

## **COPA 2014 FRAMENET BRASIL: ANÁLISE DA UNIDADE LEXICAL “VISITAR” DO *FRAME* DE TURISMO**

Maucha Andrade GAMONAL

Universidade Federal de Juiz de Fora  
mauchaandrade@gmail.com

**Resumo:** Este artigo apresenta os resultados iniciais do trabalho que está sendo desenvolvido através do projeto interinstitucional (UFJF-UNISINOS) Copa 2014 FrameNet Brasil. O objetivo do projeto é a criação de um dicionário eletrônico trilíngue (português, inglês e espanhol), voltado para os domínios do Turismo e do Futebol, que poderá ser utilizado por aqueles que estarão envolvidos com a organização da Copa do Mundo FIFA 2014, a ser sediada pelo Brasil, e também por turistas que necessitarão ou estarão interessados nesse vocabulário específico. O presente projeto estabelece interface com o Kicktionary (SCHMIDT, 2009), recurso lexical multilíngue (inglês - alemão - francês) da linguagem do Futebol, baseado em Semântica de *Frames*. A contraparte desse projeto para o português brasileiro, Kicktionary\_Br, está sendo desenvolvida pela equipe da professora Rove Chishman. A base teórica que norteia os estudos realizados na nossa pesquisa é também a Semântica de *Frames* (FILLMORE, 1982, 1985). Será a partir desse aporte teórico que a proposta do dicionário temático se distinguirá dos tradicionais, uma vez que este será baseado em cenas conceptuais, que poderão ser visualizadas a partir dos *frames*, e não apenas dos sinônimos que o lexema possui.

**Palavras-chave:** Semântica de *Frames*; FrameNet; Copa 2014 FrameNet Brasil; *frame* de Turismo.

### **Introdução:**

O presente artigo tem por objetivo apresentar os resultados iniciais desenvolvidos em torno do projeto Copa 2014 FrameNet Brasil. Projeto de cunho interinstitucional, que está sendo desenvolvido pela Universidade Federal de Juiz de Fora (UFJF) e pela Universidade do Vale do Rio dos Sinos (UNISINOS), visa à criação de um dicionário eletrônico trilíngue (português, inglês e espanhol), voltado para os domínios do Turismo e do Futebol. O intuito do projeto é criar um dicionário que parta de cenas conceptuais relevantes desses domínios específicos. Assim sendo, permitirá ao usuário desse recurso eletrônico tomar conhecimento não apenas do significado de determinado lexema, mas compreender um conceito relacionando-o com todas as partes nele envolvidas, ou seja, o dicionário terá como unidades básicas não os lexemas, como nos dicionários tradicionais, mas Unidades Lexicais, que são o pareamento de um lexema com um *frame*.

O recorte apresentado, no presente trabalho, foi a análise preliminar da Unidade Lexical (UL) “visitar”, que evoca o *frame* de Turismo (*Touring*). Tal análise foi realizada no âmbito do projeto FrameNet Brasil (SALOMÃO, 2009), que vem sendo desenvolvido na UFJF, em associação com o International Computer Science Institute (EUA) e em cooperação com a rede semântica (FrameNet), conduzida pelo Professor Charles Fillmore da Universidade da Califórnia em Berkeley. O Projeto FN-Br visa à criação de um recurso

lexical online para o Português do Brasil, baseado na Semântica de *Frames* (FILLMORE, 1982, 1985) e sustentado por evidência em corpus. O embasamento teórico da nossa pesquisa também está voltado para a semântica de *frames* (FILLMORE, 1982, 1985) em diálogo com a proposta da FrameNet (RUPPENHOFER et al., 2010) e com o trabalho realizado por Schmidt (2009) na criação de um recurso lexical multilíngue (inglês - alemão - francês) da linguagem do Futebol – o Kicktionary – conforme poderemos ver detalhadamente adiante.

Ademais, este trabalho pontua a necessidade de se criarem *frames* mais específicos para o domínio do Turismo, de modo a permitir que as descrições lexicográficas realizadas possam ser usadas no desenvolvimento de produtos de inovação tecnológica como a elaboração do dicionário. Nesse sentido, insere-se o Copa 2014, definido, de acordo com SALOMÃO et al. (2011, p.02) como um

dicionário, com foco no uso humano, [que] terá como audiência privilegiada a imprensa esportiva internacional, pessoas envolvidas na organização da Copa do Mundo FIFA 2014 e na recepção aos turistas estrangeiros, além dos próprios turistas. Parte-se da experiência do Kicktionary (SCHMIDT, 2006; 2007; 2009), cuja versão brasileira já se encontra em desenvolvimento pela equipe da UNISINOS (projeto FrameCorp, liderado pela profa. Chishman), para que se possa desenvolver um dicionário que permita ao usuário realizar uma consulta não só pela palavra desejada, mas, também, por cenas relevantes dos domínios do Futebol e do Turismo, por situações nas quais o usuário porventura esteja envolvido ou sobre as quais precise falar ou escrever.

Este trabalho está organizado da seguinte maneira: na seção 1, trataremos do aporte teórico da semântica de *frames*, do projeto da FrameNet e do Kicktionary; na seção 2, abordaremos o *corpus* que é utilizado pelo projeto FN-Br bem como a metodologia adotada; e a seção 3 se presta à análise da Unidade Lexical “Visitar”, com seus padrões de valência e tabela de sumariamento.

## 1. A Semântica de *Frames* e a FrameNet

Nos termos propostos por Fillmore (1982), a Semântica de *Frames* é um programa de pesquisa em semântica empírica que toma emprestada da Inteligência Artificial a ideia de que o conhecimento não deve ser visto como uma coleção de fragmentos simples e desconexos, mas como estruturas complexas, denominadas *frames*. Esse campo de estudos procura ilustrar o significado em termos da sua relação com a compreensão geral e não mais nos termos estabelecidos pela semântica das condições de verdade, partindo do pressuposto de que o conhecimento humano é estruturado a partir de um todo compartilhado e não por conceitos isolados como na semântica tradicional.

Dessa forma, para Fillmore, um *frame* é “qualquer sistema de conceitos relacionados de tal forma que, para entender um deles, é necessário entender toda a estrutura na qual ele se encaixa” (FILLMORE, 1982, p. 111). Um exemplo clássico do autor é o *frame* de comércio ou transação comercial: nele haverá uma cena em que está relacionada e incluída a pessoa interessada em trocar dinheiro por alguma mercadoria (o comprador), a pessoa interessada em trocar mercadorias por dinheiro (o vendedor), aquilo que o comprador poderá adquirir (as mercadorias) e o dinheiro adquirido pelo vendedor (o dinheiro). Vejam-se os exemplos discutidos por Fillmore (1987):

- (1) a. Eu comprei uma dúzia de rosas.  
b. Eu paguei a Harry cinco dólares.  
c. Eu paguei uma dúzia de rosas para Harry por cinco dólares.  
d. Eu paguei a Harry cinco dólares por uma dúzia de rosas.

Em (1a) o verbo *comprar* centra a atenção no comprador e na mercadoria, sendo que esses elementos aparecem mencionados; em (1b), há a presença do comprador e do dinheiro, aqui o verbo *pagar* foca a atenção no comprador e no dinheiro, que aparecem instanciados na sentença. Há ainda o elemento mercadorias (uma dúzia de rosas e por uma dúzia de rosas) presente em (1c) e (1d).

Como pontua Gawron (2008), a Semântica de *Frames* trata de questões que a semântica lexical, chamada por ele de visão clássica, não dá conta de responder, como o fato de a experiência humana ser vasta, relacionando-se com a abertura do significado das palavras, e também com a questão da riqueza de seu significado. Isso porque as palavras são difíceis de ser definidas visto que evocam uma grande quantidade de informações de fundo, que só podem ser compreendidas se essa rede de conceitos que as relaciona for evocada.

Nos dicionários tradicionais, por exemplo, a definição da palavra andarilho é “aquele que anda muito”. Entretanto, sabemos que não podemos caracterizar como andarilho uma pessoa que percorre longas distâncias fazendo caminhadas todos os dias, embora a definição dada no dicionário nos permita essa interpretação. O que percebemos aqui é que o *frame* de andarilho não inclui um atleta ou um indivíduo praticando atividade física.

Por isso a importância da Semântica de *Frames*, uma vez que ela permite abordar os significados de modo mais aprofundado, possibilitando que sejam feitas as devidas relações que existem entre os conceitos na produção de dado sentido.

### 1.1. Cenas e *frames*

Há distinções que Fillmore discute acerca dos conceitos de *frames* e cenas que, a posteriori, deverão ser abordadas na pesquisa, por hora, cabe-nos destacar em que as definições se diferenciam. Vejamos o que o autor elucida a esse respeito:

Eu quero dizer que as pessoas, na aprendizagem de uma língua, vêm a associar determinadas cenas com determinados *frames* linguísticos. Pretendo usar a palavra cena – uma palavra com a qual eu não estou completamente satisfeito – em um sentido geral ao máximo, para incluir não apenas cenas visuais, mas os tipos familiares de relações interpessoais, cenários padrão, layouts familiares, estruturas institucionais, experiências ativas, imagem corporal e, em geral, qualquer tipo de segmento coerente, grande ou pequeno, crenças humanas, ações, experiências, ou imaginação. Eu pretendo usar a palavra *frame* para se referir a qualquer sistema de escolhas linguísticas - os casos mais simples sendo coleções de palavras, mas incluindo também as escolhas de categorias gramaticais- que podem ser associados com instâncias prototípicas das cenas (FILLMORE, 1987, p. 82)<sup>1</sup>.

Pelas palavras do autor, a cena seria definida por uma estrutura de conhecimento e experiência do ser humano com o mundo enquanto o *frame* seria caracterizado pela manifestação linguística dessas cenas. Seguindo os estudos de Fillmore, Schmidt (2009) destaca que um *frame* é uma entidade estrutural utilizada para expressão de um grupo linguístico que compartilha uma perspectiva comum sobre uma determinada cena conceptual. Uma cena, nesse sentido, é uma construção superordenada de um *frame*, já a noção de *frame*

---

<sup>1</sup> “I want to say that people, in learning a language, come to associate certain scenes with certain linguistic frames. I intend to use the word scene – a word I am not completely happy with – in a maximally general sense, to include not only visual scenes but familiar kinds of interpersonal transactions, standard scenarios, familiar layouts, institutional structures, enactive experiences, body image; and, in general, any kind of coherent segment, large or small, of human beliefs, actions, experiences, or imaginings. I intend to use the word frame for referring to any system of linguistic choices – the easiest cases being collections of words, but also including choices of grammatical rules or grammatical categories – that can get associated with prototypical instances of scenes.”

está preocupada com as propriedades dos significados linguísticos concretos de expressar este tipo de conhecimento.

## 1.2. A FrameNet

A FrameNet é um projeto desenvolvido sob a liderança do Professor Charles Fillmore no International Computer Science Institute (ICSI), em Berkeley, na Califórnia. De acordo com Ruppenhoffer et al (2010), o intuito é criar um recurso lexical online para o inglês, baseado na Semântica de *Frames* e sustentado por evidência em *corpora*. Dessa forma, trabalha-se na construção de uma grande rede de *frames* com suas combinações semântico-sintáticas.

O banco de dados da FrameNet, disponível online (<http://framenet.icsi.berkeley.edu>), é legível, ao mesmo tempo, por humanos e por máquina. Contém mais de 11.600 unidades lexicais, mais de 6.800 das quais estão totalmente anotadas, em mais de 960 *frames* semânticos, exemplificados em mais de 170 mil sentenças, que fornecem um conjunto de dados para rotulação de papel semântico, utilizado em aplicações tais como extração de informação, tradução automática e dicionário de valências, por exemplo. Nos termos de Salomão (2009, p. 174),

na conclusão deste trabalho, estaríamos nos aproximando do sonho do “dicionário ideal”, no qual cada um de nós, ao consultar uma palavra, seríamos remetidos imediatamente para o *frame* que ela evoca, com todos os respectivos Elementos componentes; veríamos, além disso, uma listagem de todas as valências desta palavra, suas possibilidades combinatórias sintáticas e semânticas, ilustradas por exemplos correspondentes. A consulta ainda nos ofereceria um conjunto de outras palavras que evocam o mesmo *frame* e o conectaria com outros *frames* semanticamente relacionados.

O desenvolvimento inicial da FrameNet se deu na língua inglesa, mas já houve a expansão para outras línguas, como alemão, japonês, francês e espanhol. Recentemente houve um empenho na implantação também para o português: o Projeto FrameNet Brasil (SALOMÃO, 2009), que pode ser pormenorizado através do site do projeto (<http://www.framenetbr.ufjf.br/>).

Na FrameNet, o objetivo central é a descrição das Unidades Lexicais, pareamentos de um lexema a um significado, a partir dos *frames* que evocam. Partimos delas e extraímos informações sobre as suas propriedades sintático-semânticas (valência). As propriedades sintáticas disponibilizam informações sobre os tipos sintagmáticos (sintagma nominal, preposicional etc) e as funções gramaticais (argumento externo, objeto e dependente), já a valência semântica designa informações condizentes aos elementos de *frame*, que são papéis semânticos. Os resultados são sumarizados em padrões de valência.

A respeito dos elementos de *frame* (EFs), podemos dizer que eles são os participantes da cena. Um *frame* não poderia ser constituído sem a presença dos seus elementos, pois esses disponibilizam as informações necessárias a respeito de como aquele é instanciado. Os EFs podem ser nucleares, periféricos e extratemáticos.

Os nucleares são aqueles essenciais para que a cena seja constituída. No *frame* de Ataque (*Attack*), temos como elementos centrais o Assaltante e a Vítima, dois dados essenciais para que a cena se construa. Os periféricos ou não-nucleares são aqueles cujas informações trazidas são acrescidas ao que é atribuído ao EF central, como Tempo, Lugar, Duração, Maneira e Grau. No *frame* exemplificado, um EF periférico seria a Arma, um instrumento utilizado para assaltar a Vítima. Por último, os elementos extratemáticos, embora participem da cena, não pertencem a esse *frame*. No exemplo “Ele me assaltou duas vezes

nessa rua.” O número que determina quantas vezes a ação ocorreu é um EF extratemático chamado de Iteração.

Há casos em que os elementos de *frame* centrais podem não vir enunciados na sentença. Quando isso é verificado, a FrameNet propõe uma Instanciação Nula. Essa pode ser dividida em Instanciação Nula Definida (IND), Instanciação Nula Indefinida (INI) e Instanciação Nula Construcional (INC). A IND acontece quando o elemento pode ser recuperado no contexto. Por exemplo, se, em “Ele roubou sem deixar vestígio”, o elemento de *frame* Vítima, que não é explicitado na sentença, puder ser retomado pelo contexto, esse EF será marcado como IND. Se tivéssemos “Assaltaram e não foram encontrados”, e não tivéssemos um contexto para retomar quem foi assaltado, anotaríamos o Assaltante como uma INI, ou seja, essa instanciação se refere a casos indefinidos. Já em casos como na frase “A base secreta do governo foi invadida” temos uma construção passiva, em que o Assaltante não é evidenciado. Esse é um exemplo de INC bem como as construções em que o sujeito encontra-se omitido em sentenças imperativas.

Sobre a anotação na FrameNet, é importante ressaltar que essa pode ser realizada de duas maneiras: a anotação lexicográfica e a de texto corrido. Na anotação lexicográfica, o objetivo é “registrar todas as possibilidades semânticas e sintáticas (valências) de cada lexema em cada um de seus sentidos”<sup>2</sup> (RUPPENHOFER et al., 2006, p. 20). Esse é o tipo de anotação mais praticado pela FrameNet e é o adotado pela FN-Br até então. Na anotação de texto corrido, a sequência do texto deve ser seguida para a anotação, assim sendo, os *frames* vão surgindo ao decorrer da anotação, diferentemente do que acontece na análise lexicográfica, já que, com ela, partimos de um *frame*.

### 1.3 O *frame* de Turismo

A FrameNet define o *frame* de Turismo (*Touring*) como:

#### Turismo (*Touring*)

##### Definição:

Um **turista** visita ou experiencia uma **atração**, um local com uma história peculiar ou um caráter individual reconhecido socialmente, com o objetivo de ver e aprender sobre ele. Normalmente, a **atração** tem uma fonte de informação tal como um guia, panfletos ou displays.

##### Elementos do Frame:

###### Nucleares:

###### **Atração**

Tipo Semântico: Lugar

O local peculiar socialmente reconhecido que o **turista** visita.

###### **Turista**

Tipo Semântico: Consciente

O indivíduo que está buscando uma experiência em uma **atração**.

<sup>2</sup> “recording the range of semantic and syntactic combinatory possibilities (valences) of each word in each of its senses”

**Não-Nucleares:****Coparticipante**

Tipo Semântico: Consciente

O **coparticipante** é quem participa de modo coordenado, no caso em uma visita, com o **turista**.

**Estado**

O estado do **turista** ou da **atração** durante a visita.

**Duração**

Tipo Semântico: Duração

A quantidade de tempo que dura a atividade de turismo.

**Base**

O background perceptivo pela qual a **atração** é experimentada pelo **turista**.

**Modo**

Tipo semântico: Modo

Qualquer descrição dos detalhes do evento em termos de como eles se comparam com outros eventos (da mesma forma) ou em termos de como o estado do **turista** os afeta (felizmente, distraidamente).

**Meio**

Tipo semântico: Estado\_de\_Coisas

O **Meio** indica o método que o **turista** está usando para participar da visita.

**Iteração\_específica**

Expressões marcadas com este EF extratemático modificam um uso não-iterativo do alvo e indica que ele é concebido como incorporado dentro de uma série iterada de eventos similares ou estados.

**Lugar**

Tipo semântico: Relação\_locativa

O local onde o turismo acontece.

**Propósito**

Tipo semântico: Estado\_de\_Coisas

Alguma ação que o **turista** está tentando realizar com o turismo. Isso geralmente é (no mínimo) para fins de entretenimento e aprendizado.

**Tempo [Tim.]**

Tipo semântico: Tempo

O tempo que o turismo acontece.

**Relações entre frames:**Herda de: [Perception\\_active](#)

É herdado de:  
 Perspectivado em:  
 É perspectivado por:  
 Usa: [Visiting](#)  
 É usado por:  
 Subframe:  
 Tem os subframes:  
 Precede:  
 É precedido por:  
 É incoativo de:  
 É causativo de:  
 Veja também:

Dado o *frame* de Turismo, vimos que sua descrição diz que há um Turista que visita uma Atração, esta é um local com uma história, que apresenta um caráter peculiar, por vezes já reconhecido socialmente. O objetivo do Turismo, como pontuado na descrição, pode ser o simples deleite ou o aprendizado. Normalmente, a Atração apresenta fontes de informação como um guia ou panfletos, por exemplo.

Os elementos centrais do *frame* de turismo são a Atração e o Turista. A Atração é o local visitado pelo turista, e o Turista é aquele que busca uma atração oferecida por um passeio a um lugar especialmente voltado para esse fim. Os elementos de *frame* periféricos tratam, como vimos, das informações adicionais como o tempo em que o turismo acontece, o local onde o turismo ocorre; possíveis coparticipantes, a duração da atividade turística e, como EF extratemático, encontramos a Iteração.

Nas relações entre *frames*, é mostrado que o *frame* de Turismo herda relação do *frame* de Perception\_active e usa o *frame* de Visitar (Visiting). A relação de subframe não é estabelecida pela FrameNet para o *frame* de Turismo, conforme podemos ver na tabela dada. De acordo com os dados que temos levantado até agora, é possível que tenhamos que usar subframes, posto que, por exemplo, visitar um museu requer características específicas e distintas de realizar uma visita a uma cidade ou visitar uma igreja ou ainda uma biblioteca. Vejamos alguns exemplos para compreensão:

- 1) [eu<sub>TURISTA</sub>] tinha ido **visitar** [um museu<sub>ATRAÇÃO</sub>] [lá<sub>lugar</sub>].
- 2) [Você<sub>TURISTA</sub>] **visitou** [alguma igreja<sub>ATRAÇÃO</sub>] [em Brasília<sub>lugar</sub>]
- 3) [Mais da metade dos americanos<sub>TURISTA</sub>] **visitaram** [uma biblioteca<sub>ATRAÇÃO</sub>] [em 2007<sub>TEMPO</sub>]
- 4) [Ele<sub>TURISTA</sub>] veio **visitar** [Los Angeles<sub>ATRAÇÃO</sub>][com Erville<sub>COPARTICIPANTE</sub>]

As cenas construídas são muito distintas para que um mesmo *frame* seja capaz de abordar todas essas peculiaridades. Dessa forma, a partir dessas cenas, seria necessário criarmos subframes. Mas não é nossa pretensão aprofundar essa discussão no presente trabalho, já que nossos dados ainda são amostras preliminares de uma possível questão sobre a qual teremos que nos debruçar mais adiante.

Dentre as ULs que a FrameNet destaca não está incluída a que está sendo analisada no presente trabalho, a UL “visitar”. Entretanto, diferenças são percebidas na composição da cena “Fomos visitar o museu Murilo Mendes” e “João irá visitar a vó dele no feriado”, uma vez que é possível percebermos a diferença da finalidade da visita do primeiro exemplo que se configura em uma visita de cunho turístico para a visita a um parente como a do segundo

exemplo. Essa diferença se constitui então em uma distinção entre os *frames* de Turismo e o de Visitar na base de dados da FrameNet.<sup>3</sup>

#### 1.4. O Kicktionary

O Kicktionary é um recurso lexical multilíngue da linguagem do futebol que abrange as línguas inglesa, alemã e francesa. É desenvolvido pela equipe coordenada pelo Professor Thomas Schmidt da Universidade de Hamburgo.

O objetivo geral no desenvolvimento do Kicktionary era explorar como as teorias linguísticas sobre semântica lexical (especialmente as abordagens FrameNet e WordNet para lexicografia), métodos de corpus linguístico e da tecnologia hipermídia pode ajudar a construir recursos lexicais que são melhores (ou: bons de uma forma diferente dos) dicionários de papel tradicionais (SCHMIDT, 2009, p.01).<sup>4</sup>

Esse recurso se configura em uma das primeiras tentativas de construir um dicionário de domínio específico que utilize abordagem baseada em frames semânticos. As unidades mínimas de análise no Kicktionary são a cena, o *frame* e a Unidade Lexical, o que o torna diferente da FrameNet, que tem como unidades mínimas de descrição a UL e o *frame*.

Atualmente, esse recurso contém cerca de 2000 Unidades Lexicais e cerca de 8000 frases de exemplo. O material é acessível pelo site <http://www.kicktionary.de/>.

**bicycle-kick.n** Scenario Shot Frame Shot

---

**SHOOTER** [Player]

---

1. Not content with that, [Crespo]<sub>SHOOTER</sub> then **attempted** a **bicycle kick** only for Laštuvka to produce a reflex save to deny him a second goal. [1077219 / p9]
2. Cazorla shot narrowly wide from distance on the half-hour mark and Luciano saw [his]<sub>SHOOTER</sub> **bicycle-kick** saved by Vasil Khomutovski five minutes later before José Mari shot wide. [80107 / p3]
3. The Danish forward headed Pirlo's long pass into the path of Shevchenko who latched on to the ball but saw his shot cleared by [Celtic defender Dianbobo Balde's]<sub>SHOOTER</sub> spectacular **bicycle-kick**. [1077172 / p6]

---

Support	LU	SHOOTER
<i>attempted</i>	<b>bicycle kick</b>	Crespo
	<b>bicycle-kick</b>	his
	<b>bicycle-kick</b>	Celtic defender Dianbobo ...

---

**Synonyms**      Fallrückzieher.n  
                          overhead\_kick.n    bicycle-kick.n  
                          retourné.n

**Hypernyms**      Torschuss.n    Schuss.n  
 [Moving\_Balls]    shot.n    drive.n    strike.n  
                          tir.n    frappe.n

Figura 1: Apresentação no Kicktionary da Unidade Lexical “chute de bicicleta”.

<sup>3</sup> Para maior compreensão, comparar o *frame* de Touring com o *frame* de Visiting na base de dados da FrameNet.

<sup>4</sup> “The general aim in the development of the Kicktionary was to explore how linguistic theories about lexical semantics (especially the FrameNet and WordNet approaches to lexicography), corpus linguistic methods and hypermedia technology can help to build lexical resources that are better (or: good in a manner different from) traditional paper dictionaries.”

Como esse dicionário é elaborado para ser usado principalmente por seres humanos, tem-se a preocupação de como as Unidades Lexicais serão apresentadas bem como a sua organização estrutural.

Na apresentação da Unidade Lexical “chute de bicicleta” (Figura 1), a entrada indica a cena e o *frame* da unidade lexical “shooter” e lista as frases de exemplo anotadas em duas formas diferentes – em texto integral e em uma visão esquemática. Sinônimos e *synsets*<sup>5</sup> superordenados também são fornecidos. Além disso, cada componente da apresentação é um *hiperlink* para correspondentes em outras partes do dicionário. Por exemplo, clicando no nome da cena, o usuário será levado a uma descrição dessa cena. Os exemplos dados são retirados dos textos do corpus utilizado.

## 2. Corpora e métodos

O material usado para a análise foi retirado dos seguintes *corpora* do português do Brasil: ANCIB, ECI-EBR, e NILC/São Carlos, pertencentes ao site Linguateca; NURC-RJ, Legenda de Filmes, e Domínio Público, na ferramenta de busca SketchEngine.

Os *corpora* do site Linguateca são de acesso público e têm por objetivo facilitar o acesso aos dados do português já disponibilizados, “através do desenvolvimento de serviços de acesso na rede, e mantendo um portal com informação útil” (<http://www.linguateca.pt/>). O *corpus* ANCIB – Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação – foi criado a partir de mensagens enviadas para a lista homônima da ANCIB, até o mês de novembro de 2003, e para a lista abarreto-1 após esta data. Já o ECI-EBR – European Corpus Initiative –, baseado no *corpus* Borba-Ramsey, é uma seleção de fragmentos de obras brasileiras, contendo discurso literário, didático e oral cuidado (discursos políticos). Há também o NILC/São Carlos, *corpus* que contém textos brasileiros do registro jornalístico, didático, epistolar e redações de alunos.

O SketchEngine é um sistema de consulta a *corpus* que incorpora esboços de texto, uma página, ou resumo (<http://www.sketchengine.co.uk/>). O Legenda de Filmes é um dos *corpora* criado pelo Projeto FN Brasil que está alojado nessa ferramenta. Ele contém legendas de filmes em Português do Brasil cedidas pelo portal OpenSubtitles.org. Há também o NURC-RJ, *corpus* constituído por entrevistas gravadas nas décadas de 1970 e 1990, num total de 350 horas, com informantes de nível superior completo, nascidos no Rio de Janeiro e filhos de pais, preferencialmente, cariocas. Por último, o *corpus* Domínio Público é composto de obras literárias do português brasileiro do século XIX em diante e obras traduzidas para o PB presentes no site do Domínio Público (<http://www.dominiopublico.gov.br/>), totalizando mais de 500 obras literárias.

Como se pode ver, o *corpus* do projeto é diversificado, contendo obras literárias, textos de cunho didático, científico, da área tecnológica, jurídica e legendas de filmes. A predominância é de legenda de filmes e textos jornalísticos. Os *corpora* totalizam aproximadamente 136 milhões de *tokens*.

Para que a busca nos *corpora* seja feita, partimos do lexema a ser pesquisado e digitamos a sintaxe de pesquisa, no nosso caso, “(?)visit.\*”, na ferramenta Scketch Engine. Já no site Linguateca, utilizamos a fórmula [lema=“visitar”]. Após a busca das sentenças nos *corpora*, parte-se para a classificação delas em uma planilha eletrônica. Esse processo divide as sentenças retornadas em seis categorias (cf, Tabela 1), sendo que só serão anotadas aquelas sentenças nas quais o radical “visit” figurar como verbo e evocar o *frame* que se deseja anotar, no caso o de Turismo.

---

<sup>5</sup> Entende-se um *synset* como um grupo ou par de palavras sinônimas.

Código	Classificação
1	Sentido alvo
2	Sentido figurativo
3	Adjetivo
4	Substantivo
5	Contexto ambíguo ou insuficiente
6	Outros

Tabela 1: Classificação das sentenças retornadas

Faz-se necessário justificar que utilizamos a marcação “outros” quando o lexema evoca outro *frame*, quando há erros de digitação no texto de origem ou ainda quando sentenças repetidas são retornadas. Concluída essa etapa, passamos essa classificação pelo programa Tinn-R, que classifica estatisticamente os usos dos lexemas. A funcionalidade do programa se faz eficiente principalmente nos casos em que os *corpora* são extensos, facilitando nosso trabalho.

A partir daí passamos para a análise lexicográfica das Unidades Lexicais em termos de suas valências. Nessa análise, são postuladas as camadas de anotação como a camada da palavra alvo, a camada dos elementos de *frame*, as camadas da função gramatical e do tipo sintagmático. Na camada da palavra alvo, figura a sentença onde esta é encontrada. Ela é grifada por caixa alta, marcada de fundo preto e fonte branca. Veja-se o exemplo abaixo da UL “Visitar” do *frame* de Turismo (*Touring*).

Camada com palavra alvo	mais da metade dos americanos	VISITARAM	uma biblioteca	em 2007
-------------------------	-------------------------------	-----------	----------------	---------

A camada dos Elementos de *Frame* é onde os constituintes das sentenças são etiquetados. Nessa camada, tanto os elementos de *frame* nucleares quanto os não-nucleares são marcados. Os elementos são diferenciados por cores. Assim cada elemento de *frame* vai possuir uma cor se manterá ao longo da análise.

Camada com palavra alvo	mais da metade dos americanos	VISITARAM	uma biblioteca	em 2007
Elemento de Frame	Turista		atração	Tempo

Já na camada da função gramatical, teremos a função gramatical desempenhada pelos elementos do *frame* analisados. As funções gramaticais que podem ser estabelecidas na FrameNet são três: Argumento Externo (Ext), Objeto (Obj) e Dependente (Dep), sendo marcado como Dependente tudo aquilo que não pode ser incluído nas outras duas categorias<sup>6</sup>.

Camada com palavra alvo	mais da metade dos americanos	VISITARAM	uma biblioteca	em 2007
Elemento de Frame	Turista		atração	Tempo
Função Gramatical	Ext		Obj	Dep

<sup>6</sup> Para uma discussão a esse respeito, vide Ruppenhofer et al. (2010).

A Camada dos tipos sintagmáticos identifica os sintagmas como no exemplo dado abaixo.

Camada com palavra alvo	mais da metade dos americanos	VISITARAM	uma biblioteca	em 2007
Elemento de Frame	Turista		atração	Tempo
Função Gramatical	Ext		Obj	Dep
Tipo Sintagmático	SN		SN	SP

Há também a camada Outros, na qual se anotam elementos diretamente relacionados à UL sendo anotada, no nosso caso, um verbo. Estruturas com auxiliares, índices de indeterminação do sujeito, ou estratégias de relativização são marcadas nessa camada.

### 3. Análise da Unidade Lexical “Visitar”

Após realizarmos o processo de anotação, partimos para o sumariamento dos dados encontrados, que é efetuado pelo programa R. O programa nos retornou, conforme a Tabela 2, que, nos cinco *corpora* pesquisados, tivemos um total de 1390 sentenças, sendo que 99 foram consideradas válidas para nossa pesquisa, representando aproximadamente cerca de 7 % do total de ocorrências da busca. O corpus que mais retornou sentenças válidas foi o NURC – 44 sentenças.

<i>Corpora</i>	Sentido Alvo	Sentido Figurado	Adjetivo	Substantivo	Contexto Insuficiente	Outros	TOTAIS
Legenda de Filmes	18	-	1	162	-	218	399
Domínio Público	17	-	41	202	7	127	394
NURC	44	-	11	69	18	82	224
ANCIB	8	-	-	3	1	317	329
ECI-EBR	12	-	-	4	-	28	44
NILC-São Carlos <sup>7</sup>	-	-	-	-	-	-	-
Total	99	-	53	440	26	772	1390

Tabela 2: Sumariamento dos dados levantados

#### 3.1 Padrões de Valência

Finalizado o processo de anotação das sentenças, passamos para o processo de determinar quais são os padrões de valência que ocorrem com essa Unidade Lexical. Para que um padrão de valência seja postulado, consideramos apenas os Elementos de Frame que são nucleares e suas realizações sintáticas. Ou seja, se um dado elemento for instanciado como Sintagma Nominal e outro como Sintagma Preposicional na camada do tipo sintagmático, por exemplo, teremos de aloca-los em padrões diferentes. Ou, se o corpus retornou sentenças em

<sup>7</sup> O corpus Nilc de São Carlos retornou aproximadamente 3000 ocorrências. Entretanto, para este trabalho, não foi possível trabalhar com este corpus porque tivemos problemas técnicos com a ferramenta que realiza a amostragem.

que o Turista se instanciava como Externo e outras como Nulo Definido ou Indefinido teremos de criar padrões distintos para alocar esses casos. Demonstraremos abaixo os padrões que encontramos:

### 1º PADRÃO:

- **Elementos de Frame:** Turista - Atração
- **Realizações sintáticas:** Turista Ext/SN - Atração Obj/SN

Camadas	os turistas	VISITAM	nossos monumentos e outros edifícios
EF	Turista		Atração
FG	Externo		Obj
TS	SN		SN
Outros			

1. Aqui vos fala H.B. Cattenberg. Aqui na capital do país, há preocupação e inquietude, - mas não tem havido pânico. Inclusive, se está vivendo normalmente. - O clima é muito bom e todos os turistas VISITAM nossos monumentos e outros edifícios.

### 2º PADRÃO:

- Elementos de Frame: Turista - Atração
- Realizações sintáticas: Turista IND - Atração Obj/SN

Camadas		VISITAR	o Pão de Açúcar
EF	Turista =IND		Atração
FG			Obj
TS			SN
Outros		foi	

2. bem e/e o carioca é assim você vê... cê tem muita gente que já tá aqui há muitos anos e nunca foi VISITAR o Pão de Açúcar... [IND]

### 3º PADRÃO:

- Elementos de Frame: Turista - Atração
- Realizações sintáticas: Turista INI - Atração Obj/SN

Camadas		VISITAR	o prédio mais automático do mundo
EF	Turista =INI		Atração
FG			Obj
TS			SN
Outros			

3. - Nenhuma visita a Nova York é completa... - sem VISITAR o prédio mais automático do mundo. [INI]

## 4º PADRÃO:

- Elementos de Frame: Atração – Atração – Turista  
Realizações sintáticas: Atração Obj/SN - Turista Ext/SN

Camadas	os outros países	que	você	VISITOU
EF	Atração	Atração	Turista	
FG	Obj	Obj	Externo	
TS	SN	SN	SN	
Outros	Ant	Rel		

4. E os outros países europeus que você VISITOU ?

## 5º PADRÃO:

- Elementos de Frame: Turista - Atração  
Realizações sintáticas: Turista IND - Atração IND

Camadas		VISITAR	
EF	Turista=IND		Atração=IND
FG			
TS			
Verbo		fui	

5. No Canadá tinha uma, um Jardim Botânico em Montreal, fui VISITAR, e tinha o, o Jardim Japonês e tinha o Jardim Chinês. [IND] [IND]

## 6º PADRÃO:

- Elementos de Frame: Turista - Atração  
Realizações sintáticas: Turista INC - Atração IND

Camadas		VISITAR	
EF	Turista=INC		Atração=IND
FG			
TS			
Verbo			

6. Mas, eh, é cercada, é uma cidade que pode se VISITAR, visita-se e, e tem muita coisa interessante pra se ver, muita coisa ainda até bem conservada. [INC] [IND]

### 3.2. Tabelas de Sumariamento

#### 3.2.1. Elementos de Frame e suas Realizações Sintáticas

Conforme observamos acima o que define um padrão de valência são os Elementos de *Frame* nucleares e as realizações sintáticas. Na tabela dada abaixo, explicitamos como os Elementos de *Frame* tanto os nucleares como os não-nucleares se instanciaram sintaticamente e qual foi a quantidade para cada tipo de ocorrência. O resultado dos EFs nucleares deu-se da seguinte forma (conforme poderá ser verificado na Tabela 3 abaixo): ao todo foram anotados 99 elementos de *frame* nucleares. A respeito do EF Turista, ele se instanciou da seguinte forma: houve 35 sentenças com SN/ Externo, 55 sentenças como Instanciações Nulas Definidas, 7 se realizaram como Instanciações Nulas Indefinidas e 2 como Instanciações Nulas Construcionais. Sobre o EF Atração, 93 se instanciaram como Sintagma Nominal e função gramatical de Objeto, enquanto 6 apareceram como IND.

Também realizamos o sumariamento dos EFs não-nucleares. Ao todo, tivemos 24 aparições, o que mais ocorreu foi o EF Tempo.

Elemento de <i>Frame</i>	Número Anotado	Realizações
Turista	99	SN/Ext = (35) IND= (55) INI= (7) INC= (2)
Atração	99	SN/Obj = (93) IND= (6)
Coparticipante	2	SP/Dep= (2)
Descrição	-	
Duração	-	
Base	-	
Modo	1	SAdv/Dep= (1)
Meio	-	
Iteração específica	-	
Lugar	6	SP/Dep= (6)
Propósito	1	SP/Dep=(1)
Tempo	14	SS/Dep =(2) SAdv/Dep=(12)

Tabela 3: Realizações sintáticas dos Elementos de *Frame*

### 3.2.2. Padrões de Valência

Necessitamos alocar as sentenças em seis padrões de valência distintos. Isso porque as realizações sintáticas na camada do tipo sintagmático (TS) e da função gramatical (FG) se instanciaram diferentes bem como os EFs nucleares. Conforme trazemos na Tabela 4, o primeiro padrão totalizou 28 sentenças: o EF Turista foi instanciado como Sintagma Nominal na camada tipo sintagmático e Externo na camada função gramatical, e o EF Atração como Sintagma Nominal na camada TS e Objeto na FG. O segundo padrão foi o de maior representatividade, obteve 51 ocorrências, o EF Turista ocorreu como IND, e a Atração ocorreu como Sintagma Nominal na camada TS e Objeto na camada FG. O terceiro padrão aconteceu em 7 sentenças, o Turista foi instanciado como INI e a Atração como SN na camada TS e Objeto na camada FG. O quarto padrão recebeu 7 sentenças, e os EFs Turista e Atração apareceram como SN/Ext e SN/Obj, respectivamente. O quinto padrão foi instanciado como IND para ambos os elementos e obteve 4 sentenças, e o último padrão recebeu duas sentenças, que foram instanciadas como INC e IND, respectivamente.

Número Anotados	Padrões			
28 TOTAIS	Turista	Atração		
(15)	SN Ext	SN Obj		
	Turista	Atração	Coparticipante	
(1)	SN Ext	SN Obj	SP Dep	
	Turista	Atração	Tempo	
(6)	SN Ext	SN Obj	SAdv Dep	
	Turista	Atração	Tempo	
(2)	SN Ext	SN Obj	SS Dep	
(3)	Turista	Atração	Lugar	
	SN Ext	SN Obj	SP Dep	
	Turista	Atração	Tempo	Lugar
(1)	SN Ext	SN Obj	SAdv Dep	SP Dep
51 TOTAIS	Turista	Atração		
(42)	IND	SN Obj		
	Turista	Atração	Tempo	
(5)	IND	SN Obj	SAdv Dep	
(1)	Turista	Atração	Coparticipante	
	IND	SN Obj	SP Dep	
	Turista	Atração	Modo	
(1)	IND	SN Obj	SAdv Dep	
	Turista	Atração	Lugar	
(1)	IND	SN Obj	SP Dep	
	Turista	Atração	Propósito	
(1)	IND	SN Obj	SP Dep	
7 TOTAIS	Turista	Atração		
(6)	INI	SN Obj		
	Turista	Atração	Lugar	
(1)	INI	SN Obj	SP Dep	
7 TOTAIS	Atração	Turista		
	SN Obj	SN Ext		
4 TOTAIS	Turista	Atração		
	IND	IND		

2 TOTAIS	Turista	Atração		
	INC	IND		

Tabela 4: Sumariamento dos padrões de valência

#### 4. Considerações Finais

Nossa primeira iniciativa prática na pesquisa com o Copa 2014 FrameNet Brasil foi a busca da Unidade Lexical “visitar” evocando o *frame* de Turismo (Touring) nos *corpora* da FN-Br. Realizamos a busca conforme descrevemos, partindo para a anotação e seguindo com o sumariamento das valências conforme demonstrado. Essa busca retornou seis padrões, sendo que o que obteve mais ocorrências foi o padrão com 51 sentenças. E o de menor ocorrência foi com 2 sentenças. Entretanto esse dados não se encontram finalizados, uma vez que estamos em processo de busca por novos *corpora* que tenham foco maior na área do turismo.

No período de anotação dos dados, foi possível percebermos que será necessário aprofundarmos a discussão da criação de subframes no domínio do Turismo, já que há cenas que apresentam peculiaridades suficientes a ponto de criarmos subframes específicos a fim de caracterizá-las. Entretanto essa questão é apresentada nesse trabalho como ponto de partida para trabalhos posteriores.

Esse trabalho servirá para contribuir com o avanço no banco de dados de anotação da plataforma do projeto que vem apresentando grandes avanços.

#### Referência:

CIENKI, A. Frames, Idealized Cognitive Models, and Domains. In: *The Oxford Handbook of Cognitive Linguistics*. Oxford University Press, 2007

FILLMORE, Charles J. Frame semantics. In: *Linguistics in the Morning Calm*, Seul: Hanshin Publishing Co., p.111-137, 1982.

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

FILLMORE, Charles. *Fillmore's cases grammar: a reader*. In: DIRVEN, R. & RADDEN, G. (Ed). Heidelberg, 1987.

GAWRON, J.M., Frame Semantics. 2008. Disponível em: [http://www.hf.uib.no//forskingskole/new\\_frames\\_intro.pdf](http://www.hf.uib.no//forskingskole/new_frames_intro.pdf). Acesso em 14 de outubro de 2011.

RUPPENHOFER, J.et al. *FrameNet II: Extended theory and practice*. Disponível em: [http://framenet.icsi.berkeley.edu/index.php?option=com\\_wrapper&Itemid=126](http://framenet.icsi.berkeley.edu/index.php?option=com_wrapper&Itemid=126) Acesso em 3 de agosto de 2011.

SALOMÃO, M.M.M. FrameNet Brasil: um trabalho em progresso. In: *Calidoscópico*, Vol. 7.2, 2009a

SALOMÃO, M.M.M; TORRENT, T.T.; CAMPOS, F.C.A.; BRAGA, R.M.M. & VIEIRA, M.B. *Copa 2014 Framenet Brasil*. Projeto apresentado ao Conselho Nacional de

Desenvolvimento Científico e Tecnológico (CNPq), no âmbito do Edital Universal nº 14/2011, Juiz de Fora, 2011.

SCHMIDT, T. 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.