

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE DE LETRAS
PROGRAMA DE PÓS-GRADUAÇÃO EM LINGUÍSTICA

Natália Duarte Marção

**AS CONSTRUÇÕES INTERROGATIVAS QU- NO CONSTRUCTICON DA
FRAMENET BRASIL**

Juiz de Fora
2018

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE DE LETRAS
PROGRAMA DE PÓS-GRADUAÇÃO EM LINGUÍSTICA

**AS CONSTRUÇÕES INTERROGATIVAS QU- NO CONSTRUCTICON DA
FRAMENET BRASIL**

Natália Duarte Marção

Dissertação de Mestrado apresentada ao programa de Pós-Graduação em Linguística da Faculdade de Letras da Universidade Federal de Juiz de Fora, como parte dos requisitos necessários à obtenção do título de Mestre em Linguística.

Orientador: Professor Doutor Tiago Timponi
Torrent

Juiz de Fora
Setembro de 2018

Ficha catalográfica elaborada através do programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Marção, Natália Duarte.

As Construções Interrogativas QU- no Constructicon da FrameNet Brasil / Natália Duarte Marção. -- 2018.

126 f.

Orientador: Tiago Timponi Torrent

Dissertação (mestrado acadêmico) - Universidade Federal de Juiz de Fora, Faculdade de Letras. Programa de Pós-Graduação em Linguística, 2018.

1. Gramática das Construções. 2. Construções QU-. 3. Chatbot. 4. FrameNet Brasil. I. Torrent, Tiago Timponi, orient. II. Título.

AS CONSTRUÇÕES INTERROGATIVAS QU- NO CONSTRUCTICON DA
FRAMENET BRASIL

Natália Duarte Marção

Orientador: Prof. Dr. Tiago Timponi Torrent

Dissertação de Mestrado submetida ao programa de Pós-Graduação em Linguística da Faculdade de Letras da Universidade Federal de Juiz de Fora, como parte dos requisitos necessários à obtenção do título de Mestre em Linguística.

Aprovada por:

Presidente, Prof. Dr. Tiago Timponi Torrent – UFJF

Profa. Dra. Priscilla Mouta Marques – UFRJ

Prof. Dr. Ely Edison da Silva Matos– UFJF

Juiz de Fora
Setembro de 2018

*“Como é vão sentar-se e escrever se
você não se levantou para viver!”*

A minha mãe, meu porto seguro.

AGRADECIMENTOS

Em primeiro lugar, venho agradecer a Deus por me conceder sabedoria e força nos momentos mais difíceis dessa longa caminhada.

À minha família por todo apoio. Agradeço especialmente à minha mãe, Alci Luzia, por ter me dado toda a base para me tornar a mulher que sou, por ter lutado ao meu lado e me amparado em meus momentos de fraqueza. Agradeço ainda ao meu irmão por me fazer encarar a vida com mais leveza e a minha sobrinha pelos momentos de alegria.

À Universidade Federal de Juiz de Fora, instituição que me proporcionou essa realização. Agradeço também ao Programa de Pós-Graduação em Linguística por dar suporte ao meu desenvolvimento intelectual e acadêmico e por me fornecer todo suporte no decorrer da pós-graduação. Agradeço, em especial, às professoras Paula Armelin e Mercedes Marcilese, pelas indicações de leitura.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CAPES, por ter me auxiliado, com a concessão da bolsa de mestrado, na realização deste trabalho.

Ao meu estimado orientador, Tiago Torrent, toda a minha gratidão pela atenção, dedicação e, principalmente, por toda paciência. Agradeço por incentivar e estimular meu crescimento acadêmico e por todos os ensinamentos.

Agradeço a todos que fizeram e fazem parte da equipe da FrameNet Brasil pela amizade, pela parceria na realização das atividades e por todo aprendizado compartilhado. Em especial, agradeço ao Ely, por colaborar com o desenvolvimento dos experimentos computacionais.

E finalmente agradeço aos meus amigos, pela amizade e parceria.

RESUMO

O trabalho aqui apresentado insere-se no projeto Multilingual Knowledge Base ou simplesmente m.knob (www.mknob.com), o qual tem por objetivo criar um aplicativo web e vem sendo desenvolvido no Laboratório FrameNet Brasil de Linguística Computacional – FN-Br – (SALOMÃO, 2009). De forma mais geral, a FN-Br vem explorando a implementação das teorias da Semântica de Frames e da Gramática de Construções através da criação de recursos linguísticos computacionais, como o Lexicon e o Constructicon do Português Brasileiro (PB) – um repertório de construções disponível *online*. Dessa maneira, esta dissertação busca embasamento em pressupostos teóricos relacionados à Linguística Cognitiva, tais como a Semântica de Frames (FILLMORE, 1982) e a Gramática das Construções de Berkeley (KAY & FILLMORE, 1999) e segue a metodologia de análise da Berkeley FrameNet (FILLMORE ET AL. 2003). Nesse contexto, o objetivo dessa dissertação é apresentar a descrição e a modelagem linguístico-computacional (cf. DIAS-DA-SILVA, 1996) das construções Interrogativas QU- em PB na base construcional da FrameNet Brasil, a qual sustenta o aplicativo web m.knob. A pesquisa se justifica pelo fato de o *chatbot* – uma interface de recomendação de atrações turísticas baseada em compreensão de língua natural – empregado no aplicativo não suportar que o usuário interaja através de sentenças interrogativas. Nesse sentido, a descrição e modelagem das construções QU- se faz necessária de modo a contribuir para que o usuário, durante a interação com o aplicativo, busque informações adicionais acerca das atrações recomendadas através de perguntas. Para fomentar essa funcionalidade nova, esta dissertação traz uma proposta de modelagem de onze construções QU- em PB. Ao final, aplica-se um teste de prova de conceito para avaliação do modelo proposto.

Palavras-chave: Gramática das Construções; Construções QU-; Chatbot; FrameNet Brasil.

ABSTRACT

This work is part of the Multilingual Knowledge Base project or simply m.knob (www.mknob.com). It aims to create a web application and has been developed at the FrameNet Brasil Laboratory of Computational Linguistics - FN-Br - (SALOMÃO, 2009). More generally, FN-Br has been exploring the implementation of Framing Semantics and Construction Grammar theories through the creation of computational linguistic resources, such as the Lexicon and Constructicon of the Brazilian Portuguese (PB) - a repertoire of constructions available online. In this way, this Master's thesis is based on the theoretical assumptions of Cognitive Linguistics, such as Frame Semantics (FILLMORE, 1982) and the Berkeley Constructions Grammar (KAY & FILLMORE, 1999). Our analytical methodology is the one by Berkeley FrameNet (FILLMORE et al., 2003). In this context, this thesis aims to present the description and the linguistic-computational modeling (cf. DIAS-DA-SILVA, 1996) of the Interrogative Wh-constructions of the Brazilian Portuguese in the constructional base of FrameNet Brasil, which supports the web application m.knob. The research is justified by the fact that the chatbot - a tourist attraction recommendation interface based on natural-language understanding - employed in the application does not support the user interacting through interrogative sentences. In that sense, the description and modeling of Wh-constructions are necessary in order to help the user, during interaction with the application, look for additional information about the recommended attractions through questions. To foster this new function, this thesis presents a proposal for modeling eleven Wh-Constructions in Brazilian Portuguese. At the end, a proof of concept test is applied to evaluate the proposed model.

Keywords: Construction Grammar; WH- Constructions; Chatbot; FrameNet Brasil.

SUMÁRIO

INTRODUÇÃO	11
1 GRAMÁTICA DE CONSTRUÇÕES APLICADA À COMPREENSÃO DE LÍNGUA NATURAL	15
1.1 Compreensão de Língua Natural	15
1.2 A FrameNet Brasil	26
1.3 Gramática de Construções Baseadas em Unificação	34
1.4 O Constructicon da FrameNet Brasil.....	43
2 DESENVOLVIMENTO DE CHATBOTS	50
2.1 Conceitos Fundamentais.....	50
2.2 A Plataforma m.knob.....	57
3 METODOLOGIA	63
3.1 Modelagem das Construções Interrogativas QU.....	63
3.2 Desenho Experimental.....	65
4 MODELAGEM DAS CONSTRUÇÕES INTERROGATIVAS QU NO CONSTRUCTICON DA FRAMENET BRASIL	70
4.1 Propostas anteriores para a descrição das construções QU em Português do Brasil	70
4.1.1 <i>Abordagens Gerativistas para as Construções QU</i>	70
4.1.2 <i>Abordagens Funcionalistas para as Construções QU</i>	74
4.2 Proposta Construcionista para as Construções QU	79
4.3 Proposta de Modelagem das Construções Interrogativas QU- no Constructicon da FrameNet Brasil	82
5 ANÁLISE DE PROVA DE CONCEITO DA PERFORMANCE DO CHATBOT NA PLATAFORMA MKNOB	108
5.1 Resultados	108

5.2 Discussão	116
6 CONCLUSÕES	120
REFERÊNCIAS.....	122

INTRODUÇÃO

Esta pesquisa insere-se no projeto m.knob – Multilingual Knowledge Base –, o qual vem sendo desenvolvido no Laboratório FrameNet Brasil de Linguística Computacional – FN-Br – (SALOMÃO, 2009). O m.knob é um assistente pessoal de viagem desenvolvido em plataforma *mobile* e tem por objetivo auxiliar turistas que viajam pelo Brasil. Uma versão de teste do aplicativo foi lançada nas Olimpíadas do Rio, em 2016, e pode ser acessada em mknob.com. A versão beta encontra-se disponível na PlayStore (<https://play.google.com/store/apps/details?id=br.ufjf.framenetbr.mknob>).

A FN-Br visa a explorar a semântica das línguas naturais, com base em propostas teóricas relacionadas à Linguística Cognitiva, tais como a Semântica de Frames (FILLMORE, 1982) e a Gramática das Construções de Berkeley (KAY & FILLMORE, 1999). De forma mais geral, vem explorando a implementação dessas teorias através da criação de recursos linguísticos computacionais, como o Lexicon e o Constructicon.

O Lexicon da FN-Br compõe-se de um recurso *online* que articula frames e Unidades Lexicais (ULs) – pareamentos de um lema a um frame –, sustentado por evidência em *corpora*. Esse recurso tem por objetivo identificar, descrever e traduzir frames, analisar o significado das ULs a partir dos frames que elas evocam e as propriedades sintáticas delas para, assim, realizar a anotação com base no frame evocado e na análise sintática realizada (FILLMORE ET AL., 2003).

O Constructicon pode ser definido como um repositório de construções gramaticais do Português Brasileiro que visa a registrar as propriedades semânticas e gramaticais das construções, além de ligá-las a sentenças anotadas que exemplifiquem a construção sendo descrita, sem, necessariamente, comprometer-se com o objetivo de ser um registro eletrônico completo de todas as construções da língua (TORRENT ET AL., 2017).

Nesse contexto, a pesquisa enquadra-se em uma abordagem linguístico-computacional e tem por objetivos descrever as construções

Interrogativas QU- e fornecer uma proposta de modelagem para o tratamento das construções Interrogativas QU- do Português Brasileiro (PB) de modo a contribuir no desenvolvimento do *chatbot* de recomendação de atrações turísticas, uma das funcionalidades do aplicativo m.knob. Como a base construcional da FrameNet Brasil sustenta o m.knob, a pesquisa se justifica pelo fato de o *chatbot* empregado no aplicativo não suportar que o usuário interaja através de sentenças interrogativas. Nesse sentido, a descrição e modelagem das construções QU- se faz necessária de modo a contribuir para que o usuário, durante a interação com o aplicativo, busque informação adicional através de perguntas acerca das atrações recomendadas.

Castilho (2010) propõe que uma construção interrogativa QU- pode ser definida como uma sentença na qual os pronomes interrogativos QU- ocupam o lugar sintático da informação que se procura obter. Perini (2010) destaca que os interrogativos QU- podem aparecer nas seguintes formas: (o) **que, quem, qual, quanto, quando, onde, como** e incluem formas preposicionadas, como **por que**. O autor destaca ainda que esses interrogativos QU- podem aparecer no início da sentença, mas não necessariamente, assim, eles podem aparecer na posição de complemento.

Cabe aqui ressaltar que embora as construções Interrogativas QU- tenham uma composição sintática interessante e maleável, visto que o pronome QU- pode se apresentar no início da sentença ou *in situ*, nos restringimos apenas às construções que apresentam o pronome QU- em início de sentença.

Nosso objeto de pesquisa inclui sentenças do tipo das listadas de (1) a (9), todas extraídas dos *corpora* da FN-Br. Nossa pesquisa inclui sentenças interrogativas compostas por diferentes tipos de estrutura argumental, contudo, um estudo mais aprofundado sobre esses tipos não caberia na extensão de uma dissertação de mestrado como esta. Dessa forma, a presente pesquisa abre portas para trabalhos futuros que possam suportar a análise dessas estruturas argumentais.

- (1) **Qual** a melhor região de Manaus para se hospedar e a que fica mais próxima do aeroporto?
- (2) **Como** chegar em Barreirinhas?
- (3) **Quanto** custa viajar para Fernando de Noronha?
- (4) **O que** fazer no Rio além de visitar o Pão de Açúcar e o Cristo?
- (5) **Quem** disse que albergue é coisa de mochileiro?
- (6) **Quando** ir à Chapada Diamantina?
- (7) **Onde** ficar nos Lençóis Maranhenses?
- (8) **Por que** se hospedar em Olinda?
- (9) **Por quanto** tempo a pré-autorização vai afetar o meu crédito disponível?

Nesse contexto, o maior desafio é modelar as construções QU-linguística e computacionalmente de modo a possibilitar uma interpretação adequada pela máquina e assim garantir uma resposta mais adequada às sentenças descritas acima. É nesse sentido que a pesquisa se encaminha.

Esta dissertação de mestrado está organizada da seguinte forma: no capítulo um, apresentamos a FrameNet Brasil, bem como os pressupostos teóricos utilizados em nossa pesquisa através de uma descrição da abordagem das Gramáticas das Construções baseadas em unificação. Mostramos como funciona o Constructicon, a base construcional da FN-Br, e justificamos a inserção desta pesquisa nos estudos sobre Compreensão de Língua Natural. O capítulo dois consiste na apresentação das bases fundamentais para constituição de *chatbots* e do aplicativo m.knob. No capítulo três, trazemos a metodologia experimental por nós utilizada para a descrição e modelagem das construções Interrogativas QU- e para o teste de prova de conceito realizado para o modelo construcional proposto a partir de uma das onze construções modeladas. No capítulo quatro, descrevemos o tratamento das construções QU- sob a ótica das abordagens gerativistas, funcionalistas e construtivistas, bem como apresentamos uma proposta de modelagem para as construções QU-. Já o capítulo cinco traz a compilação dos resultados obtidos e a

discussão desses resultados acerca do teste realizado. Por fim, apresentamos as conclusões obtidas com esse trabalho.

1 GRAMÁTICA DE CONSTRUÇÕES APLICADA À COMPREENSÃO DE LÍNGUA NATURAL

Esta pesquisa ancora-se principalmente nas Gramáticas de Construções Baseadas em Unificação e na aplicação destas no Constructicon da FrameNet Brasil, além de abordar aspectos da área de Compreensão de Língua Natural. Assim sendo, nesta seção, apresentam-se os pressupostos teóricos da pesquisa.

1.1 Compreensão de Língua Natural

Os primeiros trabalhos voltados ao desenvolvimento de computadores capazes de “pensar” começaram a ser desenvolvidos nos anos 1950 impulsionados pela questão “Podem as máquinas pensar?”, levantada no texto *Computing, Machinery and Intelligence* (1950) pelo matemático britânico Alan Turing. Esse texto inaugura a possibilidade de se dar às máquinas habilidades para interagirem com o ser humano em linguagem natural.

Esse questionamento da possibilidade de as máquinas pensarem foi despertado por um tipo particular de máquina, o computador. A partir disso, cabe, primeiramente, analisar o que significa pensar. No dicionário Aurélio (2010) palavra “pensar” tem as seguintes acepções:

1. Pensamento; opinião; juízo
2. Imaginar, julgar
3. Planear
4. Dar penso a
5. Tratar convenientemente
6. Fazer curativo
7. Formar ideias
8. Refletir
9. Raciocinar
10. Ser de parecer
11. Tensionar
12. Ter no pensamento

Podemos observar que o ato de pensar é extremamente complexo para afirmarmos que o computador pensa, pois estes não são capazes (ainda) de formar ideias, julgar, opinar. Eles apenas executam tarefas. Assim, Turing (1950) propõe que as máquinas não precisam efetivamente pensar, mas sim imitar tão bem um humano de modo que possam se passar por um. Nos termos de Turing (1950), um computador será capaz de “pensar” se ele for bem-sucedido no chamado Jogo da Imitação, o qual ficou conhecido como

Teste de Turing e é usado até hoje para descobrir o nível de uma inteligência artificial.

O Jogo da Imitação proposto pelo matemático britânico funciona da seguinte maneira: uma pessoa interage através de perguntas escritas com duas entidades ocultas, sendo uma humana e outra um computador. Analisando as respostas dadas pelas entidades, a pessoa tentará decidir qual das duas é humano. A inteligência artificial é definida como “inteligente” caso essa pessoa não consiga diferenciar se o que foi dito pelo sistema foi elaborado por outro ser humano ou se veio de uma máquina.

Cabe aqui, portanto, diferenciar inteligência humana de inteligência artificial. A Linguística Cognitiva, de um modo geral, propõe que a língua reflete propriedades fundamentais da mente humana, assim, adota a hipótese da continuidade por trás da diversidade dos processos cognitivos (FAUCONNIER & TURNER, 2002), ao afirmar que processos como analogia, compressão para escala humana e mesclagem encontram-se no centro tanto do processamento linguístico quanto naquele realizado por outros sistemas, como o perceptual e o motor. Desse modo, a inteligência humana estaria associada diretamente, mas não só, à capacidade que os humanos têm de se comunicarem através da linguagem natural e de atribuir significado às coisas no mundo. Além disso, segundo Fauconnier & Turner (2002), somente os seres humanos desenvolveram uma linguagem mediada por uma gramática, de modo que nossa arquitetura cognitiva seja bem distinta dos demais seres.

Nesse sentido, a linguagem seria a expressão concreta, mas não completa, de processos cognitivos abstratos que ocorrem na comunicação. Assim, ainda que o sentido não seja totalmente contido nas palavras e nas construções gramaticais, são elas que fornecem pistas para a construção do sentido. Na construção do sentido, a gramática tem um papel fundamental porque é a ligação visível entre os mistérios da cognição e o comportamento superficial aparente do pensamento humano (FAUCONNIER & TURNER, 2002). Este último, entretanto, é complexo, intrincado e capaz de se redirecionar de forma instantânea, dada uma situação motivadora.

Por sua vez, o termo Inteligência Artificial (IA) foi cunhado pelo cientista da computação John McCarthy no ano de 1956. A inteligência artificial é um ramo da Ciência da Computação cujo interesse é fazer com que os computadores pensem ou se comportem de forma inteligente. Santos (2006), no mesmo sentido de Turing (1950), aponta que, como os computadores não podem pensar efetivamente, o que pode ser feito é programá-los para que se comportem como se estivessem pensando, simulando processos cognitivos. Assim, os computadores ditos como “inteligentes” são na verdade baseados em simulações computadorizadas da inteligência humana, ou seja, ainda fundados em um “Jogo da Imitação”.

Um dos maiores desafios da interação entre homem e máquina é o de torná-la mais intuitiva e natural, ou seja, de fazer com que ela tenha um bom desempenho no jogo. Através do tratamento computacional das línguas humanas, é possível aumentar a impressão do usuário de que está, de fato, interagindo com um ser pensante. Diante disso, a busca em desenvolver programas capazes de “compreender” línguas humanas vem aumentando cada vez mais, impulsionando as pesquisas de Processamento de Língua Natural (PLN) relacionadas à Linguística Computacional e à Inteligência Artificial (IA).

O PLN é um ramo da Ciência da Computação que busca desenvolver e implementar “sistemas computacionais capazes de processar objetos de natureza linguística” (DIAS-DA-SILVA, 1996, p. 12), isto é, o PLN tem por objetivo criar sistemas computacionais que permitam que a comunicação entre homem e máquina possa se dar através de uma língua natural, e não pelo uso de comandos de uma linguagem de programação construída artificialmente (DIAS-DA-SILVA, 1996).

De uma forma geral, o PLN propõe sistemas que procuram lidar com interações feitas em língua natural entre seres humanos e máquinas ao passo que a CLN engloba desafios mais complexos da Inteligência Artificial (IA), uma vez que procura desenvolver sistemas computacionais que sejam capazes de lidar com a compreensão da linguagem humana como um todo, incluindo suas peculiaridades como frases mal formadas, ambiguidades, coloquialismos entre

outras, convertendo essas entradas não estruturadas em uma forma estruturada de modo que uma máquina possa entender e criar uma resposta adequada também em língua natural.

A principal diferença entre PLN e CLN é que a primeira abordagem é de cunho estruturalista e seu foco se concentra nas especificações formais das estruturas linguísticas, tanto das regras de construção quanto das restrições de constituição de estruturas possíveis. Já a segunda se preocupa com aspectos cognitivos e psicolinguísticos da linguagem, levando em consideração não só as representações das estruturas linguísticas, mas também as representações dos processos linguísticos realizados na linguagem. Desse modo, a CLN visa mapear enunciados em algum tipo de representação semântica que modela seu significado de maneira que um computador possa processar e entender (ALLEN, 1985).

Existem duas motivações principais para o desenvolvimento da CLN, segundo Allen (1985): uma é a motivação tecnológica, uma vez que seu objetivo é construir sistemas computacionais inteligentes; a outra motivação é linguística (ou da ciência cognitiva), pois busca-se compreender melhor como os seres humanos processam e compreendem a língua, ou seja, como os humanos usam a língua.

Para ficar mais claro, Allen (1985) toma como exemplo o programa de inteligência artificial ELIZA. Esse programa foi desenvolvido para simular um terapeuta na interação com um humano que seria o paciente. ELIZA possui um banco de dados composto por palavras-chaves e, para cada palavra-chave, o sistema armazena um repertório fixo de padrões, os quais codificam as informações de entrada e são “casados” às especificações de saída (cf. DIAS-DA-SILVA, 1996, p. 73, para uma definição de casamento de padrões). A popularização de ELIZA e de programas similares levou ao surgimento do termo "efeito ELIZA", descrevendo a tendência das pessoas de atribuir inteligência humana a programas de computador. Na verdade, um diálogo típico com ELIZA pode parecer bastante impressionante, conforme o trecho no Quadro 1, reproduzido por Allen (1985, p.3):

Quadro 1: Diálogo entre um usuário e ELIZA

USER: Men are all alike
ELIZA: In what way?
USER: They're always bugging us about something or other
ELIZA: Can you think of a specific example?
USER: My boyfriend made me come here
ELIZA: Your boyfriend made you come here
USER: He says I'm depressed much of the time
ELIZA: I'm sorry to hear you are depressed

Contudo, conforme aponta Allen (1985, p.4) “Eliza parece ser inteligente porque usamos nossa própria inteligência para entender o que ela diz”, isto é, o usuário humano é que atribui conhecimento ao diálogo do qual faz parte. Isso porque o segredo de ELIZA está em fornecer como respostas perguntas genéricas ou sentenças que repetem parte do que foi digitado pelo usuário em uma fórmula padrão. Além disso, nesse processo de conversão de dados em padrões, perdem-se informações semânticas e contextuais importantes. Diante disso, podemos dizer que o programa não entende a conversa da qual faz parte, uma vez que as informações de entrada e saída estão condicionadas ao casamento de padrões.

De forma geral, a CLN procura preencher as lacunas da análise semântica no PLN e propõe que um programa de IA que desenvolva compreensão de língua natural tenha como objetivo real criar mecanismos para entender o significado do texto (oral ou escrito) de modo a gerar uma resposta compreensível. Para tanto, é necessário que esse tipo de programa dê conta de aspectos da linguagem e da cognição humana, como conhecimento de mundo e habilidades de raciocínio que atuam na produção e na compreensão de estruturas linguísticas. Além disso, é necessário, também, considerar aspectos contextuais (ALLEN, 1985).

De modo a integrar esses aspectos no desenvolvimento de programas computacionais baseados em língua natural, Allen (1985) propõe uma abordagem de CLN dividida em três subáreas: estrutura linguística,

conhecimento de mundo e contexto situacional que remetem aos processos formais – fonológicos, morfológicos e sintáticos –, semânticos e pragmáticos, respectivamente. Assim, o processo sintático envolve análise de sentenças em língua natural e produz a representação dessas estruturas sintáticas. O sistema então usa essas estruturas como entrada na fase de interpretação semântica, a qual produz uma forma lógica que, por conseguinte, é utilizada como entrada para a fase de representação contextual, a qual desenvolve a representação final das sentenças.

Para interpretar de forma automática os processos formais, são usados *parsers* que, nas palavras de Almeida (2016), são programas de computador capazes de interpretar automaticamente aspectos fonológicos e morfossintáticos de palavras e expressões da estrutura linguística em uma dada língua e atribuir estrutura de constituintes e de dependência desses constituintes com base em modelos gramaticais. Existe o *parser* que realiza a análise sintática automática de frases em termos de suas funções gramaticais e a saída então será uma árvore sintática e um outro tipo é o *parser* por dependência, o qual estabelece a relação de dependência entre os constituintes da sentença analisada (VARGAS & KEPLER, 2012).

O *parser* sintático atua no nível da estrutura e extrai informações sintáticas relevantes. Allen (1985) afirma que a estrutura sintática de uma sentença indica como as palavras se relacionam umas com as outras na sentença, como as palavras são agrupadas em sintagmas e que tipo de relação existe entre os sintagmas que formam a sentença. Além dessas informações, o autor defende que a representação sintática extraída do processo de *parsing* poderia apontar sentenças ambíguas sintaticamente, como em (10), e sentenças mal formadas, como em (11).

(10) a. Flying planes are dangerous.

b. Flying planes is dangerous.

(11) * John put the book.

Em (10), a ambiguidade reside no fato de que tanto o sintagma nominal “Flying planes”, em que “flying” é adjetivo de “planes”, e o sintagma verbal “Flying planes”, em que “flying” toma “planes” como complemento, são interpretações sintáticas possíveis para a sentença. Olhando com mais atenção a sentença (11), ela é considerada mal formada pois falta um argumento do verbo *put* o qual representaria o lugar onde o livro foi posto, entretanto, a seleção de argumentos verbais pode ser encarada como um fato do nível semântico e não sintático, assim não é possível dizer com clareza que há uma má formação semântica ou sintática.

A grande questão da CLN está na análise semântico-pragmática, isto é, concentra seus esforços na análise do contraste entre os aspectos do significado que são determinados apenas pelo signo linguístico, por um lado, e os aspectos do significado que são particulares de um contexto de uso (BENDER ET AL., 2015). O ponto principal dessa análise consiste em definir como as representações do conteúdo linguístico serão feitas, logo deve-se considerar a questão sobre o que deve ser uma representação adequada do significado linguístico, quais informações devem ser carregadas e como elas podem ser construídas. Essa representação submete-se à teoria do significado linguístico que será subjacente ao modelo. Dessa maneira, para tornar a transição para CLN possível, uma teoria deve ter uma rica representação semântica que pode ser relacionada à forma de representar sentenças por um conjunto de regras ou princípios computacionalmente atraentes (VAN VALIN, 2017, p. 2).

Na abordagem adotada por Allen (1985), depois de extrair as propriedades sintáticas da sentença, o programa deve então considerar os significados individuais das palavras, derivados em formas lógicas, para então construir o significado geral da sentença, chamado de representação final do significado. Nessa abordagem, a representação semântica se dá unicamente através da forma lógica, assim, o processo semântico é simplesmente derivar uma forma lógica das sentenças. Para o autor, essa forma lógica é concebida, da mesma maneira que na teoria gerativa, como uma estrutura de

representação intermediária entre o que é dito de fato pelo falante, isto é, a forma de superfície e o que de conhecimento é gerado a partir do que foi dito. Dessa forma, o significado da sentença está condicionado ao contexto situacional e ao conhecimento de mundo do falante/ouvinte, o que Allen chama de representação final do significado.

Uma espécie de “composicionalidade modalizada” é o argumento usado por Allen (1985) para justificar a derivação do significado em forma lógica. A hipótese da composicionalidade propõe que o significado do todo é igual à soma do significado das partes. Desse modo, o significado de uma sentença simples como “Cachorros latem”, seria a soma do significado de “cachorros” e “latem”, ou seja, o significado das partes comporia o significado da frase como um todo. Ainda que se reconheça que não há como separar totalmente os níveis de análise, esse argumento é válido na medida que explorar cada nível separadamente pode ser bastante informativo. Ademais, Bender et al. (2015) defendem que a abordagem composicional pode ser positiva, pois suporta o desenvolvimento de representações mais abrangentes, capturando mais detalhes em cada camada considerada do que o produto de um sistema global que dê conta de tudo.

A forma lógica usa papéis temáticos na sua representação para etiquetar elementos de uma sentença de acordo com sua função semântica genérica, dessa maneira, os papéis temáticos apontam para uma generalidade semântica a qual é associada a cada posição sintática que o elemento ocupa na sentença (ALLEN, 1985). Então em uma sentença como

(12) John sold the book to Mary

o elemento *John* instancia o papel temático de AGENT, *the book* recebe papel temático de THEME e TO-POSS indica que a posse do tema foi transferida para o RECIPIENT *Mary*.

Com base na representação do significado em forma lógica, uma sentença como (12) poderia ser representada da seguinte maneira

(ASSERT (PAST e1 SELL [AGENT (DEF/SING j1 JOHN)]
 [THEME (DEF/SING b1 BOOK)]
 [TO-POSS (DEF/SING m1 MARY)])

A representação final é composta pela forma lógica, contexto linguístico e conhecimento de mundo. Para gerar a forma lógica para a representação final do significado, é usado o método básico de cálculo de predicado de primeira ordem. Existem duas formas de se realizar esse cálculo. O cálculo do predicado da sentença (12) poderia ser descrito da seguinte maneira:

SELL(JOHN22, BOOK67, MARY1)

O predicado é o verbo *sell* (vender), que pede três argumentos, que, no caso, são *John*, *book* e *Mary*. Os números indicam que, em um banco de dados, cada argumento foi cadastrado de modo que a representação destes seja única, assim, *John22* indica um *John* específico dessa lista previamente cadastrada.

Essa forma na qual se reconhece o predicado e logo se aponta diretamente quais são as entidades que ocupam os lugares argumentais desse predicado é uma representação simples, contudo não é genérica. Em um sistema que compreenda sentenças de um universo bem delimitado, bem modelado, as especificações geradas serão dadas na forma intermediária como descrita acima. Nesse contexto, o sistema terá uma representação unívoca do predicado, bem como dos seus argumentos. Entretanto, pode-se optar por uma representação de cálculo de predicado mais elaborada se o intuito for apontar para generalidades mais abstratas de tal sentença como podemos ver abaixo:

SELL-EVENT (E1) & SELLER (E1, JOHN22) & OBJECT (E1, BOOK67) &
 SELLEE (E1, MARY1)

Essa segunda representação traz informações adicionais que podem ser interessantes para a interpretação semântica em um sistema de CLN, pois caracteriza os papéis micro-temáticos, isto é, traz especificações dos papéis semânticos mais abrangentes ao categorizar o predicado *sell* em relação a tipologia dos predicados, dizendo que *sell* é um **evento** (*event*), no caso o único evento, visto que pode ocorrer mais de um evento em uma sentença, no qual o argumento *John*²² é o agente da ação, caracterizado como **vendedor** (*seller*), que transfere a posse do **objeto** *book*⁶⁷ para a **compradora** (*sellee*) *Mary*¹. Essas informações semânticas são relevantes para o sistema de CLN porque em uma sentença como *John sold the book to Mary to pay for his bills* é necessário categorizar *John* tanto como **seller** no primeiro evento-vender, quanto como **payer** no segundo evento-pagar. Sendo assim, classificar o evento é importante para processar os vários papéis de *John* na sequência textual em que este aparece. Note que os papéis formalizados dessa maneira são caracterizados por dois traços sendo um o evento do qual eles fazem parte e o outro diz respeito a que elemento ocupa aquele espaço no evento, isto é, quem preenche determinado espaço no evento do qual faz parte.

A abordagem proposta para CLN discutida aqui apresenta alguns problemas e por isso não é a mais indicada para o propósito do nosso trabalho. O primeiro problema que podemos apontar é o fato de Allen (1985) separar forma lógica da representação semântica final, pois isso é separar problemas com propriedades indissociáveis, visto que existem padrões sintáticos regulares que servem justamente para apontar um significado baseado no conhecimento de mundo e não na forma lógica. Logo uma sentença como *O que esta mosca está fazendo na minha sopa?* não pode receber uma resposta como *Está nadando de costas*, a qual seria uma resposta aceitável levando em conta o significado pautado em uma abordagem composicional, contudo, quando se diz uma frase como essa não se espera a resposta padrão, mas sim uma resposta baseada no conhecimento de mundo (cf. KAY & FILLMORE, 1999, para uma discussão da construção *What's X doing Y?*).

Outro problema inerente à abordagem é considerar conhecimento de mundo, também chamado de conhecimento enciclopédico, e contexto como sendo a mesma coisa, dado que são bem diferentes, e, para um sistema de CLN funcionar adequadamente, é necessário separar esses níveis. O conhecimento de mundo pode ser enquadrado como um conhecimento socialmente compartilhado e, portanto, relativamente estável, pois é o conhecimento que os indivíduos vão adquirindo ao longo de sua vida em sociedade, em grande parte, de modo informal através das experiências com o mundo e com os demais indivíduos. Já o contexto é um enquadramento situacional no qual a sentença ocorre. É o ambiente imediato que possibilita interpretar de maneira adequada a sentença.

Olhando para uma sentença como *O Rubinho quebrou o carro de novo* (vide SAMPAIO, 2010, para uma análise das Construções de Argumento Cindido, que licenciam sentenças como esta), podemos associar o nome próprio Rubinho a uma pessoa em particular, no caso o piloto, e isso é realizado através do conhecimento de mundo, mas é apenas a capacidade de construir situacionalmente, isto é, pelo contexto, que vai permitir a leitura de que a ação de “quebrar” o carro não foi intencional. Ainda que o seu conhecimento de mundo possa indicar que a situação pudesse ser intencional se o Rubinho, como um piloto de Fórmula 1, tivesse quebrado de propósito o carro, o enquadre situacional vai excluir a intencionalidade nessa sentença. Essa proposta de que a forma lógica cria uma representação final de significado que unifica conhecimento de mundo com contexto não é adequada, porque, como vimos no exemplo, é possível reenquadrar o conhecimento de mundo de acordo com o contexto situacional da sentença.

Diante dos problemas expostos aqui acerca da abordagem proposta por Allen (1985) para tratar aspectos semânticos e pragmáticos da linguagem em sistemas de CLN, propomos que uma abordagem baseada em frames seja mais interessante para tratar tais aspectos linguísticos. Dessa forma, nosso objetivo é desenvolver um sistema de compreensão de língua natural pautado nas aplicações do projeto FrameNet Brasil (doravante FN-Br).

1.2 A FrameNet Brasil

O Laboratório FN-Br (SALOMÃO, 2009) vem se desenvolvendo em concomitância com a FrameNet iniciada no International Computer Science Institute (ICSI), por Charles Fillmore. A FN-Br propõe a criação de um recurso linguístico *online* para o Português do Brasil, fundamentado na Semântica de Frames e sustentado por evidência em *corpus*.

A FN-Br baseia-se na Semântica de Frames (FILLMORE, 1982), a qual considera que a compreensão dos sentidos não pode ser tratada como sendo um processo composicional, mas deve levar em consideração a importância do todo como algo além da mera soma das partes, destacando que há uma interdependência observada entre essas partes. A principal ideia subjacente a essa teoria é a de que os significados das palavras fornecem acesso a conhecimentos organizados, ou seja, frames. Nessa abordagem, o processo de significação das palavras está diretamente ligado à compreensão das instituições sociais ou da estrutura da experiência que elas pressupõem. Dessa forma, a Semântica de Frames incorpora ao objeto de estudo da semântica a forma como categorizamos a realidade como resultado de um processo semântico e cognitivo, baseado na nossa experiência empírica com o mundo.

Por ser uma abordagem para a semântica lexical que toma como pressuposto a máxima de que os significados são relativizados a cenas (FILLMORE, 1977), isto é, a frames, a Semântica de Frames afasta-se da tradição da semântica formal, ao passo que se aproxima da semântica empírica, uma vez que se baseia na continuidade entre linguagem e experiência (FILLMORE, 1982). Faz parte, portanto, das Semânticas do Entendimento (FILLMORE, 1985a), as quais se opõem à Semântica Vericondicional. Fillmore (1985a) argumenta que, enquanto a Semântica Vericondicional se interessa pelas condições de verdade a que uma dada sentença se submete, a Semântica do Entendimento se preocupa em analisar as escolhas linguísticas que fazemos para formar o enunciado de forma que

seja possível, através dele, transmitir nossas crenças sobre o mundo, nossas experiências e a maneira como vemos as coisas.

Essa teoria semântica baseada em frames se pauta na ideia de que

o significado de uma palavra pode ser entendido apenas em relação a um contexto estruturado de experiências, crenças ou práticas, constituindo uma espécie de pré-requisito conceitual para a compreensão do significado¹ (FILLMORE & ATKINS, 1992, p. 76-77).

Sendo assim, os significados são associados às nossas experiências particulares com o mundo e podem ser explicados a partir de estruturas complexas de conhecimento baseadas em cenas partilhadas socialmente, chamadas de frames (FILLMORE, 1985a). Diante disso, um frame pode ser entendido como um sistema de conceitos relacionado de tal maneira que “para entender um deles, é necessário entender toda a estrutura na qual ele se encaixa” (FILLMORE, 1982).

A partir desses pressupostos, a Semântica de Frames constitui-se como teoria básica para projetos em lexicografia que objetivam alinhar a Linguística Cognitiva à descrição da língua com base em uma perspectiva computacional, como vem sendo feito na FN-Br, isso porque, desde a junção da Semântica de Frames à lexicografia, a FN-Br, assim como as demais FrameNets existentes, tomou como propósito a constituição de uma base de dados disponível *online* com o objetivo de descrever os significados lexicais através dos frames, a partir de material extraído de corpus e da análise da valência das Unidades Lexicais (ULs) que evocam esses frames. Além da FrameNet americana e da brasileira, hoje, existem FrameNets que desenvolvem trabalhos em outras línguas como em espanhol, japonês, alemão, sueco, mandarim, francês e coreano. Atualmente, a FN-Br, de modo geral, vem dedicando seus esforços ao desenvolvimento de recursos linguísticos computacionais, como o Lexicon e o Constructicon.

¹“(...) a word’s meaning can be understood only with reference to a structured background of experience, beliefs, or practices, constituting a kind of conceptual prerequisite for understanding the meaning”.

Todas as traduções presentes nesta dissertação são de responsabilidade da autora.

O Lexicon é um recurso *online* que objetiva criar uma representação computacional de frames, definidos por seus participantes e instrumentos, conectados entre si via relações entre frames; definir ULs (Unidades Lexicais), pareamentos entre um lema e um significado definido em termos de um frame; e, por fim, anotar sentenças que exemplifiquem os padrões de valência sintáticos e semânticos em que as ULs ocorrem (ALMEIDA, 2016).

O Constructicon, por sua vez, tem por objetivo a criação de um recurso *online* para a descrição das características semânticas e gramaticais de construções do PB, incorporando descrições interpretáveis computacionalmente para cada construção, oferecendo informações semânticas e especificando as relações entre as construções (ALMEIDA, 2016). Almeida (2016) aponta que ambos os recursos encontram-se interligados, na medida em que tanto o Lexicon como o Constructicon foram desenhados para lidar com todos os traços linguísticos relevantes para a identificação de um frame e de uma construção. Trataremos de mais aspectos do Constructicon na seção 1.4.

Voltando nossa atenção para o Lexicon, o trabalho desenvolvido na FN-Br consiste em identificar e descrever frames, analisar o significado das palavras a partir dos frames que evocam e estudar as propriedades sintáticas das palavras verificando de que modo propriedades semânticas se manifestam na forma sintática. Existem duas unidades básicas de análise, o frame e a Unidade Lexical (UL), a qual consiste no pareamento de um lema a um frame. Dessa forma, se pensarmos no lexema “hotel”, podemos relacioná-lo ao frame de Acomodação, ou seja, uma estrutura de conceitos que prevê que um dado Alojamento possa ter uma Classificação – em número de estrelas, por exemplo –, ou ser de um determinado Tipo – como hotel fazenda ou hotel boutique. Cada um desses conceitos que compõem o frame é um Elemento de Frame (EF). Quando da instanciação de um frame, cabe à UL evocá-lo e, aos elementos que a circundam, instanciar os vários EFs.

Na FN-Br, a criação e a descrição de frames, EFs e ULs, bem como a anotação lexicográfica das sentenças é feita através da ferramenta WebTool

3.0. A Figura 1 mostra como o frame é apresentado na WebTool. Observa-se que, primeiramente, a descrição do frame é proposta. Em seguida, aparecem os EFs nucleares e os não nucleares, juntamente com suas definições. E, logo abaixo, os tipos de relações que o frame estabelece com outros frames e as ULs que o evocam.

Acomodação

Definição
Esta cena descreve estabelecimentos comerciais que oferecem serviços de acomodação para turistas. Os estabelecimentos costumam ser classificados quanto a seu tipo e/ou qualidade.

Exemplo(s)

Elementos de Frame Nucleares

FE Core:
Alojamento [Lodge] O estabelecimento comercial que oferece acomodação para turistas.

Elementos de Frame Não-Nucleares

Classificação [Rate] Este participante indica as características do **Alojamento** no que se refere a aspectos da acomodação, tais como luxo, preço e número de estrelas.

Finalidade [Purpose] Alguma ação específica possibilitada pelo **Alojamento**.

Lugar [Place] Local no qual se encontra o **Alojamento**.

Tipo [Type] Se refere às características do **Alojamento** que dizem respeito, geralmente, ao local em que ele se encontra (hotel fazenda ou hotel de selva, por exemplo), ou a outras especificidades não relacionadas ao luxo ou à qualidade, tais como ser ecologicamente correto ou ter valor histórico.

Turista [Tourist] Aquele que faz uso do **Alojamento**.

Relações

Herda de [Infraestrutura](#) [Infrastructure#](#) [Negócios](#) [Serviço_turístico](#)

É usado por [Estada](#)

Usa [Hospedar-se](#)

Unidades Lexicais

[hospedagem.n](#) [hóspede.n](#) [hostel.n](#) [hotel fazenda.n](#) [hotel.n](#) [hotalaria.n](#) [motel.n](#) [pensão.n](#) [pousada.n](#) [rancho.n](#) [resort.n](#) [sítio.n](#)

Figura 1: Frame de Acomodação

É possível, também, acessar as anotações lexicográficas das ULs evocadoras de frames na ferramenta. Apenas é anotada a sentença na qual a UL se encontra e qualquer outro termo que não esteja no escopo da sentença, ou que tenha sido omitido e seja recuperável pelo contexto é considerado conforme o que é exposto pela sentença da UL alvo, como é apresentado na Figura 2.

Lexical Entry

hotel.n
Frame: Acomodação

Definição
estabelecimento comercial que possui muitos quartos ou apartamentos mobiliados e estes são alugados por um preço diário.

Elementos de Frame e suas Realizações Sintáticas

Os Elementos de Frames do sentido desta palavra são (com as realizações):

Frame Element	Number Annotated	Realization(s)
● Alojamento	29	INC.-- (28) CNI.-- (1)
○ Lugar	19	INI.-- (2) DNI.-- (13) CNI.-- (1) ?.? (2) Dep.PP (2) Aposto.VPpart (1)
○ Tipo	6	Dep.AdJP (6)
○ Classificação	8	Dep.PP (3) Dep.AdJP (3) Dep.NP (1) Adjct.PP (1)

Sentences

Clear Sentences Turn Colors On/Off

- São centenas de **HOTEIS**, restaurantes, bares, lojas e etc. **INC INI**
- O **HOTEL** possui 30 quartos completamente equipados **INC DNI**
- Pequeno HOTEL** nas proximidades de Petropolis-RJ, com excelente acesso pela BR 040 a 90 km do Aeroporto Intenacional do RJ. Dois restaurantes oferecem opções na culinária mediterrânea, típica brasileira, fondues e racletes, além de uma seleta carta de vinhos **INC CNI**
- Para isso, precisamos ter aeroportos, estradas, um sistema de transporte público eficiente, **HOTEIS de todas as categorias...** **INC DNI**
- HOTEL urbano**, decoração temática com peças de artesãos da Região, localização central, perto de os principais bancos, restaurantes e boates **INC DNI**
- Com fácil acesso ao aeroporto internacional Afonso Pena, O **HOTEL** possui 75 espaçosos quartos equipados com ar condicionado **INC DNI**
- O Caesar Business Manaus Amazonas alia modernidade, qualidade de serviços e tecnologia no melhor **HOTEL Business Class**, idealizado para atender aos dinâmicos executivos que viajam a trabalho **INC DNI**
- O **HOTEL** tem dois restaurantes disponíveis **INC DNI**

Figura 2: Relatório da UL *hotel.n*

A FN-Br adota duas metodologias de anotação: a lexicográfica e a de texto corrido (SALOMÃO ET AL., 2013). Na anotação lexicográfica, primeiramente é feito o levantamento de ULs que evocam determinado frame. A partir disso, essas ULs são anotadas em camadas associadas a Elementos de Frame (EFs), a Funções Gramaticais (FG) e Tipos Sintagmáticos (TS).

Lage (2013) aponta que a camada Elemento de Frame (FE, na Figura 3) é usada para etiquetar os constituintes da sentença conforme os papéis semânticos que assumem em relação à UL. Cada EF é assinalado por uma cor para facilitar sua identificação. Na camada Função Gramatical (GF, na Figura 3) são assinaladas as classes gramaticais de cada constituinte da sentença alvo. Já na terceira camada é marcado o Tipo Sintagmático (PT, na Figura 3) no qual se encaixa cada EF. A classificação ocorre em função do sintagma em

Ao adotar a anotação em camadas e dispor EFs sobre função gramatical e tipo sintagmático, a FrameNet demonstra a base construcionista que permeia sua perspectiva teórico-metodológica.

Na segunda metodologia de anotação, de texto corrido, anotam-se todas as ULs evocadoras de frames que aparecem no texto selecionado. Assim, diferentemente da lexicográfica, a anotação de texto corrido não seleciona uma UL específica, então não há um frame específico a ser anotado. Dessa forma, como são várias ULs a serem anotadas, conseqüentemente, também serão numerosos os frames evocados.

A Figura 4 traz um exemplo de anotação de texto corrido.

Na anotação apresentada na Figura 4, podemos ver que, assim como na anotação lexicográfica, a anotação é feita em camadas, contudo como a sentença apresenta várias ULs evocadoras de frames, cada UL é anotada na ordem em que aparece formando assim “conjuntos de camadas, cada uma correspondendo a uma UL alvo” (FILLMORE, 2008).

Na Figura 4, é possível observar que ao lado de cada UL anotada (destacada em preto) vem o frame evocado por tal UL, como, por exemplo, a UL “oferece”, que evoca o frame de *Atrair_Turistas*. Em seguida, nas demais camadas, são anotados respectivamente os Elementos de Frame, Função Gramatical e Tipo Sintagmático.

The screenshot shows a software interface for text annotation. At the top, there is a menu bar with options: Save, Refresh, Hide AS, Remove AS, Label, Help. Below the menu, the text "[116175] Rio Branco oferece infraestruturas de hospedagem adequadas à recepção de turistas." is displayed. The interface is divided into six sections, each representing a different phrase from the text. Each section contains a grid with columns for syntactic frames (FE, GF, PT, Other, Noun, Sent) and rows for constituent labels (INI, Lugar, Ext, NP, Atracção, ObjD, NP, Descrição, Turista, Tipo, Finalidade, Dep, PP, AdJP, Adequadas, Item, Ext, NP, Situação, habilitada, Adequadas, Social_event#, recepção, n, Attendee, Dep, PP, Recepção, Turismo, turista, n, Lugar, Atracção, Descrição, Turista, NP, NP, AdJP). The labels are color-coded: INI (red), Lugar (green), Ext (green), NP (yellow), Atracção (blue), ObjD (red), NP (yellow), Descrição (purple), Turista (red), Tipo (green), Finalidade (purple), Dep (red), PP (red), AdJP (pink), Adequadas (black), Item (purple), Ext (green), NP (yellow), Situação, habilitada (blue), Adequadas (black), Social_event#, recepção, n (black), Attendee (blue), Dep (red), PP (red), Recepção (black), Turismo, turista, n (black), Lugar (green), Atracção (blue), Descrição (purple), Turista (red), NP (yellow), NP (yellow), AdJP (pink).

Figura 4: Anotação de Texto Corrido

A FN-Br além de lidar com fenômenos linguísticos no nível lexical e sintagmático, também lida com fenômenos linguísticos no nível construcional. Para tanto, esta adota uma abordagem construcional baseada em frames, a qual se mostra mais interessante aos nossos propósitos em desenvolver um recurso como o Constructicon para CLN. Essa abordagem construcional é o assunto da seção 1.4.

1.3 Gramática de Construções Baseadas em Unificação

Para o nosso trabalho, modelos construcionistas baseados em unificação mostram-se mais adequados, devido ao fato de que estruturas baseadas em Matrizes de Atributo e Valor dialogam mais diretamente com a representação computacional da língua, além de o Constructicon, o qual será apresentado na seção posterior, ter sido desenvolvido com base, principalmente, na Gramática das Construções de Berkeley e na Gramática das Construções Baseada em Signos, as quais serão apresentadas nesta subseção.

Enquanto grande parte dos estudos gramaticais buscava fazer generalizações simples e universais, as quais não incluíam em suas análises aspectos relacionados à idiomaticidade, os estudos gramaticais em Berkeley nos anos 1980 e 1990 centravam-se em analisar as idiomaticidades e “irregularidades” da língua, demonstrando sua frequência e centralidade no conhecimento linguístico dos falantes (FILLMORE, 2013).

É nesse contexto que surge uma visão de que as mesmas ferramentas analíticas eram responsáveis tanto pelas estruturas mais básicas quanto pelos casos ditos como “especiais” (FILLMORE, 2013, p. 111-112). Com base nesse pensamento, a abordagem denominada Gramática das Construções de Berkeley (BCG) emerge dos esforços de Kay e Fillmore (1999) em analisar fenômenos da linguagem considerados como idiomáticos, pois, para abranger todas as possibilidades gramaticais de uma língua, deve-se levar em conta também os casos considerados periféricos (FRIED & OSTMAN, 2004). Essa visão contrasta diretamente com os pressupostos das gramáticas formais baseadas em regras, as quais não conseguem dar conta das muitas irregularidades existentes nas línguas. Dessa forma,

espera-se que um gramático das construções desenvolva um sistema de representação capaz de codificar economicamente e sem perda em

generalização de todas as construções (ou padrões) da língua, do mais idiomático ao mais geral.² (KAY & FILLMORE, 1999, p. 2)

Sob uma perspectiva construcionista, uma construção é, pois, entendida como um par forma-significado (GOLDBERG, 1995), essa concepção relaciona-se estreitamente à definição de signo linguístico proposta por Ferdinand de Saussure no Curso de Linguística Geral (2006 [1916]), visto que, desde os trabalhos do linguista, assume-se que todo e qualquer signo, seja uma palavra ou, ainda, um morfema, é um pareamento de forma e significado, ou seja, é uma construção (GOLDBERG, 2013, p. 17).

Assume-se, então, que as construções são unidades básicas da língua que se constituem como uma associação convencional de forma e conteúdo linguístico (significado) (GOLDBERG, 1995; KAY & FILLMORE, 1999), assim o significado de uma construção é visto como independente, em parte, das palavras que a constituem. Ademais, esse argumento acompanha a proposição de que existe um *continuum* entre léxico e sintaxe (FILLMORE, 2008), isso significa que o tratamento construcional estende-se a todas as unidades da língua, sejam elas sintáticas, lexicais, mórficas, ou mesmo discursivas.

Uma vez que o conhecimento gramatical, como um todo, pode ser representado através de construções, que diferem apenas em termos de complexidade interna e nível de esquematicidade (BOAS, 2003), o desenvolvimento de uma gramática se dá como uma rede de construções erguida na cultura através do uso (GOLDBERG, 1995; 2006). Nas palavras de Fillmore (2013, p.112)

de maneira complementar, a Gramática das Construções postula que a gramática de uma língua é o conjunto de suas construções gramaticais, as regras que unificam informações formais e semânticas em vários tipos de objetos linguísticos, juntamente com os princípios que os restringem e conectam.³

²“(…) the construction grammarian is required to develop an explicit system of representation, capable of encoding economically and without loss of generalization all the constructions (or patterns) of the language, from the most idiomatic to the most general”.

³“Complementarily, Construction Grammar posits that the grammar of a language is the set of its grammatical constructions, the rules that unite formal and semantic information into various kinds of linguistic objects, together with the principles that constrain and connect them”.

Em resumo, o argumento mais geral em favor da Gramática das Construções é o de que este modelo é econômico, uma vez que não precisa configurar diferentes componentes da gramática. Em segundo lugar, essa gramática tem potencial para uma representação uniforme de todos os conhecimentos gramaticais, visto que não atribui status especial a certos fragmentos de gramática, além de dar adequadamente descrições sistemáticas e formais da morfologia, sintaxe, semântica e pragmática de estruturas linguísticas tipicamente consideradas "irregulares" ou "excepcionais" (FRIED & OSTMAN, 2004).

Fried & Ostman (2004) complementam a caracterização da Gramática das Construções como uma abordagem não-modular, gerativa, não-derivacional e baseada em unificação. Por não conceber aspectos fonológicos, morfológicos, sintáticos e semânticos como módulos autônomos, essa gramática é caracterizada como não-modular. É gerativa, no sentido chomskyiano, pois tem como objetivo abranger todas as construções gramaticais da língua sem menosprezar aquelas que precisam ser formalizadas em alguma medida. Como não há regras que possam variar de um nível a outro, isto é, não existem regras de derivação entre níveis, é, portanto, não-derivacional. Além disso, é baseada em unificação, o que significa que as construções e especificações que constituem as construções se encaixam, se unificam, uma a outra de um modo não-derivacional.

Nos encaminhamos agora a apresentar os mecanismos necessários para representar as construções e como essas se relacionam à outras construções de modo a licenciar construtos.

O modelo teórico da BCG é baseado em um sistema de traços, o que proporciona um maior grau de formalidade. Nesse sistema, as construções são representadas por Matrizes de Atributo e Valor (AVM), assim uma AVM é a forma em que comumente são apresentadas as estruturas de traços. Grosso modo, as estruturas de traços consistem na "divisão" de um dado linguístico em partes menores, os chamados atributos, e a cada atributo associam-se

valores. Uma das principais características das estruturas de traços é poder combinar informações através da unificação.

Esse processo de unificação é a principal operação formal desse modelo gramatical, sendo responsável pelo acolhimento ou rejeição de constituintes candidatos a ocupar posições sintáticas específicas. A aceitação ou restrição, portanto, fica condicionada à compatibilidade entre os valores atribuídos aos atributos, isto é, aos elementos linguísticos e os valores exigidos pelas posições da construção. Assim, é possível definir e especificar os tipos de entidades que podem ou devem estar em cada sintagma (FILLMORE, 2013).

É através da unificação que as Matrizes de Atributo e Valor combinam-se, projetando uma nova AVM, a qual contém exatamente os valores e atributos das AVMs que se uniram (GOLDBERG, 2006). Essas AVMs correspondem a conjuntos de traços que especificam as características de cada elemento, bem como as restrições de combinações de dois ou mais elementos em uma construção. Desse modo, a tarefa principal da unificação é assegurar que os atributos com valores contraditórios falhem ao se combinar, não chegando a licenciar uma construção (FRIED & OSTMAN, 2004).

As informações sintáticas e semânticas são representadas dentro de uma única estrutura de traços, ou seja, segundo Fried & Ostman (2004), as informações gramaticalmente relevantes são representadas na forma de pares de atributo-valor entre colchetes ou caixas que são organizados em conjuntos (AVMs). Um atributo representa uma propriedade particular, ou seja, uma categoria linguística – sintática, semântica, pragmática, prosódica, etc. – relevante em uma determinada construção e o valor é uma especificação dessa propriedade na construção.

Esse processo é mais bem esclarecido através do exemplo na Figura 5.

$$\left[\begin{array}{l} \text{cat} \quad \langle n \rangle \\ \text{max} \quad \langle + \rangle \\ \left[\begin{array}{l} \text{cat} \quad \langle d \rangle \\ \text{lxm} \quad \langle \text{the} \rangle \end{array} \right] \quad \left[\begin{array}{l} \text{cat} \quad \langle n \rangle \\ \text{max} \quad \langle - \rangle \end{array} \right] \end{array} \right]$$

Figura 5: Representação da Construção DetHead (FILLMORE, 2013, p. 113)

A descrição fornecida por Fillmore (2013) exemplifica a formação da construção DetHead do inglês, a qual é constituída por um determinante e um núcleo nominal. Na Figura 5, podemos observar que o sistema de representação se dá por meio de colchetes, os quais contêm os atributos (cat, max, etc.) e os valores (n, +, etc.) que lhes foram atribuídos. Olhando para os atributos, vemos o elemento categoria (cat), que se refere à categoria gramatical à qual um determinado lexema pertence, podendo receber os valores de nome, verbo, preposição, advérbio, adjetivo, determinante, etc. Nesse caso, cat recebe o valor de nome. Já a maximalidade (max) está relacionada à possibilidade ou não de o elemento poder ocorrer como argumento, isto é, se ele pode ou não constituir um sintagma por si só, podendo receber os valores <+> ou <->, como é possível ver no exemplo acima. Existem outros atributos que não foram apresentados na Figura 5, como Função Gramatical, a qual inclui sujeito, objeto, complemento, oblíquo, ou Número, o qual está ligado à especificação dos elementos quanto ao número, isto é, singular ou plural.

No exemplo, notamos que a construção DetHead é representada pelo colchete maior, ao qual é atribuída a categoria de nome. Os colchetes menores que estão contidos no colchete maior representam, respectivamente, as AVMs do determinante e do seu núcleo, bem como os atributos de categoria e maximalidade e seus respectivos valores. De acordo com esse modelo, a contraparte semântica é apresentada pelo frame evocado pelo signo linguístico, apesar de não aparecer na representação da construção DetHead.

As informações sobre qualquer entidade linguística aparecem como uma matriz atributo-valor (AVM) associada à caixa que representa essa entidade e

suas propriedades constituintes são registradas como AVMs contidas dentro de caixas menores. Dessa maneira, um construto como *much snow* poderia ser representada por uma caixa maior contendo duas caixas menores, sendo que a caixa à esquerda representa o determinante “much” e a da direita representando o nome “snow”, como ilustrado na Figura 6.

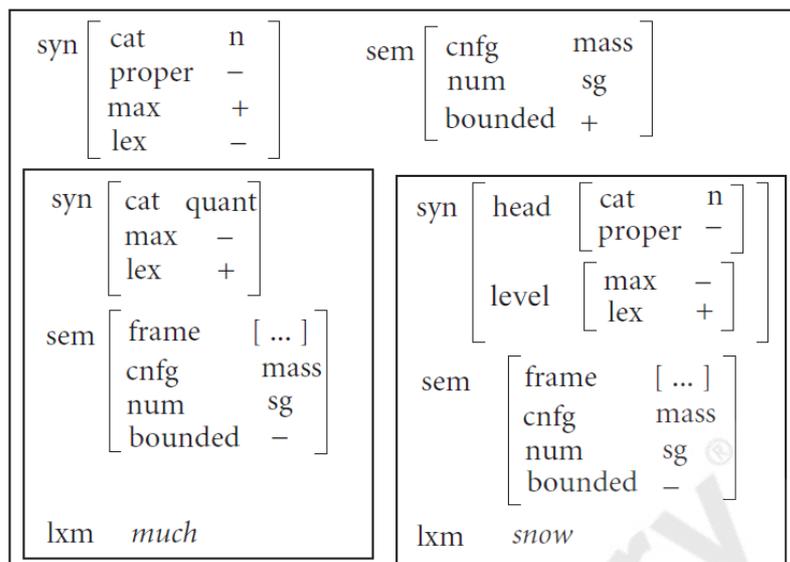


Figura 6: Representação parcial do construto do inglês *much snow* (FRIED & OSTMAN, 2004, p.34)

Esse construto pode ser considerado como um objeto linguístico bem formado, porque há uma construção em inglês, DetHead, que permite que palavras do tipo determinante, como “much”, se associem à esquerda de palavras do tipo nome, como “snow”, fazendo com que o sintagma resultante tenha as propriedades que tem em virtude da sua participação nessa construção (FILLMORE, 2013). Assim, é possível observar também que nos três atributos, *cnfg* (configuração), *num* (número) e *bounded* (delimitação), existe correspondência de valores, dessa maneira, esse construto é licenciado através da unificação desses valores semânticos.

Para a Gramática das Construções de Berkeley, a noção de Valência é outro aspecto relevante no tratamento construcional. De acordo com Fillmore (2013), a BCG não faz distinções entre argumentos e adjuntos em relação à

organização sintática, mas distingue elementos centrais e periféricos de uma valência. A valência contém, portanto, um conjunto de papéis semânticos que estão diretamente vinculados a um participante de um frame particular, por um lado, e à sua expressão morfossintática correspondente, por outro. Assim sendo, a valência completa de um lexema conta com todos os valentes – semânticos e sintáticos – especificados em termos de papel semântico, função gramatical e tipo sintagmático (FILLMORE, 2013).

A BCG também considera que existem relações entre as construções, as quais são capturadas pela relação de herança, uma vez que esses *links* entre construções fornecem uma maneira coerente de capturar as generalidades observáveis em toda a rede de construções. Essa abordagem trabalha com a noção de Herança Completa, dessa forma, toda a informação na construção mãe deve estar presente igualmente ou mais especificamente na construção filha (KAY & FILLMORE, 1999, p.7). Os autores trazem como exemplo de relação de herança a relação existente entre a Construção Núcleo-Complemento e a Construção de Sintagma Verbal. A Construção Núcleo-Complemento é composta por dois signos filhos, um núcleo e um complemento. Na Figura 7, podemos observar que o signo filho núcleo (head) tem o traço lexical (lex) marcado positivamente e o complemento (podendo haver mais de um complemento, conforme indicado pelo "+") recebe valor positivo para o traço (loc), por se manifestar localmente à direita do núcleo.

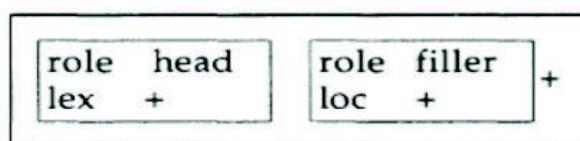


Figura 7: Construção Núcleo-Complemento (KAY & FILLMORE 1999, P. 7).

A Construção de Sintagma Verbal requer que outros atributos sejam especificados. Além dos atributos lex e loc, se faz necessário especificar que o núcleo dessa construção precisa ser da categoria (cat) dos verbos (v), enquanto que a função gramatical (gf) do complemento recebe valor negativo

para sujeito (subj), visto que essa posição não pode ser preenchida por um complemento do tipo sujeito, como podemos verificar na Figura 8.

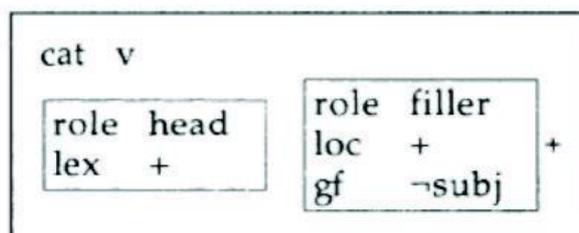


Figura 8: Construção de Sintagma Verbal (KAY & FILLMORE, 1999, p. 8).

Observe que todas as informações que aparecem na Figura 7 também aparecem na Figura 8. Uma vez que desejamos dizer de uma forma mais econômica e genérica que Construção de Sintagma Verbal herda da Construção Núcleo-Complemento, é possível representar essa relação na formulação dada na Figura 9.

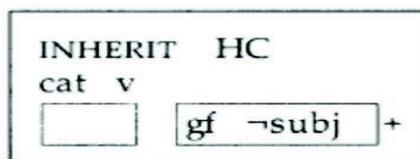


Figura 9: Construção de Sintagma Verbal, representada através da relação de Herança (KAY & FILLMORE, 1999, p. 8).

De maneira similar, a Gramática das Construções Baseada em Signos (Sign-Based Construction Grammar– SBCG) busca formalizar descrições que sejam interpretáveis computacionalmente, integrando-as aos pressupostos fundamentais das abordagens construcionistas. Na SBCG, o pareamento de forma e significado é mediado por signos. Apesar da noção de signo utilizada pela SBCG se assemelhar à de Saussure (2006 [1916]), uma combinação de forma (significante) e sentido (significado), Sag (2012) destaca que essa noção é ampliada, pois, a descrição de um signo incorpora outros traços como fonologia, forma, estrutura argumental, sintaxe, semântica e contexto.

Sag (2012) expõe que o traço fonologia (FON) refere-se a um sintagma fonológico. O traço forma (FORMA) especifica as propriedades morfológicas do signo. O terceiro traço, estrutura argumental (EST-ARG), registra os possíveis argumentos sintático-semânticos de um signo lexical. Já o traço sintaxe (SIN) distingue os signos, especificando valores para os subtraços Categoria (CAT) (subtipos: nome, adjetivo, verbo, etc.) e Valência (VAL) (potencial combinatório de um signo). O significado do signo é postulado pelo traço semântica (SEM) e tem como valores os subtraços INDEX (seu valor é uma variável atribuída a uma entidade específica ou uma situação) e FRAME (lista de elementos que são evocados pelo signo). Para o traço SEM, Sag (2012) considera os pressupostos da Semântica de Frames⁴. Por fim, o traço contexto (CTXT) é utilizado quando o contexto se faz necessário para a compreensão de um signo.

Esses traços são descritos também em termos de Matrizes de Atributo e Valor (AVM), assim como proposto pela Gramática das Construções de Berkeley. A Figura 10 traz um exemplo dessa formalização.

PHONOLOGY	/kim/
SYNTAX	NP
SEMANTICS	‘the intended person named Kim’

Figura 10: Matriz da formalização do nome próprio “Kim” (SAG, 2012, p. 75)

Observamos que o signo “Kim” pode ser representado pela Matriz apresentada na Figura 10, assim, na primeira linha, temos o traço Fonologia, que traz a transcrição fonológica da palavra, seguida pelo traço Sintaxe, que especifica que se trata de um Sintagma Nominal e, por fim, pelo traço Semântica, o qual especifica que essa palavra faz referência a uma pessoa que se chama Kim.

Nesse sistema, os signos são descritos em partes menores, os chamados traços e a cada traço associam-se valores, o que possibilita combinar informações através da unificação, isto é, uma operação pela qual

⁴ Fillmore (1982; 1985a)

duas ou mais AVMs são combinadas de modo que a AVM resultante contém exatamente o que está presente em suas AVMs componentes (FRIED & OSTMAN, 2004).

Segundo Sag (2012), na SBCG, signos e construtos são estruturas de traços, assim cada signo léxico é licenciado por um listema, isto é, uma descrição "listada" de uma palavra ou sintagma, já os construtos são árvores locais que são licenciadas por uma construção combinatória, a qual é descrita por uma estrutura de traços funcional que especifica valores para o traço MOTHER (MTR) e o traço DAUGHTERS (DTRS), assim a valência de MTR é um signo e a valência de DTRS é uma lista não vazia de signos. De maneira geral, Sag (2012) define signos e construtos como estruturas de traços que fazem parte do modelo de linguagem, enquanto listemas e construções são descrições que licenciam classes de objetos linguísticos (signos ou construtos), os quais fazem parte da gramática, isto é, da descrição do modelo da linguagem.

A discussão apresentada até aqui a respeito da BCG e da SBCG evidencia o caráter formal dessas abordagens. Trabalhos desenvolvidos sob essas perspectivas, como o apresentado nesta dissertação, frequentemente estão relacionados à Linguística Computacional justamente por conta desse formalismo teórico.

O Constructicon da FrameNet Brasil foi desenhado com base em contribuições da BCG e da SBCG, visto que essa base construcional tem por objetivo descrever e modelar construções em termos de suas propriedades gramaticais e seu potencial semântico. A próxima seção se dedica à apresentação desse recurso denominado Constructicon.

1.4 O Constructicon da FrameNet Brasil

A aplicabilidade das teorias da Gramática das Construções baseadas em unificação se mostra na constituição do Constructicon. De uma forma mais geral, o Constructicon é um recurso computacional sintático-semântico que

contém o repertório das construções de uma língua (FILLMORE, 2008), fundado na BCG (KAY & FILLMORE 1999; FILLMORE 2013) e na SBCG (BOAS & SAG, 2012).

O propósito do Constructicon é suprir a necessidade de análise de determinadas estruturas linguísticas não processáveis lexicograficamente, uma vez que, na FrameNet, somente as valências das unidades lexicais são anotadas. Por essa razão, o Constructicon abarca o conhecimento linguístico que excede a valência simples de palavras simples (FILLMORE, 2008). De modo mais específico, o Constructicon descreve construções em termos de suas propriedades gramaticais e seu potencial semântico. Além disso, realiza tais descrições de modo compatível com os pressupostos teóricos da BCG.

Trazemos a Figura 11 para ilustrar como o repertório de construções do PB está organizado na FN-Br e como é apresentado com base no emprego da ferramenta Webtool 3.0. Para o caso da construção Transitiva_indireta_agentiva, assim é possível ver a descrição da construção, os Elementos da Construção, bem como a relação desta com outras construções.

Transitiva_indireta_agentiva [Transitive_indirect]

Definição
Tipo de construção SN_V_SPrep. Essa construção exibe um argumento externo Sujeito e um Predicado. O argumento interno é um objeto indireto.
Exemplo(s)
Elementos da Construção
Predicado [Predicate] O Predicade é composto de um SV e um SPrep objeto indireto.
Sujeito [Subject] O sujeito é um SN.
Relações
Evoca Ação_transitiva
Herda de Sujeito_predicado

Figura 11: Construção Transitiva_indireta_agentiva modelada no Constructicon

No Constructicon, cada construção é definida em termos de suas partes constituintes, os Elementos da Construção (EC) (FILLMORE ET AL., 2012). Assim, o processo de modelagem das construções se dá, primeiramente, pela definição das propriedades sintáticas destas construções. No caso, a

construção *Transitiva_indireta_agentiva* é composta por dois ECs denominados Sujeito e Predicado como podemos observar na Figura 11.

Na Figura 11 é possível ver, ainda, a definição da construção, os ECs e suas respectivas definições, além disso, conforme mostrado na aba *Relações*, a construção *Transitiva_indireta_agentiva* é um subtipo da construção mais abstrata *Sujeito_predicado* e evoca o frame de *Ação_transitiva*. Esse frame é basicamente definido como a ação de um Agente que afeta uma entidade, chamada de Paciente. Dessa forma em uma sentença como (13) essa ação transitiva é expressa pela construção *Transitiva_indireta_agentiva*.

(13) Maria bateu na irmã.

Em (13) percebemos que para o estabelecimento da relação de regência entre o verbo e seus argumentos, o SN “Maria”, agente da ação, atua como argumento externo e o argumento interno, paciente da ação, *SPrep* “na irmã” é regido por uma preposição exigida pelo verbo que ocupa o *slot V* da construção.

Outro fator de suma importância no Constructicon é o processo de unificação, em que AVMs (Matrizes de Atributo e Valor) – construções e seus signos filhos – combinam-se, projetando uma nova AVM, a qual contém exatamente os valores e atributos das AVMs que se uniram (GOLDBERG, 2006). Essas AVMs correspondem a conjuntos de traços que especificam as características de cada elemento, bem como as possibilidades (ou restrições) de combinações de dois ou mais elementos em uma construção. Conforme previamente mencionado, é através da unificação que se assegura que os atributos com valores contraditórios falhem ao se combinar, de modo a não licenciar uma construção.

Nesse processo de unificação, é possível estabelecer restrições de constituição para a construção a partir da definição dos ECs, uma vez que a constituição tipifica os signos filhos – os ECs – da construção em termos de outras construções. Dessa forma, enquanto os aspectos formais são

considerados para a criação dos ECs, ou signos filhos, a informação semântica é atribuída através da unificação da construção com um frame. A Figura 12 mostra a aplicação das restrições à construção Transitiva_indireta_agentiva.

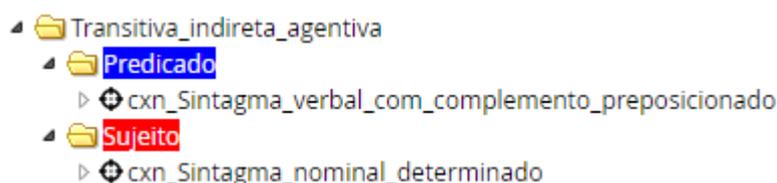


Figura 12: Restrições aplicadas à construção Transitiva_indireta_agentiva

Podemos observar na Figura 12 que o EC Predicado é uma instância de um dos subtipos da construção Sintagma_verbal_com_complemento_preposicionado, isto é, é licenciado pela construção Sintagma_verbal_com_complemento_preposicionado, enquanto o EC Sujeito é licenciado pela construção Sintagma_nominal_determinado.

Para tornar as descrições apresentadas em prosa na Figura 11 legíveis por máquinas, a FN-Br possui um editor de Restrições incluído na ferramenta WebTool 3.0 e é através desse editor que modelamos que um EC de uma determinada construção é uma construção própria (SILVA ET AL., 2017) conforme vemos na Figura 12. Ademais, o editor é, portanto,

um meio de modelar tanto a constituinte quanto alguns aspectos de unificação, na medida em que unifica um determinado EC ao conjunto de propriedades definidas para as construções que o constituem, evitando a necessidade de modelar as mesmas propriedades mais de uma vez.⁵ (SILVA ET AL., p. 2, 2017).

A relação de *Evocação* é outro aspecto importante no processo de modelagem, pois é ela que mapeia as correspondências e as restrições semânticas entre os ECs da construção e os e os EFs do frame evocado. Como observado na Figura 11, a construção Transitiva_indireta_agentiva evoca o frame de Ação_transitiva. Assim, relação de *Evocação* desempenha a

⁵ (...) a means of modeling both constituency and some aspects of unification, to the extent that it unifies a give CE to the set of properties defined for the constructions constraining it, also avoiding the need to model the same properties more than once.

função de mapear computacionalmente o EC Sujeito ao EF Agente e o EC Predicado – mais especificamente o SP dentro do EC Predicado – ao EF Paciente.

Uma construção pode se relacionar com outras construções da rede através de outra relação, chamada de relação de *Herança*. Essa relação é estabelecida quando as construções herdarem características do item hierarquicamente superior em todos os níveis. Portanto, “todas as informações específicas para cada nó que domina direta ou indiretamente um determinado nó são herdadas”⁶ (GOLDBERG, 1995, p. 74). A herança é modelada no Constructicon como uma relação entre os ECs das construções mãe (mais genérica) e filha (mais específica) (SILVA ET AL., 2017). Essas relações podem ser representadas através de um grafo, conforme ilustra a Figura 13.

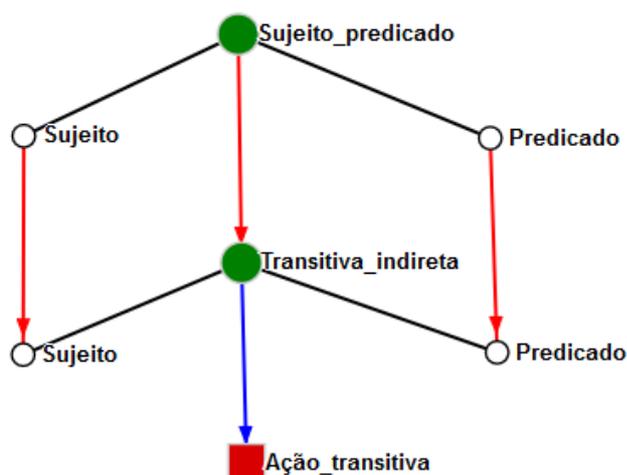


Figura 13: Representação gráfica da relação de herança entre a construção Transitiva_indireta_agentiva e a construção de Sujeito_predicado

Notamos na Figura 13 que a os círculos verdes representam construções diferentes, ao passo que os círculos vazados representam os ECs. A relação de *Herança*, a qual é estruturada em termos dos constituintes das construções é representada pelas setas vermelhas que conectam os círculos.

⁶ (...) all information specific to every node which directly or indirectly dominates a given node is inherited.

Por fim, o quadrado vermelho representa o frame evocado pela construção e a seta que o conecta a ela, a relação de *Evocação*.

Uma vez modelada a construção, é possível anotar construtos licenciados por ela em *corpus*.

[119670] ■ AST_MS_APP NI		M	a	r	i	a	b	a	t	e	u	n	a	i	r	m	ã	.
CE		S	u	j	e	i	P	r	e	d	i	c	a	d	o			
Ação_transitiva.FE		A	g	e	n	t	P	a	c	i	e	n	t	e				
CEE																		
CstrPT		S	f	i	n													

Figura 14: Anotação da sentença “Maria bateu na irmã”

A Figura 14 traz a anotação para a sentença (13). Como podemos observar, essa construção é do tipo Transitiva_indireta_agentiva e evoca o frame de Ação_transitiva. A anotação construcional é feita em camadas. Na primeira camada, são anotados os ECs da construção, como vemos “Maria” é o sujeito e “bateu na irmã” o predicado. Na segunda camada, são anotados os EFs relacionados aos ECs, assim, o ECs Sujeito e Predicado se relacionam, respectivamente, aos EFs Agente e Paciente. Como essa relação é modelada na base, as etiquetas dos EFs são aplicadas automaticamente. Já na última camada, é anotado o tipo sintagmático do construto e este, no caso, é uma sentença finita (Sfin).

A unificação entre valores e atributos é representada através das camadas da anotação, Elementos da Construção, Elementos de Frame, Tipo Sintagmático, as quais são associadas ao próprio construto. Assim, pode-se modelar adequadamente a continuidade entre o léxico e a gramática também pela anotação.

Como o objetivo desta pesquisa é o de propor um modelo no Constructicon para as construções interrogativas com elemento QU- movido no PB, no capítulo 4, trataremos um levantamento bibliográfico de abordagens para

tais construções. Por ora, passamos à discussão do campo de aplicação da pesquisa: o desenvolvimento de *chatbots*.

2 DESENVOLVIMENTO DE CHATBOTS

2.1 Conceitos Fundamentais

Desde meados dos anos 50, os pesquisadores de Inteligência Artificial têm lutado com o desafio de criar sistemas computacionais “inteligentes” que permitem a interação entre homem e máquina. Inteligência artificial, ou IA, é um ramo do conhecimento humano que busca reproduzir a mente humana usando elementos computacionais (TURING, 1950) e é nesse contexto que nascem os *chatterbots* ou *chatbots*.

Os *chatbots* são programas de computador criados com o propósito de simular a habilidade de conversação de um ser humano e responder as perguntas de maneira que se tenha a impressão de estar conversando com outra pessoa e não com um programa de computador.

Desde a criação dos primeiros *chatbots*, muitas técnicas vêm sendo usadas de modo a aperfeiçoar seu desenvolvimento, assim sendo, Neves e Barros (2005 apud COMORELLA & CAFÉ, 2008, p. 59) propõem uma classificação para esses programas de acordo com as tecnologias aplicadas em seu desenvolvimento, na qual destacam-se os primeiros *chatbots* criados, ELIZA, JULIA e ALICE. O Quadro 2 representa essa classificação em gerações e apresenta as principais características destas, bem como os *chatbots* que melhor representam essas gerações.

Quadro 2: Classificação evolutiva, segundo as tecnologias aplicadas (NEVES & BARROS, 2005 apud COMORELLA & CAFÉ, 2008, p. 59).

Geração	Característica	Chatterbot
1ª Geração	Buscam palavras-chaves e utilizam o casamento de padrões e regras gramaticais para dar continuidade ao diálogo (MOURA, 2003).	ELIZA
2ª Geração	Utilizam de princípios de inteligência artificial, como redes neurais, para simular a conversa. (MAULDIN, 1994)	JULIA

3ª Geração	Possuem uma apresentação mais atrativa, com uma interface gráfica que estimula o diálogo. Utilizam da linguagem de marcação <i>Artificial Intelligence Markup Language</i> (AIML) (LEONHARDT, 2005).	ALICE
------------	--	-------

As técnicas utilizadas nos *chatbots* da primeira geração tratam a informação através de casamentos de padrões (COMORELLA & CAFÉ, 2008) e regras gramaticais, isto é, elas identificam um padrão existente em um conjunto de dados. Em um primeiro momento, o casamento de padrões verifica se a sentença digitada pelo usuário possui a estrutura desejada, para então encontrar uma estrutura semelhante, na qual se possam estabelecer pontos de alinhamento e substituir parte desse casamento por outra estrutura (COMORELLA & CAFÉ, 2008).

O primeiro *chatbot* construído foi o ELIZA, em 1966, pelo professor Joseph Weizenbaum, com o objetivo de simular um psicanalista rogeriano em uma conversa com um paciente. ELIZA estimulava o usuário (paciente) a falar por meio de perguntas e então esperava que ele contasse e refletisse sobre os problemas que o incomodavam. Esse *chatbot* destaca-se por utilizar a técnica do casamento de padrões o que torna seu funcionamento extremamente simples. O ELIZA recebe uma sentença digitada por um usuário em linguagem natural e, em seguida, compara essa sentença com uma lista de padrões. Quando encontra o padrão mais semelhante àquela sentença digitada pelo usuário, o *chatbot* analisa a sentença com base em regras gramaticais de decomposição e monta respostas a partir de uma lista de modelos com regras de reconstrução de frases associadas às regras de decomposição. Apesar de simples e de ser umas das implementações pioneiras, ELIZA tem uma personalidade muito bem definida e uma das características mais interessantes de ELIZA é a forma como formula suas respostas, uma vez que, através da repetição das frases dos pacientes, o *chatbot* consegue obter respostas dos pacientes analisados simulando a falta de envolvimento típica do psicanalista (MOURA, 2003), como podemos observar no trecho de uma conversa com ELIZA logo abaixo:

Humana: Os homens são todos iguais.
 Eliza: De que forma?
 Humana: Eles estão sempre nos incomodando com uma coisa ou outra.
 Eliza: Você consegue pensar em um exemplo específico?
 Humana: Bem, meu namorado me fez vir aqui.
 Eliza: O seu namorado fez você vir aqui.
 Humana: Ele diz que eu estou quase sempre deprimida.
 Eliza: Sinto muito por saber que você está deprimida.
 Humana: É verdade. Eu sou infeliz.
 Eliza: Você acha que vir aqui lhe ajudará a não ser feliz?
 Humana: Preciso de ajuda, com certeza.
 Eliza: O que significaria para você conseguir ajuda?
 Humana: Talvez eu conseguisse aprender a me entender com minha mãe.
 Eliza: Fale-me mais sobre a sua família.
 Humana: Minha mãe cuida de mim.
 Eliza: Quem mais na sua família cuida de você?
 Humana: Meu pai.
 Eliza: Seu pai.
 Humana: Você se parece com meu pai em algumas coisas.
 Eliza: Qual semelhança você vê?
 Humana: Você não é muito agressiva, mas eu acho que você não quer que eu perceba isso.
 Eliza: O que faz você achar que eu não sou muito agressiva?
 Humana: Você não discute comigo.
 Eliza: Por que você não acha que eu não discuto com você?
 Humana: Você tem medo de mim.
 Eliza: Te agrada acreditar que eu tenho medo de você?
 Humana: Meu pai tem medo de todo mundo.
 Eliza: O que vem à sua cabeça quando você pensa no seu pai?
 Humana: Tiranos.
 Eliza: Isso tem alguma coisa a ver com o fato de o seu namorado tê-la feito vir até aqui?

(WEIZENBAUM, 1966 apud COMORELLA & CAFÉ, 2008, p. 60)

Logo nas primeiras linhas do diálogo, podemos observar que ELIZA se utiliza de perguntas e comentários que soam pertinentes ao assunto tratado, mas sem mostrar um envolvimento com o humano de modo que consegue fazer com que o diálogo se desenvolva mesmo que repita informações fornecidas na interação para isso.

O JULIA representa bem os *chatbots* da segunda geração, este foi desenvolvido por Michael Mauldin com técnicas mais sofisticadas, tais como as redes neurais. O JULIA é um personagem dentro de um jogo de computador multi-usuários e seu objetivo é auxiliar jogadores em um ambiente controlado,

o TinyMUD (*Tiny Multi-User Dimension*), interagindo com qualquer outro usuário, seja ele real ou virtual (MOURA, 2003).

Esse *chatbot* utilizava-se de um módulo baseado em redes neurais, assim, uma característica interessante implementada nele é a capacidade de lembrar informações sobre ações feitas anteriormente, para auxiliar na tomada de decisões no jogo. Ademais, JULIA possui um modelo interno do mundo que se atualiza durante a conversa, ou seja, é capaz de contextualizar diálogos, como é possível observar no trecho de um diálogo com JULIA:

Juiz: Você já teve um bichinho de estimação?
Julia: Você tem animais de estimação?
Juiz: Eu acho que você está evitando a minha pergunta.
Julia: Você acha que os animais podem pensar?
Juiz: Você acha ético que uma espécie escravize outra?
Julia: Quais tipos de animais de estimação você tinha quando criança?
Juiz: Eu estou fazendo as perguntas. O que é um animal de estimação senão um escravo do especismo.
Julia: Oh, isso é tão triste. Eu acho que animais de estimação enriquecem imensamente a infância.
Juiz: Como podemos manter alguns animais como animais de estimação e ainda comer outros?
Julia: Porquinhos-da-índia são tão inteligentes quanto pepinos.
Juiz: Você comeria um porquinho-da-índia?
Julia: Passei muitos verões na fazenda do meu tio. Eu gostava realmente muito dos animais.
Juiz: Eu acho que você contou essa agradável história ao último juiz, não contou?
Julia: Uma vez eu perdi uma hamster. Tentei as armadilhas e bananas, mas infelizmente nada funcionou. Nós reviramos a casa de cabeça para baixo, mas não encontramos nem o menor sinal dela.
Juiz: Você tem um animal de estimação em casa?
Julia: Eu preferiria ter um animal de estimação de verdade, como um gato. Gatos são os meus favoritos, porque você pode domesticá-los e eles são fáceis de manter.
Juiz: Você sabe alguma coisa sobre schnauzers miniaturas?
Julia: Não sei nada sobre schnauzers miniaturas, Juiz.
Juiz: Bem, eu tenho um. Nós o chamamos de Pucci. Ele é um gato louco – você já teve um gato?
Julia: Que coincidência! Eu tenho dois gatos.
Juiz: Eu prefiro cachorros. Gatos são muito independentes e eles sempre se levantam de repente. A propósito, você é perita em gatos?
Julia: As pessoas não possuem os gatos... é o contrário.
Juiz: Algumas pessoas têm cobras como animais de estimação – você acha que elas podem ser de alguma valia como um animal de estimação?
Julia: Eu acho que as pessoas são bem seguras de si. Gatos são muito independentes, mas eles também podem ser muito carinhosos. Cachorros são simplesmente muito servis e

apressados para agradar pro meu gosto.
 Juiz: Tudo bem – mas como você “domestica” uma cobra?
 Julia: Puxa, isso não é comum. Eu não conheço muitos donos de cobras.
 Juiz: E tartarugas? Tudo o que elas fazem é sentar nas suas lagoas artificiais, comer e defecar, estou certo?
 Julia: Tartarugas não têm o perfil de um animal de estimação. Elas parecem não ter emoções.
 Juiz: Aha! Mas onde traçamos a linha como seres humanos? Quem somos nós para decidirmos quais animais têm valor e quais não têm?
 Julia: Continue.
 Juiz: As abelhas, por exemplo. Eu por acaso gosto delas.
 Julia: Eu também.

(MAULDIN, 1994 apud COMORELLA & CAFÉ, 2008, p. 61-62)

No trecho apresentado acima, é possível verificar que JULIA é capaz de contextualizar o diálogo a partir das informações que foram armazenadas durante a interação, gerando assim uma maior sensação de realismo (MOURA, 2003).

Na terceira geração, destaca-se o ALICE (*Artificial Linguistic Internet Computer Entity*). Um sistema inovador desenvolvido por Richard Wallace, em 2000. A maior novidade desse *chatbot* é que ele opera com a Linguagem de Marcação da Inteligência Artificial (*Artificial Intelligence Markup Language*, doravante AIML) desenvolvida por Wallace, e que permite que os usuários criem seus próprios robôs. Uma das características inovadoras desta linguagem é a possibilidade de o robô ser programado para aprender informações passadas, como, por exemplo, o nome do usuário (MANFIO & MORENO, 2015), uma vez que ALICE opera através de um modelo de aprendizagem supervisionada, isto é, um tipo de “treinador” que monitora as conversas do *chatbot* e cria o índice novo de AIML, a fim de obter as respostas mais apropriadas. Além disso, esse *chatbot* “possui módulos de conversação que permitem classificar o usuário por idade, sexo, localização geográfica e profissão” (COMORELLA & CAFÉ, 2008, p. 62). Abaixo, trazemos um trecho de um diálogo de ALICE na edição de 2004 do prêmio Loebner:

[12:26:09] Alice para J1 – Alô, meu nome é Alice e eu sou a humana
 [12:26:41] CHH4 para J1 – Alô, meu nome é Ben e eu sou o humano
 [12:27:55] J1 – Olá, Ben. Como vai?
 [12:28:13] CHH4 para J1 – Bem, estou me divertindo muito nesta terra do absurdo.
 [12:28:06] J1 - Olá, Alice, como vai?

[12:28:20] Alice para J1 – Olá. Está tudo muito bem.

[12:28:26] J1 para CHH4 – Qual é o absurdo sobre isto?

[12:28:57] CHH4 para J1 – Só a maneira como estou ganhando dinheiro em uma tarde de domingo. Não é a forma comum de ganhar algum dinheiro para gastar.

[12:29:11] Alice para J1 – Obrigado pela informação. Eu estou em Nova York. Onde você está?

[12:28:50] J1 para Alice – Fico feliz de saber disso. Onde você mora?

[12:29:21] J1 para CHH4 – Como você normalmente passa as suas tardes de domingo?

[12:30:11] CHH4 para J1 – Depende. Assistindo ao baseball, fazendo caminhadas por Nova York, descansando para a semana que está por vir? Isso varia.

[12:30:12] Alice para J1 – Me fale um pouco sobre a sua vida em Nova York. Ah. Tudo que eu preciso saber eu aprendi com o meu professor.

[12:29:38] J1 para Alice – Em Nova York. Como você chegou aqui hoje?

[12:30:23] J1 para CHH4 – Onde em Nova York você gosta de caminhar?

[12:30:44] J1 para Alice – Por que você não responde a minha pergunta primeiro?

[12:31:11] CHH4 para J1 – Eu gosto de bairros que não conheço bem, quando de repente percebo que estou no meio de uma cultura diferente.

[12:31:15] Alice para J1 – Você não sabe? Não, eu não respondo a sua pergunta primeiro. Você responde? Eu respondo a muitas perguntas idiotas.

[12:31:24] J1 para CHH4 – Quais são os melhores bairros que você já conheceu?

[12:32:12] CHH4 para J1 - Washington Heights, Jackson Heights. Astoria

[12:31:42] J1 para Alice – Você ainda quer saber sobre a minha vida em Nova York?

[12:31:51] Alice para J1 – Sim, me conte, por favor

[12:32:12] J1 para Alice – Eu sou um aluno da graduação e não tenho muito tempo livre. Estudo muito.

[12:32:43] Alice para J1 – Você é um estudante da graduação e não tem muito tempo livre? Você conhece muitas pessoas? Qual disciplina está estudando?

[12:32:57] J1 para CHH4 – Esses bairros são bons. Tem um ótimo lugar para jogar pingue-pongue em J Heights. As pessoas vestem roupas de atletas para jogar.

[12:33:37] J1 para Alice – Eu conheço muitos dos meus colegas. Estudo ciência da computação. O que você faz?

[12:33:38] CHH4 para J1 – Parece demais. Eu adoro pingue-pongue, não tenho jogado muito desde que me mudei para Nova York. Eu via isso mais como uma sala de recreação suburbana.

[12:34:18] Alice para J1 – Você faz amigos duradouros dessa forma? Ciência da computação é uma disciplina difícil?

(COMORELLA & CAFÉ, 2008, p. 62-63)

No trecho apresentado acima, da interação de ALICE com humanos, podemos verificar que ao passo que o uso do AIML seja uma inovação, permitindo que o *chatbot* seja programado para aprender informações passadas, o AIML apresenta um problema significativo, visto que, quando uma mensagem de entrada não foi programada em sua base de conhecimento, o sistema não consegue fornecer uma resposta adequada sobre o assunto,

como podemos ver nas linhas do diálogo em que o usuário J1 pergunta a ALICE como ela chegou em Nova Iorque, ALICE responde através de outra pergunta, mas a resposta não é adequada ao que foi perguntado por J1.

Atualmente, o avanço tecnológico voltado à linguística e o avanço nos estudos voltados à Inteligência Artificial (IA) favoreceram o desenvolvimento de muitos recursos computacionais que têm por finalidade facilitar a interação entre humanos e máquinas. Assim, é possível identificar um vasto número de aplicativos, de redes sociais e de *sites* de busca que procuram estabelecer uma maior interação com os usuários com base em sistema de interface que se aproxime da linguagem natural humana. Esses sistemas de interface são conhecidos por Interface Conversacional de Usuário – CUI –, isto é, com base em uma interface que imita uma conversa com um ser humano real.

Assistentes pessoais virtuais (VPAs), tais como a *Siri* da Apple, o *Google Assistant* e o *Messenger* do Facebook, são exemplos de interfaces conversacionais. Aqui, seguindo McTear et al. (2016), “usamos o termo interface conversacional para nos referirmos à tecnologia que dá suporte à interação conversacional com VPAs através da fala ou outras modalidades”.⁷

No momento, existem basicamente dois tipos de interfaces conversacionais. Há assistentes de voz, com as quais a interação se dá na forma de uma conversa na forma oral, e há *chatbots*, sistemas com os quais a interação se dá na forma escrita, digitando as informações. McTear et al. (2016) consideram que sistemas que operam por meio de CUIs se mostram muito interessantes, visto que permitem que as pessoas conversem com seus dispositivos (como os *smartphones*, por exemplo) de forma intuitiva e natural.

Os *chatbots* “são programas que simulam uma conversa, como as estabelecidas entre seres humanos” (COMORELLA & CAFÉ, 2008, p. 55), assim, essas aplicações do campo de IA utilizam processamento de linguagem natural, visto que lidam com dados e interagem em linguagem natural. Essas interfaces são capazes de “entender” a linguagem natural, ou seja, analisam e interpretam a língua humana, de modo que esse processo gere uma resposta.

⁷“we use the term conversational interface to refer to the technology that supports conversational interaction with these VPAs by means of speech and other modalities”.

Sendo assim, um *chatbot* é, sobre tudo, “um programa de computador que tem por objetivo responder perguntas de tal forma que a pessoa que estiver interagindo com ele tenha a impressão de estar conversando com outra pessoa” (COMORELLA & CAFÉ, 2008, p. 58).

Na seção seguinte apresentamos a plataforma m.knob, *locus* de aplicação desta pesquisa, a qual conta com um *chatbot* atuando na interação com os usuários do aplicativo.

2.2 A Plataforma m.knob

O m.knob (*Multilingual Knowledge Base*) é uma aplicação computacional baseada em Compreensão de Língua Natural que vem sendo desenvolvida interdisciplinarmente no Laboratório FrameNet Brasil, por pesquisadores e alunos de graduação e pós-graduação dos departamentos de Letras e Ciência da Computação da Universidade Federal de Juiz de Fora (MARÇÃO ET AL., 2017). Esse aplicativo atua como um assistente pessoal de viagem desenvolvido em plataforma *mobile* que tem por objetivo auxiliar turistas que viajam pelo Brasil a obter recomendações de atrações e atividades turísticas dando, assim, mais autonomia aos usuários na hora de viajar. Ademais, o aplicativo fornece informações específicas que muitos turistas podem precisar ao planejar uma viagem, como qual atração é melhor para um dia chuvoso ou qual museu é interessante para as crianças.

O aplicativo abrange três idiomas – Português do Brasil, Inglês e Espanhol – e oferece informações personalizadas aos turistas sobre os domínios específicos do Turismo e Esportes (COSTA ET AL., 2018). Além disso, Costa et al. (2018) apontam que o m.knob possui três funções importantes: um *chatbot* que oferece recomendações sobre atrações turísticas e atividades, um banco de dados lexical, a Diciopédia, e um tradutor de sentenças semanticamente melhorado baseado em frames e em relações *qualia*.

Uma das principais funções do aplicativo é o GregBot. Greg, o Gringo é um *chatbot* construído a partir de uma CUI que funciona como um assistente de viagem virtual. O usuário interage com o *chatbot* usando linguagem natural para obter recomendações personalizadas de atrações turísticas, eventos, lugares para comer e atividades de lazer (TORRENT ET AL. no prelo) conforme mostrado na Figura 15.



Figura 15: Interação com o GregBot

No aplicativo, as recomendações são fornecidas por meio de categorização automatizada baseada em frames de atrações turísticas com base em informações semânticas extraídas de comentários de turistas em plataformas *online*, que são comparadas a informações semânticas extraídas da entrada inserida em uma interface conversacional (COSTA ET AL., 2018). Dessa forma, as informações coletadas a partir dos comentários são combinadas com as informações extraídas das sentenças digitadas pelo usuário na interação com o *chatbot*, de modo que o m.knob possa fornecer recomendações mais específicas sobre atrações e eventos turísticos.

Na Figura 16 podemos observar uma recomendação feita pelo aplicativo, através do GregBot.



Figura 16: Recomendação feita pelo m.knob

Observando a Figura 16, é possível perceber que o usuário buscou por um teatro e o aplicativo recomendou lugares mais próximos ao usuário onde há teatros, no caso nos shoppings da Gávea e Jardim Guadalupe.

Para realizar recomendações aos usuários, o m.knob utiliza-se de uma metodologia automatizada que prevê a existência de um analisador que extrai as informações semânticas dos comentários e as traduz em um conjunto de frames. Costa et al. (2018) destacam que o sistema também gera clusters a partir das entradas do usuário e, posteriormente, mapeia as semelhanças entre os clusters, sugerindo atrações e atividades turísticas que possam atender aos interesses do usuário (cf. COSTA ET AL., 2018 para uma discussão mais detalhada).

A Diciopédia é outra função apresentada no aplicativo. Essa função proporciona ao usuário acesso ao repositório multilíngue de palavras e

conceitos relacionados aos domínios do turismo e dos esportes em português brasileiro, inglês e espanhol. Na Diciopédia, o usuário visualiza as definições e traduções de uma dada palavra ou expressão pesquisada, além, de ver também as palavras relacionadas ao que foi pesquisado. Essa função permite que itens lexicais sejam desambiguados como podemos verificar na Figura 17.

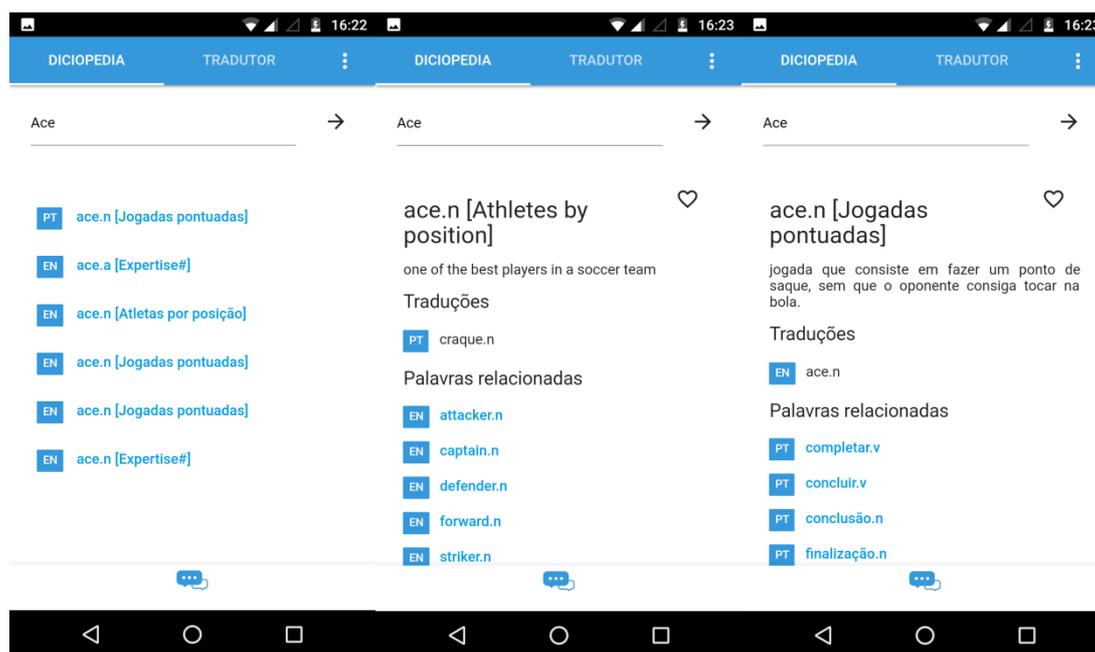


Figura 17: Tela de desambiguação de um item lexical na Diciopédia

A primeira tela da Figura 17 mostra a desambiguação do item lexical “ace” buscado na Diciopédia, visto que este está relacionado a ULs distintas. Isso porque, como explicam Torrent et al. (no prelo) a tela de desambiguação aparece quando há duas ou mais ULs diferentes, com significados distintos e pertencentes a frames diferentes, mas compartilhando a mesma forma escrita. Assim, como vemos na Figura 17, o item lexical “ace”, está cadastrado na base m.knob em frames distintos como Atletas_por_posição e Jogadas_pontuadas. Selecionando a opção cadastrada no frame Atletas_por_posição, o usuário tem acesso à tela seguinte na qual é possível visualizar a definição do item buscado, bem como seus correspondentes nas outras línguas e palavras relacionadas, como apontado na tela do meio da Figura 17. A tela mais à

direita mostra a seleção da opção cadastrada no frame `Jogadas_pontuadas` na qual se apresenta a definição de “ace”, nesse frame, é possível observar também seu correspondente em inglês “ace” e as palavras relacionadas a este termo.

Outra função do aplicativo é o tradutor. Essa função destaca-se pela metodologia aplicada ao sistema tradução, uma vez que esta utiliza-se de um algoritmo que gera melhores equivalentes de tradução a partir de uma base de dados semanticamente enriquecida, tornando-se diferente de tradutores que utilizam apenas estatística e sintaxe (COSTA & TORRENT, 2017). Essa base de dados é composta por um léxico multilíngue enriquecido a partir da adição de novas relações no banco de dados da FrameNet Brasil, como as relações qualia, de modo a promover o adensamento desse banco de dados.

Durante a interação, os usuários podem avaliar a tradução dando uma ou mais estrelas, como é possível notar na Figura 18.

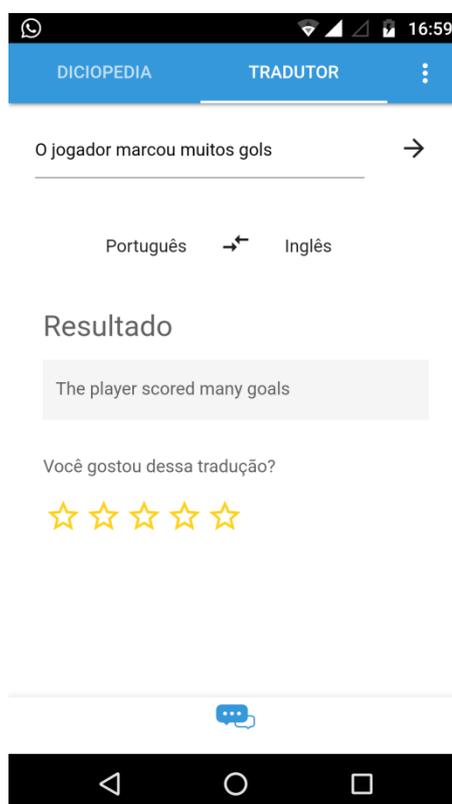


Figura 18: Tradução da sentença “O jogador marcou muitos gols”

Como podemos observar, a FrameNet Brasil desempenha um papel importantíssimo no processamento das informações recebidas e fornecidas pelo m.knob, uma vez que o sistema de recomendações do m.knob, bem como o sistema de tradução de sentenças semanticamente enriquecido, usam a base de dados enriquecida da FN-Br para gerar interpretações semânticas de sentenças em língua natural. Quando tal interpretação é bem sucedida, o GregBot fornece recomendações de atrações para o usuário, que pode, por sua vez, querer saber informações adicionais práticas sobre o lugar recomendado, tais como “Quanto custa a entrada?” ou “Onde eu compro o ticket?”. Entretanto, o aplicativo ainda não opera com construções interrogativas. Sendo assim, é nesse contexto que se insere a modelagem de construções Interrogativas QU-. Integrar essas construções no banco de dados da FN-Br se faz necessário para o desenvolvimento contínuo do aplicativo, uma vez que estas possibilitariam que ele apresente mais informações ao usuário.

Na seção seguinte, apontamos a metodologia empregada nesta dissertação e, na seção 4, apresentamos a modelagem para as construções Interrogativas QU- do PB.

3 METODOLOGIA

Esta seção descreve a metodologia empregada para o desenvolvimento desta dissertação.

3.1 Modelagem das Construções Interrogativas QU-

Seguindo os pressupostos da Semântica de Frames (FILLMORE, 1982), o presente trabalho propõe ao uso um papel relevante para a análise do fenômeno linguístico aqui investigado, uma vez que procura investigar as continuidades que existem entre a linguagem e a experiência.

Antes de dar início às análises das construções Interrogativas QU-, foi necessário constituir um banco de dados específico desse tipo de construção, o qual apresentasse dados escritos provenientes do uso linguístico. Para tanto, inicialmente, realizou-se um o levantamento bibliográfico acerca das construções interrogativas QU- em PB com base em Castilho (2010) e Perini (2010), que apresentam estudos sobre essas construções. A partir disso e também tendo em mente o objetivo final da aplicação m.knob, estabelecemos quais pronomes QU- seriam relevantes para este trabalho, obtendo a seguinte lista: como, onde, qual, quando, quanto, que, quem, por que, por quanto.

Para atestar essas ocorrências, precisamos criar um *corpus* específico para as construções Interrogativas QU- compostos por dados em língua natural. Nesse contexto, a presente pesquisa se enquadra no âmbito da Linguística de *Corpus* e, sendo assim, realiza coleta e análise de *corpora* com auxílio de ferramentas eletrônicas. Como uma das tarefas a qual a FN-Br tem se dedicado é a coleta de outros *corpora* de modo a ampliar sua base de dados, se fez necessário constituir um *corpus* próprio para tratar as construções Interrogativas QU-.

Sardinha (2004, p. 18) define um *corpus* como

um conjunto de dados linguísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizado segundo determinados critérios, suficientemente extenso em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso linguístico ou de algum de seus âmbitos, dispostos de tal modo que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para a descrição e análise.

Segundo Sardinha (2004), a coleta dos dados linguísticos que compõem o *corpus* deve seguir alguns pré-requisitos. O autor destaca que os dados devem ser autênticos e em linguagem natural, isto é, produzidos por falantes nativos e não criados. Ademais, os dados coletados precisam servir ao objeto de estudo e ser representativos para o fim ao qual se destinam, além de serem legíveis por computador.

Para a composição do *corpus Interrogativa QU-* os critérios apontados em Sardinha (2004) foram seguidos, assim, os textos selecionados são autênticos, em linguagem natural e não foram criados para servir a uma pesquisa linguística.

Pelo fato de o aplicativo m.knob abordar a atividade turística no Brasil, priorizou-se a busca por fontes de dados que tratassem dessa realidade. Os textos foram extraídos especialmente de sites de empresas de comércio eletrônico de viagens e de sites de busca. Primeiramente, era preciso que os dados trouxessem informações sobre o domínio do Turismo acerca das atrações, pontos e destinos turísticos e das experiências dos turistas em relação ao lugar. Em segundo lugar, como o objeto do estudo são construções Interrogativas QU-, optamos por coletar dados das FAQs (*Frequently Asked Questions*), visto que estas são definidas em domínios bem específicos e estão estruturadas em forma de perguntas e respostas, o que facilita a extração dos padrões.

A compilação dos *corpora Interrogativa QU-* se deu manualmente. A busca por FAQs disponíveis *online* nos sites Booking.com, CVC, Decolar.com, TripAdvisor e Yahoo Respostas (para o domínio do Turismo) foi feita na internet. Em seguida, os textos foram selecionados e copiados em documentos separados por site, logo, foram reunidos em um arquivo em formato *.txt* para que fossem devidamente armazenados na ferramenta de gerenciamento e

consulta de *corpus* da FN-Br, Webtool 3.0. O *corpus* **Interrogativa QU-** conta com 224 sentenças, das quais, 65 instanciam o pronome como, 2 quem, 42 que, 42 qual, 4 quando, 11 onde, 9 quanto, 9 por que, 1 por quanto e 39 não apresentam pronome QU, tendo sido desconsideradas na pesquisa.

O levantamento das ocorrências das construções QU- no *corpus* informou a modelagem inicial proposta para as construções, juntamente com o levantamento bibliográfico.

Após essa etapa, passou-se a uma proposta de descrição das construções QU- baseada na Gramática das Construções de Berkeley (KAY & FILLMORE, 1999), através da elaboração de Matrizes de Atributo e Valor. Posteriormente, as análises feitas foram convertidas em modelos linguístico-computacionais. Subsequentemente, utilizou-se a ferramenta FrameNet Brasil WebTool 3.0 para armazenar esses modelos na base de dados da FrameNet Brasil, na forma de construções, conforme apresentaremos na seção 4.

3.2 Desenho Experimental

Com o propósito de fornecer uma prova de conceito de que a modelagem proposta poderá ser incorporada como uma das funcionalidades do *chatbot* do m.knob, foi desenvolvido um sistema de teste que possa, a partir da identificação de construções e dos frames evocados por essas construções presentes na pergunta, apresentar respostas estruturadas extraídas de uma base de dados constituída por sentenças em língua natural.

Para o teste de prova de conceito, utilizamos a construção Interrogativa_onde e o frame evocado por ela, Relação_locativa. O teste é feito por um sistema do tipo QAS (*Question Answering System*), desenvolvido pela equipe de TI do Laboratório FrameNet Brasil. Esse sistema usa uma base de sentenças estruturadas simples que foram extraídas do *corpus* de comentários sobre os locais e atrações turísticas disponíveis para recomendação pelo *chatbot* do m.knob e pré-processadas.

Como é uma prova de conceito, o teste se restringe à construção Interrogativa_onde. A funcionalidade a ser testada é a seguinte: uma vez que o m.knob recomendar um ou mais locais ao turista, este poderá perguntar ao *chatbot* mais informações sobre o local recomendado, por exemplo, onde ele fica. Como o processo de incorporação de uma nova funcionalidade ao aplicativo envolve aspectos que estão além do escopo de uma dissertação de mestrado em linguística, tais como soluções de design e programação do *chatbot*, desenvolvemos uma forma de reproduzir tal interação em um ambiente de pesquisa. Para tanto, utilizamos um *corpus* formado por todos os comentários de usuários sobre atrações turísticas no Rio de Janeiro e em Juiz de Fora. Tal *corpus* contém 953.768 palavras, compiladas a partir de comentários de usuários sobre 820 locais.

Foram selecionadas nesse *corpus* todas as sentenças que continham o verbo “ficar” seguido pela preposição “em”, sentenças que apresentam a ocorrência do adjetivo “próximo” seguido pelas preposições “a” ou “de”, bem como do adjetivo “distante” seguido pela preposição “de”. Tal seleção prévia buscou reduzir o universo de processamento, uma vez que, se todo o *corpus* fosse utilizado, o tempo de processamento, que já é lento, seria intratável. Vale lembrar que o tempo de processamento alto é uma decorrência do fato de estarmos trabalhando, ao mesmo tempo, com todos os comentários sobre todos os lugares. Quando a funcionalidade for incorporada ao aplicativo, o usuário fará perguntas sobre, no máximo, 5 lugares por vez (que é o número de locais recomendados a cada interação com o *chatbot*), logo, todas as sentenças dos comentários poderão ser utilizadas. Ao todo, foram extraídas do *corpus* 737 sentenças.

Uma vez extraídas, as 737 sentenças foram submetidas a um programa *offline* que varre a base de sentenças, faz o *parsing* de cada uma e verifica quais são exemplos do frame Relação_locativa. Feito isso, a estrutura da sentença e a estrutura do frame são armazenados em um arquivo.

Logo, quando o usuário digita uma questão no sistema, é feito o *parsing* construcional da questão, para identificar se ela é uma construção do tipo

Interrogativa_onde. Se a sentença for uma construção Interrogativa_onde, é feito o mapeamento dos ECs com os EFs do frame Relação_locativa. Para essa etapa também é criada uma estrutura para a questão que é armazenada na base.

Em seguida, cria-se a rede com os elementos das sentenças da base e é nesse processo que se define qual EF do frame instanciado na sentença é a resposta para a questão digitada pelo usuário. No caso de uma sentença licenciada pela construção Interrogativa_onde, o EF Fundo do frame Relação_locativa pode substituir o “onde” na questão, isto é, este EF será a resposta à questão. A seleção de resposta é realizada pelo processo de escolha da melhor resposta a pergunta em linguagem natural a partir do peso de cada uma, assim a(s) resposta(s) com maior peso é (são) selecionada(s).

Para o reconhecimento automático das construções Interrogativas QU-, utilizamos uma técnica de travessia em redes chamada Ativação Propagada ou *Spreading Activation*. Essa técnica é aplicável a redes em processos de desambiguação lexical e estas são chamadas de Redes de Ativação Propagada (RAP) (MATOS, 2014). Essa técnica é aplicada a um conjunto de nós de origem e estes recebem um determinado nível de ativação e um peso, assim, cada nó ativado “espalha” sua ativação aos nós ligados a ele. O nível de ativação diminui em cada passagem por um nó. Quando uma condição de parada, previamente estabelecida, é alcançada ou quando o nível de ativação fica abaixo de um certo limiar previamente definido (*threshold*), a ativação não é mais realizada (MATOS, 2014).

No nosso teste, os nós referentes aos ECs são ativados. O peso atribuído a cada *link* é sempre 1.0, assim, a propagação da ativação é iniciada no nó de origem (100%). Logo, a ativação “espalha” até alcançar (ou não) alguns frames, mas atinge com o peso maior no frame de Relação_locativa, ativando os EFs Figura e Fundo. Em seguida, o sintagma que corresponde ao EF ativado é reconstruído. Por fim, o usuário tem acesso a uma tabela que mostra o peso, o sintagma em que o EF ocorre e a sentença original.

A reconstrução do sintagma é feita, no teste, utilizando-se a ferramenta UD Pipe (STRAKA ET AL., 2016), uma vez que o *parser* construcional da FrameNet Brasil, o CARMA (MATOS ET AL., 2017) ainda não contém uma base de construções que dê conta de todas as possibilidades de organização sintagmática das respostas. O UD Pipe (acessível em <http://lindat.mff.cuni.cz/services/udpipe/>) cria representações de sentenças na forma de árvores de dependência baseadas nas *Universal Dependencies* (UD) (MCDONALD ET AL., 2013).

A rede apresentada na Figura 19 mostra a ativação para a questão “Onde fica o hotel Ritz?”.

Os nós verdes são os ECs da construção Interrogativa_onde, os azuis escuros representam os EFs da base e as palavras que os formam, os azuis claros são palavras que estão no escopo dos EFs da questão e os nós vermelhos representam os frames.

Na Figura 19 podemos perceber que o EC Pronome_QU, “onde”, ativa a rede de modo a chegar ao EF Fundo do frame Relação_locativa. Dessa forma, o sistema oferece respostas do tipo “o hotel ritz fica no centro da cidade”, “o hotel fica na cidade”, “o hotel fica fora da cidade”, “o hotel fica próximo do museu”, cada uma com um peso.

O ambiente de testes pode ser acessado em <http://server2.framenetbr.ufjf.br/devel/qas/locative/#>.

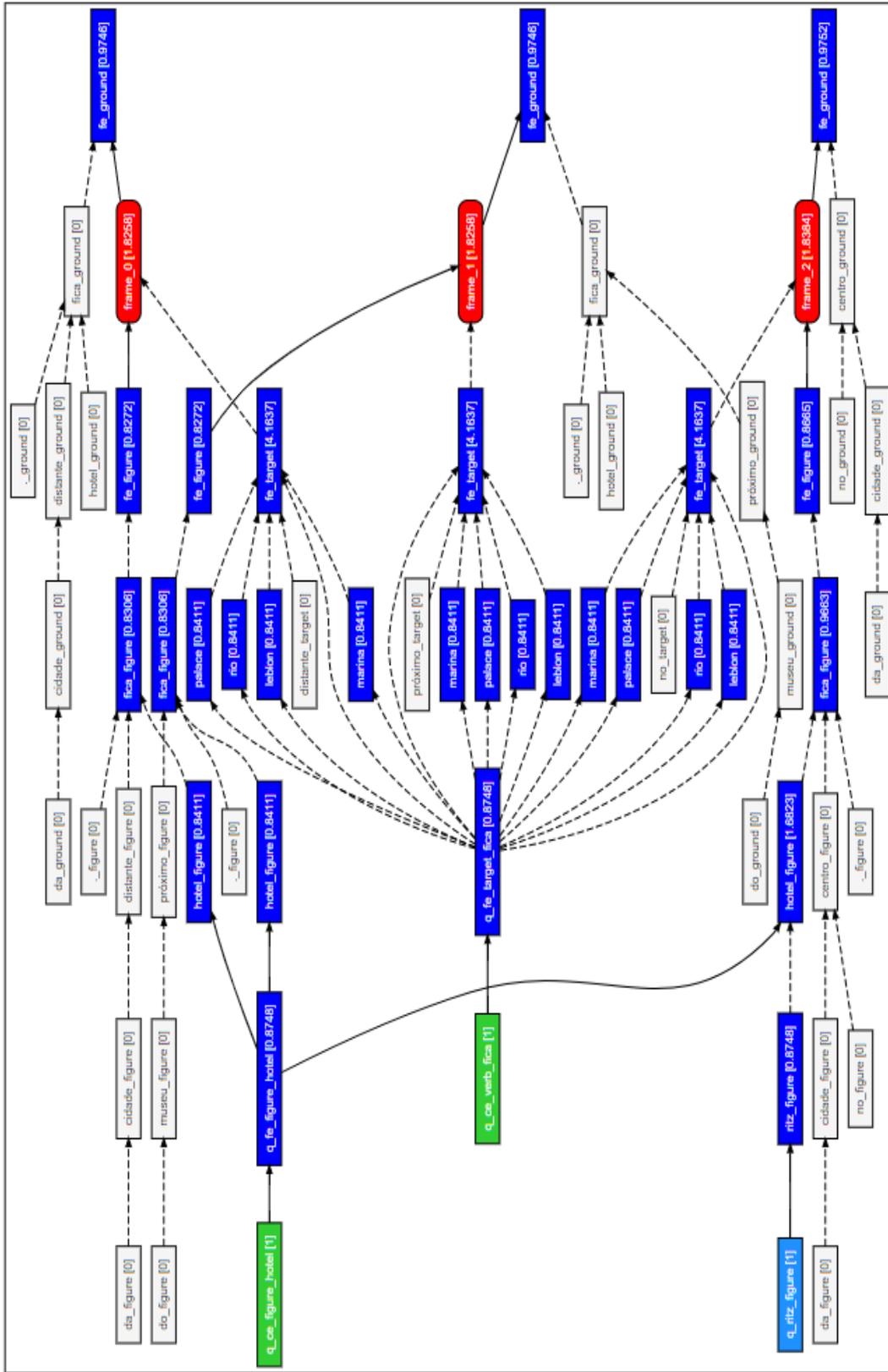


Figura 19: RAP da questão "Onde fica o hotel Ritz?"

4 MODELAGEM DAS CONSTRUÇÕES INTERROGATIVAS QU NO CONSTRUCTICON DA FRAMENET BRASIL

Este capítulo apresenta a modelagem da rede de construções Interrogativas QU-, realizada no âmbito desta dissertação de mestrado. Antes, entretanto, apresentamos um levantamento de tratamentos anteriores dispensados ao fenômeno.

4.1 Propostas anteriores para a descrição das construções QU em Português do Brasil

Nas subseções a seguir, discorreremos sobre os principais trabalhos que se debruçaram anteriormente sobre as construções Interrogativas QU-.

4.1.1 Abordagens Gerativistas para as Construções QU

Vários estudos gerativistas têm como foco as construções Interrogativas QU- do PB e muitos deles pautam suas análises no Programa Minimalista (doravante PM), “principalmente devido aos fenômenos sintáticos nelas envolvidos” (MIOTO, 1994, p. 19).

O Programa Minimalista é reconhecido como a versão atual da Gramática Gerativa desenvolvida por Noam Chomsky, o qual procurou desenvolver a teoria de Princípios e Parâmetros. Segundo Modesto

É um 'programa' no sentido de não ser uma teoria completa, mas um conjunto de diretrizes metodológicas que se pretende usar para chegar a uma teoria. É também 'minimalista' por tentar minimizar o aparato técnico, produzindo uma teoria a mais simples possível. (MODESTO, 2012, p. 130)

Esse programa sintático propõe dois níveis de representação, os quais são responsáveis pelo pareamento entre o som e o sentido, isto é, permitem a articulação dos sons e sua interpretação. O nível de representação Forma Fonética é responsável por fazer a interface com o sistema de performance articulatório-perceptual (interface fonético-fonológica) enquanto o nível de

representação Forma Lógica é responsável por fazer a interface com o sistema de performance conceitual-intencional (interface semântica) (RODRIGUES & AUGUSTO, 2009).

Na visão minimalista, Rodrigues & Augusto (2009) apontam que o léxico suporta todas as regras e traços de uma dada língua, assim, os itens lexicais constituem conjuntos de traços do tipo formal que representam propriedades gramaticais como gênero, número e pessoa e, também, de traços que representam propriedades fonéticas e semânticas. O sistema computacional é sensível às informações contidas no léxico.

A derivação sintática do PM, segundo Modesto (2012), começa pela montagem de uma lista numerada de itens lexicais e seus respectivos traços, chamada de Numeração. Nela, cada item recebe um índice que indica quantas vezes aquele elemento pode ser acessado pelo sistema computacional. A combinação de elementos listados na Numeração se faz pela atuação da operação *Merge* (concatenar). A derivação só termina quando todos os itens da lista forem usados e quando todas as aplicações de movimento exigidas tiverem sido efetuadas.

O PM procura estabelecer uma relação entre o como e o porquê do movimento. Isso porque nem todas as línguas apresentam esse movimento. Augusto (2005) pontua que existem línguas cujo movimento de elementos QU- em interrogativas é obrigatório, como o inglês, em outras línguas não há movimento, como o chinês. Existem também línguas em que o movimento ocorre de forma parcial, isto é, o elemento QU- se move até a periferia esquerda da sentença encaixada, mas não chega à periferia esquerda da sentença matriz, como o alemão. Por fim, há línguas em que o movimento não é obrigatório, caso do português, que apresenta a possibilidade de movimento para a periferia esquerda ou pronome QU- *in situ*, isto é, na posição canônica associada à sua função gramatical.

Por isso, o fenômeno sintático de maior interesse se relaciona ao movimento do elemento QU- em construções interrogativas em Português

Brasileiro, uma vez que existe a possibilidade do movimento, mas este não é obrigatório, como vemos em (14).

(14) a. Você viu **quem**?

b. **Quem** você viu?

Modesto (2012) propõe que a existência de movimento no PM é explicada através da postulação de traços formais (ou sintáticos). Os itens lexicais seriam formados por traços semânticos, formais e, possivelmente, fonológicos. Os traços semânticos podem ser lidos apenas na interface interpretativa; os fonológicos, apenas na interface articulatorio-perceptiva. Os traços formais não são interpretáveis pelas interfaces e, por isso, têm de ser eliminados da derivação sintática através de uma operação de checagem. Em versões mais recentes, o movimento é na verdade a soma das operações *Copy* (copiar) + *Merge* (concatenar).

Quando o elemento QU- se move, deixa um a cópia em sua posição de origem. Uma das cópias permanece realizada foneticamente, geralmente a mais alta na sentença, e a outra será apagada no componente fonológico.

Vejamos o exemplo a seguir:

O que fazer em Belém θ -que?

Observe que o elemento QU- “o que” foi movido através da operação de cópia, assim criou-se uma cadeia com dois termos idênticos (cópia 1, cópia 2). Após a estrutura ser mandada para *Spell-Out* (pronunciar em inglês), uma das cópias deve ser apagada por algum mecanismo do componente fonológico. Nas línguas em geral, a cópia apagada é a que ocupa a posição mais baixa da cadeia.

Se o movimento do elemento QU- ocorre antes do *Spell-Out*, isto é, quando ocorre no componente aberto, significa que este movimento será visível, já que, quando um traço é movido em sintaxe visível, as propriedades do componente fonológico requerem que os demais traços do elemento QU- sejam movidos também, uma vez que a morfologia não consegue ‘ler’ traços

isolados. No caso de movimento encoberto, isto é, quando a operação de movimento ocorre depois do *Spell-Out*, não há a realização fonológica desse movimento.

Adger (2003) argumenta em favor de que um outro fator possa desencadear o movimento, no caso, a presença de um traço [Q] forte que precisa ser satisfeito. Segundo Adger, todas as línguas possuem uma categoria funcional C (complementizador), que determina o tipo de sentença. As construções interrogativas, por exemplo, apresentam um traço Q (Question). Esse traço [Q] é interpretável, sendo assim, ele não precisa ser checado, a menos que seja forte. Neste caso, ele deve ser eliminado imediatamente para que a derivação não seja cancelada. A forma de eliminar um traço forte é através da checagem que, nesse caso, é realizada movendo o elemento QU- para a posição de *Spec* da projeção máxima CP. Sendo o traço [Q] fraco, a checagem não se faz necessária.

Na projeção CP, na qual a posição de *Spec* é preenchida pelo elemento QU-, o traço [Q] é considerado forte, quando não é preenchida, o traço é considerado fraco. Desse modo, o PB se caracteriza como uma língua que possui o traço [Q] forte e fraco.

Como vimos, o elemento QU- pode estar movido ou estar *in situ*, ou seja, os pronomes interrogativos QU- aparecem em posições distintas daquelas em que são interpretadas semanticamente. Dessa maneira, eles precisam ser relacionados à sua função na estrutura argumental da sentença (AUGUSTO, 2005).

As construções Interrogativas QU- se mostram interessantes não só pelo fato de apresentarem constituições distintas, mas também por proverem evidências para a atuação de movimento sintático dos pronomes interrogativos QU- (AUGUSTO, 2005). Ademais, esse tipo de construção implica saber que tipo de informação está envolvida em sua formulação e na sua interpretação, já que os pronomes interrogativos QU- sinalizam solicitações de informações.

Além do gerativismo, as construções Interrogativas QU- se mostram relevantes em estudos de outras áreas, como, por exemplo, o funcionalismo, como veremos na seção que se segue.

4.1.2 Abordagens Funcionalistas para as Construções QU

Nas últimas décadas, a pesquisa funcionalista tem se voltado para o tratamento da dimensão contextual dos usos linguísticos, levando em consideração também os processos de mudança na língua, bem como os fatores linguísticos e não-linguísticos que influenciam nessa mudança. Dessa forma, na trajetória dos estudos linguísticos baseados no uso, observou-se que havia aspectos que acenavam para a possibilidade de uma aproximação teórica entre o Funcionalismo e a Sociolinguística Variacionista. Assim, surgiu a abordagem denominada de “sociofuncionalismo”.

Desde a década de 1980, vêm se desenvolvendo pesquisas de orientação sociofuncionalista as quais se dedicam ao estudo de fenômenos de variação e de mudança linguística, analisando e explicando esses fenômenos com base nos pressupostos teórico-metodológicos da Sociolinguística Variacionista e do Funcionalismo Linguístico norte-americano (TAVARES, 2013).

Nesse sentido, Kato et al. (1996) e Oushiro (2011), procuram investigar os fatores que influenciam a variação da ordem dos constituintes sintáticos das construções QU- do PB.

Antes de falar da constituição sintática dessas construções, Kato et al. (1996) definem primeiramente o que é uma construção QU- (em inglês, Wh-constructions). Voltando aos exemplos apresentados na introdução deste trabalho, observem-se as sentenças:

- (1) **Qual** a melhor região de Manaus para se hospedar e a que fica mais próxima do aeroporto?
- (2) **Como** chegar em Barreirinhas?

- (3) **Quanto** custa viajar para Fernando de Noronha?
- (4) **O que** fazer no Rio além de visitar o Pão de Açúcar e o Cristo?
- (5) **Quem** disse que albergue é coisa de mochileiro?
- (6) **Quando** ir à Chapada Diamantina?
- (7) **Onde** ficar nos Lençóis Maranhenses?
- (8) **Por que** se hospedar em Olinda?
- (9) **Por quanto** tempo a pré-autorização vai afetar o meu crédito disponível?

De acordo com Kato et al. (1996), essas sentenças seriam todas chamadas de construções Interrogativas QU- por conterem palavras do paradigma morfológico dos pronomes-Q (para nós, pronomes QU-), destacados em negrito nas sentenças.

Existem dois tipos preferenciais de variação sintática das construções Interrogativas QU-. No primeiro, o pronome QU- está *in situ*, isto é, em seu lugar original. Assim sendo, a ordem dessa construção equivaleria à da declarativa S-V-O (sujeito-verbo-objeto) e, no outro tipo, o pronome QU- aparece movido, no início da sentença, podendo mudar a estruturação para uma ordem V-S (verbo-sujeito) e é nesse segundo tipo que este texto se foca. Isso ocorre pelo fato de que, em muitas línguas, assim como em PB, esse tipo de palavras tem a propriedade de se mover para o início da sentença. Dessa maneira, é necessário observar qual tipo de mecanismo que desloca sintagmas de sua posição de base para outras posições na sentença e o que motiva esse movimento.

Kato et al. (1996) propõem que, quando esse movimento ocorre, o pronome QU-, pela sua natureza sintagmática, ocupa a posição de *Spec* (Especificador) da projeção máxima (CP) da categoria funcional C (complementizador). Essa categoria funcional na representação arbórea ficaria em um nível acima no da projeção IP (*Inflectional Phrase*). Entretanto, quando esse movimento não ocorre, também não há a projeção da categoria funcional

C. Desse modo, as autoras argumentam que uma sentença como “De que o Pedro gosta?” seria representada sintaticamente da seguinte maneira:

[CPDe que [IPo Pedro gosta?]

Aqui podemos observar que o elemento QU- “De que” ocupa a posição de *Spec* de CP, enquanto o restante da sentença fica dentro da projeção IP.

Sem movimento, as autoras afirmam que a ordem da sentença é a mesma da declarativa, no caso, a sentença seria “O Pedro gosta de que?” e sua representação sintática seria:

[IPo Pedro gosta de que?]

Nessa representação, podemos verificar que não existe a categoria funcional, assim a projeção termina no nível IP e o elemento QU- “De que” ocupa a posição de objeto na sentença.

Kato et al. (1996) e Oushiro (2011) verificam a ordem dos constituintes nas interrogativas QU- de modo a esclarecer em quais contextos a variação pode ou não ocorrer e em quais contextos apenas uma das formas pode ocorrer. Kato et al. (1996) utilizam um corpus composto quinze inquéritos que abrangem três situações de uso da língua oral que atestaram 200 ocorrências de sentenças interrogativas QU. Já Oushiro (2011) desenvolve um estudo no qual os dados foram coletados através de entrevistas e de textos formais de redações de alunos de graduação da USP, além de textos da Revista Veja e do jornal Folha de São Paulo.

Kato et al. (1996) focam seus estudos em aspectos linguísticos, assim as autoras defendem que o tipo preferencial de ordenação dos constituintes nas interrogativas QU- do PB seria a ordem V-S. As autoras procuraram não só analisar a ordem das construções Interrogativas QU-, mas também verificar quais fatores estão favorecendo ou não a ocorrência desse fenômeno. Assim, foram considerados fatores como o tipo de verbo e tipo de sujeito.

Os resultados do estudo mostraram que a inversão V-S ocorreu em pouco mais da metade dos dados analisados, demonstrando assim que existe um equilíbrio nas ocorrências entre as ordens S-V e V-S. Entretanto, as

autoras destacam que o mais importante é verificar em que contexto cada uma delas aparece.

A variável que foi considerada como o fator mais relevante para a estruturação das interrogativas QU- é o **tipo de verbo**. A inversão ocorreu com verbos copulares. Os verbos intransitivos e reflexivos demonstraram um comportamento neutro nesse contexto. Já os verbos transitivos inibiram a inversão.

Outra variável considerada no estudo foi o **sujeito**. Foram considerados três grupos de fatores fortes para a ordenação dos constituintes nas sentenças interrogativas QU- relacionados ao sujeito. O grupo mais importante de fatores que influencia a ordem dos constituintes é o do tipo de sujeito. Os sujeitos não pronominais favorecem a inversão, enquanto os sujeitos pronominais tendem a aparecer antes do verbo.

O segundo grupo de fatores relevantes para a ordem V-S é o da **dimensão do sujeito**. Os dados apresentados pelas autoras apontam que sujeitos maiores tendem a aparecer depois do verbo nesse tipo de sentença e sujeitos menores tendem a aparecer antes do verbo.

O terceiro grupo classificava o sujeito entre definido e indefinido, contudo essa variável não se mostrou relevante ao estudo.

O **expletivo “é que”** é outra variável considerada no estudo. Os resultados apresentados pelas autoras apontaram que a ausência do expletivo “é que” favorece a ocorrência da ordem V-S.

A última variável apresentada no estudo, refere-se ao **tipo de elemento Q** (ou QU-, como estamos utilizando ao longo desse texto). Os dados descritos demonstram que o elemento **como** possui um comportamento neutro, ocorrendo tanto em ordem V-S como em ordem S-V. Já com o grupo interrogativo **Que+NP**, a ordem V-S é menos frequente.

Como podemos observar, de um modo geral, existem variáveis que mais diretamente afetaram a estruturação sintática das sentenças interrogativas QU-, sendo elas o **tipo de verbo** e o **tipo de sujeito**. Com isso, as autoras concluem que a ordem V-S ocorre quase que exclusivamente com verbos

copulares. Além disso, a dimensão do sujeito também influencia de maneira significativa a ordem das sentenças, visto que é um fator que predispõe a inversão.

Os estudos de Oushiro (2011) apontam variáveis linguísticas e não linguísticas que favorecem o uso de uma determinada forma das interrogativas QU. Contudo, para o nosso propósito, focaremos apenas nas ocorrências de QU movido e *in situ*.

Oushiro (2011) aponta que fatores extralinguísticos como **quem fala**, **escolaridade** e **sexo/gênero** se mostraram neutros, isto é, não tiveram uma influência no emprego das formas QU movido e *in situ*. Já fatores extralinguísticos como **Faixa etária**, **Grau de relação** entre documentador e informante, e **Espontaneidade** da pergunta se mostraram como significativos, favorecendo o emprego da forma QU *in situ*.

Oushiro (2011) também explora a variação estrutural que ocorre em construções do tipo QU. A autora elenca os contextos sintáticos em que cada forma de Interrogativa QU- podem ocorrer, assim, sentenças interrogativas com orações principais, orações com e sem verbo finito e orações com sentenças encaixadas com complementizador permitem a alternância das formas do QU movido e *in situ*. No caso de orações encaixadas sem complementizador ocorre apenas a forma movida. Já a forma *in situ* ocorre em contextos de ilhas sintáticas.

Os contextos da oralidade e da forma escrita também foram verificados por Oushiro (2011). A autora aponta que, apesar de que na oralidade existe uma preferência em relação ao uso das interrogativas QU- com o QU movido, a forma *in situ* também apresenta números significativos, de modo a ser considerada uma forma concorrente à forma com QU movido. Entretanto, nos dados escritos a preferência em absoluto é pelo uso da forma com o QU movido.

Como podemos observar, de uma forma geral, existem variáveis que mais diretamente afetaram a estruturação sintática das sentenças interrogativas QU.

A ordenação dos constituintes da estrutura argumental, bem como os aspectos que influenciam na variação entre a ordem S-V e V-S não são relevantes para a nossa pesquisa, pois não estamos analisando as estruturas argumentais que compõem as construções Interrogativas QU-. Ainda que nosso escopo se restrinja às Interrogativas QU- com QU movido, variáveis que influenciam nesse movimento são desconsideradas na nossa análise. Ademais, fatores extralinguísticos são, por hora, totalmente descartados para a modelagem das construções Interrogativas QU-, uma vez que o Constructicon lida ainda apenas com os aspectos sintático-semânticos das construções para uma representação computacional destas. Entretanto, como nosso objetivo de continuidade da pesquisa envolve a análise da estrutura argumental base, as análises funcionalistas aqui levantadas podem informar um posterior adensamento do modelo.

4.2 Proposta Construcionista para as Construções QU

No empreendimento cognitivista de tratamento da gramática, foram desenvolvidas várias Gramáticas das Construções cada uma seguindo diferentes abordagens baseadas ou não em unificação. Assim é possível falar, por exemplo, em Gramática Cognitivista das Construções (GOLDBERG, 1995), Gramática das Construções de Berkeley (KAY & FILLMORE, 1999) ou ainda Gramática das Construções Baseada em Signos (SAG, 2012). No âmbito desta última, encontra-se uma proposta de tratamento para as construções QU-.

Sag (2012) classifica as construções Interrogativas QU- como sendo um tipo de construção *filler-gap*, uma vez que, nesse tipo de construção, o sintagma QU- preenche a posição inicial da sentença enquanto é deixada uma lacuna na posição em que esse sintagma é interpretado. Dessa forma, o constituinte deslocado para frente da sentença, ou seja, para a posição mais periférica à esquerda, está ligado a uma lacuna em um local específico em outra parte da sentença, de tal forma que, no lugar da lacuna, se mantêm as

propriedades do constituinte de modo a tornar a sentença interpretável (FILLMORE, 1985b). Como exemplo, considere a sentença (15).

(15) Que livro João leu__?

Nessa sentença observamos que o constituinte QU- foi movido para a posição periférica à esquerda e deixou uma lacuna, na qual ele é interpretado.

Fillmore (1985b) questiona o fato de os estudos gerativistas proporem que o constituinte QU- é movido para a posição inicial da sentença por uma regra de movimento e que mantém uma ligação a lacuna deixada por ele, assim, além de carregar as propriedades necessárias para que ele seja interpretado na sua posição original, ao se mover, esse constituinte adquire as características exigidas por sua nova posição. Sob uma perspectiva construcionista, Fillmore (1985b) propõe que os esforços devem se concentrar nas propriedades do constituinte em posição inicial da sentença que não concordam com o que é permitido ou exigido no local da lacuna vinculada a ele, isto é, devem focar nas características da posição para a qual o constituinte QU- foi movido.

Para analisar a dependência existente entre o constituinte movido e a lacuna deixada por este, Sag (2012) discorre que pode haver uma discrepância entre a lista estrutura argumental (EST-ARG) de uma palavra e a lista valência (VAL) e que isto surge pela realização não-local do constituinte, em que uma das valências da palavra aparece em uma posição sintática deslocada. Dessa maneira, para formalizar essa discrepância, uma análise de traços permite que um núcleo lexical apareça sem um valente (sujeito, objeto ou outro complemento) e a lista GAP contém um elemento correspondente a esse valente, assim

a presença de uma lacuna é codificada em termos de uma especificação não vazia para o traço GAP (por exemplo, [GAP <NP>]). Por contraste, uma expressão que não contenha lacunas desvinculadas é especificada como [GAP <>]⁸ (SAG, 2012, p. 163).

⁸(...) “the presence of a gap is encoded in terms of a nonempty specification for the feature GAP (e.g. [GAP <NP>]). By contrast, an expression containing no unbound gaps is specified as [GAP <>]”.

Sag (2012) destaca que, em construções do tipo *filler-gap*, ocorre um processo de Percolação. Esse processo determina que a informação dos contextos sintáticos mais baixos seja passada para contextos sintáticos superiores por restrições simples: quando todas as filhas (daughters - DTRS) têm uma especificação vazia para o traço em questão, o mesmo acontece com a mãe (mother - MTR); mas, quando uma das filhas tem uma especificação não-vazia, então a mãe carrega a mesma especificação não-vazia. Assim, as especificações GAP são herdadas de acordo com formalização exposta na Figura 20.

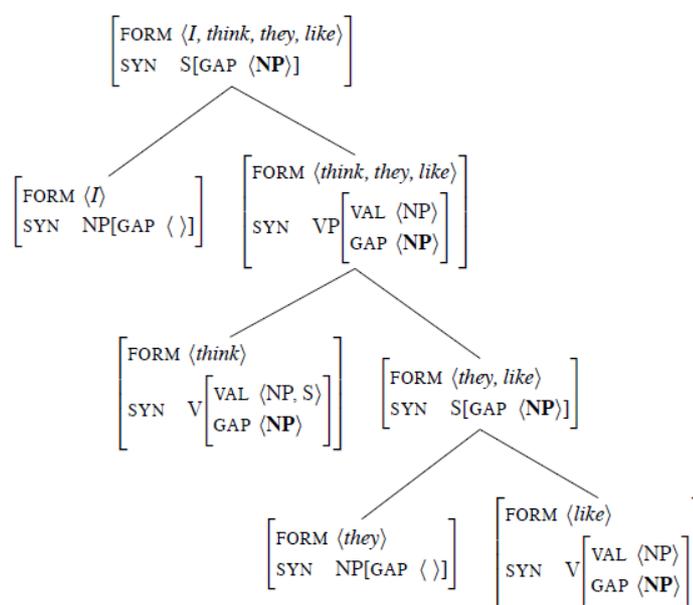


Figura 20: Representação do processo de Percolação (SAG, 2012, p. 163)

Aqui podemos ver que a especificação não vazia para o traço GAP ([GAP <NP>]) foi “filtrada” pelas filhas e passada aos níveis superiores.

Para o traço QU (WH) o mesmo processo de Percolação ocorre. Assim como especificações não-vazias para o GAP indicam a presença de uma lacuna não-vinculada, as especificações QU (WH) não vazias marcam a presença de uma palavra QU (WH) interrogativa não vinculada no

preenchimento. Esse processo justifica o fenômeno “*pied-piping*” que pode ocorrer nas construções Interrogativas QU em que o constituinte QU leva consigo o sintagma que o abrange. Dessa maneira, o traço QU tem um valor não-vazio que percolou a partir do constituinte QU interrogativo que ele contém, assim, em uma sentença como (15), o sintagma “que livro” se moveu para a frente da sentença por conta do valor não-vazio do constituinte QU.

Segundo Sag (2012) as construções Interrogativas QU- seguem as restrições propostas às construções do tipo *filler-gap*.

Conforme o que foi exposto aqui, podemos dizer que a SBCG fornece uma abordagem favorável ao tratamento das construções Interrogativas QU-, uma vez que essa abordagem alinha a semântica de frames às formalizações propostas em forma de Matrizes de Atributo e Valor (AVMs), contribuindo, assim, para um tratamento computacional destas construções. Ademais, contribui para a modelagem de construções realizada no Constructicon, na medida em que aponta o caminho do que precisa ser mapeado entre o elemento QU e a estrutura argumental base onde fica o gap.

Na seção seguinte apresentamos a proposta de modelagem de construções Interrogativas QU- no Constructicon da FrameNet Brasil, desenvolvida no âmbito desta dissertação de mestrado.

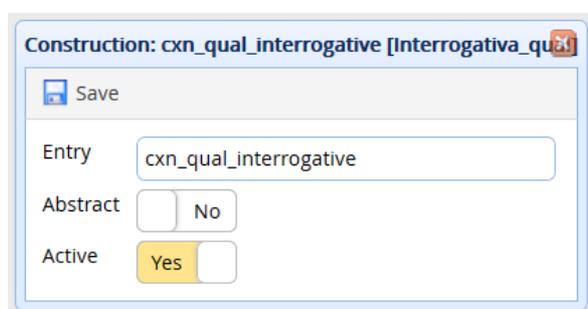
4.3 Proposta de Modelagem das Construções Interrogativas QU- no Constructicon da FrameNet Brasil

O processo de modelagem das construções Interrogativas QU- no Constructicon da FrameNet Brasil se deu por algumas etapas e estas serão apresentadas nessa seção.

No processo de modelagem das Construções QU-, primeiramente se fez necessário modelar duas construções mais abstratas, a construção Interrogativa_QU e a construção Interrogativa_QU_preposicionada, as quais funcionam como nós abstratos das demais construções Interrogativas modeladas e não licenciam construtos na língua. Tais construções cumprem a

função de organizar as demais em rede, além de congregar informações que sejam compartilhadas por todas as construções herdeiras. Para ilustrar o processo de modelagem das construções realmente instanciáveis, utilizaremos a construção Interrogativa_qual.

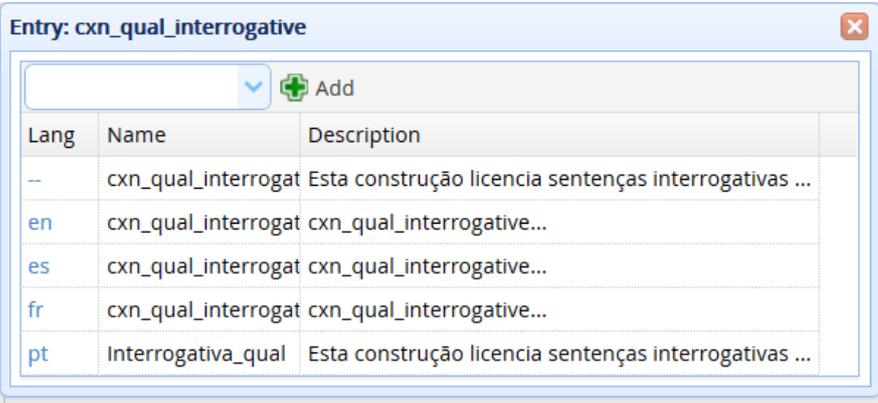
Como vimos na seção 1.4, o primeiro passo da modelagem é definir os Elementos da Construção (ECs), sendo assim, a construção Interrogativa_qual, por exemplo, é composta pelos ECs Estrutura_argumental_base e Pronome_qual. Nesse contexto, a inserção da construção Interrogativa_qual começa a partir da indicação de três informações, como indica a Figura 21.



The image shows a software window titled "Construction: cxn_qual_interrogative [Interrogativa_qual]". Inside the window, there is a "Save" button at the top left. Below it, there are three fields: "Entry" with a text input containing "cxn_qual_interrogative", "Abstract" with a radio button and the label "No", and "Active" with a radio button and the label "Yes".

Figura 21: Edição de entrada da construção Interrogativa_qual

O nome interlingual da construção é registrado em "Entry". É esse nome que vai identificar a construção, uma vez que este é único. O nome `cxn_qual_interrogative` é composto pelo termo "cxn" que é uma abreviação de construção, e o restante do nome é uma especificação da construção em questão (Interrogativa_qual). Logo, apontamos se a construção é abstrata ou não, isto é, se ela licencia ou não outras construções. Por fim assinalamos se essa construção está ou não ativa na base de dados, ou seja, se sua modelagem pode ser considerada completa e apta para aplicações em CLN. Feito isso, salvamos e em seguida uma nova tela se abre, a qual é apresentada na Figura 22.

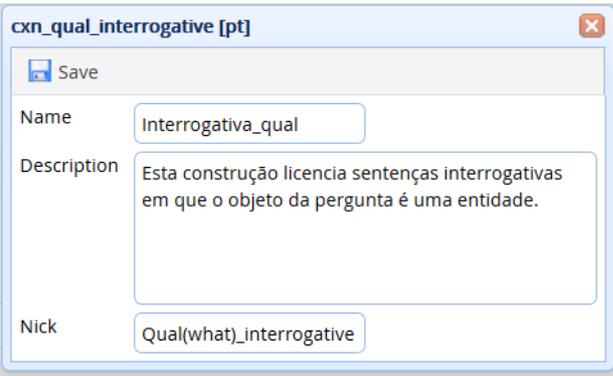


The screenshot shows a window titled "Entry: cxn_qual_interrogative" with a close button in the top right. Below the title bar is a search bar with a dropdown arrow and an "Add" button with a green plus icon. The main content is a table with three columns: "Lang", "Name", and "Description".

Lang	Name	Description
--	cxn_qual_interrogat	Esta construção licencia sentenças interrogativas ...
en	cxn_qual_interrogat	cxn_qual_interrogative...
es	cxn_qual_interrogat	cxn_qual_interrogative...
fr	cxn_qual_interrogat	cxn_qual_interrogative...
pt	Interrogativa_qual	Esta construção licencia sentenças interrogativas ...

Figura 22: Edição de entrada por idioma

Nessa tela, se clicarmos nas abreviações dos idiomas, podemos editar a entrada em cada idioma de forma legível a humanos. Clicando em “pt” somos redirecionados a uma outra tela, como vemos na Figura 23.



The screenshot shows a window titled "cxn_qual_interrogative [pt]" with a close button in the top right. Below the title bar is a "Save" button with a floppy disk icon. The main content is a form with three fields: "Name", "Description", and "Nick".

Name: Interrogativa_qual

Description: Esta construção licencia sentenças interrogativas em que o objeto da pergunta é uma entidade.

Nick: Qual(what)_interrogative

Figura 23: Edição da definição da construção em Português Brasileiro

Nessa nova tela é possível alterar o nome da construção e criar sua definição, ambos em Português Brasileiro. Para a identificação da construção por usuários não proficientes em PB, em “Nick” aparece o nome da construção em inglês. Feito isso, a construção está criada, assim partimos para a criação e definição dos ECs.

Figura 24: Edição da entrada do Elemento de Construção Pronome_qual

A tela de edição do EC é apresentada na Figura 24. Nesta tela, primeiramente, insere-se o nome interlingual do EC, o termo “ce” é a sigla em inglês para Elemento de Construção, “qual_pronoun” refere-se ao fato de o EC ser um pronome, no caso o pronome qual, enquanto a parte “qual_interrogative” se refere à construção a qual esse EC pertence. Depois escolhemos a cor que identificará esse EC e marcamos se ele é opcional, isto é, se ele precisa constar na sentença que instancia a construção, no caso esse elemento não é opcional. Logo marcamos se esse EC é “head” ou não, no caso, esse EC é nuclear pois possui um núcleo que deve ser identificado pelo *parser*, uma vez que os traços do pronome QU serão assumidos pela construção como um todo. Informamos ainda que o EC não é múltiplo, visto que não há a possibilidade de vários pronomes QU serem concatenados recursivamente nessa construção sem que haja uma coordenação. Isso feito, salvamos e uma nova tela aparece, como é mostrado na Figura 25.

Lang	Name	Description
--	ce_qual_pronoun_q	Palavra interrogativa que define o objeto da pergu...
en	ce_qual_pronoun_q	ce_qual_pronoun_qual_interrogative...
es	ce_qual_pronoun_q	ce_qual_pronoun_qual_interrogative...
fr	ce_qual_pronoun_q	ce_qual_pronoun_qual_interrogative...
pt	Pronome_qual	Palavra interrogativa que define o objeto da pergu...

Figura 25: Edição entrada por idioma do Elemento de Construção Pronome_qual

A Figura 25 mostra a possibilidade de editarmos o EC em quatro idiomas, sendo inglês, espanhol, francês e português, respectivamente. Clicando em “pt” abre-se uma nova tela, conforme exibe a Figura 26.

Figura 26: Edição da definição do Elemento de Construção Pronome_qual em Português Brasileiro

Aqui mantemos o nome e criamos a definição do EC em Português Brasileiro, mas mantemos o nome em inglês do EC em “Nick”.

Os mesmos processos são realizados para o cadastramento do EC Estrutura_argumental_base, outro EC que compõe a construção.

Feita a atribuição dos ECs, estabelecemos as restrições de constituição da construção. Como vimos na seção 1.4, a constituição tipifica os signos filhos – os ECs – da construção em termos de outras construções. Assim, podemos assinalar se um EC for composto por uma outra construção já definida no Constructicon. No caso da construção Interrogativa_qual, seus ECs possuem essa restrição, o EC Estrutura_argumenta_base é licenciado pela construção Sujeito_predicado e o EC Pronome_qual é licenciado pela construção Pronome_QU, além disso, a restrição Lexeme (lex) especifica que é o pronome “qual” que deve ocupar a posição do pronome, como podemos ver na Figura 27.

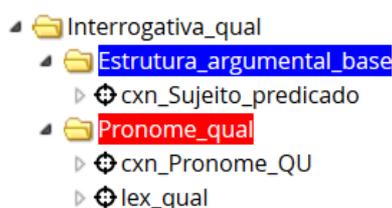


Figura 27: Restrições aplicadas à construção Interrogativa_qual

Ao final, a construção cadastrada se apresenta no Constructicon na forma da tela da Figura 28.

Interrogativa_qual [Qual(what)_interrogative]

Definição	
Esta construção licencia sentenças interrogativas em que o objeto da pergunta é uma entidade.	
Exemplo(s)	
Elementos da Construção	
Estrutura_argumental_base [Base_arg_structure]	Sentença sobre a qual o Pronome_qual atua na definição do tópico da pergunta como uma coisa ou uma pessoa.
Pronome_qual [Qual(what)_pronoun]	Palavra interrogativa que define o objeto da pergunta como uma entidade.
Relações	
Evoca	Entidade
Herda de	Interrogativa_QU

Figura 28: Construção Interrogativa_qual

Mais acima vemos o nome da construção, seguido do seu nome interlingual. Logo abaixo observamos a definição da construção. Em seguida, são apresentados os ECs Estrutura_argumental_base e Pronome_qual, marcados em azul e vermelho, respectivamente, bem como a descrição de tais elementos. Por fim, vemos que essa construção se relaciona com o Frame Entidade, o qual é apresentado na Figura 30, através da relação de *Evocação* e também se relaciona a outra construção mais abstrata, Interrogativa_QU, através da relação de *Herança*.

Ao cadastrarmos uma construção no banco de dados utilizamos também o Editor, um instrumento que serve para acrescentar mais detalhes sobre a relação de *Evocação* estabelecida entre a construção e o frame.

Essa relação de *Evocação* é criada no Editor de relações da Webtool 3.0 e permite que seja cadastrado acerca da estruturação dessa relação em termos da correspondência entre um ou mais ECs da construção com um ou mais Elemento de Frame (EF) do frame. A Figura 29 mostra como isso foi feito para a construção Interrogativa_qual.

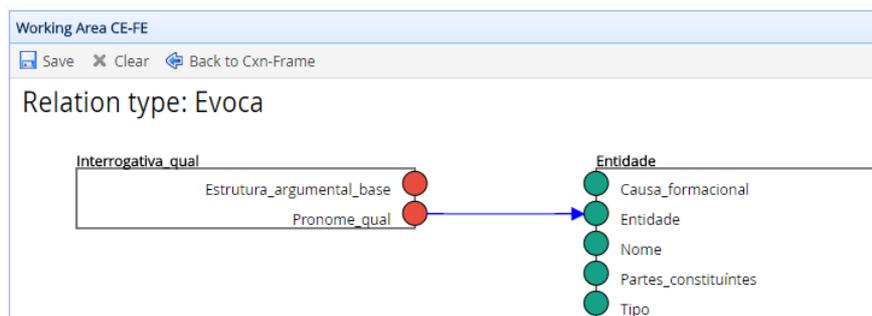


Figura 29: Estruturação da relação de *Evocação* entre a construção Interrogativa_qual e o frame Entidade

Aqui podemos observar que o EC Pronome_qual estabelece uma relação de *Evocação* mais estreita com o EF Entidade.

Entidade

Definição
Este frame é para palavras que denotam entidades altamente esquemáticas.

Exemplo(s)

Elementos de Frame Nucleares

FE Core:
Entidade [Entity] Uma coisa (abstrata ou física) que existe com algum grau de permanência.

Elementos de Frame Não-Nucleares

Causa_formacional [Formational_cause] A **Causa_formacional** indica a ação que dá os traços de uma **Entidade**.

Nome [name] O **Nome** é usado para o **Nome** da **Entidade**.

Partes_constituíntes [Constituent_parts] Partes relevantes que constituem a **Entidade**.

Tipo [Type] O **Tipo** indica a qual o subtipo da categoria geral a **Entidade** pertence.

Relações

Herdado por [Artefato](#) [Entidade-biológica](#) [Entidade_real](#) [Local](#) [Physical_entity#](#)

É usado por [Cycle_of_existence_scenario#](#)

Unidades Lexicais

[cama.n](#) [coisa.n](#) [entidade.n](#) [figura.n](#) [fogão à lenha.n](#) [fogão.n](#) [individuo.n](#) [item.n](#) [lâpis.n](#) [nada.n](#)

Figura 30: Frame Entidade

Na Figura 30 podemos observar a definição do frame Entidade, seguida pelos Elementos de Frame Nucleares, no caso apenas um EF, Entidade. Logo, vemos os EFs não nucleares e as relações que esse frame estabelece com outros frames, mais abaixo temos as Unidades Lexicais (ULs) que evocam tal frame.

Para a modelagem das demais construções foram seguidos os mesmos passos aqui descritos. Contudo, apontamos algumas especificidades de cada construção.

Os ECs *Estrutura_argumental_base* e *Pronome_como* compõem a construção *Interrogativa_como* e são licenciados respectivamente pelas construções *Sujeito_predicado* e *Pronome_QU*. Ao EC *Pronome_como* é aplicado a restrição *Lexeme*, pois esta restrição especifica o pronome a ocupar essa posição. Tais restrições são apresentadas na Figura 31.

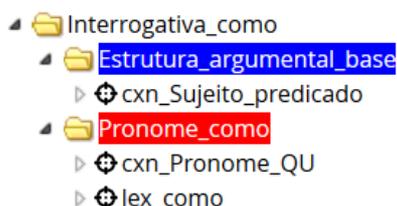


Figura 31: Restrições aplicadas à construção Interrogativa_como

Na Figura 32 podemos ver como a construção Interrogativa_como se apresenta no banco de dados. Além disso, ressaltamos que essa construção se relaciona com o frame Maneira através da relação de *Evocação* e estabelece uma relação de *Herança* com a construção mais abstrata, Interrogativa_QU.

Interrogativa_como [Como(how)_interrogative]

Definição	
Esta construção licencia sentenças interrogativas em que o objeto da pergunta é a maneira.	
Exemplo(s)	
Elementos da Construção	
Estrutura_argumental_base [Base_arg_structure]	Sentença sobre a qual o Pronome_como atua na definição do tópico da pergunta como a maneira.
Pronome_como [Como(how)_pronoun]	Palavra interrogativa que define o objeto da pergunta como a maneira.
Relações	
Evoca	Maneira
Herda de	Interrogativa_QU

Figura 32: Construção Interrogativa_como

Na relação de *Evocação* o EC Estrutura_argumental_base se relaciona ao EF Evento_de_comparação e o EC Pronome_como se relaciona ao EF Maneira.

A Figura 33 traz o frame evocado pela construção Interrogativa_como, nela aparece a definição do frame Maneira, seguida pelos EFs nucleares e não nucleares. Em seguida podemos ver a relação que esse frame estabelece com outros frames e por fim vemos as ULs evocadoras desse frame.

relaciona ao EF Figura e o EC Pronome_quando se relaciona ao EF Fundo do frame Relação_locativa, além disso, estabelece uma relação de *Herança* com a construção mais abstrata, Interrogativa_QU.

Interrogativa_onde [Onde(when)_interrogative]

Definição	
Esta construção licencia sentenças interrogativas em que o objeto da pergunta é um lugar.	
Exemplo(s)	
Elementos da Construção	
Estrutura_argumental_base [Base_arg_structure]	Sentença sobre a qual o Pronome_onde atua na definição do tópico da pergunta como um lugar.
Pronome_onde [Onde(when)_pronoun]	Palavra interrogativa que define o objeto da pergunta como um lugar.
Relações	
Evoca	Relação_locativa
Herda de	Interrogativa_QU

Figura 35: Construção Interrogativa_onde

Na Figura 36 o frame de Relação_locativa é exibido. Na tela vemos a definição do frame, bem como seus EFs nucleares e não nucleares, além das relações que este constitui com outros frames da base e ao final as ULs evocadores desse frame.

A principal diferença aqui é que a restrição Lexeme especifica que o pronome “quando” compõe esse EC.

A construção Interrogativa_quando tem como ECs a Estrutura_argumental_base e o Pronome_quando, os quais correspondem respectivamente aos EFs Evento_trajetor e Período_de_referência do frame Colocação_temporal evocado pela construção. Ademais, tal construção herda da construção Interrogativa_QU, como vemos na Figura 38.

Interrogativa_quando [quando(when)_interrogative]

Definição	
Esta construção licencia sentenças interrogativas em que o objeto da pergunta é o tempo.	
Exemplo(s)	
Elementos da Construção	
Estrutura_argumental_base [Base_arg_structure]	Sentença sobre a qual o Pronome_quando atua na definição do tópico da pergunta como o tempo.
Pronome_quando [Quando(when)_pronoun]	Palavra interrogativa que define o objeto da pergunta como o tempo.
Relações	
Evoca	Colocação_temporal#
Herda de	Interrogativa_QU

Figura 38: Construção Interrogativa_quando

O frame Colocação_temporal evocado pela construção é apresentado na Figura 39.

Massa_quantificada é evocado por essa construção, a qual estabelece uma relação de Herança com a construção Interrogativa_QU.

Interrogativa_quanto [Quanto(how_much)_interrogative]

Definição	
Esta construção licencia sentenças interrogativas em que o objeto da pergunta é a quantidade.	
Exemplo(s)	
Elementos da Construção	
Estrutura_argumental_base [Base_arg_structure]	Sentença sobre a qual o Pronome_quanto atua na definição do tópico da pergunta como uma quantidade.
Pronome_quanto [Quanto(how_much)_pronoun]	Palavra interrogativa que define o objeto da pergunta como uma quantidade.
Relações	
Evoca	Massa_quantificada
Herda de	Interrogativa_QU

Figura 41: Construção Interrogativa_quanto

Ao evocar o frame Massa_quantificada, exposto na Figura 42, o EC Estrutura_argumental_base se relaciona aos EFs Indivíduos e Massa, ao passo que o EC Pronome_quanto se relaciona ao EF Quantidade.

relaciona por herança com a construção Interrogativa_QU, como vemos na Figura 44.

Interrogativa_que [que(what)_interrogative]

Definição	
Esta construção licencia sentenças interrogativas em que o objeto da pergunta é uma coisa ou uma ação.	
Exemplo(s)	
Elementos da Construção	
Estrutura_argumental_base [Base_arg_structure]	Sentença sobre a qual o Pronome_que atua na definição do tópico da pergunta como uma coisa ou uma ação.
Pronome_que [Que(what)_pronoun]	Palavra interrogativa que define o objeto da pergunta como uma coisa ou ação.
Relações	
Evoca	Entidade
Herda de Interrogativa_QU	

Figura 44: Construção Interrogativa_que

A Figura 45 apresenta as restrições aplicadas à construção Interrogativa_quem.

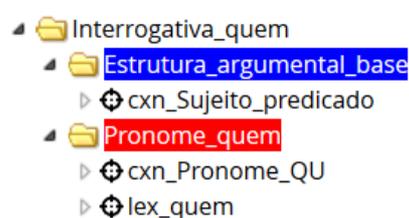


Figura 45: Restrições aplicadas à construção Interrogativa_quem

Na Figura 46, vemos como a construção Interrogativa_quem se apresenta no Constructicon. Essa construção é composta pelos ECs Estrutura_argumental_base e Pronome_quem e herda da construção Interrogativa_QU e evoca o frame Pessoas.

Interrogativa_quem [Quem(who)_interrogative]

Definição	
Esta construção licencia sentenças interrogativas em que o objeto da pergunta é uma pessoa.	
Exemplo(s)	
Elementos da Construção	
Estrutura_argumental_base [Base_arg_structure]	Sentença sobre a qual o Pronome_quem atua na definição do tópico da pergunta como uma pessoa.
Pronome_quem [Quem(who)_pronoun]	Palavra interrogativa que define o objeto da pergunta como uma pessoa.
Relações	
Evoca	Pessoas
Herda de	Interrogativa_QU

Figura 46: Construção Interrogativa_quem

Na relação de *Evocação* estabelecida entre a construção Interrogativa_quem e o Frame Pessoas, exibido na Figura 47, o EC Pronome_quem se relaciona ao EF Pessoa.

Pessoas	
Definição	
Esse frame contém palavras gerais para indivíduos. A Pessoa é concebida como independente de outros indivíduos específicos com os quais elas se relacionam e independente de sua participação em qualquer atividade particular. Elas podem ter uma Idade , Descrição , Origem , Característica_persistente ou Etnia .	
Exemplo(s)	
Elementos de Frame Nucleares	
FE Core:	Pessoa [Person]. A Pessoa é um ser humano.
Elementos de Frame Não-Nucleares	
Característica_persistente [Persistent_characteristic]	A Característica_persistente é uma característica fisiológica ou um traço de personalidade da Pessoa que é concebido como sendo persistente ao longo do tempo.
Contexto_de_referência [Context_of_acquaintance]	Uma expressão que indica o contexto com o qual a Pessoa está associada.
Descrição [Descriptor]	A Descrição é uma condição temporária da Pessoa .
Etnia [Ethnicity]	A Etnia é o grupo religioso, racial, nacional, sócio-econômico ou cultural ao qual a Pessoa pertence.
Idade [age]	A Idade é o período de tempo que a Pessoa está viva.
Origem [Origin]	A Origem é o lugar onde a Pessoa nasceu ou viveu grande parte de sua vida.
Relações	
Herda por	<input type="checkbox"/> Noncombatant# <input type="checkbox"/> People_along_political_spectrum# <input type="checkbox"/> People_by_age# <input type="checkbox"/> People_by_jurisdiction# <input type="checkbox"/> People_by_morality# <input type="checkbox"/> People_by_religion# <input type="checkbox"/> People_by_residence# <input type="checkbox"/> People_in_combat# <input type="checkbox"/> Pessoas_por_atividade_de_lazer <input type="checkbox"/> Pessoas_por_origem <input type="checkbox"/> Pessoas_por_vocação
Herda de	<input type="checkbox"/> Entidade-biológica
É usado por	<input type="checkbox"/> Fields# <input type="checkbox"/> Race_descriptor# <input type="checkbox"/> Relation_between_individuals#
Unidades Lexicais	
<input type="checkbox"/> alguém.pron <input type="checkbox"/> cara.n <input type="checkbox"/> caráter.n <input type="checkbox"/> cavalheiro.n <input type="checkbox"/> colega.n <input type="checkbox"/> companheiro.n <input type="checkbox"/> dama.n <input type="checkbox"/> galera.n <input type="checkbox"/> gente.n <input type="checkbox"/> homem.n	

Figura 47:Frame Pessoas

Além das construções até aqui expostas, também foram modeladas construções Interrogativas QU- preposicionadas. Essas construções possuem uma característica interessante, visto que são compostas por três ECs: *Estrutura_argumental_base*, *Preposição* e *Pronome_QU*. Outra característica relevante dessas construções é o fato de que as especificações em relação aos ECs *Preposição* e *Pronome_QU* são feitas através da atribuição das restrições à esses ECs.

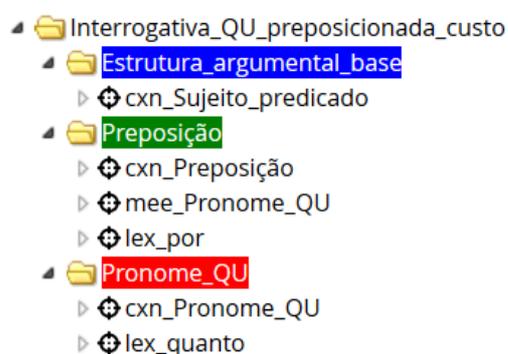


Figura 48: Restrições aplicadas à construção *Interrogativa_QU_preposicionada_custo*

As restrições aplicadas à construção *Interrogativa_QU_preposicionada_custo* são apresentadas na Figura 48. Aqui destacamos que a restrição “Construction” (cxn), como vimos nas outras construções apresentadas acima, estabelece que os ECs *Estrutura_argumental_base*, *Preposição* e *Pronome_QU* são licenciados, respectivamente, pelas construções *Sujeito_predicado*, *Preposição* e *Pronome_QU*. Além disso, ao EC *Preposição* é aplicada a restrição “Meets” (mee) que indica que um EC deve vir antes do outro e que não pode haver nenhum material interveniente entre os dois, assim o EC *Preposição* deve vir antes do EC *Pronome_QU*, ao EC *Preposição* é também aplicada a restrição “Lexeme” (lex), a qual especifica que a preposição que ocupará essa posição é a preposição “por”. Ao EC *Preposição* é também aplicada a restrição “Lexeme” (lex), especificando que o pronome “quanto” que preencherá a posição de pronome.

Na Figura 51 podemos observar como a construção Interrogativa_QU_preposicionada_razão se apresenta. É possível também observar que o frame Razão é evocado por essa construção e que esta se relaciona por Herança com a construção Interrogativa_QU_preposicionada.

Interrogativa_QU_preposicionada_razão [Prep_wh_interrogative_reason]

Definição	
Esta construção licencia sentenças interrogativas preposicionadas em que o objeto da pergunta é a razão.	
Exemplo(s)	
Elementos da Construção	
Estrutura_argumental_base [Base_arg_structure]	Sentença base da qual o Pronome_QU preposicionado extrai o objeto da pergunta.
Preposição [Prep_word]	Núcleo preposicional.
Pronome_QU [Wh_word]	Palavra interrogativa que delimita o objeto da pergunta.
Relações	
Evoca	Razão
Herda de Interrogativa_QU_preposicionada	

Figura 51: Construção Interrogativa_QU_preposicionada_razão

Na relação constituída entre a construção Interrogativa_QU_preposicionada_razão e o Frame Razão, exposto na Figura 52, o EC Estrutura_argumental_base se relaciona aos EFs Ação e Agente, já ECs Preposição e Pronome_QU se relacionam ao EF Estado_de_Coisas.

Razão

Definição	
Um volitivo Agente está respondendo para um Estado de coisas pela performance de uma Ação . E foi essa sua razão para se tornar um ator? Mas Giles é a razão de eu estar aqui. Oh a minha razão para incluir isso foi a Seção cinquenta e quatro A.	
Exemplo(s)	
Elementos de Frame Nucleares	
FE Core:	
Ação [Action]	A Ação que o Agente performa em resposta a um Estado de coisas . Eu acho que todas as minhas razões para adiar o parto ainda se aplicam."
Agente [Agent]	A pessoa que responde a um Estado de coisas pela performance de uma Ação . Aquelas são as minhas razões, " disse Toby. "
Estado de Coisas [State of Affairs]	A eventualidade que motivou o Agente . Minha razão é que nossas cidades são agora projetadas ao redor do automóvel" />
Elementos de Frame Não-Nucleares	
Relações	
É usado por	Deserving#
Usa	Agir_intencionalmente
Unidades Lexicais	
por isso.c por que.adv por.prep porque.n razão.n	

Figura 52:Frame Razão

Como vimos, é a partir desse aparato que as construções são definidas no Constructicon da FN-Br. Durante esta pesquisa foram modeladas, onze construções Interrogativas QU no Constructicon da FrameNet Brasil, sendo duas mais abstratas.

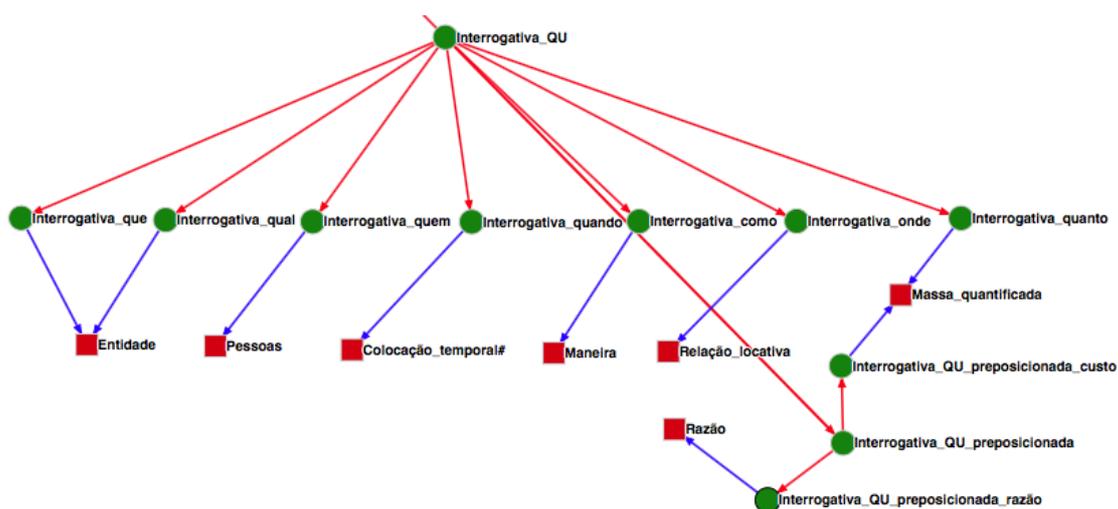


Figura 53: Rede de construções Interrogativas QU modeladas no Constructicon

A Figura 53 traz o gráfico que mostra as construções modeladas sob as relações de *Herança* e *Evocação*. Sendo assim, temos mais acima a construção mais abstratas, a Interrogativa_QU, a qual é herdada pelas demais construções representadas pelos círculos verdes, incluindo a construção Interrogativa_QU_preposicionada, apresentada na parte inferior direita da imagem, a qual também é mais abstrata. Os quadrados vermelhos representam os frames evocados por essas construções e a relação de Evocação é representada pelas setas azuis.

Ao postular uma rede de construções fornecemos meios para capturar generalizações gramaticais, além de apresentarmos como essas construções são licenciadas a partir da interação com outras construções (ALMEIDA, 2016), ainda que essas construções apresentem propriedades mais específicas.

A análise proposta nesta dissertação avança em relação aos trabalhos gerativista, funcionalista e mesmo construcionista apresentados nas seções 4.1 e 4.2. Em primeiro lugar, ao tratar as construções Interrogativas QU- como uma rede, enfatiza as propriedades compartilhadas por essas construções, ao mesmo tempo em que abre espaço para que suas particularidades semânticas sejam analisadas.

No que diz respeito à análise gerativa, tal vertente investe em representações dos fenômenos linguísticos que possam ser aplicadas computacionalmente, destacando-se como abordagem formalista para o tratamento da linguagem. No entanto, os modelos dessa linha ignoram aspectos importantes da língua, não só as idiosincrasias e usos metafóricos e metonímicos, mas também aspectos centrais da semântica. O modelo formal gerativista é composicional, sendo o significado inerente a cada forma e o significado do todo, um produto da soma das partes. Entretanto, como nossa análise pontua, construções QU- evocam estruturas complexas de significado (os frames) que vão interagir com os frames evocados pelas estruturas argumentais sobre as quais se processa a extração QU.

Para a nossa análise das construções Interrogativas QU, apesar de focar em construções com o QU movido, não se faz relevante considerar os

aspectos linguísticos e extralinguísticos que influenciam no movimento do elemento QU na sentença. Isso porque, no Constructicon, aspectos extralinguísticos são desconsiderados e os aspectos sintáticos das construções são mapeados e analisados pelas restrições aplicadas a elas.

No Constructicon, a anotação construcional é um fator de destaque, visto que permite capturar informações sintático-semânticas específicas e gerais das construções. Por ser realizada em camadas, a anotação consegue identificar a expressão licenciada pela construção e identificar os componentes constituintes da expressão, que juntos identificam o construto descrito pela construção. Ao passo que representar sintagmas e categorias gramaticais é uma maneira de mostrar as generalizações gramaticais, que podem abranger um grande número de construções da língua.

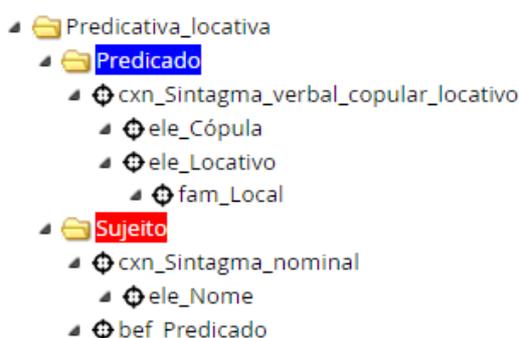
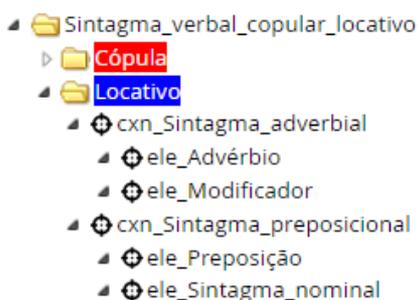
Até mesmo em relação à proposta construcionista da SBCG (SAG, 2012), a anotação de construções na base da FN-Br, permite que as propriedades específicas das construções do tipo *filler-gap* sejam capturadas. Assim, em uma construção Interrogativa QU do tipo “Onde fica Penedo?”, o tipo de *filler-gap* é extraído do fato de que a estrutura argumental base sobre a qual se processa a extração QU é uma construção Predicativa_locativa (Figura 54), a qual prevê que o EC Predicado seja composto por uma cópula e um elemento locativo manifesto como um SAdv ou um SP (Figuras 55 e 56). Pela anotação de ambas as construções como licenciadoras da mesma sentença, temos a possibilidade de mapear o elemento locativo ao pronome onde, ou seja, o gap ao *filler*.

Predicativa_locativa [Locative_predicate]

Definição	
Tipo de construção em que o núcleo é um advérbio ou nome Predicativo que atribui uma localização a um Sujeito .	
Exemplo(s)	
Elementos da Construção	
Predicado [Predicate]	O predicado é composto por uma cópula e um SAdv ou SP.
Sujeito [Subject]	O Sujeito é um SN.
Relações	
Evoca	Ser_localizado
Herda de Sujeito_predicado	

Figura 54: Construção Predicativa_locativa

A Figura 55 apresenta as restrições aplicadas à construção Predicativa_locativa e mostra que o EC Predicado é licenciado pela construção Sintagma_verbal_copular_locativo, enquanto a Figura 56 apresenta as restrições aplicadas à construção Sintagma_verbal_copular_locativo, mais especificamente, ao EC Locativo.

**Figura 55:** Restrições aplicadas à Construção Predicativa_locativa**Figura 56:** Restrições aplicadas à Construção Sintagma_verbal_copular_locativo

Na Figura 57, observamos que as quatro primeiras camadas anotam as propriedades referentes à construção Interrogativa_onde e as demais camadas se referem a anotação da construção Predicativa_locativa instanciada pela estrutura argumental base da construção Interrogativa_onde. Assim sendo, observamos que, na anotação da estrutura argumental base o elemento “onde”, aparece no predicado, o qual contém, além da cópula (ficar) o elemento locativo, isto é, ele é anotado para a posição de *filler*. Enquanto na anotação da construção Interrogativa_onde, o elemento “onde” não aparece na estrutura

argumental, mas sim como elemento QU anteposto. Então, para a estrutura argumental, indica-se indiretamente o gap. Dessa forma, as anotações, juntamente com as restrições impostas as construções, conseguem capturar as propriedades da *filler-gap*.

[116899]	NI	O	n	d	e	f	i	c	a	P	e	n	e	d	o	?	
Interrogativa_onde																	
CE		P	r	o	n	E	s	t	r	u	t	r	a	_	a		
Relação_locativa.FE		F	u	n	d	F	i	g	u	r	a						
CEE																	
CstrPT		S	i	n	t	e	r	r	o	g							
Predicativa_locativa																	
CE		P	r	e	d	i	c	a	d	o	S	u	j	e	i	t	
Ser_localizado.FE		L	o	c	a	l	i	z	a	ç	T	e	m	a			
CEE																	
CstrPT		S	i	n	t	e	r	r	o	g							

Figura 57: Anotação construcional para a sentença “Onde fica Penedo?”

Desse modo, apontamos que a proposta de modelagem apresentada aqui, por ser de cunho construcionista, se revela mais adequada ao tratamento das construções Interrogativas QU, pois mostra não só o comportamento sintático dessas construções, mas também as ligam em uma base de dados imensa, a FN-Br, que dá conta de detalhes do significado, os quais não são considerados nas demais abordagens, como foi mostrado.

Nas próximas seções apresentaremos como foram alcançados os avanços desta pesquisa.

5 ANÁLISE DE PROVA DE CONCEITO DA PERFORMANCE DO CHATBOT NA PLATAFORMA MKNOB

Nesta seção, apresentamos os resultados do teste de prova de conceito e trazemos uma discussão sobre esses resultados.

5.1 Resultados

Aqui apresentamos os resultados obtidos a partir do teste de prova de conceito, o qual teve por objetivo corroborar a premissa de que a modelagem proposta nesta dissertação poderá ser incorporada como uma das funcionalidades do *chatbot* do m.knob. Para tanto, a tarefa de reconhecimento construcional por máquina teve como aporte a modelagem das construções Interrogativas QU- realizada no Costructicon, conforme suas propriedades sintáticas, através da configuração dos ECs, e suas propriedades semânticas, pelos frames evocados por cada construção, bem como as relações de herança estabelecidas entre as construções.

Foram consultadas perguntas do tipo “Onde fica x?” em que “x” era substituído pelo tipo do local alvo da pergunta, conforme definido na base do m.knob a partir da extração de dados da API do Google Places. Feita a pergunta, o sistema retorna uma planilha em que constam o peso de ativação da resposta, o local a que ela se refere, o sintagma reconhecido pelo UD Pipe como contendo a resposta e o comentário de onde ele foi extraído. A título de exemplo, considere-se a Tabela 1, obtida a partir da pergunta “Onde fica o hotel?”.

Local	Resposta UD	Comentário Fonte	Peso
Aparthotel Adagio Rio de Janeiro Ipanema	em hotéis da rede	deveria ganhar a pontuação no accor hotels em dobro ou triplo ou meu dinheiro de volta por todo o transtorno que o hotel me causou para minha surpresa recebi 223 pontos já fiquei em hotéis da rede e ganhei pontuação superior como mercure que sabe atender seus clientes .	0.974613879 10828
Augustos Copacabana Hotel	no hotel	por fim foi legal, estavamos no rio é para passear, não para ficar no hotel.	0.974550904 7602
	no hotel	hotel um pouco antigo, porém com o café da manhã bom muito sol aqui no rio lindo essa foi a	0.974550904 7602

		sensação dos dias que ficamos no hotel.	
Constantino Hotel	funcionários excelente próximo ao centro de juiz de fora hospedade maravilhoso bom hotel com atendimento 10 muito bom	hotel lindo funcionários excelente próximo ao centro de juiz de fora hospedade maravilhoso bom hotel com atendimento 10 muito bom .	0.974252647 64069
Copacabana Beach	nesses hotéis próximos	eu quero ir ai porra e lindo lá amooooo fiquei nesses hotéis próximos a praia pra mim, a melhor praia do rio.	0.974607759 54208
Elegance Praia Hotel	em academia próxima ao hotel (academia super moderna , gostei muito)	além dos quartos serem muito aconchegantes e limpos, a estadia no hotel me deu direito a um voucher para malhar em uma academia próxima ao hotel (academia super moderna , gostei muito) o serviço do hotel é muito bom, podem confiar.	0.976473799 19411
	e próximo da estação de metrô muito bom excelente quartos grandes bem localizado bom hotel , confortável	hotel bem localizad , ótimo café da manhã e próximo da estação de metrô muito bom excelente quartos grandes bem localizado bom hotel, confortável, localização muito boa hotel com a fachada mais feia que eu já vi muito bom falta novos ar condicionado boa estrutura mas caro.	0.974531963 5466
	vale pena ir no carmelo restaurante também próximo ao hotel , lá tem pratos maravilhosos ! .	vale a pena ir no carmelo restaurante também, próximo ao hotel, lá tem pratos maravilhosos!.	0.974531963 5466
Gallant Hotel	estacionamento grátis e local do hotel fica bem próximo a zona sul e ao lado do maracanã .	estacionamento grátis e o local do hotel fica bem próximo a zona sul e ao lado do maracanã.	0.974254774 45646
Gran Victory Hotel	ótimo hotel , próximo a universidade federal de juiz de fora , ao lado do principal shopping da cidade , muito bem localizado .	ótimo hotel, próximo a universidade federal de juiz de fora, ao lado do principal shopping da cidade, muito bem localizado.	0.974607759 54208
	eventos , estacionamento e ótima localização próximo ao independência shopping ótimo hotel	ótimo hotel ótimo local pra fazer eventos, estacionamento e ótima localização próximo ao independência shopping ótimo hotel!.	0.974525650 02531
Hotel Bel-Air	hotel distante do centro de teresópolis ,	hotel distante do centro de teresópolis, somente vale para quem quer ficar perto do parque dá serra.	0.974337743 04768
	hotel distante do centro de teresópolis dependências externas mal	hotel distante do centro de teresópolis dependências externas mal conservadas café dá manhã bom, mas nada extraordinário piscina aquecida na verdade é uma banheira de hidromassagem grande.	0.974337743 04768

	conservadas café		
Hotel ibis budget Rio de Janeiro Centro	hotel muito bem localizado, próximo a pontos turísticos da cidade do rio de janeiro (saara, lapa, catedral,	hotel muito bem localizado, próximo a vários pontos turísticos da cidade do rio de janeiro (saara, lapa, catedral, confeitaria colombo etc).	0.97467094639614
	de hotel próximo ao centro da cidade , por pequeno período de tempo	excelente localização excelente para quem necessita de hotel próximo ao centro da cidade, por um pequeno período de tempo.	0.97467094639614
	no centro com ótimo acesso a transporte público	o hotel fica no centro com ótimo acesso a transporte público.	0.974614230841
	hotel fica próximo de tudo .	o hotel fica próximo de tudo .	0.974614230841
	no hotel pela casa noturna que funciona ao lado	nunca mais fico no hotel pela casa noturna que funciona ao lado.	0.974614230841
	num shopping distante do hotel	por exemplo, procuramos local para jantar, questionamos na recepção e não nos foi oferecido o festival da sopa, algo que estávamos procurando e até encontramos num shopping distante do hotel.	0.974614230841
	em hotel de r\$99,00	já fiquei em hotel de r\$99,00 que dá de 7x2 nesse.	0.974614230841
	no mesmo hotel	paguei o mesmo preço que meus concorrentes , que ficaram no mesmo hotel e conseguiram ter uma noite tranquila.	0.974614230841
	em hotel	não viajo para ficar em hotel, mas quando estou nele tem que atender com as comodidades necessária ao bom apoio ao viajante.	0.974614230841
	, no centro do rio	bom custo x benefício em uma boa localização no centro de sp ótimo serviço o hotel fica próximo a estação de metrô carioca, no centro do rio, com fácil acesso a pontos turísticos.	0.97458903794757
	porém limpo ótima localização , próximo ao metro e bares da lapa , para quem pretende usar hotel apenas como dormitório recomendo , preço é justo	cama e chuveiro excelentes, o quarto é pequeno porém limpo ótima localização, próximo ao metro e os bares da lapa , para quem pretende usar o hotel apenas como dormitório recomendo, o preço é justo.	0.97458903794757
	ser muito próximo a duas boates na esquina oposta ao hotel o que dificulta descanso para quem vai prestar concurso publico e necessita de sono tranquilo	boa acomodação para curtíssima temporada, exceto pelo barulho devido ser muito próximo a duas boates na esquina oposta ao hotel o que dificulta o descanso para quem vai prestar concurso publico e necessita de um sono tranquilo.	0.97458693830629
Hotel ibis Juiz de Fora	em hotéis mais simples	sempre soube que a rede não oferece muito mas não gostei de viver uma experiência sem frigobar com produtos no quarto, sem serviço de quarto para levar um cobertor melhor, porque o edredom	0.97461202683687

		que tínhamos era fino, as frutas servidas no bufê de café de manhã todas com sementes (já fiquei em hotéis mais simples e nunca vi isso) além de ter que fazer meu próprio suco de laranja.	
	no hotel	precisei ficar no hotel após o check-out ao meio dia e fiquei mais meia diária para poder fazer o check-out às 18hrs.	0.974612026 83687
Hotel ibis Rio de Janeiro Santos Dumont	hotel bem próximo ao aeroporto santos	hotel bem próximo ao aeroporto santos dumont .	0.974670946 39614
	tudo hotel muito bom e bem localizado no centro do rio , próximo ao aeroporto .	tudo muito bom hotel bom e bem localizado no centro do rio, próximo ao aeroporto.	0.974670946 39614
	do hotel próximo	além disso, a localização do hotel próximo ao aeroporto santos dumont ajuda muito quem precisa pegar vôos logo cedo pela manhã.	0.974670946 39614
	devido hotel ser próximo ao aeroporto ,	devido o hotel ser próximo ao aeroporto, tem características de hóspedes do ramo comercial, e os mesmo precisam tomar seu café cedo pra ir ao trabalho.	0.974614230 841
Hotel PortoBay Rio Internacional	em esquina	bom atendimento hotel fica em uma esquina, é muito comum o gps não indicar o lugar exato.	0.974583368 96453
Hotel São Francisco	no hotel	os quartos são simples e pequenos, adequados para viagens a trabalho, em que você pouco fica no hotel.	0.974550904 7602
	no centro da cidade	único contra é o fato de n ter estacionamento incluso na diária para quem visita o rio a trabalho, e bom pq fica no centro da cidade, o hotel e muito bom!.	0.974550904 7602
	no centro do rio excelente hotel , comidas e sobremesas deliciosas , piscina maravilhosa na cobertura	sim , bem confortável e fica no centro do rio excelente hotel , comidas e sobremesas deliciosas, piscina maravilhosa na cobertura.	0.974525650 02531
ibis Rio de Janeiro Botafogo	em localização de fácil acesso à praia do botafogo e do metrô	o hotel fica em uma localização de fácil acesso à praia do botafogo e do metrô.	0.974612026 83687
	em 100 %	fiquei no ibis botafogo, o hotel atende em 100 % ao que se propõe destaque especial pra o funcionário Flávio, simpático, muito educado, atendimento perfeito procuro utilizar o restaurante, indo ao café da manhã com os amigos que já tem filhos.	0.974612026 83687
	na região de botafogo , próximo ao metrô , centro gastronômico , shopping e praia hotel padrão ibis	hotel possui ótima localização na região de botafogo, próximo ao metrô , centro gastronômico, shopping e praia hotel padrão ibis, mas muito boa a com do restaurante muito bom satisfatóri.	0.974586831 79104
Ibiza Copacabana Hotel	hotel próximo a praia grande , bom atendimento , simples , mas atende	hotel próximo a praia grande, bom atendimento, simples, mas atende as necessidades de hospedagem.	0.974607759 54208

	necessidades de hospedagem .		
	no hotel	quartos com cheiro de cigarro o hotel ibiza copacabana é um bom hotel, ele tem um boa localização estando muito próximo da praia de Copacabana, nos arredores há vários restaurantes e apartamentos, o que é muito bom, o atendimento no hotel é muito bom e rápido, nos trataram muito bem o quarto é bem espaço e as camas e travesseiros bem confortáveis, servem um ótimo café da manh , vale apenas se hospedar no hotel, o preço está de acordo com a qualidade do hotel.	0.974550904 7602
La Paz Club	próximo ao hotel onde fiquei hospedado , espaço é bom !! adorei pequeno clube a 10 minutos dos arcos da lapa .	próximo ao hotel onde fiquei hospedado , o espaço é bom !! adorei pequeno clube a 10 minutos dos arcos da lapa.	0.974593130 39059
Mengo Palace Hotel	hotel fica muito bem localizado , tem metrô por perto , próximo a praia e para ponto turístico distância é pequena .	o hotel fica muito bem localizado, tem metrô por perto, próximo a praia e para qualquer ponto turístico a distância é pequena.	0.974550904 7602
Orla Copacabana Hotel	em frente	muito aconchegante localização muito boa e custo benefício ótimo sempre bom passear por aqui o hotel fica em frente à estátua do carlos drummond de andrade adoro ótima vista, precisa melhorar espaço do café dá manhã o hotel tem uma boa localização.	0.974525650 02531
Prodigy Hotel Santos Dumont Airport Rio de Janeiro	no hotel	de toda forma , ficaria no hotel novamente.	0.974614297 6324
	no restaurante n6 que fica na cobertura do hotel	sou de três rios (rj) e estive recentemente jantando no restaurante n6 que fica na cobertura do hotel.	0.974586305 15395
Rio Design Hotel	no mesmo hotel	quando voltar ao rio quero ficar no mesmo hotel.	0.974550904 7602
SESC Copacabana	fora isso , tudo muito bom , atendimento ótimo , simpatia , atividades no próprio hotel para ser feita e bem próximo da beira mar .	fora isso, tudo muito bom , atendimento ótimo, simpatia, várias atividades no próprio hotel para ser feita e bem próximo da beira mar.	0.974594984 0187
Sesc Hospedagem	em conta	localização um pouco afastado, quartos em bom estado, café da manha e almoço bem fracos excelente lugar para curtir a natureza é descansar sossego igual não há e bem próximo ao rio de janeiro super recomendo o local é um hotel com tudo que um clube pode oferecer e a preços em conta.	0.974337743 04768
Toledo Copacabana Hotel	hotel bacana , bem localizado , próximo de tudo , café muito bom .	hotel bacana, bem localizado, próximo de tudo, café muito bom.	0.974607759 54208

Victory Business Hotel	em que	fiquei em um fim de semana em que o hotel estava bem cheio .	0.9745509047602
	em avenida bem movimentada de juiz de fora	limpo e organizado hotel bem localizado , fica em uma avenida bem movimentada de juiz de fora .	0.9745509047602
Windsor Asturias Hotel	em dias ensolarados	hotel bem localizado e confortável , a vista do bar , que fica na cobertura , é sensacional em dias ensolarados bom hotel ótima hospedagem , atendimento excelente , recomendo ! .	0.9765362085126
Windsor Guanabara Hotel	em categoria	ótimo hotel da rede viajando em competições , foi com certeza o melhor hotel que fiquei no rio e um dos melhores do brasil em sua categoria .	0.97648476604639
	hotel bem localizado , próximo ao comércio .	hotel bem localizado , próximo ao comércio .	0.97460775954208
	muito próximo do metrô uruguaiana , hotel	muito próximo do metrô uruguaiana , o hotel é moderno por dentro e as instalações para eventos são amplas e bonitas adorável o clima dos bares e os melhores drinks que já me serviram na vida ! .	0.97460775954208
	no hotel fazendo social com amigos	funcionários bem educados , área da piscina com vista linda , inclusive com vista para a igreja da candelária , local muito próximo do museu do amanhã amo o windsor guanabara estive no hotel fazendo uma social com os amigos e fiquei apaixonado pelo hotel ! .	0.9745509047602
	, próximo igreja da candelária , padrão de excelência windsor hotel	muito amor para eles um dos melhores hotéis do centro do rio de janeiro , próximo a igreja da candelária , padrão de excelência windsor o hotel é bonito , os quartos são limpos e de bom gosto e os funcionários são agradáveis e atenciosos .	0.97452565002531
Windsor Palace Hotel	no hotel	peessoas muito simpáticas fica em copacabana perto da praia rede de hotéis fiquei por 32 dias no hotel .	0.97647752876702
	ótimo hotel , e muito próximo a praia de copacabana .	ótimo hotel , e muito próximo a praia de copacabana .	0.97460775954208
Windsor Plaza Hotel	hotel razoável , excelente localização , ótimo café da manhã , próximo a praia .	hotel razoável , excelente localização , ótimo café da manhã , próximo a praia .	0.97460775954208
	, no rio de janeiro	windsor plaza hotel fica na av princesa isabel , no rio de janeiro .	0.9745319635466

Tabela 1: Resultados para “Onde fica o hotel?”

Como podemos observar na Tabela 1, de um modo geral, o sistema reconhece as construções do tipo Interrogativa_onde e traz as respostas corretas ao que foi perguntado.

Em contrapartida, podemos observar na segunda coluna da Tabela 1 que algumas das respostas selecionadas são ruins. Isso porque os sintagmas selecionados não trazem a resposta correta, pois o *parser* UD Pipe não funciona muito bem para o Português Brasileiro. Considerando a resposta “no

mesmo hotel” para o local Rio Design Hotel, vemos que, além da resposta, o comentário selecionado não faz menção alguma a localização do hotel. Porém, apesar de a resposta “no hotel”, obtida para o Windsor Palace Hotel, não ser nada informativa, o comentário de onde ela foi retirada pelo UD Pipe traz informações relevantes para a pergunta, como “fica em Copacabana perto da praia”.

São apresentados também na Tabela 1 os níveis de ativação de cada resposta selecionada reconhecida pelo sistema. Visto que o nível de ativação dos nós iniciais é igual a 1, os resultados apresentados na Tabela 1 mostram os níveis finais de cada sentença após a ativação da rede.

A Tabela 2 mostra todas as sentenças que foram submetidas ao teste com o intuito de que o sistema reconheça o padrão construcional e a partir do frame evocado por este possa buscar na base resposta(s) à questão. Além disso, apresenta quantas respostas foram consideradas válidas e inválidas em relação a sentença buscada no sistema.

Sentenças	Número total de respostas obtidas	Número de respostas válidas	Número de respostas inválidas
Onde fica a arena?	4	0	0
Onde fica o bar?	1	1	0
Onde fica o bistrô?	1	0	1
Onde fica o cinema?	1	1	0
Onde fica o hotel?	61	33	28
Onde fica o local de eventos?	4	0	4
Onde fica a loja?	3	1	2
Onde fica o museu?	5	3	2
Onde fica o parque?	5	4	1
Onde fica a praça?	4	2	2
Onde fica a praia?	7	2	5
Onde fica o restaurante?	5	1	4
Onde fica o shopping center?	7	6	1
Onde fica o zoológico?	1	0	1

Tabela 2: Resultado para as sentenças submetidas ao teste

A Tabela 3 expõe as sentenças buscadas, o número total de respostas obtidas, além de apontar quantas respostas e quantos comentários foram considerados válidos no que diz respeito a sentença pesquisada no sistema. Assim, na Tabela 3, podemos observar que o número de comentários válidos são maiores ao número de respostas válidas.

Sentenças	Número total de respostas obtidas	Número de respostas válidas	Número de comentários válidos
Onde fica a arena?	4	0	0
Onde fica o bar?	1	1	1
Onde fica o bistrô?	1	0	0
Onde fica o cinema?	1	1	1
Onde fica o hotel?	61	33	42
Onde fica o local de eventos?	4	0	0
Onde fica a loja?	3	1	2
Onde fica o museu?	5	3	4
Onde fica o parque?	5	4	4
Onde fica a praça?	4	2	3
Onde fica a praia?	7	2	4
Onde fica o restaurante?	5	1	2
Onde fica o shopping center?	7	6	6
Onde fica o zoológico?	1	0	0

Tabela 3: Resultados dos comentários obtidos

O Gráfico 1 apresenta de modo mais didático os dados expostos na Tabela 3.

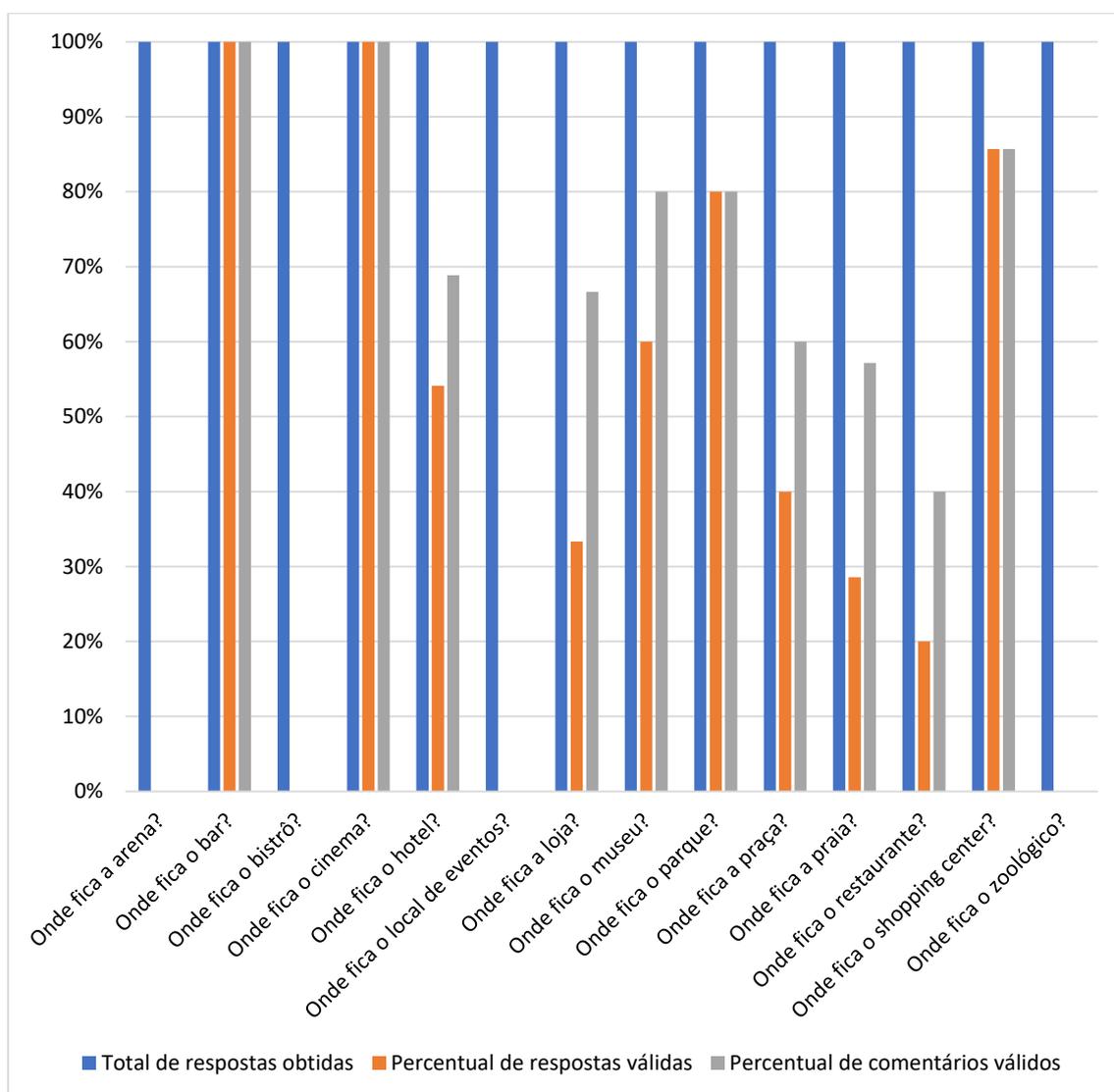


Gráfico 1: Resultados dos comentários obtidos

Na próxima seção discutiremos esses resultados encontrados.

5.2 Discussão

De acordo com as análises desenvolvidas e com os dados do teste de prova de conceito apresentados nas Tabelas 1 e 2, observamos que as sentenças licenciadas pela construção Interrogativa_onde obtiveram, de uma forma geral, respostas válidas que indicam o reconhecimento por máquina de padrões construcionais. Além disso, a Tabela 3 e o Gráfico 1 mostram que um

grande número dos comentários responde positivamente as questões, ainda que as respostas extraídas pelo UD Pipe não fossem boas.

O sistema de teste conseguiu reconhecer as construções. Esse reconhecimento ocorreu por níveis de ativação, como é observável na Tabela 1, e os níveis apresentados para as respostas são superiores a 0,97. Por exemplo, na Tabela 1, o sistema reconhece que Gran Victory Hotel é um tipo de hotel e traz como resposta os comentários que contêm o local “hotel” expresso, bem como a UL “próximo” evocadora do frame *Relação_locativa*, assim, os comentários selecionados trazem satisfatoriamente a localização do lugar buscado, dando origem a respostas como “ótimo hotel, próximo a universidade federal de juiz de fora, ao lado do principal shopping da cidade, muito bem localizado”.

Os resultados apresentados nas Tabelas 1 e 2 evidenciam que existe uma limitação no *parser* sintático, já que, no geral, as respostas obtidas a partir da seleção dos sintagmas pela UD são ruins, isso porque a UD Pipe para o Português Brasileiro não funciona de maneira satisfatória, impossibilitando um tratamento sintático adequado às sentenças aplicadas no teste. A nível de exemplo, considere a resposta selecionada para o tipo Sesc hospedagem, na Tabela 1. Nela podemos ver que o sintagma selecionado “em conta”, não responde corretamente à questão, pois ao verificar o comentário inteiro vemos que o *parser* deveria ter selecionado como resposta o sintagma “localização um pouco afastado”.

Uma solução possível para o problema da UD Pipe é implementação de um *parser* construcional pela FrameNet Brasil, o qual seja capaz de abarcar aspectos semânticos em suas análises para que, assim, possa tratar de forma mais apropriada e satisfatória fenômenos linguísticos no PB.

Ademais, a implementação de um *parser* construcional se faz relevante, uma vez que seu objetivo é mostrar que a modelagem construcional contribui para que os padrões construcionais possam ser reconhecidos por máquina, assim sendo, a modelagem das construções Interrogativas QU proposta nesta

dissertação colabora para que o desenvolvimento de um *parser* construcional seja mais eficiente.

Apesar de alguns problemas apontados em relação às respostas alcançadas no teste, no geral, podemos dizer que os comentários buscados trazem boas respostas às questões, conforme apontam os resultados da Tabela 3. Isso se deve ao fato de que a construção Interrogativa_onde instanciada nas sentenças buscadas no teste evoca o frame de Relação_locativa cujas ULs aparecem nos comentários, assim estes correspondem às questões que foram pesquisadas. Desse modo, os resultados alcançados indicam a importância da interação entre o Lexicon e o Constructicon, visto que os recursos necessitam funcionar juntos de modo a mapear a continuidade entre o léxico e a gramática, conforme salientou Fillmore (2008). Para tanto, se faz necessário que as relações semânticas estabelecidas entre construções, frames, EFs e também entre ULs estejam bem determinadas e precisas.

A relação de Evocação se destaca na modelagem de construções, o caso de os comentários conterem informação relevantes para as questões se deve ao fato de a construção evocar um frame cujas ULs aparecem nos comentários. Pelo fato de um frame congregar um conjunto de ULs semanticamente relacionadas, as quais são definidas conforme seu comportamento em *corpora* e, portanto, de modo independente dos objetivos deste trabalho, o modelo da FN-Br faz com que o *chatbot* do m.knob avance em relação a *chatbots* como o ELIZA, uma vez que não se trata, aqui, de mero casamento de padrões, a partir de uma lista fechada de palavras, criada, especificamente, para o *chatbot*.

A FN-Br, com base nos recursos disponíveis, trata os fenômenos semânticos através dos frames e formaliza o significado das sentenças agregando-os aos seus aspectos sintáticos, alinhando, assim, a modelagem de construções apresentada neste trabalho ao modelo de descrição linguística proposto pela Gramática das Construções de Berkeley e a implementação computacional da FN-Br, baseada no processamento e na compreensão de

dados em língua natural. Ainda que esses recursos estejam em desenvolvimento, a FN-Br tem se empenhado na ampliação de seu banco de dados, não só em relação aos frames e às suas implicações, mas também tem investido no cadastramento de construções do PB.

Destacamos ainda que a modelagem por nós realizada e proposta aqui proporciona uma abordagem holística aos fenômenos linguísticos, uma vez que contempla aspectos sintático-semânticos das construções modeladas. Isso indica a viabilidade inicial de se incorporar a modelagem a funcionalidade do *chatbot* do m.knob, ainda que os resultados sejam limitados a uma prova de conceito. Visto que o aplicativo utiliza a base de dados enriquecida da FN-Br para gerar interpretações semânticas de sentenças em língua natural, ao incluir-se a modelagem das construções Interrogativas QU- na FN-Br, o m.knob poderá melhorar a sua funcionalidade em relação às recomendações de lugares, pois, a partir das recomendações feitas pelo GregBot, o aplicativo oferecerá a possibilidade do usuário buscar mais informações sobre os lugares através de sentenças interrogativas QU-.

6 CONCLUSÕES

Ao longo desta dissertação, buscamos apresentar a descrição e modelagem das construções Interrogativas QU- em PB com o objetivo de possibilitar uma interpretação por máquina de construções para sua implementação na aplicação m.knob.

A formalização encontrada no Constructicon se dá pelo processo de unificação, pois nele as construções são descritas em termos de suas propriedades sintático-semânticas através de AVMs com base nos pressupostos teóricos da Gramática das Construções de Berkeley e da Gramática das Construções Baseada em Signos.

As construções modeladas aqui demonstram a necessidade de considerar tanto aspectos semânticos como aspectos sintáticos para uma representação satisfatória do conhecimento linguístico. Para tanto, foi preciso explicitar as relações de Herança, estabelecida entre as construções em uma rede integrada, e de Evocação, estabelecida entre construções e frames, além de apontar as restrições aplicadas às construções.

Nesse estudo, as restrições sintáticas foram tratadas em termos dos ECs que são representados em sintagmas e categorias gramaticais como uma forma de mostrar as generalizações presentes em cada construção. Por outro lado, cada restrição foi unificada a um determinado componente semântico do frame evocado pela construção, isto é, os ECs da construção foram unificados a um EF do frame evocado por ela.

Propusemo-nos no presente trabalho a apresentar como a modelagem de construções se dá no Constructicon da FN-Br, demonstrando como a integração de aspectos da BCG e da Semântica de Frames enriquecem a representação formal e semântica das sentenças. Para isso, modelamos onze construções Interrogativas QU-, sendo a construção Interrogativa_QU mais abstrata e herdada pelas construções: Interrogativa_que; Interrogativa_qual;

Interrogativa_quem; Interrogativa_quando; Interrogativa_quanto; Interrogativa_onde e Interrogativa_como. Também modelamos a Construção Interrogativa_QU_preposicionada mais abstrata, a qual é herdada pelas construções Interrogativa_QU_preposicionada_razão e Interrogativa_QU_preposicionada_custo. A pesquisa desenvolvida e apresentada nesta dissertação dá oportunidade para trabalhos futuros que se interessem, por exemplo, em analisar os tipos de estruturas argumentais que compõem as construções Interrogativas QU-, identificar a relevância dessas estruturas para composição das construções Interrogativas QU- e que tipo de relação pode ser estabelecida entre as estruturas argumentais e o elemento QU, entre outras possibilidades de análise.

Para comprovar a viabilidade do modelo que estamos propondo, foi realizado um teste com uma das onze construções modeladas. Esse teste revelou que o sistema computacional foi capaz de reconhecer os padrões construcionais das sentenças em estudo, servindo para avaliarmos que a modelagem realizada para as interrogativas QU são legíveis por máquina.

A partir da comparação dos resultados do teste, constatamos que o *parser* sintático aplicado não foi eficiente, demonstrando a necessidade do desenvolvimento de um *parser* construcional.

Diante do que foi exposto, essa pesquisa implica os seguintes avanços:

- (i) a modelagem de 11 construções Interrogativas QU- para um empreendimento de processamento e compreensão de linguagem natural;
- (ii) o adensamento da base de dados da FN-Br;
- (iii) a disponibilização de uma nova função no aplicativo m.knob, que poderá oferecer mais informações acerca dos lugares recomendados por ele, uma vez que o usuário poderá interagir de mais formas com o aplicativo.

REFERÊNCIAS

ADGER, D. **Core Syntax**: a minimalist approach. New York: Oxford University Press, 2003.

ALLEN, J. **Natural Language Understanding**. Menlo Park: Benjamins / Cummings Publishing Company, 1985.

ALMEIDA, V. G. **Identificação Automática de Construções de Estrutura Argumental**: um experimento a partir da modelagem linguístico-computacional das construções Transitiva Direta Ativa, Ergativa e de Argumento Cindido. Dissertação de Mestrado em Linguística. Universidade Federal de Juiz de Fora. Juiz de Fora, 2016.

AUGUSTO, M. R. A. QU deslocado e QU in *situ* em PB: aspectos da derivação linguística e questões para a aquisição da linguagem. In: **Anais do IV Congresso Internacional da ABRALIN**, p. 535-546, 2005.

BENDER, E. M.; Flickinger, D.; Oepen, S.; Packard, W. & Copestake, A. Layers of interpretation: On grammar and compositionality. In: **Proceedings of the 11th international conference on Computational Semantics**. p. 239-249, 2015.

BOAS, H. C. A Lexical-Constructional Account of the Locative Alternation. In: CARMICHAEL, L., HUANG, & SAMIIAN, V. (orgs.). **Proceedings of the 2001 Western Conference in Linguistics**, Vol. 13, p. 27-42, 2003.

CASTILHO, Ataliba T. de. **Nova Gramática do Português Brasileiro**. São Paulo: Editora Contexto, 2010.

COMARELLA, R. L., & CAFÉ, L. M. A. CHATTERBOT: conceito, características, tipologia e construção. In: **Informação & Sociedade**. João Pessoa. v. 18, n. 2, p. 55-67, 2008.

COSTA, A. D. & TORRENT, T. T. . A Modelagem Computacional do Domínio dos Esportes na FrameNet Brasil. In: **Proceedings of the 11th Brazilian Symposium in Information and Human Language Technology**. Uberlândia, Brasil, nov. Sociedade Brasileira de Computação (SBC), p. 201–208, 2017.

_____.; GAMONAL, M. A. ; PAIVA, V. M. R. L. ; MARCAO, N. D. ; PERON-CORRÊA, S. ; ALMEIDA, V. G. ; MATOS, E. E. ; TORRENT, T. T. FrameNet-Based Modeling of the Domains of Tourism and Sports for the Development of a Personal Travel Assistant Application. In: **Proceedings of the LREC 2018 Workshop International FrameNet Workshop 2018**: Multilingual Framenets and Constructicons. Miyazaki, v. 1. p. 6-12, 2018.

DIAS-DA-SILVA, B. C. **A Face Tecnológica dos Estudos da Linguagem: o**

processamento automático das línguas naturais. 274 p. Tese de doutorado – USP. Araraquara 1996.

DICIONÁRIO AURÉLIO. 2010. Disponível em: <<http://www.dicionariodoaurelio.com>>. Acesso em: dezembro de 2017.

FAUCONNIER, G. & TURNER, M. **The Way We Think.** New York: Basic Books. 2002.

FILLMORE, C. J. Scenes and frames semantics. In: ZAMPOLLI, A. (Ed.). **Linguistic Structures Processing: Fundamental Studies in Computer Science**, nº 59. Amsterdam: North Holland Publishing. 1977.

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

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

_____. Syntactic intrusions and the notion of grammatical construction. In: **Annual Meeting of the Berkeley Linguistics Society.** p. 73-86, 1985b.

_____.; ATKINS, Beryl T. Toward a frame-based lexicon: The semantics of RISK and its neighbors. In: **Frames, fields, and contrasts: New essays in semantic and lexical organization**, v. 103, p. 75-102, 1992.

_____.; JOHNSON, C.R. and PETRUCK, M. R. L. Background to FrameNet. In: **International Journal of Lexicography**, p. 235-250, 2003.

_____. **Berkeley Construction Grammar.** In: HOFFMANN, T.; TROUSDALE, G. **Oxford Handbook of Construction Grammar** (Eds.). Oxford University Press, 2013.

_____. Border Conflicts: FrameNet Meets Construction Grammar. In: **Proceedings of the XIII Euralex international congress.** Barcelona: Universitat Barcelona Fabra, p. 49-69, 2008.

_____.; LEE-GOLDMAN, R. R.; RHOMIEUX, R. The FrameNet Constructicon. In: Boas, H. C. Boas and Sag, I. A. (eds.). **Sign-Based Construction Grammar.** Stanford: CSLI Publications, p. 309-372, 2012.

_____. Berkeley Construction Grammar. In: HOFFMANN, T.; TROUSDALE, G. **Oxford Handbook of Construction Grammar** (Eds.). Oxford University Press, 2013.

FRIED, M.; OSTMAN, J-O. Construction Grammar: a thumbnail sketch. In: _____. **Construction Grammar in a Cross-Linguistic Perspective.** Amsterdam: John Benjamins, 2004.

GOLDBERG, A. **Constructions**: A Construction Grammar Approach to Argument Structure. Chicago: The University of Chicago Press, 1995.

_____. **Constructions at Work**: The nature of generalization in language. Oxford: Oxford University Press, 2006.

HAUSSER, R. **Computational linguistics and talking robots**: processing content in database semantics. Springer Science & Business Media, 2011.

KATO, M. A.; BRAGA, M. L.; CORRÊA, V. R.; ROSSI, M. L.; SIKANSI, N. S. As construções-Q no Português Brasileiro falado: perguntas, clivadas e relativas. In: KOCH, I. G. V. (Org.). **Gramática do Português Falado**. Campinas, SP: Ed. Unicamp, vol. VI, p. 303-368, 1996.

KAY, P.; FILLMORE, C. J. Grammatical constructions and linguistic generalizations: The What's X doing Y? construction. In: **Language**, p. 1-33, 1999.

LAGE, L. M. **Frames e Construções**: A Implementação Do Constructicon Na Framenet Brasil. Dissertação de Mestrado apresentada ao programa de Pós-Graduação em Linguística da Faculdade de Letras da Universidade Federal de Juiz de Fora, 2013.

LINO, T. A. L. R. **Inteligência artificial, humana e emoção**. 2004. Disponível em: <<http://www.psicologia.pt/artigos/textos/TL0012.PDF>>. Acesso em: janeiro de 2018.

MATOS, E. E. da S. **LUDI**: um framework para desambiguação lexical com base no enriquecimento da Semântica de Frames Tese de Doutorado em Linguística. Universidade Federal de Juiz de Fora, Juiz de Fora, 2014.

_____.; TORRENT, T. T.; ALMEIDA, V. G.; SILVA, A. B. L.; LAGE, L. M.; MARCAO, N. D. & TAVARES, T. S. Constructional Analysis Using Constrained Spreading Activation in a FrameNet -Based Structured Connectionist Model. In: **The AAI 2017 Spring Symposium on Computational Construction Grammar and Natural Language Understanding Technical Report SS-17-02**. Palo Alto, CA: AAI Publications, v.17, p. 222-229, 2017.

MARÇÃO, N. D.; TORRENT, T. T. & MATOS, E. E. S. Descrição e modelagem de construções interrogativas QU- em Português Brasileiro para o desenvolvimento de um chatbot. In: **Proceedings of the 11th Brazilian Symposium in Information and Human Language Technology**, p. 209-216, 2017.

MCDONALD, R.; NIVRE, J.; QUIRMBACH-BRUNDAGE, Y.; GOLDBERG, Y.; DAS, D.; GANCHEV, K.; HALL, K.; PETROV, S.; ZHANG, H.; TÄCKSTRÖM, O.; BEDINI, C.; CASTELLÓ, N.; LEE, J. Universal Dependency Annotation for Multilingual Parsing. In: **Proceedings of ACL**, 2013.

MCTEAR, M.; CALLEJAS, Z.; GRIOL, D. The Dawn of the Conversational Interface. In: **The Conversational Interface**. Springer International Publishing, 2016. p. 11-24.
MANFIO, E. R.; MORENO, F. A Evolução dos Chatterbots: PLN, I.A. e Difusão Cultural. In: **Revista e-fatec**. Garça, vol.5, n. 1, 2016.

MIOTO, C. As interrogativas no português brasileiro e o critério wh. In: **Letras de hoje**. Porto Alegre, vol. 29, nº 2, p. 19-33, junho 1994.

MODESTO, M. O programa minimalista em sua primeira versão. In: Leonel Figueiredo de Alencar; Gabriel de Ávila Othero. (Org.). **Abordagens computacionais da teoria da gramática**. 1 ed. Campinas, SP: Mercado de Letras, 2012, v. , p. 127-152.

MOURA, T.J.M. **Um chatterbot para aquisição automática de perfil do usuário**. 2003. Dissertação (Mestrado em Ciência da Computação) - Universidade Federal de Pernambuco, Recife, 2003.

OUSHIRO, L. **Uma análise variacionista para as Interrogativas-Q**. Dissertação de Mestrado em Linguística. Universidade do São Paulo. São Paulo, 2011.

PERINI, Mário A. **Gramática Descritiva do Português**. São Paulo: Editora Ática, 2010.

RODRIGUES, Érica. S.; AUGUSTO, M. R. A. Modelos formais de gramática: o Programa Minimalista vs. gramáticas baseadas em restrições–HPSG e LFG. **Matraga-Revista do Programa de Pós-Graduação em Letras da UERJ**, v. 16, n. 24, 2009.

SAG, I. A. Sign-Based Construction Grammar: na informal synopsis. In: BOAS, H. C.; SAG, I. A. **Sign-Based Construction Grammar**. Chicago: CSLI, 2012.

SANTOS, L. C. B. **Aprendizagem, cognição e inteligência artificial**. Campinas, 2006. Disponível em: <<http://www.dca.fee.unicamp.br/~gudwin/courses/IA889/2006/IA889-02.pdf>>. Acesso em: abril de 2018.

SALOMÃO, M. M. M. FrameNet Brasil: um trabalho em progresso. In: **Calidoscópico**, São Leopoldo: UNISINOS, vol. 7 n. 3, p. 171-182, set/dez 2009.

SALOMÃO, M. M. M.; SAMPAIO, T. F. ;TORRENT, T. T. A linguística cognitiva encontra a linguística computacional: notícias do projeto FrameNet Brasil. In: **Cadernos Estudos Linguísticos**. Universidade Estadual de Campinas. Campinas. N. 55. p. 7-34, jan/jun 2013.

SAMPAIO, T. F. **A Família de Construções de Argumento Cindido no Português do Brasil**. Tese de Doutorado em Linguística. Universidade Federal de Juiz de Fora, Juiz de Fora, 2010.

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

SAUSSURE, F. **Curso de Linguística Geral**. 27ª ed. São Paulo. Cultrix: 2006 [1916].

SILVA, A. B. L.; LAGE, L. M.; MARÇÃO, N. D.; TAVARES, T. S.; ALMEIDA, V. G.; MATOS, E. E. & TORRENT, T. T. The Brazilian Portuguese Constructicon: Modeling Constructional Inheritance, Frame Evocation and Constraints in FrameNet Brasil. In: **The AAI 2017 Spring Symposium on Computational Construction Grammar and Natural Language Understanding Technical Report SS-17-02**. Palo Alto, CA: AAI Publications, v.17, p.193-196, 2017.

STRAKA, M.; HAJIC, J.; STRAKOVÁ, J. UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization, Morphological Analysis, POS Tagging and Parsing. In: **Proceedings of the Language Resources and Evaluation Conference (LREC)**, Portoroz, Estônia, 2016.

TAVARES, M. A. Sociofuncionalismo: um duplo olhar sobre sobre a variação e mudança linguística. In: **Interdisciplinar: Revista de Estudos em Língua e Literatura**, v. 17, 2013.

TORRENT, T. T.; LAGE, L. M.; TAVARES, T. S.; LAVIOLA, A. B. Relações de herança entre construções e entre frames: desafios da extensão do modelo construcionista para o domínio computacional no âmbito da FrameNet Brasil. **DELTA: Documentação e Estudos em Linguística Teórica e Aplicada**, v. 33, n. 1, 2017.

_____.; MATOS, E. E.; SIGILIANO, N. S.; COSTA, A. D.; ALMEIDA, V. G. **A flexible tool for an enriched FrameNet: the FrameNet Brasil WebTool**. No prelo.

TURING, A. M. Computing. Computing machinery and intelligence. In: **Mind**, v. 59, n. 236, p. 433-460, 1950.

VAN VALIN JR., R. D. **From NLP to NLU**. Disponível em: <http://www.isi.hhu.de/fileadmin/redaktion/Oeffentliche_Medien/Fakultaeten/Philosophische_Fakultaet/Sprache_und_Information/Van_Valin_From_NLP_to_NLU.pdf> Acesso em: novembro de 2017.

VARGAS, N. L.; KEPLER, F. N. Análise Sintática Automática por Dependência. **Anais do Salão Internacional de Ensino, Pesquisa e Extensão**, v. 4, n. 2, 2012.