

Diretrizes para a criação de um recurso lexical multilíngue a partir da semântica de *frames*: a experiência turística em foco

Guidelines for the creation of a multilingual lexical resource based on Frame Semantics: the tourist experience in focus

Maucha Andrade Gamonal*
Tiago Timponi Torrent**

RESUMO: Este artigo apresenta as diretrizes utilizadas para o desenvolvimento do Dicionário FrameNet Brasil da Copa do Mundo, dicionário eletrônico trilingue (Português – Inglês – Espanhol) para os domínios da Copa, do Futebol e do Turismo. Caracterizada como uma teoria linguística que enfatiza a estreita relação entre sistema linguístico e experiência humana, possibilitada através de *corpus*, a Semântica de *Frames* desenvolve molduras que são evocadas pelas palavras na construção de seus significados. A FrameNet, rede semântica em constante desenvolvimento para a língua inglesa no International Computer Science Institute, em Berkeley, e em processo de extensão para outras línguas em diversos países, como o Brasil, fornece a metodologia necessária para a criação deste recurso. O produto desenvolvido a partir deste aporte teórico-metodológico, disponível online gratuitamente, é aqui apresentado através do domínio turístico.

PALAVRAS-CHAVE: FrameNet Brasil. Semântica de *Frames*. Lexicografia Computacional. Dicionários Eletrônicos. Experiência turística.

ABSTRACT: This paper presents the guidelines used for the development of the FrameNet Brazil World Cup Dictionary, a trilingual electronic dictionary (Portuguese – English – Spanish) for the domains of the World Cup, Football and Tourism. Emphasizing the close relation, attested by corpora, between language systems and human experience, Frame Semantics develops frames that are evoked by words for constructing their meanings. FrameNet, a semantic network being developed for English at the International Computer Science Institute in Berkeley, and in process of extension to other languages in several countries, such as Brazil, provides the methodology necessary for the creation of this resource. The product developed from this theoretical and methodological basis, freely available online, is presented here from the perspective of the tourist domain.

KEYWORDS: FrameNet Brazil. Frame Semantics. Computacional Lexicography. Electronic Dictionaries. Touristic Experience.

1. Introdução

Uma imensa quantidade de pesquisas é financiada mundo afora com o objetivo de desenvolver mecanismos computacionais inteligentes capazes de manipular a linguagem

* Doutoranda em Linguística pela Universidade Federal de Juiz de Fora. Bolsista de Doutorado Sanduíche do Programa Ciência sem Fronteiras junto ao International Computer Science Institute e à University of California Berkeley.

** Docente do Programa de Pós-Graduação em Linguística da Universidade Federal de Juiz de Fora.

humana. Por mais que a diferença linguística não deva ser fator que limite a comunicação entre pessoas de todo o mundo, este ainda é o panorama. Tendo isso em vista, o linguista que orienta seus estudos através de *corpus* e apresenta interesse em contribuir para os estudos voltados ao Processamento de Linguagem Natural assume papel importante neste cenário.

A rede de *frames* FrameNet¹ (BAKER, 2008; RUPPENHOFER *et al*, 2010) é um projeto que utiliza *corpus* como fonte de respostas empíricas para as propriedades lexicais da língua inglesa. Em desenvolvimento desde 1997 no International Computer Science Institute, na cidade de Berkeley, Califórnia, a FrameNet explora o conceito de *frames* na construção de um recurso lexical para a língua inglesa através da Semântica de *Frames* de Charles Fillmore (1982). Para Fillmore (1985), os sentidos são possibilitados através de palavras inseridas em contextos, os *frames*, ou seja, molduras cognitivas que nos permitem fazer as devidas correlações de sentidos.

Com intuito de estender tal recurso para o português brasileiro, a professora Margarida Salomão investiu na implementação da FrameNet Brasil – <http://www.ufjf.br/framenetbr/> – (SALOMÃO, 2009), que vem sendo desenvolvida desde então na Universidade Federal de Juiz de Fora. O retorno instigante possibilitado pelos *insights* da Semântica de *Frames* e a tentativa de criar um recurso lexical nos termos de tal teoria levou a FrameNet Brasil a investir na criação de um recurso lexical trilingue – Português, Inglês, Espanhol – para os domínios do Turismo, do Futebol e da Copa do Mundo, o Dicionário FrameNet Brasil da Copa do Mundo, disponível para consulta através do endereço <http://dicionariodacopa.com.br/>.

O presente trabalho tem como intuito apresentar, através do domínio do turismo, as principais decisões teórico-metodológicas adotadas para este dicionário. Assim, as duas principais questões que orientam a estruturação deste texto são: como a Semântica de *Frames* e a FrameNet² podem atuar no desenvolvimento de dicionários eletrônicos para usuários não especializados e qual é a vantagem de se utilizar *corpus* para viabilizar tal empreendimento.

2. A Semântica de *Frames* no desenvolvimento de dicionários eletrônicos

Inserida nos estudos da Linguística Cognitiva, a Semântica de *Frames* surgiu através de Charles J. Fillmore a partir do desenvolvimento da Gramática de Casos (FILLMORE, 1968a,

¹ <https://framenet.icsi.berkeley.edu>

² Quando o intuito for fazer referência ao projeto mãe desenvolvido em Berkeley, apenas o nome Framenet com a inicial maiúscula será utilizado. Ao se referir a suas extensões, será acrescentado o país de origem.

1968b), hipótese para representação semântica tendo por base as relações existentes entre predicador e seus complementos, o que ele chamou de “casos”. Tanto casos sintagmáticos como semânticos foram abordados por esta investigação e a regularidade em tais combinações foi o que Fillmore considerou como os *case frames*. O desenvolver da pesquisa mostrou que os casos atribuídos não eram suficientes para dar conta de diferenças semânticas importantes, o que o fez optar por funções microtemáticas (FILLMORE, 2003), desenvolvendo, assim, uma Semântica de *Frames*, que, nas palavras de seu próprio criador pode ser assim definida:

um programa de pesquisa em linguística empírica e uma metodologia descritiva para apresentar os resultados de tal pesquisa (...) pelo termo *frame*, eu tenho em mente qualquer sistema de conceitos relacionado de tal forma que, para entender qualquer um deles, você tem de entender toda a estrutura na qual ele se encaixa; quando um dos conceitos em dada estrutura é introduzido dentro de um texto ou de uma conversa, todos os outros são automaticamente disponibilizados (...) (FILLMORE, 1982, p.111).³

Um clássico exemplo é o *frame* que trata da Transação_comercial (FILLMORE, 1977). Há uma pessoa – o VENDEDOR – interessada em repassar mercadorias em troca de dinheiro para outra – o COMPRADOR – que aceita trocar DINHEIRO por MERCADORIAS⁴. Destacando tal experiência em termos verbais, podemos dizer que “comprar”, “pagar” e “vender”, sem dúvida, fazem referência a tal evento, mas perfilam perspectivas distintas, uma será a do comprador sobre a mercadoria, outra do comprador sobre o dinheiro necessário para a obtenção de mercadorias e a outra será do vendedor sobre a mercadoria. Vejam-se, na Tabela 1, exemplos do *corpus* Copa 2014 FRAMENET BRASIL, constituído para o desenvolvimento do dicionário.

³ Texto original: “(...) a research program in empirical semantics and a descriptive framework for presenting the results of such research (...) By the term 'frame' I have in mind any system of concepts related in such a way that to understand any one of them you have to understand the whole structure in which it fits; when one of the things in such a structure is introduced into a text, or into a conversation, all of the others are automatically made available”.

⁴ Ao longo do texto, os nomes dos *frames* aparecem com a fonte Courier enquanto os elementos que os compõem, em letra maiúscula. Já os predicadores considerados potenciais evocadores de *frames* são destacados em negrito e letra maiúscula.

Tabela 1. Exemplos de sentenças que instanciam os diferentes *frames* da transação comercial.

<i>Frame</i>	Exemplo
Comércio_comprar	COMPRAMOS uma bacia de camarão por cinco reais.
Comércio_vender	As barracas VENDEM comida, bebida e artesanato típico de cada região.
Comércio_pagar	Teríamos pago menos da metade do que nós PAGAMOS .

Fillmore quis salientar que a compreensão de uma palavra só é possível caso o *frame* seja mentalmente acessado. Compreender, por exemplo, o sentido de “comprar”, “vender” ou “pagar” nos exemplos dados significa conhecer molduras de conhecimento como Comércio_comprar, Comércio_vender e Comércio_pagar. Por tal motivo, o significado linguístico é relativizado a *frames* (FILLMORE, 1977, p.59).

Enquanto isso, em outras áreas, pesquisadores também escolheram o mesmo termo para desenvolverem seus estudos. Na sociologia, Goffman explorava a palavra *frame* para enfatizar a moldura de conhecimento necessária para a compreensão de intenções, perspectivas, rituais e padrões que os indivíduos estabelecem na interação cotidiana. Com a publicação do livro “Frame Analysis: An Essay on the Organization of Experience”, em 1974, ele discorreu sobre a organização social das experiências humanas, revelando como as interações sociais são definidas em torno de molduras específicas que orientam as ações dos indivíduos na sociedade.

Dentre suas variadas analogias, ele considerava que a atuação do Homem na sociedade pode ser comparada a uma peça de teatro, em que vários papéis sociais são assumidos a depender da função da interação estabelecida. Bastante ilustrativa, nesse sentido, é a ideia do uso de máscaras no legado teatral grego. Elas eram utilizadas para representar personagens, trocar de máscara era trocar de papel. Na vida cotidiana, num mesmo dia, o comportamento varia diversas vezes, ora somos filhos, ora pais; ora estudantes, ora professores; podemos ser vendedores, mas, certamente, também somos consumidores. Esse fluxo constante nos mostra que ocupamos diversos papéis, a troca de “máscaras” é uma necessidade diária, não uma escolha, pois o “cenário” e “os personagens” variam.

Já Minsky, um dos fundadores do laboratório de Inteligência Artificial do Instituto de Tecnologia de Massachusetts, demonstrou interesse pelos estudos da cognição humana com a proposição de estruturas de dados estereotipadas que representam as situações. Segundo ele, o conhecimento não deve ser visto como uma coleção de fragmentos simples e desconexos, e, sim, como estruturas complexas, denominadas *frames*, definidas, por ele, como

uma estrutura de dados para a representação de situações estereotipadas, tais como estar em certo tipo de sala de estar ou ir a um aniversário de criança. Anexos a cada *frame* existem diversos tipos de informações. Algumas dessas informações dizem respeito ao modo de uso do *frame*. Algumas concernem ao que se espera que aconteça em seguida. Algumas tratam do que fazer caso essas expectativas não se confirmem. (MINSKY, 1975, p.1)⁵

Um exemplo utilizado é o da festa de aniversário, com ele, o autor pondera que as definições de dicionário nunca dizem o suficiente. Por mais que qualquer pessoa saiba que esse tipo de evento envolve mais que um encontro para comemorar mais um ano de vida completado por alguém, nenhuma definição breve é capaz de mostrar a complexidade desse evento. No Brasil, como em vários outros países, há o costume de acender as velas postas no bolo durante a música de comemoração. Dessa forma, se, antes deste momento, o anfitrião lamenta que se esqueceu da vela, dificilmente, alguém irá questionar se a luz acabou, o que aponta para o fato de que todos compartilham das mesmas expectativas no que diz respeito à experiência com a festa de aniversário.

O resultado comum pelo termo “*frame*” não foi uma mera coincidência dos diferentes objetos de estudo no final do século XX, mas confirma a necessidade de investigar tais estruturas e a necessidade de investir mais no diálogo científico. O interesse pelo comportamento social e pelas maneiras de conceitualizar a linguagem humana em termos linguísticos e computacionais corrobora a nossa opção pela Semântica de *Frames*, abordagem que inspirou e baseou o desenvolvimento do Dicionário FrameNet Brasil da Copa do Mundo. A pretensão esteve em testá-la como mecanismo único de organização de um dicionário para não especialistas (TORRENT et al, 2014), diferentemente, por exemplo, do recurso lexical *online* do domínio do Futebol, o Kicktionary (SCHMIDT, 2006; 2007; 2008; 2009), disponível *online* em <http://www.kicktionary.de>, que, apesar de fazer uso também de aspectos da Semântica de *Frames*, confia mais nas WordNets para seu desenvolvimento.

Nesse sentido, passamos, a seguir, à apresentação da metodologia adotada para o desenvolvimento do Dicionário FrameNet Brasil da Copa do Mundo.

⁵ Texto original: “A frame is a data-structure for representing a stereotyped situation like being in a certain kind of living room or going to a child's birthday party. Attached to each frame are several kinds of information. Some of this information is about how to use the frame. Some is about what one can expect to happen next. Some is about what to do if these expectations are not confirmed”.

3. A metodologia da FrameNet adaptada aos propósitos de um dicionário eletrônico para usuários não-especialistas

Considerada uma extensão teórica da Semântica de *Frames* para o domínio da lexicografia computacional, o objetivo inicial da FrameNet de criar um recurso lexical para descrever a língua inglesa num viés semântico e sintático passou também a ser útil para profissionais ligados à lexicografia, na estruturação de dicionários, e para a Linguística Computacional, em tarefas ligadas ao Processamento de Linguagem Natural (PLN).

Como extensão à ideia original, atualmente, a *framenets* estão sendo desenvolvidas em outros lugares do mundo, além do Brasil, como Alemanha, Coreia do Sul, Espanha, Japão e Suécia. O intuito é desenvolver bancos de dados em suas próprias línguas e, futuramente, viabilizar empreendimentos multilíngues. No Brasil, a FrameNet vem sendo explorada desde 2007, na Faculdade de Letras da Universidade Federal de Juiz de Fora. O sítio da FrameNet Brasil, disponibiliza, com acesso livre, a rede semântica para o Português do Brasil, tanto de vocabulário específico como de vocabulário genérico.

3.1. A FrameNet

Tendo todas as análises lexicais ancoradas em evidências em *corpora*, a estrutura de uma *framenet* que siga os princípios do projeto-mãe se alicerça em três grandes tipos de dados: *frames*, unidades lexicais e sentenças anotadas.

Os *frames* são vistos como uma modelagem com interesse computacional de uma estrutura de conhecimento reconhecível em uma dada cultura. São definidos em torno de seus constituintes, os Elementos de *Frame* (EF), que podem ser atores, ferramentas ou circunstâncias, por exemplo.

As unidades lexicais (ULs) são entendidas como o pareamento de uma forma, com todas as suas flexões, a um significado específico, são essas palavras associadas a sentidos específicos que evocam os *frames*. Os verbos assumem destaque especial, pois são predicadores por natureza, mas nomes, adjetivos, advérbios e também preposições são evocadores de *frames*.

As anotações lexicográficas giram em torno de sentenças, elas fornecem evidência empírica para as análises que levaram à constituição dos *frames* e à definição das ULs. São analisadas tanto sintaticamente quanto semanticamente. Assim, dada uma UL, as sentenças em que ela se instancia têm seus constituintes sintagmáticos anotados tanto para os EFs que circundam a UL (anotação semântica), quanto para as funções gramaticais e tipos sintagmáticos

que caracterizam o material linguístico que manifesta os EFs (anotação sintática). As funções gramaticais e os tipos sintagmáticos utilizados para o português do Brasil foram definidos por Torrent & Ellsworth (2013).

Como pode se ver, nesse contexto, a semântica é o centro para a explicação da gramática e as regularidades combinatórias abstraídas das anotações, ou seja, os padrões de valência, assumem destaque primordial. Considerando-se as três estruturas de dados principais apresentadas, *framenets* se caracterizam como esforços de lexicografia prática que têm como tarefas, segundo Fillmore (2008):

- i) descrever Unidades Lexicais a partir dos *frames* evocados, bem como descrever os respectivos *frames*;
- ii) descrever os Elementos de *Frame* que compõem cada *frame*;
- iii) extrair sentenças de *corpora* para validar as análises das Unidades Lexicais;
- iv) selecionar, dentre as sentenças extraídas, aquelas que sejam representativas das diversas possibilidades de valência das Unidades Lexicais;
- v) disponibilizar os resultados na forma de entradas lexicais que resumem os padrões de valência sintático-semântica das Unidades Lexicais;
- vi) definir uma rede de relações entre *frames* e apresentá-la graficamente.

Na subseção seguinte, as etapas descritas acima serão ilustradas com os dados do Copa 2014 da FrameNet Brasil através de adaptações necessárias no que diz respeito à metodologia.

3.2. Tarefas linguísticas para a constituição de um dicionário baseado em *frames*

A primeira tarefa realizada no projeto do dicionário foi a compilação de *corpora* específicos. Nessa tarefa, foram fundamentais as contribuições de Sardinha (2004), no que tange aos requisitos para a criação e caracterização dos *corpora*, e de Calvi (2010), no que tange aos gêneros textuais a serem incluídos neles.

Começando pelas primeiras, Sardinha (2014) propõe que textos a serem incluídos em *corpora* devem ser autênticos, em linguagem natural e não desenvolvidos com o propósito de servir a uma pesquisa linguística. Numa perspectiva multilíngue, *corpora* autênticos de mais de uma língua podem ser comparáveis ou paralelos. Enquanto estes se caracterizam por conter versões traduzidas de um mesmo texto, aqueles são compostos por textos de um mesmo gênero, porém, sem o compromisso de representação de um mesmo conteúdo. O Dicionário FrameNet Brasil da Copa do Mundo explorou ambos os tipos de *corpora*, uma vez que o tamanho dos

textos paralelos encontrados (c. 375.000 tokens por idioma) não se mostrou suficiente para atestar adequadamente as ULs levantadas.

Além desse aspecto, vários outros critérios presentes na literatura da Linguística de *Corpus* (modo, tempo, seleção, conteúdo, autoria e finalidade) foram especificados no que tange aos *corpora* levantados para o domínio do turismo. Assim, os *corpora* coletados sobre turismo são: escritos (modo); sincrônicos/contemporâneos, por designarem o período corrente, atual (tempo); dinâmicos, pois podem ser aumentados e/ou diminuídos (seleção); de domínios especializados (conteúdo); e, por último, são textos de falantes nativos (autoria).

Já no tangente aos gêneros textuais selecionados para compor os *corpora*, seguiram-se as considerações de Calvi (2010, p.19), para quem os gêneros textuais cujos objetivos comunicativos são os de descrever e promover destinos turísticos são os mais representativos do vocabulário turístico. Assim sendo, os *corpora*, nas três línguas alvo do dicionário, são compostos por guias de turismo (textos paralelos com c. 375.000 tokens por idioma), *sites* governamentais de fomento a atividades turísticas (textos comparáveis com c. 585.000 tokens por idioma) e *blogs* de viagem (textos comparáveis, com c. 40.000 tokens por idioma), totalizando cerca de 1.000.000 de tokens por idioma.

Todos os *corpora* foram pré-processados sintaticamente, utilizando-se os *parsers* PALAVRAS (BICK, 2000), para o português, e TreeTagger (SHMID, 1994), para o inglês e o espanhol. Posteriormente, foram armazenados, compilados e acessados através da ferramenta SketchEngine (<http://sketchengine.co.uk>). Criada a infraestrutura para o trabalho com os *corpora*, a próxima tarefa centrou-se na criação dos *frames*.

Na etapa de estruturação de *frames*, priorizou-se o método *bottom-up*, que partia do mais básico, as ULs, para aquilo mais abstrato, os *frames*. Assim, o anotador parte dos dados para criar o *frame*. A primeira tarefa é selecionar, intuitivamente, um agrupamento de lexemas que se relacionam semanticamente, por exemplo, *visitar*, *turista*, *apreciar*, *visitante* e *atração*. Estes lexemas, ainda que apresentem particularidades quanto ao sentido e não pertençam todos às mesmas classes de palavras, participam de um mesmo domínio semântico, o das atividades turísticas.

Num segundo momento, parte-se para a pesquisa dos itens lexicais selecionados nos *corpora*, tendo em mente um possível *frame* para investigação. Após esse estudo, selecionam-se algumas sentenças com esses itens lexicais com o objetivo de analisar o comportamento desses predicadores no que tange à valência sintático-semântica. Com isso em mãos, o analista

procura regularidades tanto semânticas quanto sintáticas que permitam a estruturação de uma situação específica, essa etapa é a definição do *frame*. Nela, são determinadas as ULs, previamente selecionadas pelo analista, são especificadas a nuclearidade dos Elementos de *Frame* (EFs) bem como as relações entre os EFs. Quando o *frame* estiver estruturado, é possível relacioná-lo com outros, caso existam.

É importante ressaltar que, por serem funções microtemáticas, EFs – e, por consequência, os *frames* que deles se compõem – podem ser bastante específicos em sua definição. A título de exemplo, observem-se (1) e (2):

- (1) [Jô Soares VISITANTE] VISITA [a presidente Dilma Rousseff ENTIDADE], [em Brasília LUGAR].⁶
- (2) [Dilma TURISTA] VISITA [praia deserta ATRAÇÃO] [na Ilha dos Frades, na Bahia LOCAL].⁷

Veja-se que, tanto em (1) quanto em (2), poderíamos assumir a mesma estrutura sintática NP V NP. Todavia, os dois enunciados referem-se a experiências distintas. Em um caso, há elementos que sugerem um tipo de visita que não se adequa ao esperado para atividades turísticas. Assim, as duas sentenças evocariam *frames* de *background* distintos na construção do sentido de *visitar* em cada uma: o de Visitar, para (1) e o de Turismo_por_turista, para (2). Porém, como a atividade turística não deixa de compartilhar traços genéricos de uma visita, na base de dados, esses dois *frames* estão relacionados entre si.

A análise de dados acontece pela anotação lexicográfica da FrameNet Brasil, majoritariamente, em três camadas, Elemento de Frame (no inglês FE, Frame Element), Função Gramatical (no inglês GF, Gramatical Function), e Tipo Sintagmático (no inglês PT, Phrase Type). A Figura 1 mostra a anotação de uma sentença que instancia a UL *visitar.v*, evocadora do *frame* Turismo_por_turista.

⁶ Manchete acessada em 23 set. 2015. <http://www.ofuxico.com.br/noticias-sobre-famosos/jo-soares-visita-a-presidente-dilma-rousseff-em-brasilia/2015/05/18-239130.html>

⁷ Manchete acessada em 23 set. 2015. <http://correiodopovo.com.br/Noticias/545102/Dilma-visita-praia-desertana-Ilha-dos-Frades,-na-Bahia>

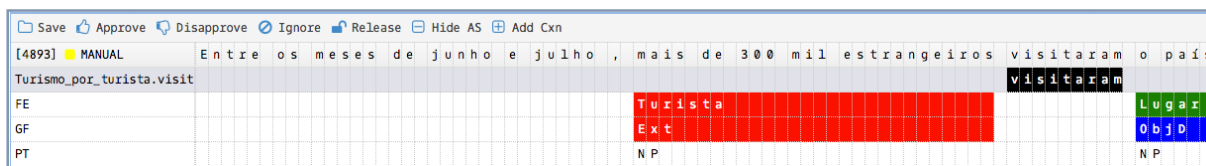


Figura 1. Anotação de uma sentença na FrameNet Brasil

Como as palavras que evocam *frames* são predicadores, como verbos, substantivos e adjetivos, especifica-se, a partir de um conjunto de anotações, a valência desses itens lexicais tanto em relação à sintaxe quanto à semântica.

É válido destacar que, diferentemente da WordNet (FELLBAUM, 1998), recurso lexical que organiza o léxico por uma perspectiva paradigmática, através dos chamados *synsets*, a rede semântica FrameNet e o recurso ora apresentado não sistematizam as informações de maneira semelhante. Ainda que haja o empenho em descrever as informações combinatórias dos itens lexicais, não se pode afirmar que a FrameNet assumam um viés apenas sintagmático. Ao conceber o conhecimento lexical através de molduras intrinsecamente relacionadas às experiências humanas, os *frames* são os responsáveis pelo agrupamento. É possível analisar os dados por um viés paradigmático, por mais que isso não esteja sistematizado no recurso. Por outro lado, as relações combinatórias entre os participantes, os Elementos de *Frame*, desfrutam de grande atenção. Todas as sentenças são analisadas sintaticamente e, depois, são fornecidos os padrões de tais combinações.

As Tabelas 2 e 3 destacam os resultados do chamado processo de anotação, fazendo referência à palavra *apreciar*. Os *integrantes* da cena a que está vinculada (os Elementos de *Frame*) aparecem em destaques coloridos. A Tabela 2 enfatiza os dados em relação aos participantes e as suas realizações sintáticas, já a Tabela 3 mostra os padrões de valência. Ou seja, a partir das realizações sintáticas encontram-se os padrões combinatórios das análises feitas em *corpora*.

Tabela 2. Padrões de realização dos EFs – *apreciar*

Elemento de Frame [Frame Element]	Número Anotado [Number Annotated]	Realizações [Realization(s)]
Atração	(17)	NP.Ext (1) NP.ObjD (16)
Lugar	(11)	NP.Dep (3) 2nd.-- (1) DNI.-- (7)
Maneira	(4)	PP.Dep (3) NP.Dep (1)
Turista	(17)	CNI.-- (6) NP.Ext (10) DNI.-- (1)

Tabela 3. padrões de valência - *apreciar*

Número Anotado [Number Annotated]	Patterns				
3 TOTAL	Atração	Lugar	Lugar	Turista	
(1)	NP Ext	NP Dep	NP Dep	CNI --	
(1)	NP ObjD	NP Dep	NP Dep	CNI --	
(1)	NP ObjD	NP Dep	NP Dep	NP Ext	
1 TOTAL	Atração	Lugar	Maneira	Turista	
(1)	NP ObjD	DNI --	PP Dep	NP Ext	
1 TOTAL	Atração	Lugar	Maneira	Turista	Turista
(1)	NP ObjD	2nd --	PP Dep	NP Ext	NP Ext
4 TOTAL	Atração	Lugar	Turista		
(1)	NP ObjD	DNI --	CNI --		
(1)	NP ObjD	DNI --	DNI --		
(2)	NP ObjD	DNI --	NP Ext		
2 TOTAL	Atração	Lugar	Turista	Turista	
(2)	NP ObjD	DNI --	NP Ext	NP Ext	
2 TOTAL	Atração	Maneira	Turista		
(1)	NP ObjD	NP Dep	CNI --		

No Dicionário FrameNet Brasil da Copa do Mundo, três conjuntos diferentes de etiquetas foram utilizados para a camada GF e outros três para a camada PT, uma vez que nessas camadas registram-se as especificidades morfossintáticas dos idiomas cobertos pelo recurso. Já os EFs foram traduzidos do português para os dois outros idiomas (inglês e espanhol) para fins de interface, sendo os *frames* os mesmos para as três línguas⁸.

⁸ Vide Gamonal e Torrent (GAMONAL;TORRENT, 2014) para uma discussão aprofundada desta questão, à qual este artigo retornará na próxima seção.

4. O Dicionário FrameNet Brasil da Copa do Mundo: o Turismo em foco

O Dicionário FrameNet Brasil da Copa do Mundo encontra-se disponível gratuitamente, na forma de um *web app*, tendo sido lançado em junho de 2014, semanas antes da Copa do Mundo FIFA Brasil 2014. Por ser voltado a não especialistas, no dicionário, a terminologia adotada pelas framenets sofreu pequenas alterações: *frames* passaram a ser chamados de cenas, EFs de participantes, ULs de palavras e assim por diante.

A Figura 2 apresenta a interface inicial ao usuário, na qual ele deve selecionar o idioma através do qual pretende interagir com o aplicativo. O recurso pode ser explorado através dos seguintes comandos: *buscar por palavra*, *digitar texto*, *ver significado* e *explorar a rede*, mostrados na tela de acesso principal, na Figura 3.



Figura 2. Seleção do idioma de interface



Figura 3. Tela de acesso aos sistemas de busca

Clicando em *Buscar palavra*, o usuário é levado a uma lista de palavras na língua de sua escolha. Ao clicar sobre qualquer uma delas, por exemplo, *apreciar.v*, o aplicativo apresenta, na primeira tela de resultados – Figura (4) –, a cena evocada pela palavra: Fazer Turismo. Em seguida, há uma definição do item lexical (glosa) bem como equivalências para as demais línguas do dicionário. Tais equivalências são calculadas automaticamente pelo próprio aplicativo, com base nos padrões de valência armazenados na base de dados.

Arrastando a tela de resultados para o lado – Figura (5) –, o usuário pode visualizar os participantes envolvidos na cena evocada pela palavra, os quais, através do código de cores, serão marcados na tela que traz as sentenças de exemplo – Figura (6).

A Figura (5) mostra os participantes da cena Fazer Turismo. A compreensão desta experiência está vinculada, necessariamente, à existência do Turista e também da Atracção. Outro participante é o Acompanhante, aquele que compartilha a experiência com o Turista, o que não significa que não possa assumir o papel de turista, mas, nas evidências em *corpora*, não é o participante em destaque. Veja que a nomenclatura escolhida tem o intuito de ser transparente ao usuário.

Na Figura (6), as sentenças que exemplificam os itens lexicais advêm de *corpora* coletados durante a elaboração do dicionário. O usuário pode optar por acessar o sítio no qual as sentenças foram encontradas ao clicar nelas.

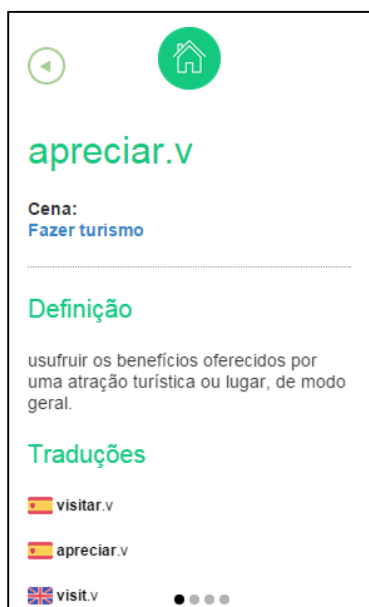


Figura 4. Verbete: cena, glosa e traduções

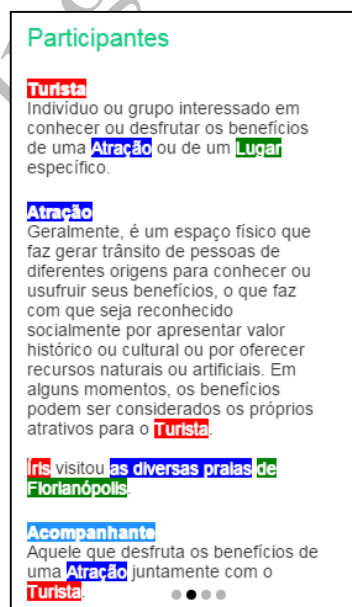


Figura 5. Verbete: participantes da cena

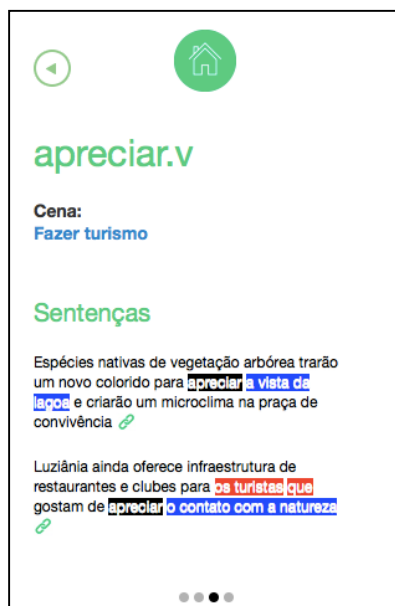


Figura 6. Verbetes: sentenças de exemplo



Figura 7. Verbetes: mais palavras

Por fim, na última tela de resultados, Figura (7), são mostradas as demais palavras que evocam a mesma cena, tais como *conhecer.v*, *desfrutar.v* e *tour.n*. Após o recurso *Digital frase*, ferramenta de busca para acessar o significado das palavras no contexto de uso, a próxima opção é *Ver significado*. Nela, o usuário encontra várias cenas vivenciadas na atividade turística. A Figura (8) ilustra essa ferramenta de busca através da cena Fazer Turismo. Além da definição e apresentação dos participantes, há descrição das palavras que evocam tal experiência tanto na língua portuguesa quanto nas outras duas línguas.

Um recurso muito importante no desenvolvimento da rede semântica FrameNet é a designação das relações estabelecidas entre os *frames*, tais relações projetam como as experiências e os eventos se conectam no estabelecimento das relações de sentido. Objetivou-se reproduzir esse propósito no recurso ora apresentado. A Figura (9) apresenta as relações descritas. A Figura (10) destaca a partir do Cenário da Chegada as relações entre as cenas descritas no dicionário.

Fazer turismo

Definição

Um **Turista** visita ou experiencia o contato com uma **Atração**, reconhecida socialmente, geralmente, por apresentar valor histórico ou cultural e/ou oferecer recursos naturais ou artificiais.

Participantes

Turista
Indivíduo ou grupo interessado em conhecer ou desfrutar os benefícios de uma **Atração** ou de um **Lugar** específico.

Atração
Geralmente, é um espaço físico que faz gerar trânsito de pessoas de diferentes origens para conhecer ou usufruir seus benefícios, o que faz com que seja reconhecido socialmente por apresentar valor histórico ou cultural ou por oferecer recursos naturais ou artificiais.

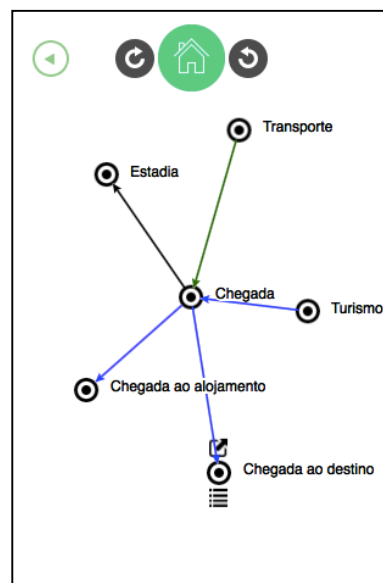


Figura 8. Ver significado - resultado

Figura 9. Explorar rede – busca

Figura 10. Explorar rede: grafo

Ao primeiro trabalho concluído em torno desse projeto, coube formular as diretrizes para a constituição do dicionário, o que foi feito com a criação de *frames* para o domínio do Turismo, bem como através das análises semânticas e sintáticas dos itens lexicais incluídos no recurso. Várias perguntas surgiram no início do estudo e, logo, se tornaram o objeto de pesquisa, dentre elas: em que medida os *frames* do domínio turístico modelados com *corpora* compilados da língua portuguesa do Brasil servem para representar os *frames* do Turismo para as demais línguas do dicionário?

Responder essa pergunta considerava a reiteração de que *frames* podem atuar como modelagens da conceptualização humana. Através de leituras na literatura da área, mas centrando-se na experiência individual e, ao mesmo tempo, coletiva, manifesta na linguagem, conclui-se que o evento turístico pode ser considerado dotado de uma estrutura transcultural. Qualquer pessoa com o intuito de se juntar à prática turística domina todo o *background* envolvido. Tal fato pôde ser comprovado na medida em que a estrutura de modelagem aplicada ao Cenário do Turismo, mostrado parcialmente na Figura 10, para o português brasileiro serviu igualmente para o tratamento lexicográfico do inglês americano e do espanhol europeu (variantes dos textos que compõem os *corpora* dos demais idiomas do dicionário).

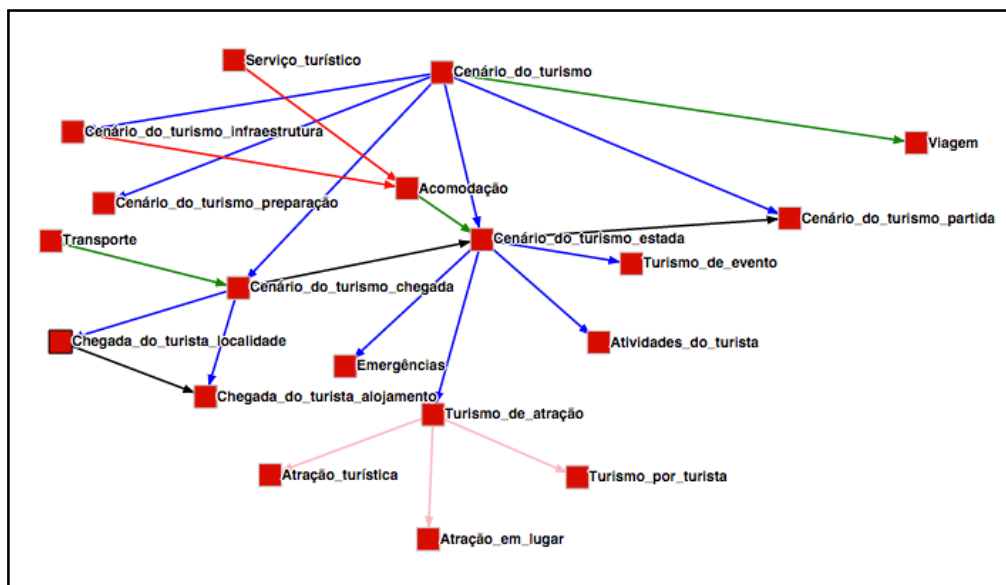


Figura 10. Os *frames* no Cenário do Turismo⁹

A afirmação de que se trata de uma estrutura de *frames* transcultural significa que tal mapeamento é reconhecível por todos, independentemente da diversidade cultural/linguística. Há o conhecimento compartilhado por todos os envolvidos, de que se trata de uma atividade com data marcada, ainda que não haja um dia definido, em que o retorno acontecerá, normalmente, ao local de origem. E assim, os *frames* em tal domínio vão de moldando e sendo esmiuçados.

Não se pode, entretanto, estender essa mesma afirmação para os correspondentes de tradução. Não necessariamente, as palavras do dicionário terão um correspondente perfeito em todas as línguas. Em Peron-Corrêa (2014), isso é detalhado. Uma de suas importantes contribuições neste sentido foi mostrar como aqueles considerados os melhores equivalentes de tradução por diversos dicionários conceituados, na verdade, não evocam o mesmo *frame*. No caso do *frame* de Turismo_por_turista, cujos dados são mostrados no Quadro (1), enquanto algumas Unidades Lexicais cognatas apresentam-se como equivalentes de tradução para o domínio do turismo, *apreciar/apreciar*, *desfrutar/disfrutar* e *visitar/visitar*, outras, como *conhecer*, em espanhol, *conocer*, não permitem correspondência direta.

⁹ Para mais informações sobre o a estruturação deste cenário, ver Gamonal (2013).

Quadro 1. Sumariamento dos usos de Unidades Lexicais nas línguas portuguesa e espanhola
 Fonte: Peron-Côrrea (2014)

UL em Português	Frame evocado	Frame secundário	UL em Espanhol	Frame evocado	Frame secundário
conhecer	Turismo_por_turista	Conhecimento	conocer	SEM EQUIVALÊNCIA	
apreciar	Turismo_por_turista	Exp._em_foco	apreciar	Turismo_por_turista	Exp._em_foco
desfrutar	Turismo_por_turista	Exp._em_foco	disfrutar	Turismo_por_turista e/ou Atração_em_lugar	Exp._em_foco
visitar	Turismo_por_turista	Visitante	visitar	Turismo_por_turista	Visitante

Entretanto, o fato de não haver equivalência entre palavras cognatas para o domínio do turismo entre duas línguas não invalida a proposta de que se use a Semântica de *Frames*, através da FrameNet, como princípio organizador de dicionários multilíngues. Pelo contrário, isso reitera a importância da Semântica de *Frames* como meio para tratar adequadamente das implicaturas culturais envolvidas no processo tradutório. Uma vez que os equivalentes de tradução propostos pelo Dicionário FrameNet Brasil da Copa do Mundo se baseiam na comparação das valências sintático-semânticas dos itens lexicais, as quais, por sua vez, são oriundas de *corpora*, tal propositura pode se relevar muito mais ancorada na realidade do uso linguístico do que aquelas feitas *ad-hoc* ou com base exclusivamente na etimologia.

5. Considerações finais

A Semântica de *Frames* se coloca como sendo a hipótese que fundamenta as principais abordagens circunscritas na Linguística Cognitiva. Dessa forma, nada mais plausível do que explorar métodos de aplicação dessa teoria. O desenvolvimento do Dicionário FrameNet Brasil da Copa do Mundo correspondeu ao grande objetivo do projeto: construir um recurso lexical online multilíngue a partir da Semântica de *Frames* e da metodologia da FrameNet. A atividade turística apresentada em diversos *frames* ilustra um domínio transcultural, o que não significa que haverá sempre equivalentes perfeitos de tradução. Por outro lado, o recurso possibilita encontrar palavras dentro de um mesmo campo semântico com base em evidências oriundas de *corpora*.

Tendo isso em vista, percebe-se que a teoria linguística e a metodologia explorada tanto para a Lexicografia Computacional quanto no desenvolvimento de tarefas no âmbito do Processamento de Linguagem Natural podem gerar diferentes contribuições. A continuação deste trabalho acontecerá através da abordagem de outros domínios do conhecimento, que nos permitam explorar mais fenômenos linguísticos.

Referências Bibliográficas

BAKER, C. FrameNet, Present and Future. In: WEBSTER, J.; IDE, N. & CHENGUY FANG, A. (Eds.). **The First International Conference on Global Interoperability for Language Resources**. Hong Kong: City University, 2008.

BICK, E. **The parsing system PALAVRAS**: automatic gramatical analysis of Portuguese in a constraint grammar framework. 2000. 505p. Tese de Doutorado em Filosofia, Aarhus University, Aarhus, 2000.

FELLBAUM, C. (Ed.). **WordNet: An Electronic Lexical Database**. Cambridge, MA: MIT Press, 1998.

FILLMORE, C. J. The case for case. In: BACH, E. & HARMS, R. T. (Eds.). **Universals in linguistic theory**. New York: Rinehard and Winston, 1968a, p. 1-88.

_____. Lexical Entries for Verbs. **Foundations of Language**, v.4, n.4, 1968b, p. 373-393.

_____. The case for case reopened. **Syntax and semantics**, v. 8, 1977, p. 59-82.

_____. Frame semantics. In: THE LINGUISTICS SOCIETY OF KOREA (Ed.). **Linguistics in the Morning Calm**. Seul: Hanshin Publishing Co., 1982, p.111-137.

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

_____.; CALLEJAS, C. M. B. Entrevista a Charles J. Fillmore. **Odisea**, n. 4, 2003, p. 41-48.

_____. Border Conflicts: FrameNet Meets Construction Grammar. In: **Proceedings of EURALEX 13**. Barcelona, 2008, p. 49-68.

GAMONAL, M. A. **Copa 2014 Framenet Brasil: diretrizes para a constituição de um dicionário eletrônico trilingue a partir da análise de frames da experiência turística**. 2013. 146p. Dissertação de Mestrado em Linguística, Universidade Federal de Juiz de Fora, Juiz de Fora, 2013.

_____.; TORRENT, T. T. Frames como Interlíngua na Estruturação de Dicionários Eletrônicos Multilíngues de Domínios Especializados. **Revista da ANPOLL**, n. 37, 2014, p. 247-261.

GOFFMAN, E. **Frame Analysis: An Essay on the Organization of Experience**. New York: Harper & Row, 1974.

MINSKY, M. A framework for representing knowledge. In: WINSTON, P. (Ed.). **The Psychology of Computer Vision**. New York: McGraw-Hill, 1975, p. 211-277.

PERON-CORRÊA, S. R. **Copa 2014 FrameNet Brasil: frames secundários em unidades lexicais evocadoras da experiência turística em português e em espanhol**. 2014. 147p. Dissertação de Mestrado em Linguística, Universidade Federal de Juiz de Fora, Juiz de Fora, 2014.

RUPPENHOFER, J.; ELLSWORTH, M.; PETRUCK, M. R. L.; JOHNSON, C & SCHEFFCZYK, J. **FrameNet II: Extended theory and practice**. Berkeley: International Computer Science Institute, 2010.

SARDINHA, T. B. **Linguística de Corpus**. São Paulo: Manole, 2004.

SALOMÃO, M. M. M. FrameNet Brasil: um trabalho em progresso. **Calidoscópio**, São Leopoldo: UNISINOS, v. 7 n. 3, 2009. p. 171-182.

SCHMID, H. Probabilistic Part-of-Speech Tagging Using Decision Trees. In: **Proceedings of International Conference on New Methods in Language Processing**. Manchester, UK, 1994.

SCHMIDT, T. Interfacing Lexical and Ontological Information in a Multilingual Soccer FrameNet. In: **Proceedings of OntoLex 2006 – Interfacing Ontologies and Lexical Resources for Semantic Web Technologies**. Itália, 2006.

_____. The Kicktionary: A Multilingual Resource of the Language of Football. In: REHM, G., WITT, A. & LEMNITZER, L. (Eds.). **Data Structures for Linguistic Resources and Applications**. Tübingen: Gunter Narr, 2007.

_____. The Kicktionary: Combining *Corpus* Linguistics and Lexical Semantics for a Multilingual Football Dictionary. In: LAVRIC, E. et al. (Eds.). **The Linguistics of Football**. Tuebingen: Gunter Narr, 2008, p. 11–23.

_____. The Kicktionary – a multilingual lexical resource of football language. In: BOAS, H. (Ed.). **Multilingual FrameNets – Methods and Applications**. Berlin/New York: Mouton de Gruyter, 2009, p. 101-132.

TORRENT, T. T.; ELLSWORTH, M. Behind the labels: criteria for defining analytical categories in FrameNet Brasil. **Veredas**, v.17, n.1, 2013, p. 44–65.

_____.; SALOMÃO, M. M. M.; CAMPOS, F. C.; BRAGA, R. M.; MATOS, E. E.; GAMONAL, M. A.; GONÇALVES, J.; GOMES, D. S.; SOUZA, B. C. P. & PERON-CORREA, S. R. Copa 2014 FrameNet Brasil: a frame-based trilingual electronic dictionary for the Football World Cup. In: **Proceedings of the 25th International Conference on**

Computational Linguistics (COLING 2014) – System Demonstrations. Dublin, 2014, p. 10-14.

Artigo recebido em: 30.03.2015

Artigo aprovado em: 23.06.2015

Domínios de Lingu@gem