodologia de anotação, como ocorrência de mais de uma cena na mesma sentença, a tradução das metáforas em inglês, além de profundas marcas da oralidade em construções com, por exemplo, os verbos-suporte.

PALAVRAS-CHAVE: Semântica de Frames; Anotação de Corpus; Léxicos Computacionais.

ABSTRACT: This paper has the objective of discussing the relevant results of the Kicktionary_Br project, which proposes the construction of a soccer themed *corpus*, in Brazilian Portuguese, semantically annotated and based on Frame Semantics (FILLMORE, 1982; 1985). This project was inspired by the lexical resource The Kicktionary, created by Thomas Schmidt, from the FrameNet project. The main results of this investigation include the emergence of cases that present challenges to the annotation methodology, as the occurrence of more than one scene in the same sentence, the translation of metaphors in English, as well as deep oral marks in constructions like, for example, support-verbs.

KEYWORDS: Frame Semantics; Corpus Annotation; Computational Lexicons.

Introdução

Este trabalho tem como objetivo discutir os principais resultados do projeto Kicktionary_Br, que propõe a construção de um *corpus* sobre a linguagem futebolística do português brasileiro semanticamente anotado com base em *frames*. Para tal, consideram-se questões que antecedem a anotação semântica do *corpus* do projeto, como a compilação, o armazenamento desse *corpus* e as escolhas metodológicas que assumimos para realizarmos esta investigação. A tarefa de anotação propriamente dita levou em conta as categorias propostas por Schmidt (2009) na ferramenta denominada The Kicktionary, uma base de dados lexical multilíngue representativa do léxico futebolístico dos idiomas inglês, francês e alemão baseada no princípio dos *frames* semânticos de Charles Fillmore (1977). Baseamo-nos em categorias propostas para a língua inglesa disponibilizadas nessa plataforma e, assim, por meio de seus respectivos equivalentes de tradução em Português brasileiro, realizamos a anotação de nosso *corpus*, que possui as mesmas características (um *corpus* comparável, portanto) dos *corpora* oferecidos pelo referido pesquisador alemão em sua ferramenta.

Organizamos este artigo da seguinte forma: na primeira seção, apresentaremos os princípios teóricos que embasaram esta investigação, sejam eles da Linguística Cognitiva – subárea da Linguística que abrange a teoria dos *frames* semânticos de Fillmore, segundo elemento de nossa revisão bibliográfica – e, por fim, apresentaremos a FrameNet (FILLMORE et al, 2003), recurso computacional que implementa os *frames* semânticos e que é central para o entendimento da ferramenta de Schmidt – que, grosso modo, poderia ser resumida como “uma FrameNet voltada ao futebol”.

Na segunda seção, dedicada à metodologia, primeiramente, na subseção 2.1, apresentamos a ferramenta multilíngue *The Kicktionary* (SCHMIDT, 2009), central para o entendimento deste artigo, uma vez que nos propomos à construção de uma ferramenta homônima voltada à descrição do Português brasileiro. A subseção seguinte, 2.2, aborda a construção de nosso tutorial de anotação, um recurso baseado nas categorias em inglês oferecidas pelo projeto pioneiro, com vistas à aplicação na anotação de nosso *corpus*. Na subseção 2.3, relatamos o processo de coleta, compilação e armazenamento de nosso *corpus* comparável. O último item da seção, dedicado à metodologia, 2.4, versa sobre o processo de anotação semântica e suas respectivas etapas.

Na terceira seção, apresentamos a análise dos dados, bem como os principais resultados deste empreendimento.

Por fim, na quarta seção, elencamos algumas considerações sobre a realização desta investigação: nele discutimos a validade deste estudo, as dificuldades encontradas durante o processo e apontamos para futuras contribuições no que tange à anotação semântica de *corpora* baseada em *frames*.

1. Questões introdutórias à Linguística Cognitiva

A Linguística Cognitiva é uma subárea da Linguística relativamente nova. Ela surgiu no fim da década de setenta, no oeste dos Estados Unidos, e consolidou-se no fim da década seguinte, tendo, como marcos formais, a criação do ICLA (*International Cognitive Linguistics Association*), em 1989, a publicação da revista *Cognitive Linguistics*, em 1990 (EVANS, 2006; SILVA, 2007), bem como o lançamento da coleção *Cognitive Linguistics Research* (SILVA, 2007).

Em pouco mais de três décadas, desde sua concepção, as investigações empreendidas pela Linguística Cognitiva têm-se interessado pela linguagem enquanto um meio de conhecimento relacionado à experiência humana no mundo (SILVA, 2007), ou, mais precisamente, pela investigação da relação entre a linguagem humana, a mente e a experiência sócio-física dos falantes (EVANS, 2006). A experiência sócio-física contempla a experiência humana sob duas perspectivas inter-relacionadas: primeiramente, pela perspectiva social, no que tange ao convívio e à interação de falantes por meio da linguagem em uma dada comunidade, e pela perspectiva individual, que parte da fisiologia do ser humano, da constituição de seu corpo, para explicar como ele percebe o mundo que o rodeia – esse processo dá-se por meio dos cinco sentidos, e, por isso, de natureza física.

O estudo cognitivo da linguagem demanda uma abordagem interdisciplinar, uma vez que ele coteja outras áreas do conhecimento: sua origem perpassa a filosofia e chega às ciências cognitivas da década de setenta, principalmente a Psicologia Cognitiva (EVANS, 2006). Essa interdisciplinaridade motiva os adeptos da vertente cognitivista a caracterizarem-na como um empreendimento, ou um movimento, que justapõe diferentes saberes com a finalidade de investigar os fundamentos da linguagem humana.

Para os cognitivistas, a linguagem humana não é um mero código voltado à emissão e à recepção de mensagens: essa é uma de suas facetas, cujo foco é estritamente seu aspecto funcional. Ao invés disso, interessa à vertente cognitiva abordar a linguagem enquanto representação do conhecimento humano.

Na qualidade de paradigma, a Linguística Cognitiva tem dois compromissos fundamentais (LAKOFF, 1990) – o compromisso com a generalização e o compromisso cognitivo.

O compromisso com a generalização não se restringe à abordagem cognitiva da linguagem, mas perpassa o pensamento científico de um modo geral: a partir do

estabelecimento de conclusões gerais acerca de um objeto de estudo – neste caso, a linguagem – é possível obter-se, então, particularizações. Trata-se, portanto, de um processo vertical, que parte daquilo que é mais abrangente, do todo, do conjunto, para chegar às partes que o compõem. Esse tratamento dispensado à linguagem justifica-se na medida em que uma abordagem segmentada permite poucas generalizações, uma vez que não inter-relaciona os diferentes níveis que a integram – sintático, semântico e fonológico – como postulam pensadores gerativistas, em especial, Chomsky (1965, 1981, 1985).

O segundo compromisso fundamental da Linguística Cognitiva é o compromisso cognitivo. Seu foco recai sobre a relação estabelecida entre os princípios gerais da linguagem e aquilo que é sabido sobre o cérebro humano – neste ponto, cabe mencionar as contribuições advindas de outras áreas do conhecimento, especialmente das ciências que estudam a cognição, como a Psicologia Cognitiva. Como afirma Evans, por meio de um curioso jogo de palavras, “é esse compromisso que faz da Linguística Cognitiva cognitiva” (p. 4). O compromisso cognitivo, portanto, procura verificar em que medida a cognição humana é refletida pelas estruturas linguísticas: os modelos linguísticos devem refletir aquilo que se sabe sobre a linguagem humana, e não ter como base formalismos fundamentados na estética.

Como teorias representativas dessa subárea, podemos citar a teoria dos *Frames* semânticos, de Fillmore (1982, 1985) – fundamental para este trabalho – e os trabalhos de linguistas como Lakoff, Langacker, Talmy, Goldberg e Croft.

Ainda em relação aos *frames* semânticos, cabe ressaltar que assumimos a definição adotada por Schmidt em seu projeto pioneiro: o pesquisador alemão vale-se de uma primeira definição postulada por Fillmore (1977), a qual evidencia o *frame* como uma categoria semântica subjugada à cena. De acordo com essa definição, os *frames* seriam considerados entidades linguísticas relacionadas ao campo semântico da cena, que seria uma representação abstrata de um evento ou instituição social, como afirma Fillmore. Em trabalhos subsequentes (1982, 1985), o linguista estadunidense abandonou a categoria cena, estando ao *frame* o papel de categoria superordenada – não necessariamente linguística – em relação às unidades lexicais, que, por sua vez, referem-se às realizações linguísticas de um dado *frame*.

Vale ressaltar que a Semântica de Frames teve seus conceitos básicos fundamentados em *An Alternative to Checklist Theories of Meaning*, de 1975. Na obra, Fillmore evidencia a relevância de tais conceitos para a teoria semântica. Bertoldi (2011, p.21) nos explica que, para Fillmore (1975), “os *frames* são uma estrutura de conceitos que se conectam entre si, dando coerência a alguma área do conhecimento [...]”.

A Semântica de Frames, neste ponto, trabalha com os conceitos de *cena* e de *frame*. Uma cena seria mais do que uma simples representação visual, sendo, portanto, “um conjunto de tipos familiares de transações interpessoais, cenários padrões definidos culturalmente, estruturais institucionais, experiências nativas, imagem corporal, crenças humanas, ações, experiências e imagens” (BERTOLDI, 2011). Os *frames* estariam mais relacionados com as escolhas linguísticas relacionadas a uma cena.

Fillmore (1975), através da distinção entre *cen*as e *frames*, diferencia as estruturas cognitivas das estruturas linguísticas. Ao ter conhecimento de uma cena, o falante tem a possibilidade de escolher as palavras mais apropriadas para descrevê-la, ao passo em que o conhecimento das palavras permite acionar a estrutura cognitiva e compreender o significado das palavras. Assim, o pesquisador (1975) cria o termo *cen*as-e-*frames*, a fim de referir-se à interação entre as estruturas cognitiva e linguística.

Fillmore (1976) divide os *frames* em duas categorias:

- *Frames* interacionais: são as escolhas linguísticas adotadas para uma determinada situação. Saber qual tipo de cumprimento usar em situações diversas, formais e informais, também faz parte dos *frames* interacionais.
- *Frames* cognitivos: demonstra o conhecimento da cena relacionada a um determinado evento, os seus participantes e o papel desempenhado por cada um deles no evento (BERTOLDI, 2011).

O termo *Semântica de Frames* foi utilizado pela primeira vez no artigo intitulado *Frame Semantics*. A Semântica de Frames realiza pesquisas de forma empírica e não formal, uma vez que objetiva “encontrar as formas detalhadas em que expressões específicas se encaixam em suas situações” (FILLMORE, 1985, p.231).

Em obras de 1982 e 1985, Fillmore abandona a distinção entre cenas e *frames*, sem relacionar mais o conceito de cena com estruturas cognitivas e o de *frame* com estruturas linguísticas. Os *frames* englobam as dimensões cognitivas e linguísticas simultaneamente, pois itens lexicais são vistos como representantes de estruturas cognitivas. Tal afirmação quebra a tradicional diferenciação entre conhecimento linguístico e conhecimento enciclopédico (BERTOLDI, 2011, p.24).

Em 1992, Fillmore e Atkins, na obra *Toward a Frame-Based Lexicon: The Semantics of Risk and its Neighbors*, apresentam o primeiro estudo semântico baseado em *frames*. Estudando os itens lexicais relacionadas a *risco* (como *risk*, *danger* e *hazard*), os autores propõem onze categorias para descrever os participantes do *frame Risk*. São elas: *chance*, *harm*, *victim*, *valued object*, *risky situation*, *deed*, *actor*, *intended gain*, *purpose*, *beneficiary* e *motivation*. Tais categorias são, assim, elementos de *frame*, denominação que seria dada aos papéis semânticos no contexto do projeto FrameNet.

A FrameNet é um projeto computacional lexicográfico que objetiva a coleta de informação a respeito das propriedades semânticas e sintáticas de itens lexicais da língua inglesa e que teve seu nome inspirado pela WordNet (FELLBAUM, 1998). Essa coleta é feita através de um vasto número de *corpora* eletrônicos através de procedimentos tanto manuais quanto automáticos. A FrameNet baseia-se nos postulados da Semântica de *Frames* e preocupa-se com as redes de significados em que as palavras atuam (FILLMORE et al, 2003, p.235).

Dentre os trabalhos desenvolvidos no âmbito da FrameNet, estão: a identificação e a descrição de *frames* semânticos, a análise dos sentidos das palavras dos *frames* em que elas residem, e, através do estudo das estruturas sintáticas, o estudo do modo com que as estruturas semânticas se expressam através da linguagem. O trabalho desempenhado pela FrameNet ocorre a partir de dois segmentos chave: os *frames* e as unidades lexicais. De acordo com Cruse (1986), esses elementos representam o emparelhamento entre a palavra e o sentido. Segundo Fillmore et al (2003, 236), “os sentidos separados de uma palavra correspondem aos diferentes quadros semânticos em que a palavra pode participar”. Quando o sentido de uma palavra fundamenta-se em um determinado *frame*, dizemos que essa palavra *evoca o frame*.

A FrameNet, segundo Fillmore et al (2003), pode ser caracterizada em duas partes: uma base de dados lexical com informações sobre os *frames* e os elementos de *frame*, e uma base de dados de anotação, onde ficam as sentenças anotadas semanticamente. A metodologia de anotação da FrameNet segue cinco passos, sendo eles:

1. Caracterização do *frame*;
2. Descrição e nomeação dos elementos de *frame*;
3. Seleção das unidades lexicais;
4. Anotação de sentenças;
5. Geração automática de entradas lexicais.

2. Kictionary_Br: questões metodológicas

Conforme referimo-nos anteriormente, nosso intuito é aplicar a teoria dos *frames* semânticos ao domínio do futebol, visando à construção de um *corpus* anotado com base em textos legítimos que versem sobre o tema em nossa língua, o português brasileiro. Para tal, tomamos como referência o projeto do pesquisador alemão Thomas Schmidt, pioneiro na aplicação da teoria dos *frames* ao léxico futebolístico. Apresentaremos aqui, em linhas gerais, esse recurso, bem como as escolhas metodológicas que assumimos a partir daquilo que o projeto pioneiro propõe – começando pela adaptação das categorias do Kicktionary a nossa língua – além da compilação de nosso próprio *corpus*, cujas características obedecem aos *corpora* do projeto pioneiro, e as etapas de sua anotação.

2.1. O Kicktionary

O Kicktionary é um recurso lexical multilíngue *online* voltado à descrição semântica da linguagem futebolística: por meio de onze cenas, o pesquisador alemão Thomas Schmidt sistematiza todos os eventos envolvidos com uma partida de futebol e vale-se de *corpora* paralelos em inglês, francês e alemão, para ilustrar esses eventos.

Cada cena é composta por um dado número de *frames* semânticos (FILLMORE, 1977, 1982, 1985), que representam as diferentes perspectivas da cena em questão: na cena “chute”, por exemplo, é possível observar *frames* que privilegiam, não apenas o próprio chute, mas os outros possíveis eventos relacionados a ele, como a defesa, a intervenção, além de outros nove *frames*.

Além da organização em cenas e *frames*, o projeto pioneiro também oferece uma relação de unidades lexicais que ilustram esses eventos, em sua maioria, nominais eventivos (*partida, falta, cobrança, chute*, além de outros menos típicos).

2.2. O manual de anotação semântica

O chamado “manual de apoio para a anotação semântica” ou tutorial de anotação consiste em um arquivo de texto contendo as descrições das cenas e as traduções dos *frames* e das unidades lexicais propostos por Schmidt no Kicktionary. Realizamos, em grupo, a tradução desses dados para identificarmos os possíveis equivalentes de tradução em português brasileiro. Essa atividade revelou-se deveras proveitosa, uma vez que não houve problemas em relação a vocábulos como *chute, defesa, finalização, intervenção*, entre outros. O desafio mostrou-se na maneira como o português brasileiro lexicaliza os eventos do futebol: por exemplo, *bicicleta*, que é o *tipo de chute* (é esse, inclusive, o nome do *frame* ao qual essa unidade lexical remete), cujo movimento se assemelha ao de um ciclista, com as costas voltadas para o chão. *Chutar de bicicleta* é uma metáfora para a qual não há correspondente direto na língua inglesa – a segunda língua tomada como referência não apenas para este projeto, uma vez que não prevemos o uso de dados em francês ou alemão.

Algumas dificuldades foram encontradas durante a construção do manual devido ao fato de que algumas unidades lexicais em inglês não apresentam equivalentes de tradução direta em português brasileiro. Durante o processo de anotação do *corpus*, percebemos que, em português brasileiro, existem unidades lexicais que o projeto pioneiro de Schmidt não

apresenta, e vice-versa. Portanto, fez-se necessária uma adaptação do referencial do projeto pioneiro para que fosse possível contemplar os dados em português: por exemplo, na cena *chute*, em relação ao *frame* *Chute*, além do equivalente de tradução do substantivo *shot* (em inglês, *chute*), há o substantivo *over-head kick*, cujo equivalente de tradução em português é uma expressão metafórica: a *bicicleta*. Em contrapartida, não há *bicycle* no referencial teórico do projeto pioneiro. Portanto, pretendemos, a partir dos dados e resultados da anotação, complementar este manual com a proposição de novos *frames* e unidades lexicais correspondentes ao léxico do futebol em português brasileiro. Dessa maneira, teremos um guia de anotação baseado no projeto pioneiro e, simultaneamente, adaptado às construções futebolísticas de nossa língua materna.

O manual está organizado da seguinte forma:

- 1) Sumário;
- 2) Listagem das unidades lexicais em português, seus respectivos *frames* e número da página para consulta;
- 3) Listagem dos cenários do Kicktionary e de suas cenas;
- 4) Apresentação das cenas, suas descrições, *frames* em inglês e em português e unidades lexicais em inglês e em português.

Voltaremos a nos referir ao manual de anotação (também denominado de “kit de anotação”) na última subseção desta seção, informando, de forma breve, a forma como procedemos em seu uso.

2.3. Coleta e compilação

Uma atividade central do projeto Kicktionary_Br foi a compilação de um *corpus* comparável próprio em português brasileiro. Essa compilação fez-se necessária uma vez que não havia um *corpus* especializado sobre a linguagem futebolística para o português do Brasil – o projeto pioneiro, Kicktionary, embora multilíngue, não contempla dados em português brasileiro. Cabe lembrar que os *corpora* disponibilizados na base de dados alemã são *paralelos*, ou seja, cada um dos textos é construído a partir da tradução de outro. Em nosso caso, pela ausência de textos em português brasileiro, tivemos de compilar um *corpus comparável*: textos com as mesmas características daqueles oferecidos por Schmidt. Em relação aos critérios de compilação empregados no projeto Kicktionary-Br, destacam-se a representatividade, a autenticidade e a escolha criteriosa, conforme postula Berber Sardinha (2004).

A representatividade diz respeito à temática sobre a qual versam os textos que compõem esse *corpus*: nesse caso, interessam ao projeto textos que tratam dos momentos de uma partida de futebol – tal qual fazem os textos do projeto pioneiro.

O critério autenticidade prima pela existência desses textos em ambientes alheios à pesquisa, ou seja: tais textos não foram criados com o propósito de ilustrar as cenas de uma partida de futebol; são textos autênticos que tratam do tema.

A escolha criteriosa é um fator importante, na medida em que os textos devem possuir certas características para que cumpram sua função: como o objetivo do projeto Kicktionary-Br é construir um *corpus* representativo da linguagem futebolística do Pt-br, interessam ao projeto textos que tratem exclusivamente dos momentos de uma partida de futebol.

Após a escolha dos critérios de compilação, procedemos à primeira coleta de textos de maneira automatizada. Através do software HTTrack, foram coletados diversos textos do *site* do Sport Club Internacional. Porém, essa coleta não se mostrou eficiente: grande parte desses textos, embora versassem sobre o clube, não puderam ser usados, uma vez que, além de descrições das partidas disputadas pelo referido clube, havia também notícias institucionais

sobre o Internacional. Sentenças como “A delegação colorada anuncia que...”, bem como diversas outras, não descrevem propriamente uma partida de futebol, mas sim, um evento extra-campo que envolve o clube gaúcho, e, portanto, não se adequam ao projeto Kicktionary-Br.

A primeira coleta foi invalidada, pois seria deveras trabalhoso excluir as notícias que não serviam ao propósito do projeto. Dessa maneira, optou-se por um processo manual de coleta, o que garantiu que apenas textos pertinentes ao jogo em si fossem obtidos.

A segunda teve como base *sites* de nove equipes brasileiras engajadas em três competições distintas: Campeonato Brasileiro, Copa Santander Libertadores e Copa do Brasil. Essas equipes são: Flamengo, Corinthians, Atlético Mineiro, Atlético Paranaense, Fluminense, Internacional, São Paulo, Palmeiras e Santos. Os respectivos *sites* foram acessados em dias subsequentes às partidas disputadas pelos times citados. Os textos selecionados consistem em resumos das partidas, o que garantiu, desta vez, a escolha criteriosa dos textos que viriam a compor o *corpus* do projeto. Esses textos totalizam pouco mais do que 150.000 palavras, o que, de acordo com Berber Sardinha (2004), corresponde a um *corpus* de tamanho médio.

Porém, a compilação não se resume à mera coleta de textos. A compilação do *corpus* do projeto Kicktionary-Br envolveu, da mesma maneira, critérios de conservação, armazenamento e distribuição propostos por Wynne (2004), que sucedem a etapa de compilação. Esses critérios asseguraram que, uma vez compilado esse *corpus*, ele seria acessível a potenciais usuários dentro e fora do ambiente de pesquisa. Wynne cita elementos como formato de arquivos a serem armazenados, ambientes virtuais onde os dados serão dispostos e os veículos pelos quais o produto será disponibilizado: em nosso caso, a plataforma *online* Framecorp Tools, atualmente em fase de desenvolvimento pelos pesquisadores da área de Tecnologia de Informação. Após a coleta, esses textos foram salvos em formato .doc (*Microsoft Word*) – para a anotação manual – e, em seguida, em formato .txt – requerido pelo parser PALAVRAS (BICK, 2000) . Esses arquivos, uma vez inseridos nesse *software*, geram outro tipo de arquivo, de extensão .xml , requerido pela ferramenta de anotação semântica SALTO (BURCHARD et. Al, 2003). Uma vez anotadas manualmente, as informações dos textos são inseridas nos arquivos .xml – arquivos para a leitura estritamente computacional – através do SALTO, sendo essa a última etapa do processo de anotação. O passo seguinte foi o armazenamento desses dados. Cópias compactadas contendo 9 pastas – uma para cada clube – foram distribuídas entre os anotadores e, em seguida, armazenadas em discos virtuais como o DropBox. O *corpus* é composto de 11 pacotes, com 100 arquivos que totalizam 3.307 sentenças.

2.4. O processo de anotação semântica

Em um primeiro momento, optou-se por anotar parte do *corpus* com o intuito de familiarizar os anotadores com o processo de anotação semântica: através dessa anotação inicial, foi possível mapear algumas características dos textos, como os tipos de construções neles presentes (metáforas, verbos-suporte, marcas de oralidade etc.). Para esse primeiro experimento, foram selecionadas as primeiras 1007 sentenças do *corpus*.

Esse primeiro contato com o *corpus* recém-compilado foi fundamental na escolha da metodologia de anotação. Em primeiro lugar, tivemos contato com as cenas mais incidentes, as quais julgamos serem as mais representativas do esporte, pois envolvem momentos decisivos em uma partida: *passe*, *gol*, *chute*. É plausível admitir que o momento do gol, por exemplo, seja mais valorado do que uma mera *substituição* – que, em oposição, incidiu menos vezes do que as supra referidas. Desse modo, optamos por anotar primeiro as cenas mais

recorrentes em nosso *corpus*. Em um segundo momento, através da análise do referencial teórico de Schmidt, foi possível ter uma breve ideia dos desafios a serem enfrentados no processo de anotação de nossos textos: a existência (ou a falta) de equivalentes de tradução entre os dados disponíveis no projeto pioneiro e os dados coletados pelo grupo, a escolha dos constituintes oracionais a serem anotados ou, ainda, o tratamento dispensado às relações de sentido como a metáfora e a metonímia entre as duas línguas..

Seguindo a metodologia *running-text* de anotação proposta pelo projeto FrameNet, os anotadores foram divididos em quatro duplas, onde cada uma delas responsabilizou-se por uma cena. As cenas escolhidas foram: Goal (Gol), Pass (Passe) e Shot (Chute) – a disparidade entre o número de duplas e o número de cenas é justificada pela atuação de um dos analisadores em *duas* duplas. O segundo processo de anotação ocorreu com as mesmas 1007 sentenças. Dessa vez, no entanto, foram escolhidas as cenas menos recorrentes em nossa compilação: One_on_one (Um contra um), Mix_up (Confusão), Substitution (Substituição), Motion (Movimento), Move (Jogada), State_of_match (Estado da partida), Foul (Falta) e Chance (Chance).

De acordo com essa metodologia, cada anotador deve trabalhar de forma independente. Primeiramente, cada anotador constrói um *subcorpus* a partir das 1004 sentenças. Nesse *subcorpus*, devem constar as sentenças que se enquadram na cena de responsabilidade do anotador. Ao término da compilação, os anotadores comparam os *subcorpora* com seus pares, com a finalidade de chegarem a uma versão final, que será anotada com base nos dados do Kicktionary.

Esses dados consistem no já referido kit de anotação semântica. De posse dele, cada um dos anotadores procede da seguinte maneira: primeiramente, identifica-se a palavra *target*, ou seja, aquela que evoca, que *ativa*, um *frame*. Geralmente esse *target* é um verbo pleno ou um nominal eventivo, como *chutar* ou *falta*. A partir disso, faz uma busca nas unidades lexicais presentes no kit, já devidamente traduzidas para nossa língua. É nesse momento que o anotador localiza a cena e o *frame* pertencente àquela cena, por exemplo, *cena* Chute, *frame* Chute. Identificados cena e *frame*, o passo seguinte é identificar os elementos de *frame*, também dispostos no kit, aos constituintes oracionais das sentenças de nosso *corpus*: [Ricardo tardelli jogador_que_chuta] **chutou...** Nesse caso, o elemento de *frame* *jogador_que_chuta* é representado na sentença pelo sintagma nominal sujeito *Ricardo Tardelli*. Caso a tradução da unidade lexical do português brasileiro já conste na lista prevista no kit, como é o caso dos equivalentes diretos *falta/foul*, o anotador procede para o caso seguinte. Caso essa unidade lexical seja exclusiva de nosso idioma – como *bater*, sinônimo de *chutar*, e para a qual não há equivalente direto na língua bretã – o anotador complementa o material adicionando-a à lista. Dessa maneira, obtivemos nossa própria versão do referencial, que obedece à disposição que o projeto pioneiro propõe, e, em segunda instância, capitaliza padrões linguísticos exclusivamente ligados ao português brasileiro.

Ao término do processo, cada dupla deve conferir as anotações com a finalidade de eliminar as dúvidas em relação ao processo de anotação – a interpretação dos *frames* é subjetiva, portanto é mais uma questão interpretativa do que uma questão de anotar de maneira certa ou errada.

Uma vez conferidas as anotações, o próximo passo foi a inserção dessas sentenças manualmente anotadas no anotador SALTO, para que, assim, obtivessem-se os arquivos em formato .xml, requeridos por qualquer sistema que processe dados referentes a linguagens naturais.

3. Análise e discussão dos dados

Esta seção discutirá algumas questões oriundas da anotação semântica de nosso *corpus*, partindo de dados quantitativos (expressos entre parênteses de forma contígua às categorias, no que tange à incidência das categorias), e posteriormente, abordando questões de cunho qualitativo, como o paralelismo entre as duas línguas, a lexicalização dos momentos de uma partida de futebol em português brasileiro e sua expressão através dos *frames* propostos pelo Kicktionary alemão. Conforme já nos destacamos, as cenas mais incidentes – Gol, Chute e Passe – tiveram prioridade em nosso estudo. Feitas as considerações sobre esse grupo de sentenças representativas das cenas mencionadas, abordaremos, em tom mais generalista, alguns resultados obtidos na análise das cenas cuja incidência foi menos perceptível.

A cena Chute totalizou 203 sentenças, ocorrendo, em nosso *corpus*, 302 vezes. É válido mencionar que, em algumas sentenças, foi comum a coocorrência de *frames* como Chute e Defender, o que justifica a discrepância numérica entre o número de sentenças anotadas e a presença de mais de um frame em uma única frase. Essa cena contém os seguintes frames: Chute (147), Defender (72), Finalizar (48), Intervenção (17), Rebote (9), Tipos_de_chute (4), Acertar (2), Chutar_a_gol (2) e Tiro_de_meta (1).

Em relação a essa cena, não encontramos dificuldades na localização dos *frames* na base de dados do projeto pioneiro. Cremos que tal fato deva-se à organização proposta por Schmidt, a qual obedece a um padrão lógico: por exemplo, eventos como *chute*, *defesa* e *cabeçada*, para citar alguns, sempre apresentam correspondentes diretos – equivalentes de tradução – se comparados aos dados em inglês da base Kicktionary.

As diferenças deram-se não no nível categorial, mas no lexical. Conforme já nos referimos, itens lexicais como *bicicleta* não possuem correspondência direta no inglês. No caso de *chute por cobertura*, ocorre o mesmo, conforme mostra o exemplo a seguir, exclusivo do português brasileiro:

[Chute/Tipos _de_chute]

294 - *No minuto seguinte, [Júnior jogador_que_chuta] aproveitou a falha da defesa catarinense na saída de bola e [tocou chute] por cobertura para fazer um belo gol*

Outro resultado que chamou nossa atenção foi o grande número de evocadores metafóricos, que totalizaram 55 ocorrências em relação à cena *chute*, conforme notamos nos exemplos a abaixo:

[Chute/Rebote-Finalizar/Defender]

285 - *O próprio [Tardelli jogador_que_chuta] pegou o rebote e, mesmo caído, finalizou para nova intervenção d[do arqueiro da Chapecoense goleiro].*

[Chute/Tipos_de_chute]

349 - *Mas em um contra-ataque que iniciou com um ótima reposição do goleiro Neto, [Alex Mineiro jogador_que_chuta] tabelou com Wagner Diniz, [o lateral jogador_que_chuta] fez grande jogada individual ao invadir a área e fuzilou.*

[Chute/Tipos_de_chute/Intervenção]

14 - *Na seqüência da jogada, [Tardelli jogador_que_chuta] tentou de bicicleta e a [bola bola] pegou no [zagueiro alvo_da_intervenção].*

Na sentença 285, nota-se a referência metafórica ao chute através do vocábulo *finalizou*, ou seja, o chute representa o fim da jogada, ou ainda, um chute é uma finalização. No exemplo seguinte, percebemos uma metáfora deveras peculiar: “um chute é um tiro”, ou a “a bola é um projétil”, e, até mesmo, o “pé é uma arma”. A sentença 14 apresenta outro evocador metafórico: *bicicleta*, target que evoca o *frame* Tipos_de_chute. O referencial teórico do projeto pioneiro não prevê essa unidade lexical, embora preveja uma outra unidade que alude ao mesmo evento, porém de forma literal: *overhead-kick*, ou, *chute por cima da cabeça*, que, por sua vez, não está registrada em nosso *corpus* de estudo.

Também percebemos casos em que determinados elementos de *frame*, em nossas sentenças, apresentam estrutura metonímica, como no exemplo a seguir:

[Chute/Chute/Intervenção]

619 - *Elias e Jorge Henrique tabelaram, o atacante se enrolou, mas recuperou-se e cruzou; [Ronaldo jogador_que_chuta] dominou [no peito parte_do_corpo] [dentro da área origem] e armou o chute, mas [a defesa colombiana interceptador] travou na hora certa.*

Percebemos o elemento de *frame interceptador* expresso metonimicamente pelo SN *a defesa colombiana*: não é toda a defesa que efetua a jogada, mas apenas um de seus integrantes, haja vista que *defesa*, em português brasileiro, é sinônimo de *zaga*, que nada mais é que o grupo de jogadores que, justamente, defendem o goleiro dos jogadores adversários. Essa sentença, embora não seja de difícil compreensão para alguém familiarizado com o esporte, representa um desafio no que tange ao Processamento da Linguagem Natural, pois está diretamente relacionada à maneira como computadores leem – a chamada “leitura de máquina” – a metáfora.

A palavra *defesa* não é, porém, em nossa língua, apenas associada à referida *zaga*, mas também, ao movimento efetuado pelo goleiro quando esse detém a bola chutada por um jogador adversário. Trata-se, portanto, do fenômeno linguístico conhecido como polissemia, também percebido em nosso estudo de *corpus*, conforme ilustramos com os exemplos abaixo:

[Chute/Defender]

7 - *Aos 21 minutos, [Aranha goleiro] fez boa defesa e [chute cruzado chute] [pela direita origem].*

Esses são apenas alguns resultados relacionados à cena *Chute*. Esses e outros dados representam, de forma singular, por um lado, peculiaridades inerentes ao léxico futebolístico do português brasileiro, e, por outro, aspectos comuns aos léxicos do inglês, francês e alemão, providos por Schmidt, conforme demonstramos de maneira breve.

A cena *Passe* e seus respectivos *frames* apareceram em 159 sentenças de nosso *corpus*. Essa cena, juntamente com a cena *Chute* e *Gol*, foi uma das mais recorrentes nos textos em português brasileiro. Os *frames* mais incidentes em relação a essa cena foram *Passe* (96), *Dominar* (48), *Interceptar* (39), *Combinação_de_passe* (16), *Estar_livre* (06), *Rolar* (05) e *Fornecer_passe* (02). Os *frames* *Marcar*, *Conectar* e *Devolver_o_passe* não foram registrados em nosso *corpus* de estudo. Em relação aos resultados ligados à anotação dessa cena, podemos citar, inicialmente, a alta incidência de evocadores de *frames* metafóricos, conforme as sentenças abaixo:

[Passe/Interceptar/Estar_livre]

802 – [Camacho _{interceptor}] **roubou** grande bola [no meio _{local_da_intervenção}] e deixou [Diego Maurício _{jogador_que_recebe}] **livre** para chutar colocado, mas a bola passou raspando a trave.

[Passe/Interceptar]

888 - Mas a [Zaga _{interceptor}] **cortou**.

[Passe/Fornecer_passe]

952 – [O autor do primeiro gol _{jogador_que_passa}] **deu** [passe milimétrico _{passe}] para [Adriano _{jogador_que_recebe}] apenas tocar na saída do goleiro chileno e fazer 2 a 0 aos 13.

Na sentença 802, relacionada ao frame *Interceptar*, notamos o sentido metafórico do verbo *roubar*, em relação à tomada de bola performatizada pelo jogador Camacho. O exemplo seguinte diz respeito ao mesmo *frame* *Interceptar* evocado por outra metáfora: cortar no sentido de interceptar a bola – a bola não é cortada literalmente, porém, o movimento usado para a interceptação é assim chamado por lembrar o movimento de uma lâmina cortando algo. Por fim, notamos o sentido metafórico do verbo “dar” na expressão “dar um passe”, cujo sujeito é *o autor do primeiro gol*. Essa metáfora, em português brasileiro, conceptualiza o *frame* *Fornecer_passe* e é exclusiva do léxico do futebol em nossa língua materna.

Outro tipo de resultado evidenciado através da anotação da cena *Passe* foi a incidência de evocadores de *frame* polissêmicos, como mostram os seguintes exemplos:

[Passe/Dominar/]

212 - Aos sete minutos, [Tardelli _{jogador_que_recebe}] **recebeu** [ótimo passe _{passe}] de [Muriqui _{jogador_que_passa}] e **tocou** na saída do goleiro para fazer o terceiro gol: Galo 3 x 1.

[Passe/Passe]

548 – [O Furacão _{jogador_que_passa}] tentou **tocar** mais a [bola _{bola}] para tirar a velocidade do Timão, mas não adiantou.

Na sentença 212, observamos o verbo *tocar* como metáfora para *chutar* – evocando, nessa sentença, o próprio *frame* *Chute*, subordinado à cena homônima. O mesmo verbo aparece, na sentença 548, evocando o *frame* *Passe*: tocar a bola é passá-la.

É importante destacar também a presença de construções com verbo-suporte entre as sentenças anotadas.

[Chute/Defesa]

495 - O Timão partiu para cima para tentar reverter o placar, mas [Felipe _{jogador_que_defende}] foi obrigado a **fazer defesa** logo aos 2min, após cabeceio de Leonardo.

A construção com verbo-suporte poderia ser substituída pelo verbo pleno *defender*. De acordo com Neves (2000), as construções com verbo-suporte oferecem ao falante a possibilidade de conferir efeitos diversificados ao discurso. Ao optar pela construção *fazer defesa*, o usuário está acessando uma gama maior de construções sintáticas, além de, através do verbo *fazer*, estar reforçando o papel semântico de ação da construção com verbo-suporte.

A presença dessas construções representa uma característica importante do *corpus* do projeto *Kicktionary_Br*, uma vez que são construções que refletem um uso menos formal da

língua e estreitamente ligado à oralidade. Para um trabalho voltado à anotação semântica com base em *frames*, esse tipo de resultado demonstra que os complementos dos verbos-suportes atuam como agentes na evocação dos *frames*. Fillmore (2007) mostra que uma das funções dos verbos-suporte seria possibilitar a identificação de subeventos em um frame maior.

A cena *Gol* (*Goal*) totalizou 155 sentenças em nosso *corpus* de estudo. Essa cena possui grande representatividade no *corpus* do projeto. O frame *Gol* foi o *frame* com maior incidência, totalizando 138 sentenças anotadas. O *frame* *Comemorar_Gol* apareceu em seis sentenças, contra cinco do *frame* *Vencer_Goleiro* e três do *frame* *Sofrer_Gol*, enquanto os *frames* *Conceder_Gol*, *Gol_Contra* e *Preparar_Gol* estiveram presentes em apenas uma sentença cada.

Um caso muito interessante observado na anotação da Cena *Gol* refere-se ao *frame* *Comemorar_Gol*, em que se percebeu que há uma diferença na perspectiva do agente da comemoração. No projeto pioneiro *Kicktionary*, a comemoração parte do jogador ou do time que realizou um gol (exemplo 1 e 2 abaixo), enquanto no *corpus* do *Kicktionary-Br* a comemoração é feita pela torcida (exemplos 3 e 4).

- (1) [Aleksei Eremenko Jr scorer] **celebrates** [his goal goal].
- (2) [Netherlands scorer_team] **celebrate** after scoring their second goal.
- (3) Dois minutos depois, Leandro recebeu passe de Ricardinho e cruzou na medida para [Tardelli jogador_que_fez_o_gol] marcar de cabeça e **explodir** de vez [o Mineirão torcida].
- (4) No lance, Carlos Alberto cruzou rasteiro pela direita e [o artilheiro jogador_que_fez_o_gol] completou para o gol, **incendiando** de vez [o Mineirão torcida].

Além da divergência na perspectiva da comemoração, pode-se apontar também o fato de que, nos exemplos tirados do *Kicktionary_Br*, os evocadores de *frame* são metafóricos. *Explodir* e *incendiar* não estão em seu sentido literal ou mais comum, mas sim representando a ação de comemorar da torcida. Isso demonstra a forte marca da oralidade presente na versão em português do Brasil do *Kicktionary*. O uso de evocadores com sentido metafórico também traz à tona a necessidade de se pensar na inclusão de elementos de *frame* próprios do *Kicktionary_Br*, uma vez que a linguagem do *corpus* do projeto abarca expressões e unidades lexicais que não são reconhecidas pelo projeto do alemão Thomas Schmidt.

Assim como ocorreu com a cena *Chute*, na cena *Gol* apareceram construções com verbo-suporte.

232 - Aos 38 minutos, [Tardelli jogador_que_marca] recebeu passe de Correa pela esquerda, driblou seu marcador e invadiu a área para tocar rasteiro na saída do goleiro e **fazer um belo gol**.

Na sentença 232, percebe-se que a construção com verbo-suporte é mais adequada. Primeiramente, temos a presença do adjetivo *belo* que impede a troca da construção por um verbo pleno. Além disso, deve-se mencionar que essa é uma construção que não possui uma contraparte constituída de verbo pleno, já que o verbo *golear* não possui o mesmo sentido da construção com verbo-suporte.

A cena *Falta* fez-se presente em 117 sentenças. O *frame* mais expressivo foi o *frame* *Jogada_de_bola_parada*, com 47 ocorrências, seguido pelos restantes: *Dar_cartão* (2), *Decisão_do_juiz* (14), *Falta* (20), *Impedimento* (5), *Penalização* (24) e

Receber_cartão (5). Não houve ocorrências dos *frames* Conceder_compensação e Dar_vantagem.

Um caso interessante aconteceu na anotação do *frame* Falta. No Kicktionary, a unidade lexical *foul* aparece de duas maneiras: como substantivo e como verbo (exemplo 5). Entretanto, no Kicktionary_Br, a unidade lexical *falta* aparece apenas como substantivo (exemplo 6).

(5) [Elin Topuzakov offender] **fouled** [substitute Christian Wilhelmsson offended_player] [in the penalty area area] and the Arsenal FC midfield player chipped in the resulting spot-kick.

(6) Após boa jogada do Timão, [Dentinho jogador_que_sofre_infração] **sofreu falta** [pela esquerda origem].

Isso ocorre porque o uso verbo *faltar* não é recorrente na linguagem contemporânea do universo do futebol, embora apareça na fala dos mais velhos (grifo nosso). Isso, mais uma vez, demonstra as diferenças entre os *corpora* dos dois projetos.

Abordaremos agora as cenas Chance (70), Falta (117) Um_contra_um (42), Confusão (8), Substituição (64), Movimento (50), e Estado_da_partida (60), as menos recorrentes em nosso *corpus* de estudo. Aqui, comentaremos quatro delas.

Quanto à cena Chance, 70 sentenças foram anotadas, sendo que o *frame* Chance foi o mais recorrente, com 57 casos. O *frame* Criar_chance apareceu em 11 sentenças e o *frame* Perder_chance em apenas 2. A cena Chance é um exemplo de cena que não apresentou divergências na relação entre os *corpora* dos dois projetos. Algo que foi possível perceber, entretanto, foi a grande incidências de adjetivos relacionados à unidade lexical “chance”, exemplificado na sentença (7).

(7) As modificações surtiram efeito, o [Flamengo time] **criou boas chances** e conseguiu, aos 41 minutos e aos 43, marcar os dois gols da vitória.

Cena Um_contra_um (one_on_one): A cena Um_contra_um possui seis *frames*: Superar (20), Desarmar (6), Perder_a_bola (5), Disputa (4), Um_contra_um (3) e Desafiar (4). Em relação a essa cena, podemos citar sentenças em nosso *corpus* que apresentam evocadores exclusivos do português brasileiro, como observamos a seguir:

[Um_contra_um/Desarmar]

141 - O lance começou com a **roubada de bola** d[o volante Zé Luis jogador_com_a_bola] [na intermediária área] pela esquerda e se desenvolveu na troca de passes envolvendo Tardelli, Correa e Muriqui.

[Um_contra_um/Disputa]

79 - Após bola recuada, [o atacante da equipe baiana jogador_adversário] **chegou solando** para [ganhar a jogada disputa] d[o goleiro Marcelo jogador_com_a_bola] e tocar para o gol.

A sentença 141 evidencia o *target* *roubada de bola* como evocador do *frame* desarmar. É, portanto, um evocador de *frame* metafórico: “tomar a bola é roubá-la”. No exemplo seguinte, percebe-se, em oposição ao supracitado, a pura descrição da maneira como um dos jogadores disputa a bola: “chegar solando” é valer-se da sola da chuteira para obter a bola – que se encontra em poder do adversário – uma maneira, no mínimo, rude, de disputar a posse de bola.

Cena Substituição (Substitution): A cena Substituição contém os *frames* Substituir (39), Retirar (8) e Colocar (16). Essa cena apresentou resultados notáveis, se comparadas as suas ocorrências em nosso *corpus* aos exemplos oferecidos pelo projeto pioneiro na língua inglesa: em português brasileiro, diversas sentenças onde o *frame* Retirar foi computado apresentam o próprio jogador como sujeito da sentença – em oposição aos exemplos do projeto alemão, onde o sujeito dessas sentenças seria o *técnico*, uma vez que é ele quem promove as modificações na equipe:

[Substituição/Retirar/Colocar]

155 - Aos 32 minutos, foi a vez de [Ricardinho jogador_substituído] **deixar** o campo para a entrada de [Júnior substituto].

[Substituição/Retirar/ Colocar]

218 - Em seguida, foi a vez de [Fabiano jogador_substituído] **deixar** o campo para a entrada de [Renan Oliveira substituto].

No caso das sentenças em português, o técnico, embora efetue a alteração, não é mencionado na sentença. Tal resultado evidencia outra peculiaridade da linguagem futebolística corrente no Brasil, o que não se verifica nos exemplos disponíveis na plataforma Kicktionary.

Também foram anotadas sentenças semelhantes aos exemplos contidos na base do projeto pioneiro, como o que vemos a seguir:

[Substituição/Retirar/Colocar]

637 - Aos 28min, [Mano técnico] mexeu novamente na equipe, **tirando** [Ronaldo jogador_substituído] e **colocando** [Souza substituto] em campo.

Essa sentença explicita a ação do técnico, em oposição às duas sentenças anteriores. Sentenças como essa apresentam uma estrutura comum nas quatro línguas – inglês, francês, alemão e português brasileiro – conforme demonstramos neste estudo.

Os resultados expostos e comentados aqui revelam que o *corpus* do Kicktionary_Br possui sua própria identidade. A presença de construções com verbo-suporte, por exemplo, demonstram que, como já foi mencionado, a linguagem do *corpus* da versão em português do Brasil do Kicktionary carrega muitos traços da oralidade, o que oferece ao projeto uma inestimável riqueza linguística.

4. Considerações finais

O presente artigo buscou descrever o processo de desenvolvimento do projeto Kicktionary_Br, cujo objetivo foi construir, para o português brasileiro, um recurso lexical monolíngue baseado nos princípios da semântica de frames. Retomamos desde as questões preliminares à anotação semântica até os primeiros resultados obtidos através dessa investigação, os quais foram divulgados em diversos eventos e trabalhos ao longo do projeto.

É importante mencionar o pioneirismo do estudo em questão, o qual, inspirado no projeto do Dr. Thomas Schmidt, desenvolve a primeira base de dados lexical monolíngue representativa do léxico futebolístico do português brasileiro. Observamos que a linguagem específica do futebol, esporte símbolo da identidade nacional, é um campo profícuo no que

concerne às investigações linguísticas, revelando importantes traços da língua através desse domínio linguístico bastante popular.

Quanto ao processo de anotação semântica das construções presentes no *corpus*, destacamos o surgimento de casos especiais, que apresentam desafios para a metodologia de anotação. O primeiro se refere à ocorrência de mais de uma cena na mesma sentença, caso constatado por todos os anotadores. Para a continuidade do trabalho, situações como essa implicarão uma análise mais minuciosa, que resultará em tomada de decisão.

Outro desafio diz respeito à tradução das metáforas em inglês, visto que esses fenômenos da linguagem são culturalmente construídos, variando de acordo com as experiências físicas e culturais de cada grupo social (LAKOFF; JOHNSON, 1980/2002). Nesses casos, uma tradução meramente literal inviabilizaria a preservação do sentido da expressão original.

Também observamos que a compilação de um *corpus* comparável próprio consistiu num desafio, já que o Kicktionary desenvolvido por Thomas Schmidt trabalhou com *corpora* paralelos do inglês, francês e alemão. Portanto, anteriormente ao nosso estudo, não havia um *corpus* representativo do português brasileiro.

Como etapas futuras do projeto, ressaltamos a compilação do *corpus* em inglês para o projeto Kicktionary_Br Copa 2014, que tem como proposta a construção de um dicionário eletrônico bilíngue para o léxico do futebol, o qual será divulgado durante a Copa de 2014, no Brasil, em formato de aplicativo para celular. Esse recurso lexical terá como público-alvo o turista que visitará o país no período da Copa, e seu sistema de pesquisa permitirá associações entre unidades lexicais relacionadas a uma mesma cena.

Vale destacar que este estudo teve como foco a anotação semântica de *corpus*, propósito um pouco distinto do projeto alemão. Elencamos também como trabalho futuro a construção de recurso lexicográfico seguindo os moldes do projeto pioneiro. Em consequência desta nova demanda, nos afastaremos da chamada anotação no modo *running text* e focalizaremos a metodologia lexicográfica de anotação.

É importante assinalar que este trabalho, naturalmente, apresenta limitações, visto que o estudo que compreendemos relaciona-se ao domínio específico do futebol, ou seja, a investigação não se estende a outros domínios. Acrescentamos, ainda, que dados específicos do projeto poderão ser acessados através de consulta ao material anotado, assim que este for disponibilizado.

Agradecimentos

A autora agradece o apoio financeiro da bolsa PQ do CNPq e dos auxílios do CNPq, CAPES e FAPERGS.

Referências

BERBER SARDINHA, A. P. *Linguística de Corpus*. São Paulo: Manole, 2004.

BERTOLDI, A. *Semântica de frames e recursos lexicais jurídicos: um estudo contrastivo*. 2011, 136 f. Tese (Dourado em Linguística Aplicada) – Programa de Pós-Graduação em Linguística Aplicada. Universidade do Vale do Rio dos Sinos, São Leopoldo, RS, 2011.

BICK, E. *The parsing system PALAVRAS: automatic gramatical analysis of Portuguese in a constraint grammar framework*. Arhus: Arhus University, 2000/

BIDERMAN, M. T. Unidades complexas do léxico. In: Rio-Torto, G.; Figueiredo, O. M.; SILVA, F. (Org.) *Estudos em homenagem ao professor doutor Mário Vilela*. 1ª ed. Porto: Portugal: Faculdade de Letras da Universidade do Porto, v. II, p.747-757, 2005.

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

EVANS, V. The Cognitive Linguistics enterprise: an overview. In: *The Cognitive Linguistics Reader*. Equinox Publishing Company, 2006.

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

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

FILLMORE, C.J. An alternative to checklist theories of meaning. *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society*. Berkeley, CA: Berkeley Linguistics Society, p. 123-131, 1975.

FILLMORE, C. J.; JOHNSON, C.; PETRUCK, M. Background to FrameNet. *International Journal of Lexicography*, vol.16, n. 3, p.235-250, 2003.

LANGACKER, R. *Foundations of the Cognitive Grammar*. Stanford: Stanford University Press, vol. 1, 1987.

LAKOFF, George. JOHNSON, Mark. *Metaphors we live by*. Chicago: The University of Chicago, 1980.

NEVES, M. H. M. *Gramática de usos*. São Paulo: Unesp, 2000.

SCHMIDT, Thomas. The Kicktionary – a multilingual lexical resource of football language. In: BOAS, Hans. (Ed.) . *Multilingual FrameNets - Methods and Applications*. Berlin/New York: Mouton de Gruyter, 2009.

WYNNE, Martin (editor). 2005. *Developing Linguistic Corpora: a Guide to Good Practice*. Oxford: Oxbow Books. Disponível online em <http://ahds.ac.uk/linguistic-corpora/> [Acesso em 27/5/2010].

Received: 18/02/2013

Revised: 26/05/2013

Accepted: 24/06/2013

Published: 31/10/2013