

## **FRAMES COMO INTERLÍNGUA NA ESTRUTURAÇÃO DE DICIONÁRIOS ELETRÔNICOS MULTILÍNGUES DE DOMÍNIOS ESPECIALIZADOS**

## **FRAMES AS INTERLINGUA IN THE DEVELOPMENT OF A DOMAIN-SPECIFIC TRILINGUAL ELECTRONIC DICTIONARY**

*Maucha Andrade Gamonal\**  
*Tiago Timponi Torrent\*\**

RESUMO: A proposta deste trabalho é apresentar a potencialidade de *frames*, constructos mentais que organizam o conhecimento, como interlíngua para o alinhamento de bancos de dados lexicais de dicionários eletrônicos multilíngues de domínios especializados (GAMONAL, 2013). Essa proposta está sendo explorada pelo projeto Copa 2014 FrameNet Brasil (SALOMÃO ET AL, 2011), no desenvolvimento de um dicionário eletrônico trilingue – Português, Inglês, Espanhol – estruturado em *frames* para os domínios da Copa do Mundo, do Futebol e do Turismo.

PALAVRAS-CHAVE: *Frames*. Interlíngua. Dicionários eletrônicos multilíngues. Copa do Mundo FIFA 2014. Turismo.

ABSTRACT: This work aims to present the deployment of frames, mental constructs that represent knowledge, as interlingual devices for the alignment of lexical databases in domain-specific multilingual electronic dictionaries (Gamonal, 2013). Such a proposal is being developed within the Copa 2014 FrameNet Brasil Project (Salomão et al, 2011), which is building a frame-based trilingual - Portuguese, English, Spanish - electronic dictionary for the domains of the World Cup, Soccer and Tourism.

KEYWORDS: Frames. Multilingual Electronic Dictionaries. FIFA 2014 World Cup. Tourism.

---

\* Universidade Federal de Juiz de Fora (UFJF), Juiz de Fora, Estado de Minas Gerais, Brasil. Doutoranda em Linguística pelo Programa de Pós-Graduação em Linguística da UFJF. Email: mauchaandrade@gmail.com.

\*\* Universidade Federal de Juiz de Fora (UFJF), Juiz de Fora, Estado de Minas Gerais, Brasil. Professor do Departamento de Letras e do Programa de Pós-Graduação em Linguística da UFJF. Email: tiago.torrent@ufjf.edu.br.

## FRAMES COMO INTERLÍNGUA NA ESTRUTURAÇÃO DE DICIONÁRIOS ELETRÔNICOS MULTILÍNGUES DE DOMÍNIOS ESPECIALIZADOS

### INTRODUÇÃO

A estruturação de dicionários envolve tanto um estudo especializado do léxico quanto o suporte de uma teoria linguística que seja capaz de responder aos desígnios esperados pelo recurso lexical a ser elaborado. Dessa maneira, Borba (2003) apresenta a lexicografia a partir de um duplo escopo: técnica e teoria. Acerca da técnica, o autor evidencia questões como seleção de entradas, sistemas definitórios e estruturação de verbetes. Sobre a teoria, é especificado o conjunto de princípios que regem a descrição do léxico. Quanto a esse aspecto, a contribuição de uma teoria linguística é crucial, pois desenvolverá, como Borba assinala, uma metalinguagem capaz de manipular e apresentar as informações que se julguem pertinentes.

Exposto esse panorama, neste artigo, apresentamos uma das aplicações da Semântica de *Frames* (FILLMORE, 1982, 1985) como teoria linguística para a tarefa lexicográfica. A apresentação desta proposta será discutida a partir da apresentação do *frame* Cenário\_do\_turismo (GAMONAL, 2013), desenvolvido para o dicionário Copa 2014 FrameNet Brasil – doravante Copa 2014.

Para tanto, a organização deste trabalho está disposta da seguinte maneira: no capítulo 1, apresenta-se o dicionário Copa 2014 FrameNet Brasil como exemplo de dicionário eletrônico multilíngue de domínio especializado que utiliza a Semântica de *Frames* como teoria linguística para a descrição do léxico; no capítulo 2, discute-se como se deu a modelagem do *frame* Cenário\_do\_turismo para o dicionário Copa 2014; e, no capítulo 3, apresentam-se as bases teóricas para os *frames* serem utilizados como interlíngua.

## 1 DICIONÁRIOS ELETRÔNICOS MULTILÍNGUES E A SEMÂNTICA DE FRAMES

Diante das exigências do mundo atual, as fronteiras linguísticas não se colocam mais como empecilho para a comunicação entre pessoas de diferentes sistemas linguísticos. Essa realidade está colocada para todos os domínios das relações humanas, e, nesse sentido, os estudos da linguagem exercem função essencial no sentido de explorar e desenvolver novas possibilidades teóricas e práticas para contribuir com esse novo cenário mundial.

Uma demanda para o ramo lexicográfico é o desenvolvimento de recursos lexicais multilíngues de domínios especializados. Entretanto, sua composição impõe diversos desafios, sendo um deles o de garantir que os equivalentes de tradução sejam representativos da realidade de cada comunidade linguística.

De acordo com Boas (2009), os padrões de lexicalização, bem como os padrões sintático-semânticos, são componentes importantes para a criação de recursos multilíngues. Talmy (2000 apud BOAS, 2009) salienta que as línguas especificam preferências acerca da lexicalização de seus componentes semânticos, o que sugere que analisar a língua em uso permite conhecer as escolhas dos falantes no processo de atribuição de sentido.

A busca por equivalentes de tradução pode ser mais bem sucedida se esses critérios forem levados em conta. Para isso, Boas também discute a importância das paráfrases. Se se conhecem bem palavras e expressões capazes de ser substituíveis entre si, assegura-se, com mais precisão, quais seriam os melhores correspondentes entre línguas. Nesse sentido, utilizar as ferramentas da Linguística de *Corpus* é uma excelente opção, pois se tem a oportunidade de embasar as escolhas em fontes empíricas, visando a garantir a representatividade das decisões tomadas para cada entrada lexical.

Some-se a isso a alternativa teórica proposta pela Semântica de *Frames*, teoria que concebe o sentido lexical como relativizado a cenas, ou *frames* (FILLMORE, 1982), que se organizam em redes de relações representáveis computacionalmente, as *FrameNets* (FILLMORE ET AL., 2003). Nessas redes, além da definição dos elementos que compõe os *frames* que são evocados pelos itens lexicais, registram-se também, através da anotação de *corpora*, as possibilidades sintático-semânticas de tais itens lexicais, suas valências, o que constitui um valioso banco de dados para o trabalho de atribuição de correspondências interlinguísticas.

É nesse contexto que surge o projeto Copa 2014 FrameNet Brasil, sobre o qual se discorre na seção 1.1.

### 1.1 Copa 2014 *framenet* Brasil: dicionário eletrônico trilingue baseado em estudo de corpus a partir da semântica de frames

O Copa 2014 é um projeto da FrameNet Brasil<sup>1</sup> de criar um dicionário eletrônico *online* trilingue – Português, Inglês, Espanhol – para a Copa do Mundo, que será sediada pelo Brasil em 2014. A proposta, que está sendo desenvolvida em parceria com a Berkeley FrameNet (FILLMORE ET AL., 2003; RUPPENHOFER ET AL., 2010) e o projeto FrameCorp (CHISHMAN ET AL., 2008), é verificar em que medida o uso de uma teoria linguística que procura sistematizar a relação língua e experiência pode contribuir para a modelagem de dicionários.

---

<sup>1</sup> Para mais informações: <<http://www.framenetbr.ufjf.br/>>

A Berkeley FrameNet<sup>2</sup>, na Califórnia, é um projeto que desenvolve uma base lexical computacional para a língua inglesa a partir de *frames*. Esse termo emprestado da Inteligência Artificial (MINSKY, 1987) e da Sociologia (GOFFMAN, 1974) sugere a existência de estruturas de dados estereotipadas para representar situações, enfatizando rituais e papéis sociais empregados no processo da interação humana. Para os estudos linguísticos, o pesquisador Charles J. Fillmore dialoga com esses autores ao defender que o processo de significação das palavras está diretamente ligado à compreensão dessas instituições sociais ou da estrutura de experiência que elas pressupõem. Dessa forma, desenvolve a Semântica de *Frames*, que em suas palavras se assume como

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).<sup>3</sup>

Com esse escopo teórico, o intuito é investigar o comportamento das palavras na tentativa de sistematizar os participantes e as circunstâncias que possibilitam a estruturação de um dado evento. A metodologia adotada para garantir a utilização dessa teoria se dá em torno de *frames*, Elementos de *Frame* e Unidades Lexicais. Em linhas gerais, os *frames* são compreendidos como modelagens – no caso, computacional – de estruturas de conhecimento reconhecível culturalmente. Os atores, as ferramentas e as circunstâncias que modelam e constituem esse constructo mental são os Elementos de *Frame*. Por sua vez, as Unidades Lexicais são o pareamento de uma forma, com todas as suas flexões, a um significado específico. A análise de dados acontece em torno delas, já que são itens lexicais predicadores que evocam *frames*, como os nomes e os verbos.

Os critérios de análise dos dados contemplam um viés semântico e sintático, corroborando os pressupostos da Linguística Cognitiva ao considerar que a forma não apresenta significado, mas escolhe as regularidades que funcionam durante todo o processo de significação (FAUCONNIER, 1997). Desse modo, é válido explorar informações semânticas e sintáticas, o que destaca a eficácia dessa plataforma para a lexicografia, pois, como Borba (2003) afirma, melhor é se o lexicógrafo consegue utilizar uma teoria integrada sintático-semântica para subsidiar a estruturação de dicionário e assim enriquecer o arcabouço de informações disponibilizadas.

Nesse sentido, a Semântica de *Frames* integrada à rede semântico-lexical FrameNet apresenta uma proposta de dicionário inovadora. Além de ser um recurso lexical representativo das realidades linguísticas por utilizar *corpus* para subsidiar as escolhas e análises pelas entradas lexicais, um dicionário estruturado com a Semântica de *Frames* expõe a indissociabilidade entre conhecimento linguístico e conhecimento

---

<sup>2</sup> Para mais informações: <<https://framenet.icsi.berkeley.edu/>>

<sup>3</sup> 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".

enciclopédico. A relação entre experiência e/ou conhecimento de mundo de forma geral mantém uma relação intrínseca com a construção de significado linguístico. Dessa forma, se se tenta separar um do conhecimento do outro, o recurso por si só já será limitado.

O propósito do Copa 2014 é explorar a palavra a partir de um dos seus sentidos e, posteriormente, definir a qual *frame* ela se vincula para então especificar os itens lexicais que participam da estruturação de determinado conceito. Assim sendo, essa proposta se aproxima dos *thesauri* ou dicionários analógicos, por organizarem os itens lexicais pelas relações semânticas estabelecidas entre si.

A Figura 1 ilustra as possibilidades de busca do dicionário. O usuário pode realizar a pesquisa, nas línguas portuguesa, inglesa e espanhola, por palavra, digitar a frase, ver o significado e explorar a rede de relações entre os sentidos.

**Figura 1** - Apresentação do Copa 2014



Fonte: <http://dicionariodacopa.com.br/>

Nas Figuras 2, 3 e 4, a busca por palavra, no caso, *abrigar*, possibilita ao usuário encontrar, além da definição do item lexical (Fig. 1), informações sobre os participantes da cena ou *frame*, Potencial\_turístico também chamado de Atração\_em\_Lugar, ver Figura 11, a que a palavra *abrigar* está relacionada, (Fig.3) e também sentenças de exemplo (Fig.4).

Figura 2 - Busca pela palavra *abrigar* – Definição

**abrigar.v**

**Cena:**  
Potencial turístico

---

**Definição**

possuir ou alojar atrações que promovem deslocamento de turistas de diferentes origens para conhecer ou usufruir seus benefícios.

**Traduções**

 house.v

Figura 3 - Busca pela palavra *abrigar* – Participantes

**abrigar.v**

**Cena:**  
Potencial turístico

---

**Participantes**

**Atração**  
Recursos naturais e/ou artificiais, que apresentam algum tipo de valor, como cultural, histórico ou ecológico.

**Lugar**  
Lugar que abriga ou possui uma **Atração**.

**Razão**  
Motivo que faz com que o **Lugar** ou a **Atração** seja interessante para turistas.

**Duração**  
Período que especifica o tempo que o **Lugar** hospeda ou possui uma **Atração**.

Figura 4 - Busca pela palavra *abrigar* – Sentenças

abrigar.v

Cena:  
Potencial turístico

---

Sentenças

A região ainda abriga grutas de grande beleza. [link](#)

O sul, por sua vez, abriga bairros que lembram vilarejos interioranos, onde os descendentes de açorianos fazem questão de manter vivas as tradições culturais de seus antepassados. [link](#)

Fundada em 1719, durante a Corrida do Ouro no Brasil, a cidade ainda abriga diversas construções históricas que foram declaradas patrimônio nacional em 1992. [link](#)

Se a opção for ver o significado, o usuário terá acesso a informações sobre as cenas envolvidas nos domínios do Turismo, Futebol e Copa. A Figura 5 mostra a cena Potencial\_Turístico com sua definição e seus participantes.

Figura 5 - Busca por ver significado Potencial\_Turístico – Cena

Potencial turístico

---

Definição

Um Lugar possui ou abriga uma Atração. Em geral, são recursos naturais ou artificiais, com valor histórico ou cultural, que promovem o movimento de pessoas para conhecer e explorar seu potencial.

Participantes

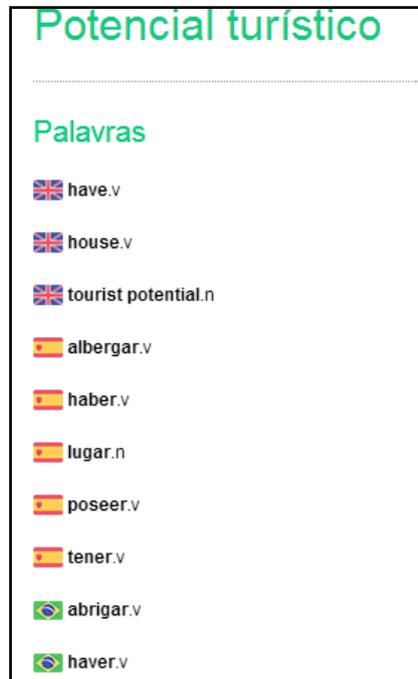
**Atração**  
Recursos naturais e/ou artificiais, que apresentam algum tipo de valor, como cultural, histórico ou ecológico.

**Lugar**  
Lugar que abriga ou possui uma Atração

**Razão**  
Motivo que faz com que o Lugar ou a Atração seja interessante para...

Ainda nessa ferramenta de busca, encontra-se a descrição das Unidades Lexicais incluídas como evocadoras da cena. Figura 6.

**Figura 6** - Busca por ver significado – Potencial\_Turístico – Palavras



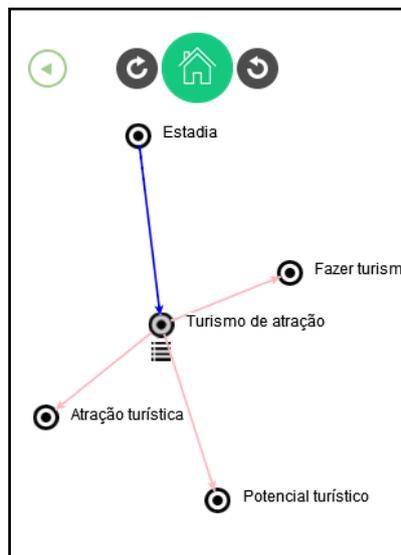
Por último, o usuário pode também conhecer as relações entre as cenas com opções que permitem especificar a busca, Figura 7. Essa ferramenta também inclui o domínio da Copa, do Futebol e do Turismo, e a busca pode ser feita nas três línguas.

**Figura 7** - Busca por Explorar Rede



A pesquisa por Potencial\_Turístico, Figura 8, mostrou que esta cena atua em relação de Perspectiva com Fazer\_Turismo (ou Turismo\_por\_turista) e Atração\_Turística. Além disso, mantém relação de Subframe com a cena Estadia (ou Cenário\_do\_Turismo\_Estada). Tais relações serão detalhadas na subseção a seguir.

**Figura 8 - Busca por Explorar Rede – Potencial\_Turístico**



## 1.2 A compilação de corpora para subsidiar as análises do dicionário

A necessidade de garantir que as decisões analíticas do dicionário estivessem amparadas por informações empíricas fez com que *corpora* específicos fossem criados para cada domínio do Copa 2014. A respeito do domínio do Turismo para a língua portuguesa do Brasil, o processo de compilação esteve focado em textos disponibilizados eletronicamente que abordassem a atividade turística no Brasil, país sede da Copa do Mundo de 2014.

Os *corpora* são compostos de textos advindos de sites de órgãos governamentais de fomento à atividade turística e *blogs* de viajantes. Por considerar o dicionário Copa 2014 como um guia de uso para turistas e demais pessoas vinculadas ao evento, as informações deveriam advir de fontes seguras, tanto acerca das atrações como em questões de infraestrutura, e também era importante valorizar as experiências dos turistas sobre cada lugar visitado. Dessa forma, a opção por essas duas fontes de dados foi positiva.

Para a estruturação dos *corpora*, vários critérios devem ser respondidos, como: modo, tempo, seleção, conteúdo, autoria e finalidade. Atendendo a essa tipologia discutida em Sardinha (2004), os *corpora* coletados para o domínio turístico são, respectivamente: escritos; sincrônicos/contemporâneos; dinâmicos, pois podem ser aumentados e/ou diminuídos; de domínios especializados; e, por último, são textos de falantes nativos.

A respeito do critério de autoria, quando colocado numa perspectiva multilíngue, *corpora* autênticos de mais de uma língua podem ser considerados comparáveis, diferentemente daqueles nomeados como paralelos, compostos por um texto na língua fonte e os demais sendo versões traduzidas. Para o domínio do turismo,

a opção foi utilizar *corpora* comparáveis, uma vez que as fontes de dados multilíngues encontradas não eram representativas das realidades linguísticas em questão. Entretanto, o projeto como um todo vem explorando tanto textos comparáveis como textos paralelos.

Dessa forma, a ideia de disponibilizar as mesmas entradas lexicais ou as mesmas sentenças ilustrativas não acontecerá em todos os casos para os dados do Turismo. O propósito será investir nas paráfrases, assim como pondera Boas (2009), com o intuito de garantir que o recurso tenha um mesmo conteúdo semântico, mas não, necessariamente, equivalentes estritos de tradução.

### 1.3 A análise lexicográfica dos corpora

A análise dos dados coletados segue a metodologia adotada pela Berkeley FrameNet (RUPPENHOFER ET AL., 2010) com as adaptações necessárias para as especificidades da língua portuguesa. O procedimento segue um modelo de análise conhecido como sendo de três camadas, no qual são especificadas informações semânticas e sintáticas sobre a valência das Unidades Lexicais analisadas.

**Figura 9** - Sentença com Unidade Lexical *conhecer*

Layer	O	t	t	r	i	s	t	a	p	o	d	e	c	o	n	h	e	c	e	r	a	s	b	e	l	e	z	a	s	l	o	c	a
FE	T	u	r	i	s	t	a													A	t	r	a	ç	ã	o							
GF_br	E	x	t																	O	b	j	D										
PT_br	N	P																		N	P												

A sentença “*o turista pode conhecer as belezas locais*”, ilustrada pela Figura 9, tem *conhecer* como Unidade Lexical (LU). A análise notifica os Elementos de *Frame*, Função Gramatical e Tipo Sintagmático (no inglês, FE, GF, PT, respectivamente) que acompanham esse predicador verbal. Cada sintagma é analisado levando em conta esses três critérios. Na sentença, *o turista* é classificado como FE Turista; GF, Externo; e PT, Sintagma Nominal. O sintagma *as belezas locais* é marcado como FE Atração; GF, Objeto Direto; e PT, Sintagma Nominal.

Diversas sentenças são analisadas com essa LU. Finalizada essa etapa, parte-se para a frequência de ocorrência dos padrões encontrados na análise dos dados. Essa etapa permite conhecer o comportamento do predicador em questão, especificando suas preferências sintáticas e semânticas. Um *frame* é estruturado a partir de um conjunto de LUs que participam de um mesmo campo semântico perfilando características familiares. Por exemplo, *conhecer* se une a outras, que não, necessariamente, serão da mesma classe de palavras. Por exemplo, *visitar*, *fazer tour*, *turista* e *visitante*, ainda que estejam incluídos verbos e nomes, compartilham traços similares, pois todos evocam o *frame* Turismo\_por\_turista, que trata da experiência turística a partir do olhar do ator principal dessa atividade, o turista.

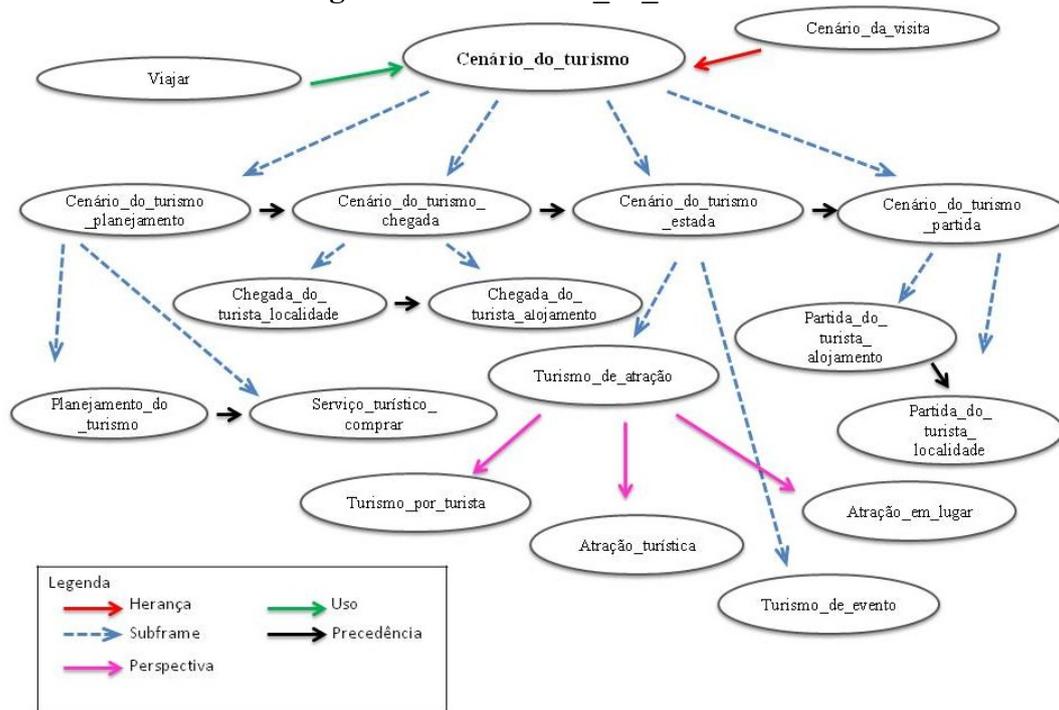
## 2 DOMÍNIO DO TURISMO: FRAMES DE EXPERIÊNCIA TURÍSTICA

A diversidade cultural encontrada no mundo favorece a variedade de opções turísticas. Entretanto, por mais que existam diferentes culturas e diferentes perfis de turistas, seja no oriente ou no ocidente, o evento turístico se modela da mesma maneira

em qualquer localidade. Isso se justifica porque qualquer povo, ao assumir o papel de turista, insere-se num evento transcultural. Por tal motivo, todos compreendem o funcionamento das regras e é capaz de fazer projeções acerca das etapas que constituem a experiência turística.

Para modelar essa experiência em *frames*, foi utilizado o *frame* cenário. Sua função principal é marcar eventos e experiências que se construam de forma complexa e/ou produtiva, ou seja, *frames* que mantenham relação entre si e que sejam capazes de se dividir em outros mais específicos.

**Figura 10 - O Cenário\_do\_turismo<sup>4</sup>**



O *frame* Cenário\_do\_turismo, ilustrado na Figura10, mostra como essa experiência pode ser modelada. Cada cor de seta especifica uma diferente relação entre *frames*. O Cenário\_da\_visita é marcado com a seta vermelha, indicando que mantém uma relação de herança com o Cenário\_do\_turismo. Dessa forma, o Cenário\_do\_turismo detalha um tipo específico de visita, aquela com fim de entretenimento. Pela literatura, isso aponta que os Elementos de *Frame* do mais amplo serão, de alguma forma, mantidos no *frame* mais específico. Assim, enquanto no Cenário\_da\_visita, há o EF Agente e o EF Entidade, no Cenário\_do\_turismo, esses EFs serão o Turista e a Atração, mais específicos que aqueles. Já a seta verde marca outra relação, a de uso. Ela indica que, para compreender o turismo, é necessário ter em mente o *frame* Viajar.

Além de detalhar as relações com outros domínios, destacam-se também como o Cenário\_do\_turismo é produtivo. A relação de subframe, marcada pela seta azul, aponta que essa atividade pode ser dividida em etapas, a partir de uma ordenação temporal: planejamento, chegada, estada e partida. No Cenário\_do\_turismo\_planejamento, são inseridas questões como pesquisas para

<sup>4</sup> O *frame* Cenário\_do\_turismo está pormenorizado em GAMONAL, 2013.

conhecer o local de destino e compras antecipadas que se fizerem necessárias. Concluída essa etapa, parte-se para o *frame* *Cenário\_do\_turismo\_chegada*, que trata da chegada do turista ao destino almejado e, conseqüentemente, ao local escolhido como alojamento. Após a chegada, segue-se com a parte mais esperada, descrita no *frame* *Cenário\_do\_turismo\_estada*. Dentro desse *frame*, além de poder especificar os variados tipos de turismo, serão descritos os acontecimentos e experiências durante a estada do turista. A última etapa desse evento está estruturada no *Cenário\_do\_turismo\_partida*, que descreve a partida do turista do local onde a atividade turística aconteceu para o seu destino de origem.

A descrição dos *frames* conta com a análise de variadas Unidades Lexicais, as quais comporão as entradas do dicionário. Por exemplo, o *Turismo\_de atração*, subframe do *Cenário\_do\_turismo\_estada*, agrega diversas Unidades Lexicais divididas em três perspectivas, *Turismo\_por\_turista*, *Atração\_turística* e *Atração\_em\_Lugar*. Cada um perfilará um EF específico para se referenciar ao *frame* *Turismo\_de atração*. No *Turista\_por\_turista*, como já dito, o enfoque será dado ao *Turista*, com ULs como *fazer tour*, *turista* e *apreciar*. O *frame* *Atração\_turística* perfila o EF *Atração* com ULs, como *oferecer*, *atrair* e *apresentar*. E, por último, o *frame* *Atração\_em\_lugar*, especifica o *Lugar* que possui ou hospeda atrações, as ULs *abrigar*, *possuir* e *haver* suscitam a perspectiva dada a esse *frame*.

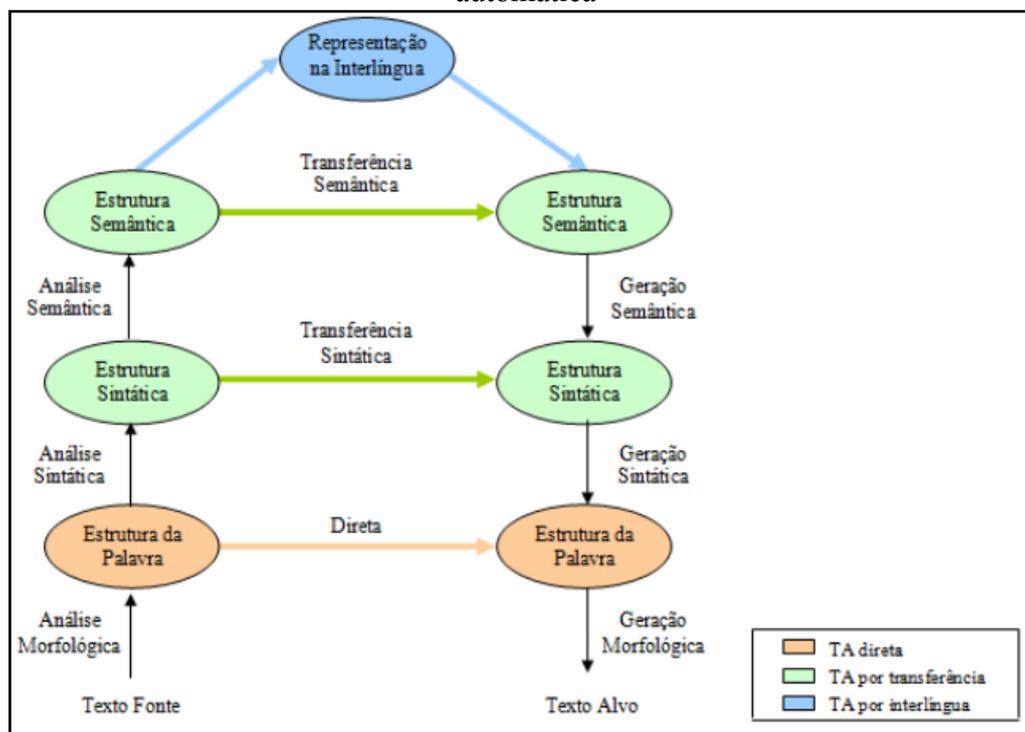
Em linhas gerais, a modelagem desse domínio permitiu chegar à conclusão de que os *frames* modelados com *corpora* do português do Brasil representará a mesma estrutura dos *frames* para as outras línguas do dicionário. Isso porque a atividade turística, assim como também o são a Copa do Mundo e a modalidade de futebol desse evento retratam experiências transculturais. Por mais que isso não invalide a necessidade de compilação de *corpora* específicos para registrar as lexicalizações dos conceitos desses eventos, a modelagem dos *frames* para a língua portuguesa do domínio turístico é capaz de projetar um molde desse *frame*, que pode ser considerado universal, possibilitando assim utilizar *frames* como interlíngua.

### 3 FRAMES COMO INTERLÍNGUA

Dicionários eletrônicos multilíngues demandam algumas especificidades diante das versões impressas de mesma finalidade, pois, neles, é necessário que o sistema computacional no qual esses materiais estarão armazenados seja capaz de comunicar na busca por traduções. Uma das formas de realizar esse alinhamento de dados é através do método indireto de interlíngua. O conceito de interlíngua assume vários vieses, sendo explorado com diferentes enfoques teóricos. Na Linguística Computacional, a interlíngua pode ser compreendida como um sistema linguístico de representação, responsável por garantir o elo de comunicação necessário entre diferentes sistemas linguísticos.

Dias da Silva et al. (2007) avaliam que o objetivo desse método é capturar o significado a ser transmitido, independentemente da língua a ser traduzida. Não há aqui o objetivo de encontrar equivalências estritas na língua alvo, o que, na avaliação dos autores, torna a interlíngua mais próxima de uma paráfrase do que de uma tradução, propriamente dita, reiterando o objetivo inicial.

**Figura 11** - Níveis de profundidade do conhecimento nos sistemas de tradução automática



Fonte: (Cf. DIAS-DA-SILVA ET AL., 2007, p.62)

Além da representação na interlíngua, a Figura 11 ilustra outros métodos de tradução. A direta é o método mais simples, já que utiliza o mínimo possível de processamento linguístico. O mais comum nesse método é o contraste de dois sistemas linguísticos, pois possibilita o uso de um dicionário bilingue para viabilizar a tradução automática, que, geralmente, acontece via palavra. Assim sendo, Dias da Silva et al. (2007) elencam várias dificuldades proporcionadas por esse tipo de tradução: o banco de dados pode não ter o equivalente buscado, pode não reconhecer as construções gramaticais utilizadas ou mesmo a estrutura dos constituintes, problemas que, certamente, influenciarão na qualidade do produto final.

Em contraste com o método de tradução direta, os métodos indiretos proporcionam análises mais refinadas, pois, conforme os autores evidenciam, reconhece-se a necessidade de entender o sistema linguístico e também extralinguístico das línguas envolvidas. No método indireto de transferência, Dias da Silva et al. (2007) destacam as etapas que comumente são seguidas, como a necessidade de criar uma representação intermediária na língua fonte para se comunicar com a representação intermediária da língua alvo e, assim, gerar a sentença nessa língua. Todavia, com a dificuldade de conseguir criar as regras necessárias para transferir dados linguísticos, o método de interlíngua é uma opção que consiste na análise completa do texto na língua fonte para que o significado seja extraído e representado na interlíngua para, posteriormente, ser gerado o texto na língua desejada.

O procedimento de tradução indireta da representação na Interlíngua é uma proposta que se alia à noção de *frame* sugerida pela Semântica de *Frames*. Os *frames* recuperam noções semânticas importantes, que não são possíveis através de relações lexicais, como sinonímia e antonímia. A relação da palavra *turista* com *check-in*, por

exemplo, é muito intuitiva nesse domínio, mas não é resgatada pelos dicionários tradicionais. Afirmar a existência de *frames* transculturais ou universais elimina as divergências socioculturais, o que faz com que os *frames* tenham o mesmo status de nuclearidade<sup>5</sup>. Assim sendo, as diferenças podem ser apresentadas nos padrões de realização, que podem se distinguir quanto às valências semânticas e sintáticas, mas não quanto à modelagem do evento.

Acerca das possíveis diferenças de lexicalização e, conseqüentemente, a busca pelos equivalentes de tradução, os padrões de valência proporcionados pela análise em três camadas serão explorados na tentativa de que contribuam para escolher os correspondentes de tradução.

#### 4 CONCLUSÃO

A utilização de *frames* como índice interlingual é uma alternativa que está sendo explorada pelo dicionário Copa 2014 FrameNet Brasil para garantir o alinhamento entre os três sistemas linguísticos abrangidos pelo dicionário. A inserção de novas teorias linguísticas que sejam capazes de contribuir no trabalho em Lexicografia é importante para garantir novas possibilidades na atual “era do dicionário eletrônico”.

#### REFERÊNCIAS BIBLIOGRÁFICAS

BOAS, H. Semantic frames as interlingual representations for multilingual lexical databases. In: BOAS, H. C. (Ed.). *Multilingual FrameNets in computational lexicography: Methods and applications*. Berlin/New York: Mouton de Gruyter, 2009.

BORBA, F. S. *Organização de dicionários: uma introdução à lexicografia*. São Paulo: Ed. UNESP, 2003;

CHISHMAN, R. L. O.; BERTOLDI, A.; LERMEN, L.; PADILHA, J. G. Corpus e Anotação Semântica: um Experimento para a Língua Portuguesa a partir da Semântica de Frames. In: *WebMedia 2008 - XIV Simpósio Brasileiro de Sistemas Multimídia e Web - V Workshop de Tecnologia e Informação*, 2008, Vila Velha. Anais - Webmedia, v. II. 2008, p. 321-325;

DIAS-DA-SILVA, B. D. et al., Introdução ao processamento das línguas naturais e Algumas Aplicações. In: *Série de Relatórios do NILC*. NILC-TR-07-10. São Carlos-SP, Agosto, 2007, p. 58-77;

FAUCONNIER, G. *Mappings in Thought and Language*. Cambridge: Cambridge University Press, 1997;

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

---

<sup>5</sup> Os Elementos de *Frame* apresentam diferentes *status*. Há EFs nucleares, aqueles essenciais para a modelagem do *frame*, e os não nucleares, que adicionam traços importantes, mas que não são indispensáveis para sua estruturação.

\_\_\_\_\_. ; JONHSON, C.; PETRUCK, M. Background to FrameNet. In: *International Journal of Lexicography*. Oxford University Press, v. 16 nº 3, 2003, p. 235-250;

RUPPENHOFER, J. et al., *FrameNet II: Extended theory and practice*. Berkeley: International Computer Science Institute, 2010. Disponível em: <<http://framenet.icsi.berkeley.edu/>>. Acesso em 28/02/13;

GAMONAL, M. A. *Copa 2014 Framenet Brasil: Diretrizes Para A Constituição De Um Dicionário Eletrônico Trilíngue A Partir Da Análise De Frames Da Experiência Turística*. Dissertação de mestrado. Juiz de Fora: Universidade Federal de Juiz de Fora, Faculdade de Letras, Programa de Pós-Graduação em Linguística, 2013.

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

MINSKY, M. *The Society of Mind*. London: Willian Heinemann Ltd, 1987;

SALOMÃO, M. M. M. ET AL. *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;

TALMY, L. Toward a cognitive semantics. V. I: *Concept structuring systems*. Cambridge, Mass.: MIT Press, 2000.

Recebido em: 16 de fevereiro de 2014.  
Aceito em: 24 de abril de 2014.