

## A modelagem do Cenário do Turismo

Maucha Andrade Gamonal  
Universidade Federal de Juiz de Fora  
mauchaandrade@gmail.com

**RESUMO:** Este artigo traz os resultados iniciais do desenvolvimento do Cenário\_do\_Turismo para o dicionário temático trilingue (português – inglês – espanhol), COPA 2014 FrameNet Brasil. O objetivo é apresentar como a experiência turística pode ser modelada se estruturada com base na teoria linguística da Semântica de *Frames*.

**PALAVRAS-CHAVE:** Semântica de *Frames*; *Frame*; FrameNet; Cenário\_do\_Turismo

**ABSTRACT:** This article presents initial results of the development of the Tourism\_scenario to the electronic thematic trilingual dictionary (Portuguese, English, Spanish), Copa 2014 FrameNet Brasil Project. The aim is to present how the tourist experience can be modeled is structured based on the linguistic theory of frame semantics.

**KEY-WORDS:** Frame Semantics; Frame; FrameNet; Tourism\_scenario

### Introdução

Esse trabalho se insere no projeto Copa 2014 FrameNet Brasil (doravante Copa 2014), um desdobramento da FrameNet Brasil, que está desenvolvendo um dicionário temático trilingue (português, inglês, espanhol), em meio eletrônico, para a Copa do Mundo de 2014. Abrangendo os domínios do futebol e do turismo, pretende-se que o dicionário seja útil para todos aqueles que estarão assistindo aos jogos e, provavelmente, visitarão as principais atrações turísticas do Brasil e também para aqueles envolvidos na recepção dos turistas, bem como na organização do evento.

O que torna esse dicionário diferente dos demais é a base teórica da Semântica de *Frames* (FILLMORE, 1982). Ela possibilita que o usuário, ao realizar uma busca, localize não apenas o significado de determinado lexema, mas tenha acesso a uma gama de informações que só estaria disponível nos dicionários tradicionais através das inferências realizadas pelo usuário. No domínio do futebol, o torcedor, por exemplo, encontrará o significado do termo que se refere a um drible, mas também terá acesso ao contexto ao qual aquela informação está vinculada.

Já em relação ao turismo, o dicionário servirá como um guia para aqueles que, entre um jogo e outro, conhecerão as belezas do Brasil. Além dos correspondentes de tradução, os usuários encontrarão, nas sentenças ilustrativas, as principais atrações nos locais em que se encontram e obterão diversas informações, como gastronomia típica, formas de hospedagem, opções de comércio e diversão noturna.

Tendo isso em vista, o objetivo deste artigo é investigar os elementos que compõem o cenário turístico, a fim de desenvolver um recurso que contenha informações que estejam direta ou indiretamente relacionadas a ele.

Nas linhas que se seguem, serão apresentados a metodologia para a construção do dicionário, o aporte teórico que alicerça a pesquisa, bem como recursos fornecidos pela FrameNet, rede semântico-lexical na qual os dados linguísticos são analisados. Com essas informações, o cenário em desenvolvimento para o turismo e os primeiros *frames* modelados serão analisados.

## 1. Materiais e metodologia

Pelo fato de o produto final ser um aplicativo eletrônico, é necessário realizar procedimentos computacionais para que os dados se tornem legíveis por máquinas. Para este trabalho, o *corpus* utilizado foi o Cetenfolha, compilação de textos eletrônicos extraídos do jornal Folha de São Paulo, disponível para acesso ao público pelo site: <http://www.linguateca.pt>. A FrameNet Brasil acessa esses dados através do Sketch Engine, sistema de consulta a *corpus*, que contém ferramentas úteis para análise de dados, como o Word Sketch, responsável por evidenciar o comportamento gramatical da sentença, através da frequência dos principais sintagmas contidos no material analisado.

Sketch Engine

corpus: Cetenfolha, Cetempublico

Concordance  
Word List  
Word Sketch  
Thesaurus  
Find X  
Sketch-Diff  
?

Save  
Change options  
Clustering  
Sorting  
Gramrels  
MW links  
More data  
Less data

Menu position

## visitar

Cetenfolha, Cetempublico freq = 6315 (95.2 per million)

modifier	976	1.2	objeto	1951	5.9	sujeito	1066	3.0	pp_por	146	3.4	pp_até	51	18.6
oficialmente	38	8.75	túmulo	19	8.09	turista	35	8.35	turista	4	5.66	31	4	6.18
anualmente	18	8.35	feira	32	7.7	papa	13	7.32	milhar	4	5.11	30	4	4.16
regularmente	14	8.15	exposição	55	7.56	comitiva	9	7.3	mil	7	4.02	dia	15	3.37
diariamente	20	8.02	instalação	47	7.52	embaixador	14	7.19	vez	4	4.0	pp_de	32	0.1
recentemente	31	7.5	museu	27	7.31	cônsul-geral	4	6.91	milhão	6	3.09	terça	4	8.46
ontem	247	6.98	folha	88	7.26	primeiro-ministro	26	6.72	peessoa	7	2.75	segunda	5	7.53
semanalmente	4	6.8	cidade	135	7.18	vale	10	6.67	grupo	4	1.9	pp_de+o	23	0.1
brevemente	5	6.77	bienal	13	7.13	reportagem	10	6.67	pp_por+o	119	2.6	cidade	5	2.52
gratuitamente	4	6.6	fábrica	22	6.97	chanceler	7	6.62	turista	5	6.0	mundo	4	2.5
amanhã	15	6.46	salão	13	6.95	secretária=de=Estado	4	6.57	folha	7	3.75	pp_durante	20	5.1
frequentemente	5	6.33	guiada	7	6.87	delegação	14	6.49	ministro	7	3.29	dia	5	1.79
quando	45	6.14	comitê	13	6.87	secretário=de=Estado	11	6.32	presidente	5	2.04	pp_entre	18	1.6
hoje	56	5.87	estande	8	6.87	candidato	43	6.25	público	5	1.99	14h00	4	10.0
jamais	4	5.69	local	65	6.58	prefeito	9	6.19	hospital	5	4.33	pp_com	14	0.3
após	4	5.66	pavilhão	11	6.52	príncipe	5	6.19	prisão	4	4.04	frequência	7	7.39
anteontem	9	5.64	cemitério	8	6.47	chefe=de=Estado	4	6.15	dia	4	1.46	pp_há	9	2.3
onde	23	5.42	ruína	7	6.43	pessoa	70	6.03	ano	4	0.61	dia	4	1.47
enquanto	4	5.26	castelo	7	6.39	ministro	47	5.98						
depois	15	5.09	santuário	6	6.38	presidente	76	5.93						
também	68	4.96	monumento	8	6.33	ministra	4	5.93						
mais	32	4.71	obra	57	6.32	rei	6	5.91						
que	6	4.66	terreiro	5	6.27	governante	4	5.73						
antes	6	4.57	igreja	17	6.21	ex-governador	4	5.72						
ainda	38	4.49	jerico	5	6.2	vereador	8	5.7						
nunca	9	4.23	ilha	13	6.18	governador	11	5.59						

Figura 1: Visualização da ferramenta Word Sketch no sistema de busca Sketch Engine

Esse recurso auxilia no processo de análise, pois, como pôde ser visto na Figura 1, é possível visualizar diferentes sintagmas que acompanham o lexema *visitar*. Dessa forma, a informação de que, em 35 das 1066 sentenças encontradas, o sujeito é o sintagma nominal *turista* facilita a pesquisa, pois indica ser uma visita para fins de entretenimento diferentemente das 76 sentenças cujo sujeito é *presidente*, pois estas apresentam grande possibilidade de se tratarem de visitas com fins oficiais. Para a construção deste dicionário, essa ferramenta é de grande relevância, pois, como o recurso é de domínio específico, não é necessário trabalhar com frequência, e sim por uma amostragem dos dados relevantes.

Após essa primeira análise, os procedimentos adotados pela FrameNet Brasil são os mesmos que a FrameNet (FN) realiza para a anotação lexicográfica<sup>1</sup>. A ferramenta Desktop, utilidade pela FN, está sendo aprimorada para o português. Trata-se de um *software* utilizado para facilitar o processo de anotação dos dados. Em linhas gerais, ela registra, através de camadas, o comportamento semântico e sintático dos sintagmas que acompanham a Unidade Lexical – que é o pareamento de uma forma com um sentido específico –, identificando os

<sup>1</sup> Cf. RUPPENHOFER ET AL., 2010



possui para aprender uma língua. Pretendia-se demonstrar quais símbolos e regras eram responsáveis pela capacidade da linguagem, além de formular sobre seu aspecto criativo – aquele que permite aos seres humanos criar sentenças infinitas a partir de regras finitas –, “o problema de Descartes”, como apontava Chomsky (1987). Dessa forma, não eram abordados aspectos sociais e culturais, já que se considerava que estes não interferiam na explicação dos fenômenos estudados.

Já nos anos setenta e oitenta, uma forte reação sobre as postulações da primeira geração do cognitivismo linguístico foi iniciada. Pesquisadores com amplo conhecimento sobre o gerativismo, como George Lakoff, Charles Fillmore, Ronald Langacker passaram a reivindicar a importância da semântica / pragmática para os estudos cognitivos. Foi apenas no final dos anos oitenta que houve de fato alguma formalização sobre esse movimento de cisão. Com a publicação de trabalhos desses e de diversos outros pesquisadores e a consequente difusão desses estudos principalmente pela Europa, uma nova linha de pesquisa estava surgindo, nas palavras de Salomão (2010), uma irreversível mudança de paradigma na forma como a linguagem era compreendida. A Linguística Cognitiva se torna uma rica nova vertente para os estudos da linguagem.

Não há como resumir o campo de atuação dessa nova vertente em poucas linhas, o que se pode afirmar é que o alvo de interesse passa a ser todos aqueles fenômenos linguísticos que eram, até então, desconsiderados. A relação do Homem com diversos fatores, como social, cultural, biológico, histórico e situacional, passa a ser central para entender o funcionamento da linguagem.

É nesse ambiente que surge a Semântica de *Frames*. Uma abordagem que enfatiza a estreita relação existente entre a linguagem e a experiência, propondo-se, por isso, ser um programa sobre semântica empírica para análise e descrição dos dados da língua.

O pesquisador Charles J. Fillmore, fundador dessa teoria, sugere uma nova perspectiva para analisar o sentido das palavras, bem como para caracterizar os princípios que são utilizados para a criação de novas palavras e a construção de novos sentidos. Para o autor, o processo de significação das palavras está diretamente ligado à compreensão das instituições sociais ou da estrutura da experiência que elas pressupõem. Assim sendo, para que se possa entender um conceito, é necessário compreender toda a estrutura que está ao redor dele. Conhecendo-se um, acessam-se os demais. *Frame* foi o termo escolhido para representar tal estrutura de conhecimento.

Uma ótima analogia, para confrontar a visão tradicional com a nova teoria apresentada, foi a que Fillmore (1982) fez entre a gramática e um conjunto de ferramentas.

Conhecer uma ferramenta pressupõe entender como ela é e de que é feita, assim o autor aponta ser a morfologia e a fonologia. Entretanto não basta isso, já que é importante saber por quais motivos ela é usada e quem a utiliza, por exemplo. O que o autor pretende apresentar com essa relação é que o texto, entendido aqui como qualquer manifestação linguística, não é um aglomerado de significados prontos, mas sim o registro das “ferramentas” utilizadas; sendo papel do interlocutor descobrir em quais “atividades” o locutor estava envolvido. Na expressão de Fauconnier (1994: xxii), “a linguagem não porta o sentido, mas o guia”.

Uma demonstração de que a palavra evoca um *frame*, e o *frame*, por sua vez, estrutura o significado da palavra pode ser percebida pela noção de fim de semana. Fillmore (1982) ressalta a existência de um contexto motivador para que uma categoria se torne compreensível. A organização do tempo está diretamente ligada a fatores culturais dos povos. Na cultura cristã, percebe-se a existência de um período de tempo mais extenso para a prática de atividades remuneradas e um curto período para o descanso e lazer. O autor pondera que, se houvesse apenas um dia para tal prática ou se os dias de descanso fossem superiores àqueles para se dedicar ao trabalho, provavelmente, esse termo não seria utilizado. A cultura muçulmana é uma demonstração disso, já que constrói um calendário semanal distinto do cristão. Não havendo, pois, uma restrição aos momentos de lazer como propõe o calendário cristão, para eles, pode-se optar por descansar entre a quinta e a sexta ou entre a sexta e o sábado, por exemplo.

Uma demonstração mais completa sobre aquilo que a Semântica de *Frames* se propõe a fazer será visto adiante com a rede semântica FrameNet.

### 3. O projeto FrameNet e os *frames* de cenário<sup>2</sup>

#### 3.1 FrameNet: uma base de dados lexicográficos fundamentada em *frames*

O projeto FrameNet (FN) surgiu na década de noventa no Instituto Internacional de Ciência da Computação, na Califórnia, quando o professor Charles J. Fillmore se aposentou na Universidade de Berkeley e foi então convidado por Jerome Feldman para se juntar ao instituto. A proposta submetida era construir uma base lexical para a língua inglesa através de *frames*. As características principais do projeto são: i) o compromisso de as análises, tanto

---

<sup>2</sup> Para maiores informações, seguem as referências desta seção: FILLMORE, 1998; RUPPENHOFER ET AL. 2010; <https://framenet.icsi.berkeley.edu/> .

semântica quanto sintática, serem evidenciadas em *corpora*; ii) analisar o *frame* a partir das Unidades Lexicais (UL) que o evocam, como os verbos e os substantivos; iii) representar, em camadas, o comportamento semântico e sintático dos elementos que acompanham a palavra-alvo (UL), chamados de Elemento de *Frame*, Tipo Sintagmático e Função Gramatical; iv) possibilitar que esse material seja também legível por máquinas a fim de contribuir com o Processamento de Linguagem Natural.

Os dados da FrameNet podem ser pesquisados através do site <https://framenet.icsi.berkeley.edu/>. Há várias ferramentas a que o usuário pode recorrer para aprofundar a pesquisa lexicográfica, seguem algumas: demonstração das relações estabelecidas entre os *frames*, como herança, subframe e perspectiva, encontradas na seção FrameGrapher; opção para o usuário pesquisar tanto por lemas quanto por *frames*; banco de dados contendo descrições dos *frames*, dos papéis semânticos dos argumentos que acompanham a UL, bem como dos seus comportamentos sintáticos; e informação acerca da frequência dos padrões sintáticos e semânticos dos Elementos de *Frame*.

Tudo isso acompanhado de um representativo banco de dados que conta com mais de 1000 *frames* descritos, ligados a mais de 170 mil sentenças anotadas manualmente, que fornecem um rico material para o treinamento de rotulagem de papel semântico. Esse levantamento de dados permite aplicações – e aprimoramentos – de ferramentas, como tradução automática e extração de informações, além de ser um dicionário de valência, com dados detalhados sobre informações semânticas e sintáticas das estruturas linguísticas da língua inglesa.

### 3.2 *Frames* de cenário na FrameNet<sup>3</sup>

Para o desenvolvimento da proposta deste trabalho, faz-se relevante destacar os *frames* de cenário que são descritos pela FrameNet. *Visiting\_Scenario*, *Employment\_Scenario* e *Crime\_Scenario* são alguns exemplos desses *frames*, que são criados para se conectar semanticamente com outros. O objetivo é demonstrar como alguns eventos complexos acontecem. A relação de precedência, por exemplo, é importante para tentar sistematizar as etapas em que a experiência ocorre.

---

<sup>3</sup> Para a compreensão destas relações, é importante estudar as relações entre *frames* que a FN propõe. As que foram representadas nas figuras foram: Subframe: traçado azul; Herança: linha vermelha; Perspectiva: linha rosa; Uso: traçado verde; Precedência: linha preta (Cf. RUPPENHOFER ET AL. 2010 ).

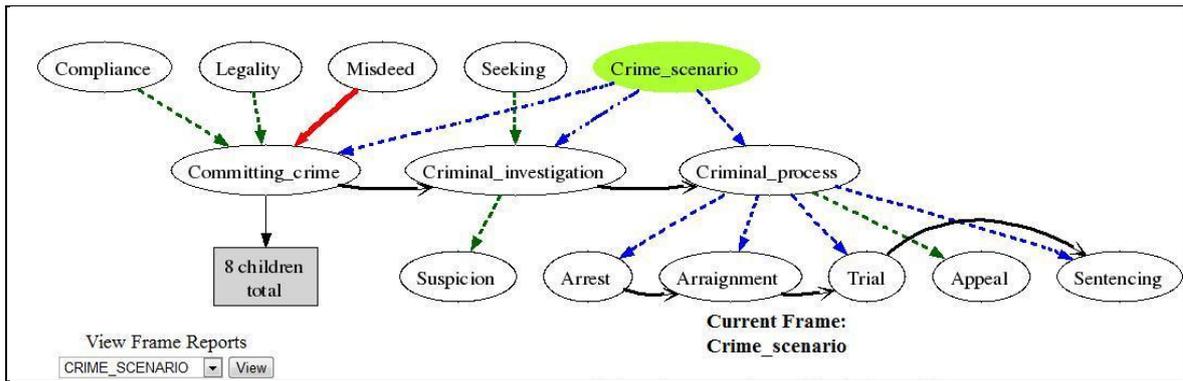


Figura 3: *Crime\_Scenario* apresentado pelo FrameGrapher

O *frame Crime\_Scenario*, na Figura 3, procura mostrar como é a organização desse evento. Sua descrição tem o objetivo de ser abrangente o bastante para considerar as subpartes que possui. Os *frames Committing\_crime*, *Crime\_investigation* e *Criminal\_process* são seus subframes. Essa relação prevê que os Elementos do *frame* mais amplo, chamado *frame* mãe, sejam encontrados também nos subframes, os *frames* filhos. Entretanto, isso não é obrigatório, pois pode haver EFs no *frame* mãe que não se instanciam no filho e vice-versa.

Um bom exemplo da relação de subframe é apresentado pelo *frame Criminal\_process*. Como se sabe, o processo criminal é compreendido de várias partes, de modo geral, ocorre a detenção do suspeito (*Arrest*); uma acusação formal é apresentada ao réu (*Arraignment*); há o julgamento (*Trial*); pode ser solicitado um recurso (*Appeal*); e, por último, há a sentença final da acusação (*Sentencing*). Além da noção de precedência, que mostra que os *frames* obedecem a uma sequência temporal, com a ressalva de o recurso (*Appeal*) não ser obrigatório para as outras etapas acontecerem, todos são subpartes da cena de processo criminal (*Criminal\_process*), que é parte de Cenário\_criminal (*Crime\_scenario*).

Outro evento complexo que a FrameNet analisa é *Visiting\_Scenario*, ilustrado na Figura 4<sup>4</sup>. Conforme sua definição, um agente e uma entidade passam a estar no mesmo local por algum propósito específico. Geralmente, a visita dura um curto espaço de tempo, logo após, o agente retorna para o ponto de partida. A organização desse cenário é semelhante ao *frame Crime\_scenario*, pois é dividido em partes para representar esse evento. Os subframes *Visiting\_scenario\_arrival*, *Visiting\_Scenario\_stay* e *Visiting\_scenario\_departing* assinalam isso ao mostrar que a cena pode ser subdividida em etapas conforme a sequência temporal das ações: a chegada, a permanência e o retorno daquele que realiza a visita.

<sup>4</sup> Conferir em <https://framenet.icsi.berkeley.edu/fndrupal/FrameGrapher> todas as relações que *Visiting\_scenario* estabelece com outros *frames*.



Com o intuito de obter mais informações sobre a organização do cenário turístico, vários dados fornecidos pelos *corpora* estão sendo estudados. As observações que já podem ser sistematizadas mostraram a necessidade de se criarem outros *frames*, já que *Touring* parte do ponto de vista do turista, o agente da ação.

Sentenças como:

- (6) [No Rio de Janeiro LUGAR], [a paisagem belíssima ATRAÇÃO] atrai [turistas de todo o mundo TURISTA]
- (7) [In Rio de Janeiro PLACE], [the beautiful landscape ATRACTION] attracts [tourists from all over the world TOURIST]
- (8) [En Río de Janeiro LUGAR], [la belleza del paisaje ATRACTION] atrae [a turistas de todo el mundo TURISTA],

evidenciaram que, apesar de o turista ser inferido, ele não é o foco do evento. A perspectiva desse tipo de sentença está na atração turística. Outras ULs que reforçaram essa possibilidade foram *atrair*, *oferecer* e *apresentar*. Para esse *frame*, convencionou-se, então, *Atração\_turística* (*Tourist\_attraction*). Outro conjunto de dados mostrou que o turista sequer é inferido. Sentenças que revelaram isso são do tipo:

- (9) [Olinda LUGAR] possui [museus, maravilhosos monumentos e igrejas históricas ATRAÇÃO].
- (10) [Olinda PLACE] has [museums, beautiful monuments and churches ATRACTION].
- (11) [Olinda LUGAR ] tiene [museos, hermosos monumentos e iglesias ATRACCIÓN].

Esses dados também devem ser destacados, já que foram encontrados com representatividade nos *corpora*. Esse *frame* foi definido como *Atração\_em\_lugar* (*Attraction\_in\_place*). Outra UL que o evoca é *possuir*.

A criação desses dois *frames*, somados ao *Touring*, *Turismo\_pelo\_turista* como será chamado para o português, previamente descrito na FN, mostrou que as atrações turísticas podem ser apresentadas de formas distintas para o turista, ou seja, através de diferentes perspectivas.

A metodologia demonstrada acima para a construção de *frames* é *bottom-up*, ou seja, parte-se da análise dos dados para a constituição dos *frames*. A análise do *corpus* é crucial para que um *frame* seja proposto, pois possibilita conhecer o comportamento semântico e sintático dos elementos que acompanham a palavra-alvo antes que se façam generalizações.

## Considerações

O evento turístico exhibe grande complexidade, mantém contato com áreas distintas e também complexas como o comércio e a infraestrutura. As análises apresentadas se

encontram em estágio inicial. O objetivo deste trabalho foi mostrar a proposta de sistematização que está sendo realizada. Pretende-se para as próximas pesquisas obter mais avanços no desenvolvimento do Cenário\_do\_Turismo (*Tourism\_Scenario*).

Referências:

CHOMSKY, N. Language and Problems of Knowledge. The Managua Lectures, 1987

FAUCONNIER, G. *Mental Spaces*. Cambridge: Cambridge University Press, 1994.

FILLMORE, C. J.. Frame semantics. *Linguistics in the Morning Calm*. Seoul, South Korea: Hanshin Publishing Co. 1982

KOCH, I. V.; CUNHA LIMA, M. L. Do Cognitivismo ao Sóciocognitivismo. In: Anna Christina Bentes; Fernanda Mussalin. (Org.). *Introdução a Linguística – Fundamentos Epistemológicos*. Vol 3.1 ed. São Paulo, 2004, v. 3, p. 251-300.

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 15/06/12;

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