

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

Vânia Gomes de Almeida

**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**

Juiz de Fora  
2016

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

**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**

Vânia Gomes de Almeida

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  
Julho de 2016

**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**

Vânia Gomes de Almeida

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. Thaís Fernandes Sampaio – UFJF

---

Prof. Dr. Bento Carlos Dias da Silva – UNESP

Juiz de Fora  
Julho de 2016

*“A verdadeira viagem de descobrimento não consiste em procurar novas paisagens, mas em ter novos olhos.”*

*Marcel Proust.*

*Aos meus sobrinhos Alan, Andressa,  
Maria Eduarda e Meirielen – os  
sorrisos mais sinceros.*

## AGRADECIMENTOS

Venho agradecer a Universidade Federal de Juiz de Fora, instituição que vem me proporcionando grande aprendizado. Agradeço também ao Programa de Pós-Graduação em Linguística, que, juntamente com a Coordenação de Aperfeiçoamento de Pessoal do Ensino Superior, CAPES, me forneceram todo suporte no decorrer da pós-graduação.

Ao meu orientador, Tiago Torrent, agradeço imensamente por toda atenção, disponibilidade e paciência. Obrigada por incentivar meu crescimento, compartilhar ensinamentos e estimular minha ambição acadêmica. O resultado desse trabalho não seria o mesmo sem a sua dedicação.

Agradeço a todos do projeto FrameNet Brasil por todo aprendizado compartilhado. Em especial, agradeço ao Ely, pela grande colaboração com os experimentos computacionais; ao Alexandre, pela amizade e parceria, e às meninas do Constructicon, Adrieli, Ludmila e Tatiane, pelo companheirismo e convivência

Não posso deixar de agradecer aos meus amigos encapsulados, Cristina, Késsia, Flávio e Allan, que foram verdadeiros amparos durante esses dois anos. Sem vocês tudo seria mais difícil.

Quero dedicar um eterno agradecimento a Prof.<sup>a</sup> Patrícia Cunha por plantar em mim a semente da pesquisa. Muito obrigada por segurar a minha mão e guiar meus primeiros passos na Linguística.

Cabe agradecer também às minhas amigas Malvina e Eulália que me acompanham desde a graduação. Vocês foram refúgio em momentos de angústia e desânimo. Obrigada pela amizade e por todos os momentos compartilhados.

Agradeço à minha família pelo apoio incondicional. Meus pais Rosalice e Expedito, que se dedicaram sempre em fazer o melhor para mim e me ensinaram que tudo é possível. Agradeço a vocês pelo carinho e amor a mim dedicados. À minha irmã Vanessa, por ser também amiga e incentivar todos os meus sonhos. Aos meus irmãos Luis Gustavo e Luiz Eduardo, por participarem da minha vida. Aos meus sobrinhos Alan, Andressa, Maria Eduarda e Meirielen, por tornarem meus dias mais felizes.

Finalmente, agradeço a Deus por acreditar em mim mais do que todos, me amparar nos momentos de fraqueza e insegurança e concretizar esse sonho.

Almeida, Vânia Gomes de.

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 / Vânia Gomes de Almeida. – Juiz de Fora: UFJF / FALE, 2016.

x, 93f.:il.; 2,0cm.

Orientador: Tiago Timponi Torrent

Dissertação (mestrado) – UFJF / Faculdade de Letras / Programa de Pós-Graduação em Linguística, 2016.

Referências Bibliográficas: f.91-93.

1. Gramática das Construções. 2. Semântica de Frames. 3. Análise Linguístico-Computacional. 4. Parsing. I. Torrent, Tiago Timponi *et al.*. II. Universidade Federal de Juiz de Fora, Faculdade de Letras, Programa de Pós-Graduação em Linguística. III Título.

## RESUMO

Este trabalho se enquadra dentro das discussões teórico-metodológicas da FrameNet Brasil (FN-Br) e do Constructicon do português brasileiro e tem por objetivo (i) fornecer a modelagem linguístico-computacional necessária para o tratamento das construções Transitiva Ativa, Ergativa e de Argumento Cindido no Constructicon do Português do Brasil; (ii) apresentar e discutir os resultados de uma tarefa de descoberta de construções por máquina cujo objetivo é o de identificar instâncias das construções acima; (iii) discutir as implicações desta tarefa para o desenvolvimento de constructicons. Dessa maneira, esta dissertação busca embasamento na abordagem construcional da Berkeley Construction Grammar proposta por Fillmore, Kay e O'Connor (1988), por Kay e Fillmore (1999) e Fillmore (2013). A nossa proposta se justifica pela escassez de estudos que integrem descrições linguísticas com modelos de representação computacionais cognitivamente plausíveis, possibilitando o tratamento de questões que ultrapassam os modelos de representação centrados exclusivamente em aspectos sintáticos . Nesse sentido, este trabalho se insere no âmbito da Linguística Computacional, que, conforme afirma Dias da Silva (1996), envolve a integração de questões em três domínios: linguístico, representacional e implementacional. Esta pesquisa se enquadra na fase representacional, uma vez que a fase linguística já foi proposta por Sampaio (2010), Ferreira (2009), Castilho (2010) e Perini (2010). Dado esse enquadramento, nosso objetivo foi modelar computacionalmente as construções Transitiva Ativa, Ergativa e CAC na base construcional da FrameNet, para que, posteriormente, na fase implementacional – não desenvolvida nesta dissertação, mas cujos resultados serão aqui discutidos – pudessem ser utilizadas em uma tarefa de descoberta construcional por máquina através do parser em desenvolvimento pela FN-Br, o CARMA.

**Palavras-chave:** Gramática das Construções. Semântica de Frames. Análise Linguístico-Computacional. Parsing.



## ***ABSTRACT***

This work is included in the theoretical and methodological discussions of FrameNet Brasil (FN-Br) and the Construction of Brazilian Portuguese. It aims to (i) provide a linguistic-computational modeling necessary for the treatment of the Active Transitive, Ergative and Split Object Constructions (CAC, in Portuguese) in the Construction of Brazilian Portuguese; (ii) present and discuss the results of a discovery-task of constructions by machine whose purpose is to identify instances of the constructions mentioned above; (iii) discuss the implications of this task for the development of constructions. Thus, this dissertation is based on the constructional approach of Berkeley Construction Grammar proposed by Fillmore, Kay and O'Connor (1988), by Kay and Fillmore (1999) and Fillmore (2013). Our proposal is justified by the lack of studies that integrate linguistic descriptions with models of computational representation cognitively plausible, enabling the treatment of issues that surpass the representation models centered exclusively on syntactic aspects. Hence, this work falls within the framework of Computational Linguistics which, according to Dias da Silva (1996), involves the integration of issues in three domains: Linguistic, Representational and Implementational. This research is inserted in the representational stage, once the linguistic stage has already been proposed by Sampaio (2010), Ferreira (2009), Castilho (2010) and Perini (2010). Given this framework, our objective is to computationally model the Active Transitive, Ergative and CAC into the constructional base of FrameNet, so that, later, in the Implementational stage – not developed in this dissertation, but whose results are discussed here – could be used in a constructional discovery-task by machine through the parser under development by FN-Br, the CARMA.

**Keywords:** Construction Grammar. Frame Semantics. Linguistic-Computational Analysis. Parsing.

## SUMÁRIO

<b>INTRODUÇÃO</b>	11
<b>1 CONSTRUÇÕES E PROCESSAMENTO DE LINGUAGEM NATURAL</b>	15
<b>1.1 A Gramática das Construções</b>	15
<b>1.2 O Constructicon</b>	27
<b>1.3 Processamento de Língua Natural</b>	39
<i>1.3.1 Bases do Processamento de Língua Natural</i>	39
<i>1.3.2 Parsing</i>	44
<b>1.4 CARMA</b>	51
<b>2 CONSTRUÇÕES DE ESTRUTURA ARGUMENTAL DO PORTUGUÊS DO BRASIL</b>	56
<b>2.1 A Construção Transitiva Direta Ativa</b>	56
<b>2.2 A Construção Ergativa</b>	58
<b>2.3 A Construção de Argumento Cindido</b>	59
<b>2.4 Rede Mínima de Construções Necessária ao Modelo</b>	62
<b>3 METODOLOGIA EXPERIMENTAL</b>	69
<b>3.1 Aplicação de Teste de Julgamento Semântico para Estabelecimento de Gold Standard</b>	69
<b>3.2 Desenho Experimental</b>	72
<b>4 TAREFA DE DESCOBERTA CONSTRUCIONAL POR MÁQUINA</b>	75
<b>4.1 Resultados</b>	75
<b>4.2 Discussão</b>	80
<b>5 IMPLICAÇÕES DO EXPERIMENTO PARA O DESENVOLVIMENTO DE CONSTRUCTICONS</b>	86
<b>6 CONCLUSÕES</b>	89
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b>	92

## INTRODUÇÃO

Este trabalho se enquadra nas discussões teórico-metodológicas da FrameNet Brasil (FN-Br) e do Constructicon do português brasileiro (PB). Nossa equipe de trabalho é composta por linguistas e cientistas da computação que estão envolvidos em vários campos de pesquisa, desde a construção de recursos lexicais e dicionários eletrônicos, identificação e catalogação de construções do PB, tradução por máquina e processamento da linguagem natural.

Nesse sentido, a FN-Br ocupa-se do desenvolvimento de dois grandes recursos computacionais: um lexicon e um constructicon. O lexicon tem por objetivos: criar uma representação computacional de frames, definidos por seus participantes e instrumentos e conectados entre si via relações entre frames; definir Unidades Lexicais (ULs), pareamentos entre um lema e um significado definido em termos de um frame; anotar sentenças que exemplifiquem os padrões de valência sintáticos e semânticos em que as ULs ocorrem.

Já o constructicon visa à 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. 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.

Diante dessa gama de recursos, este trabalho tem por objetivo (i) fornecer a modelagem linguístico-computacional necessária para o tratamento das construções Transitiva Direta Ativa, Ergativa e de Argumento Cindido no Constructicon do PB; (ii) apresentar uma tarefa de descoberta de construções por máquina cujo objetivo é o de identificar instâncias das construções em questão; (iii) discutir as implicações dos resultados desta tarefa para o desenvolvimento de constructicons.

A partir das descrições realizadas por Ferreira (2009), Castilho (2010), Perini (2010), que apresentaram estudos sobre as Construções Transitiva Direta Ativa e Ergativa, e de Sampaio (2010) para as Construções de Argumento Cindido, nossa proposta é modelar essas construções linguístico-computacionalmente com o intuito de contribuir para trabalhos futuros na área de Linguística Computacional.

Perini (2010) define construção como uma estrutura gramatical composta por seus constituintes sintáticos e pela relação semântica que cada um deles tem com verbo da oração. Nesse sentido, a construção Transitiva Direta Ativa é definida pela sequência [SN [V SN]] em que o primeiro sintagma nominal designa quem praticou uma ação e o segundo sintagma nominal designa quem sofreu a ação como em (1). Já a construção Ergativa tem como característica o fato de o sujeito ser paciente, ou seja, não pratica nenhuma ação e sim sofre o efeito do evento expresso pelo verbo. Pode ser definida como a sequência [SN [V]], como em (2). Por fim, as construções de Argumento Cindido (CACs) são construções que apresentam dois argumentos sintáticos e apenas um argumento semântico, com função paciente. Assim, temos um desencontro (*mismatch*) entre estruturas sintáticas e estruturas semânticas como em (3).

- (1) Laura comeu a pizza
- (2) O carro estragou
- (3) O celular quebrou a tela

A nossa proposta se justifica pela escassez de estudos que integram descrições linguísticas com modelos de representação computacionais cognitivamente plausíveis, possibilitando o tratamento de questões que ultrapassam os modelos de representação centrados exclusivamente em aspectos formais. No decorrer do texto, apresentamos as características dos modelos cognitivamente plausíveis. Então, a pergunta que nos motiva a investigar essas construções é:

Qual seria o caminho para a modelagem computacional dessas construções, que possibilite uma interpretação adequada pela máquina?

Para responder essa questão temos que considerar: (i) os verbos que aparecem nessas construções; (ii) os frames evocados por eles; (iii) a relação entre os elementos desses frames e os argumentos do verbo; (iv) o frame evocado pelas construções; (v) a relação entre os elementos desse outro frame e os argumentos das construções; (vi) possíveis relações de herança entre as construções. Este último ponto se justifica pelo fato de que, apesar de (4) parecer ser, sintaticamente, uma construção transitiva como em (5) poderia ser parafraseado por uma construção ergativa, como em (6).

- (4) O celular trincou a tela
- (5) A faxineira trincou o copo.
- (6) A tela do celular trincou.

Nesse sentido, este trabalho se insere no âmbito da Linguística Computacional, que, conforme afirma Dias da Silva (1996), envolve a integração de questões em três domínios: **linguístico**, **representacional** e **implementacional**. Esta pesquisa que se enquadra dentro da fase representacional, uma vez que a fase linguística já foi proposta por Sampaio (2010), Ferreira (2009), Castilho (2010) e Perini (2010). Dado esse enquadramento, nosso objetivo foi modelar computacionalmente as construções Transitiva Direta Ativa, Ergativa e CAC na base construcional da FN-Br, para que, posteriormente, na fase implementacional – não desenvolvida nesta dissertação, mas cujos resultados serão aqui discutidos – pudessem ser utilizadas em uma tarefa de descoberta construcional por máquina através da técnica de travessia em redes chamada Ativação Propagada (AP) ou *Spreading Activation*.

Os resultados da tarefa indicam que a representação proposta para as construções permitiu que o sistema identificasse, satisfatoriamente, 29 das 30 sentenças utilizadas no experimento, aproximando-se do *gold standard* definido a partir do julgamento semântico das mesmas 30 sentenças por falantes nativos do PB.

Esta dissertação está organizada da seguinte forma: no capítulo um, apresentamos os pressupostos teóricos utilizados em nossa pesquisa através de uma descrição das principais abordagens da Gramática das Construções, mostramos como funciona a base construcional da FN-Br, o Constructicon, e justificamos a inserção desta pesquisa nos estudos sobre o Processamento da Língua Natural. O capítulo dois consiste na apresentação das descrições linguísticas para as construções Transitiva Direta Ativa, Ergativa e CAC, na modelagem dessas através do Constructicon e da apresentação de uma rede mínima de construções necessárias para o modelo aqui proposto. No capítulo três, descrevemos a metodologia experimental por nós utilizada a partir da aplicação de um teste de julgamento semântico com a finalidade do estabelecimento de um *gold standard* e da exibição do desenho experimental. O capítulo quatro traz a compilação dos resultados obtidos e a discussão desses resultados para uma tarefa de descoberta construcional por máquina. Já o capítulo cinco expõe as implicações do experimento para o desenvolvimento de Constructicons. Por fim, apresentamos as conclusões obtidas com esse trabalho.

## 1 CONSTRUÇÕES E PROCESSAMENTO DE LÍNGUA NATURAL

Nesta seção, apresentaremos os pressupostos teóricos que orientaram nossa pesquisa. Começaremos pela apresentação das abordagens sobre a Gramática de Construções, passaremos pelo Constructicon e, por fim, apresentamos o campo de estudos em Processamento de Língua Natural.

### 1.1 A Gramática das Construções

O paradigma denominado Gramática das Construções, de um modo geral, assume que construções são unidades básicas da língua que se constituem em correspondências entre forma e significado (GOLDBERG, 1995; KAY & FILLMORE, 1999). Essa premissa é acompanhada pela postulação da existência de um *continuum* entre léxico e sintaxe (FILLMORE, 2008). Dessa forma, não se trata de haver módulos rigidamente separados, mas, sim, um *continuum* de construções que partem de elementos mais específicos para padrões mais abstratos. Sendo assim, qualquer tipo de expressão linguística, desde as mais simples às mais complexas, pode ser considerada uma unidade composta de uma correspondência entre forma e significado ou função.

A abordagem de Goldberg (1995) defende que estruturas argumentais são instâncias de construções que existem independentemente de verbos particulares. Baseada nessa afirmação, a autora postula a Hipótese Fraca da Composicionalidade, que preconiza que as construções de uma língua podem não apresentar um significado que se construa apenas pela função das partes que compõem uma dada sentença. Logo, essas unidades apresentam um significado que se dá por meio de uma visão processual, multidirecional e construcional e não como partes isoladas que compõem um todo. Goldberg (1995) priorizou a extensão da abordagem construcional das expressões “irregulares” para as construções “regulares”, pois os princípios explicativos das construções irregulares seriam capazes de dar conta dos fenômenos regulares.

Para isso, analisou as chamadas Construções de Estrutura Argumental, que licenciam sentenças compostas por um verbo e seus argumentos. Através dessa análise, propõe-se uma distinção entre os papéis argumentais, que são

associados às construções e os papéis participantes, associados aos verbos, ou seja, o verbo perfila os papéis participantes, já os papéis argumentais são as possibilidades previstas pela construção, mas não necessariamente pelo verbo. De acordo com a notação proposta por Goldberg (1995), as construções podem ser formalizadas conforme a Figura 1.

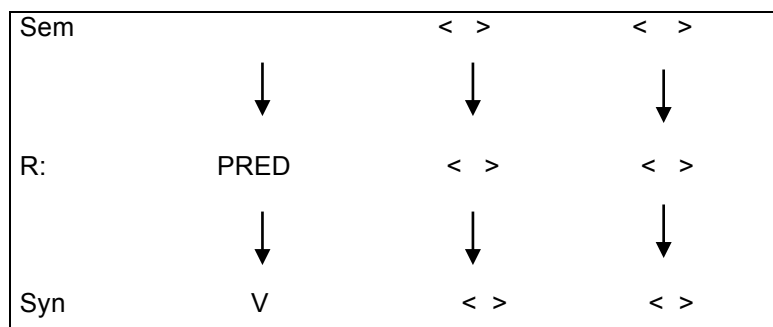


Figura 1: Modelo de formalização das construções de Goldberg (1995)

A Figura 1 mostra, na primeira linha, a semântica da construção (Sem) com os papéis semânticos dos argumentos da construção; na segunda (R), o predador, no caso, um verbo, que perfila os papéis participantes; e, na terceira linha, as funções gramaticais atribuídas à fusão de participantes e argumentos na expressão sintática (Syn).

Para uma análise construcional, a autora propõe um modelo baseado na interação entre semântica verbal e semântica construcional. Com esse modelo, não é necessário apresentar vários sentidos verbais para explicar a diferença entre as construções. Vejam-se, por exemplo, as sentenças com o verbo *quebrar*:

- (7) O vaso quebrou
- (8) Maria quebrou o vaso
- (9) O martelo quebrou o vaso

Temos que o verbo *quebrar* mantém o mesmo sentido em (7), (8), e (9), a diferença de significado resulta de semânticas construcionais distintas. Por isso, cada construção tem um significado próprio, convencional e esquemático que é independente dos verbos e de outros itens lexicais que a compõe. O verbo *quebrar* é tradicionalmente classificado como transitivo como em (8) e



(9), porém pode ocorrer em construções como em (7) em que temos *quebrar* como ergativo. As três sentenças possuem significados diferentes, oriundos de distintas combinações entre a semântica do verbo e as das construções, conforme mostram as Figuras 2, 3 e 4.

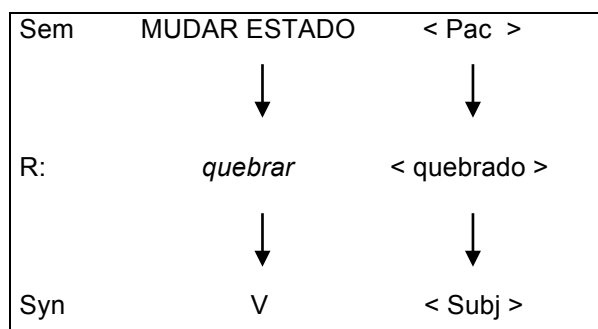


Figura 2: Construção Ergativa – *O vaso quebrou*

Na Construção Ergativa, Figura 2, a construção apresenta uma semântica em que o único argumento, o *vaso*, apresenta função gramatical de sujeito com o papel semântico de paciente configurado cognitivamente pela construção e a especificação de *quebrado* fornecida pelo predicador. Nesse exemplo, observamos que o sujeito da construção possui características semânticas semelhantes ao objeto direto em uma sentença transitiva, como mostra a Figura 3:

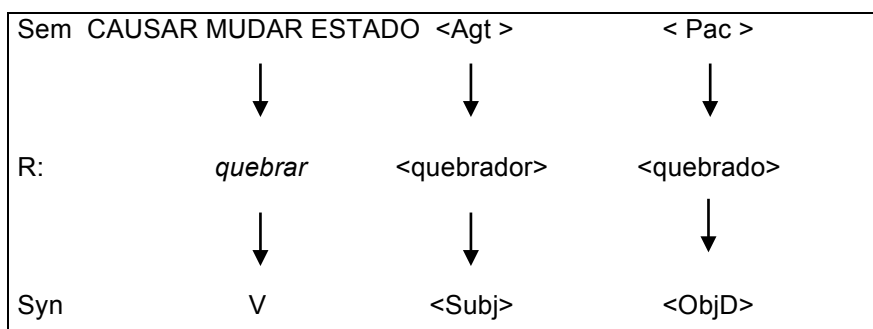


Figura 3: Construção Transitiva Direta Ativa – *Maria quebrou o vaso*

No caso da Construção Transitiva Direta Ativa, encontramos dois argumentos, *Maria* e *vaso*, com funções gramaticais de sujeito e objeto direto respectivamente. O papel semântico de *Maria* é de agente e do *vaso* de

paciente, com papéis de *quebrador* e *quebrado* especificados pelo predicador. Note que essa construção apresenta a semântica de CAUSAR MUDAR ESTADO com uma predicação agentiva não encontrada na Ergativa, mas que pode ser observada também, na Figura 4.

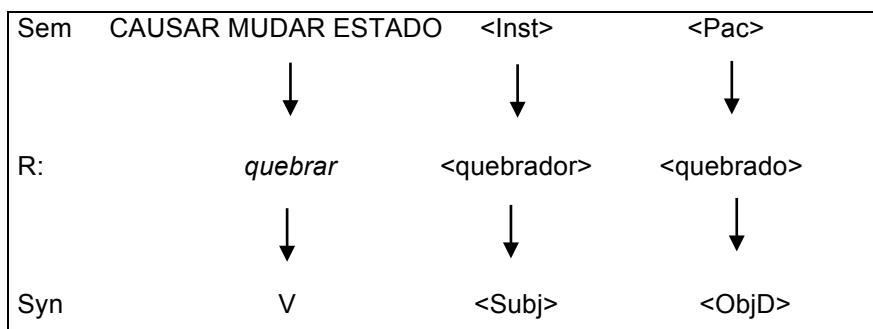


Figura 4: Construção Transitiva Direta Ativa de Instrumento – *O martelo quebrou o vaso*

Nesse exemplo, temos uma construção transitiva diferente da Figura 3: em *O martelo quebrou o vaso*, ainda encontramos uma predicação causativa, porém o sujeito e objeto direto não são mapeados como agente e paciente e sim como instrumento e paciente, que mantêm os papéis participantes de quebrador e quebrado evocados pelo predicador, como na sentença (8).

Essas diferenças construcionais, em muitos casos, podem ser motivadas por relações de herança, já que o repertório de construções gramaticais não é um conjunto desestruturado. Goldberg (1995) afirma que construções formam uma rede e estão ligadas por relações de herança que motivam muitas das propriedades de construções particulares. Essas relações de herança seguem alguns princípios psicológicos para a organização da rede.

Um dos princípios propostos por Goldberg (1995) é o *Princípio da Motivação Maximizada*, que assume que, se duas construções são sintaticamente relacionadas, tais construções são motivadas semântica ou pragmaticamente. Outro princípio que estrutura a relação entre construções é o *Princípio da Não Sinonímia*, que afirma que, se duas construções são sintaticamente distintas, tais construções devem ser também distintas semântica ou pragmaticamente (GOLDBERG, 1995, p.67).

A situação da CAC em relação à Ergativa é exatamente aquela prevista no segundo princípio. CAC e Ergativa são sintaticamente diferentes, mas

semanticamente sinônimas, assim sendo, devem ser pragmaticamente distintas. Logo, a diferença entre as construções certamente passará pela questão do uso, da estrutura informacional.

Nas sentenças Ergativas o argumento único dos verbos exibe propriedades típicas do objeto dos verbos transitivos. Castilho (2010) argumenta que essas sentenças têm um sujeito como um argumento não controlador do estado de coisas e que elas não explicitam o causador da ação. Isso fica bastante evidente, se considerarmos, por exemplo, o uso das duas estruturas, tendo os mesmos referentes, como é possível observar nas sentenças (10) e (11).

(10) A panela soltou o cabo

(11) O cabo da panela soltou

Aqui temos duas estruturas sintáticas diferentes para expressar o mesmo evento – o cabo da panela que soltou – uma instanciação da CAC (10) e uma Ergativa (11). Isso nos mostra que a diferença sintática entre a CAC e a Ergativa se reflete na diferença pragmática dessas construções, por isso, para analisarmos essas construções adequadamente, devemos levar em consideração os contextos preferenciais de uso para cada construção.

Por outro lado, a CAC é semelhante à Transitiva Direta Ativa sintaticamente, portanto, mesmo sendo semanticamente distintas, deve haver algum grau de motivação entre elas. Como se verá mais adiante, essa relação de motivação se manifesta, na rede, pelo fato de que ambas as construções herdam de um tipo específico da construção Sujeito-Predicado, que especifica a transitividade do núcleo do Predicado como Direta.

O modelo de Goldberg (1995) para a Gramática das Construções desenvolveu-se aplicado tanto a um conjunto de estudos descritivos em diversas línguas, dentre as quais o Português Brasileiro (vide MIRANDA & SALOMÃO, 2009), quanto a uma produtiva interface com a área de aquisição da linguagem (GOLDBERG, 2006). Esse modelo, entretanto, possui limitações de formalização, que tornam difícil sua aplicação para além da fase linguística do processo de modelagem computacional. Nesse sentido, o modelo da Gramática de Construções proposto por Fillmore, Kay e O'Connor (1988) e por

Kay e Fillmore (1999), a Gramática de Construções de Berkeley (BCG), revela-se mais adequado.

A BCG apresenta um modelo baseado em unificação que se fundamenta matematicamente e busca incorporar práticas computacionais e proporcionar um maior formalismo à base teórica da Gramática das Construções. Uma gramática baseada em regras não consegue dar conta de muitas irregularidades existentes nas línguas, já que o significado de uma construção é visto como independente, em parte, das palavras que a constituem. Assim, é necessário desenvolver um sistema de representação capaz de codificar economicamente e sem perda de generalização todas as construções (ou padrões) da língua, do mais idiomático ao mais geral (KAY & FILLMORE, 1999).

A unificação entre os níveis de representação, na BCG, apresenta uma relação estreita entre significado e significante, porém, essa noção é ampliada ao incluir associações entre fonologia, morfologia, sintaxe, semântica e contexto. O pareamento entre forma e sentido está relacionado com o que foi proposto por Fillmore (1968), pois temos uma relação sintático-semântica subjacente não apenas no nível lexical, mas também no nível construcional que envolve os outros componentes da gramática.

O que proporciona um maior formalismo a essa base teórica é um sistema baseado em traços em que as construções são representadas por Matrizes de Atributo Valor (AVM), em um pareamento de forma e sentido mediado por signos. Fried e Ostmann (2004) mostram essa estrutura como uma forma conveniente de organizar toda a informação necessária para dar uma descrição adequada do fato linguístico. As informações sobre qualquer entidade linguística aparecem como uma AVM, as caixas representam a estrutura de constituintes, a estrutura de traços detalha a informação gramatical e a semântica, e a coindexação mantém o controle da unificação de relações.

Nesse modelo, a unificação é a principal operação formal em que AVMs são combinadas de acordo com os valores atribuídos aos seus atributos, uma vez que o resultado de uma AVM é o que está contido nas suas AVMs componentes. Fillmore (2013) afirma que os valores de cada atributo são compatíveis aos valores requeridos por cada posição sintagmática, já que,

assim, é possível definir e especificar os tipos de entidades que podem ou devem estar em cada sintagma.

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

Figura 5: Representação Preposição + Complemento (PrepComp)

Como podemos observar na Figura 5 (FILLMORE, 2013, p.3), o sistema de representação se dá através de colchetes, dentro dos quais estão contidos os atributos – cat, max etc. –, seguidos dos valores – p, + etc. – a eles atribuídos. Os atributos carregam as informações relacionadas aos signos linguísticos. Por exemplo, a maximalidade (max) possui valores polares <+> e <-> e está relacionada com a possibilidade ou não de aquela categoria constituir por si só um sintagma. Já categoria (cat) apresenta valores de *nome*, *verbo*, *preposição*, *advérbio*, *adjetivo*, *determinante*. Outros atributos ausentes na Figura 5 incluem Número, com os valores de *singular* ou *plural*; Função Gramatical, que inclui *sujeito*, *objeto*, *complemento*, *oblíquo*; e Tipo Sintagmático, que apresenta os valores de *NP*, *VP*, *PP* etc.

No exemplo da Figura 5, vemos que o colchete maior representa a construção *PrepComp*, com a categoria de preposição podendo constituir o núcleo de um sintagma próprio. As AVMS limitadas pelos conjuntos de colchetes paralelos representam, respectivamente, a preposição e seu complemento com os seus valores de categoria e maximalidade. A contraparte semântica da construção *PrepComp* não aparece nessa representação, porém, de acordo com esse modelo, essa informação é explicitada pelo frame evocado por cada signo linguístico, sendo que os elementos que o constituem serão unificados aos signos que compõem a construção.

Nesse sentido, a BCG possui estreita relação com o que propôs Fillmore (1968; 1977) na Gramática de Casos. Com o intuito de tratar das relações de dependência sintático-semânticas entre os constituintes que estão intuitivamente relacionados ao modo como as pessoas conceptualizam as

experiências e eventos em suas línguas por meio das sentenças, Fillmore (1968) postulou o *Case frame* que buscava ligar essas descrições de situações a representações sintáticas subjacentes para encontrar um caminho para uma sintaxe baseada em signos e não só na forma, uma vez que esta manifesta linguisticamente uma cena cognitiva.

Dessa forma, a Gramática de Casos forneceu as bases para o desenvolvimento da Semântica de Frames (FILLMORE, 1982), pois o princípio norteador de ambas teorias é o de que os significados são relativizados às cenas. Assim, a Semântica de Frames elege como tarefa tratar dessas relações combinatórias (sintático-semânticas) previstas pelo modelo gramatical de Fillmore (1968), colocando a semântica no centro do debate. A noção de *frame*, que emerge na semântica linguística pela insuficiência da descrição do significado lexical em termos de uma lista de condições necessárias e suficientes, é entendida como uma estrutura conceptual complexa, organizada de tal modo que, para compreender qualquer de suas partes é imprescindível o conhecimento do todo, e, para a total compreensão dessa cena, é necessário saber os papéis semânticos que cada constituinte assume em relação à unidade lexical evocadora do frame (FILLMORE, 1982).

Nesse sentido, os casos semânticos propostos pela Gramática de Casos de Fillmore evoluem para os Elementos de Frame (EFs), que contribuem com informação a respeito do frame que é evocado pelos materiais linguísticos que compõem um enunciado. Além disso, o modelo baseado em *frames* introduz a ideia de perspectiva, já que podemos encontrar uma mesma cena com perspectivas diferentes.

Temos como exemplo os itens lexicais *comprar* e *vender*, que ativam a mesma cena de evento comercial, mas sob perspectivas diferentes. Enquanto a primeira perfila a perspectiva do comprador, a segunda perfila a do vendedor. O conhecimento linguístico do falante sobre o verbo inclui o conhecimento sobre as diversas formas de realizar as diferentes partes do evento na forma de um enunciado. A ideia é que sempre que entendemos uma expressão linguística de qualquer tipo, temos, simultaneamente, uma cena (completa) de fundo e uma perspectiva sobre aquela cena.

Em resumo, os modelos fillmoreanos são complementares, uma vez que a gramática de uma língua consiste de um conjunto de construções que

licenciam combinações de signos, ou *constructos*, cuja semântica pode ser analisada em termos dos *frames* que são, por sua vez, herdeiros da Gramática de Casos.

Para uma descrição gramatical completa é necessário representar, portanto, não só os atributos morfossintáticos, mas, também, as informações sobre os argumentos semânticos e suas propriedades. As possibilidades combinatórias dos itens lexicais e suas características surgem quando estão associadas a outros itens lexicais. Assim, podemos definir, de acordo com a valência de cada termo, sua posição na estrutura sintagmática e seu papel semântico nas construções.

A valência está relacionada com as possibilidades combinatórias de uma entidade linguística. Através dela, entendemos a estrutura da construção, identificamos as posições que cada termo ocupa com suas propriedades sintáticas e semânticas. Fillmore (2013) argumenta que a BCG faz uma distinção entre os elementos nucleares e periféricos, ou seja, aqueles que são obrigatórios e outros que podem ou não ser introduzidos por falantes em uma cena. Dessa forma, argumentos e adjuntos recebem o mesmo tratamento, uma vez que a valência nos permite descrever construções como expressões linguísticas em que os sentidos são construídos seguindo as formas de realização dos valentes em papéis semânticos, função gramatical e tipo sintagmático.

A valência contém um conjunto de papéis semânticos que estão diretamente ligados a um participante de um frame com a sua expressão morfossintática correspondente e é representada como uma lista de elementos, em que cada um deles possui uma estrutura de traços, especificando, assim, a relação entre o verbo e seus argumentos.

Fried e Ostman (2004, p.45) apresentam a valência do verbo *carry* *carregar* em Inglês, reproduzida na Figura 6. A representação mostra que os falantes de Inglês devem saber que, ao usar *carry* em uma sentença agentiva, o argumento agente da construção deve estar unificado ao EF Carregador do frame Carry e deve ter a forma e o comportamento gramatical de sujeito. Por outro lado, o argumento paciente deve estar unificado ao EF Carga e apresentar a forma gramatical de objeto direto.

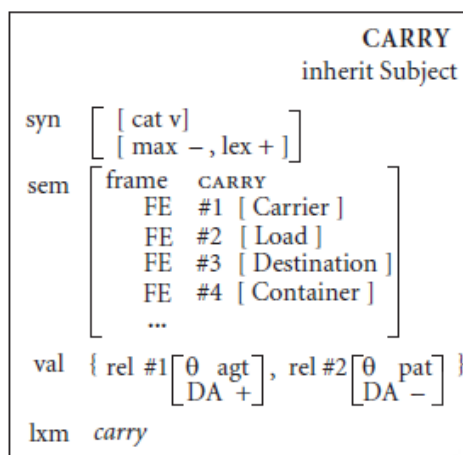


Figura 6: Representação da valência do verbo *carry*

Como o estudo dos signos através do seu padrão de valência, unificados a cenas cognitivas, fornece um modelo vantajoso para os estudos na área da Linguística Computacional e Processamento de Língua Natural – uma vez que há recursos computacionais como a FrameNet com larga trajetória de desenvolvimento – elegemos o modelo da BCG para formalizar as análises que realizamos para as construções objeto deste trabalho.

Um exemplo de formalização de construções seguindo os princípios dessa gramática pode ser observado na Figura 7. Nesse exemplo, temos a formalização do construto *A panela quebrou o cabo*, licenciado pela CAC. Cada caixa na Figura 7 indica um componente na construção. Três tipos de informações podem ser fornecidas para cada constituinte: suas características sintático-semânticas (ss), seus requisitos de valência (Val) e sua forma lexical (lform). As propriedades de cada componente são especificadas em termos de uma matriz de atributo e valor. Os valores de cada atributo podem ser expressos como binários (+/-), itens em uma lista fechada (v, det, subj, Whole...), ou não ser especificados (...). Os números de 1 a 14 são índices de unificação. Quando precedidos por uma seta para baixo, esses números indicam que algumas características semânticas dos constituintes são projetadas até um signo filho ou para toda a construção.



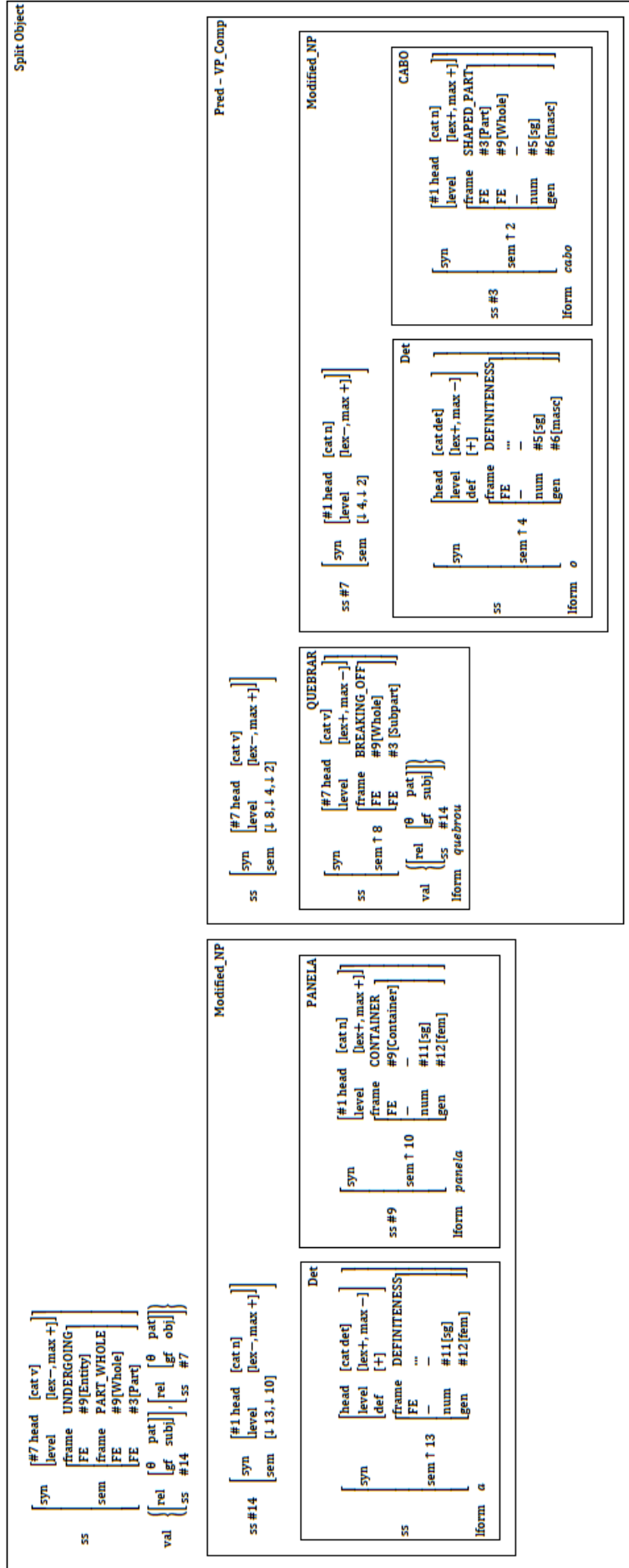


Figura 7: AVM do Construto “A panela quebrou o cabo”, licenciado pela CAC

Assim, ao se ler a Figura 7, pode-se afirmar que: o construto licenciado por esta construção é composto por dois signos filhos, um Pred\_VP\_Comp e um Modified\_NP. O Pred\_VP\_Comp é descrito a partir do verbo *quebrar* cuja valência mínima prevê o argumento sujeito paciente *a panela*, licenciado pela construção Modified\_NP. O constituinte da construção Modified\_NP, *o cabo*, está unificado semanticamente à construção Pred\_VP\_Comp, ele corresponde aos EFs #3 (Parte) / (Subparte) e #9 Todo) dos frames Artefato\_subparte<sup>1</sup> e Quebrar (). A sintaxe da construção Pred\_VP\_Comp é a mesma sintaxe do verbo e está representada pelo #7

O outro signo filho, Modified\_NP, que possui o constituinte *a panela*, unifica semanticamente o EF *Container* (#9) ao EF *Todo* (#9) do verbo. A sintaxe dessa construção (#14) será unificada à sintaxe da construção de Argumento Cindido.

Por sua vez, a construção Argumento Cindido, que evoca os frames Ser\_afetado (Undergoing) e Parte\_todo (Part\_Whole) traz a unificação entre os EFs dos frames dos dois signos filhos, e a sua valência mostra a relação sintático-semântica entre eles. Enquanto Pred\_VP\_Comp apresenta o comportamento de um objeto direto paciente, Modified\_NP se comporta como um sujeito paciente.

Notamos que a apresentação dessa CAC nos informa que a construção tem um significado próprio e um frame diferente do evocado pelo verbo. A semântica do verbo evoca o frame Quebrar (Breaking\_off) definido por uma subparte que se quebra de um todo, em que seus EFs, Todo e Subparte, são mapeados dentro do único argumento necessário na valência mínima, já a valência construcional nos permite identificar uma relação parte-todo entre as duas construções Modified\_NP representadas pela *panela* e o *cabo* respectivamente. Essa informação é encontrada apenas na descrição construcional, uma vez que a valência verbal não mostra que o cabo consiste em uma parte da panela. A construção evoca dois frames próprios: o frame Parte\_todo, que traz a informação da relação metonímica entre seus argumentos, e frame Ser\_afetado, que fornece o significado de uma entidade que foi afetada por um evento.

---

<sup>1</sup> Todos os frames mencionados neste trabalho podem ser consultados em:

Através da representação em caixas, a realidade cognitiva das construções é apresentada para demonstrar o conhecimento linguístico dos falantes e como esse conhecimento se relaciona com as várias formas com as quais as unidades linguísticas podem ser cognitivamente processadas.

A base construcional da FN-Br, o Constructicon, usada para a formalização das construções em estudo foi desenhada conforme os pressupostos da BCG, assim, passamos agora a apresentar as características dessa ferramenta, que permite a modelagem de construções no PB.

## 1.2 O Constructicon

O Constructicon constitui-se em um subprojeto da FrameNet Brasil (FN-Br) (SALOMÃO, 2009) e visa à criação de um recurso *online* para a descrição das características semânticas e gramaticais de construções em Português Brasileiro (PB). Da mesma forma em que são feitas anotações lexicográficas para a banco de dados da FN-Br, o Constructicon foi desenhado, nos moldes da FrameNet, para lidar com todos os traços linguísticos relevantes para a identificação de uma construção (SALOMÃO ET AL., 2013).

O projeto FrameNet foi iniciado no International Computer Science Institute (ICSI), por Charles Fillmore, com o intuito de fornecer, através da exposição de ULs, os *frames* evocados por essas ULs, identificados pelos Elementos de *Frames* que os constituem. As análises realizadas para as ULs, portanto, nos fornecem a descrição de suas propriedades de valência sintática (funções gramaticais e tipos sintagmáticos que co-ocorrem na localidade sintática do item lexical) e semântica (elementos de *frames* instanciados por esses valentes sintáticos). Já no *Constructicon*, encontramos a descrição de construções, seus elementos componentes e a informação semântica que podem vir a indicar.

De acordo com Salomão *et al* (2013), duas metodologias de anotação são utilizadas na FrameNet e, por consequência, na FN-Br, a saber, a lexicográfica e a de texto corrido. A anotação lexicográfica é realizada a partir do levantamento de ULs que evocam um determinado *frame*. Diante desse levantamento, faz-se uma anotação em camadas que associa os Elementos de *Frame* (EFs) a Funções Gramaticais (FG) e Tipos Sintagmáticos (TS).

A camada de EF tem o intuito de etiquetar os constituintes das sentenças conforme os papéis semânticos que estes assumem em relação à UL, assinalando os nomes dos EFs nas cores correspondentes às que foram marcadas quando da criação dos *frames*. Essas cores são mantidas no decorrer da análise para facilitar sua identificação. Os EFs contribuem com informação a respeito do frame que é evocado e podem ser classificados como nucleares, periféricos ou extra-temáticos. Os primeiros instanciam um componente conceitualmente necessário ao *frame*; os periféricos são aqueles que trazem informação adicional ao frame, dependente do tipo de entidade cognitiva representada por ele, se um evento, uma entidade, ou uma relação; já os extra-temáticos são aqueles que podem figurar numa dada sentença em que o frame se manifesta, porém não pertencem a tal cena, sendo evocados por construções próprias, e não pela UL.

A camada de FG é usada para indicar as funções gramaticais de cada constituinte da sentença alvo. As funções gramaticais codificadas pela FN-Br são de externo (Ext), que atua como sujeito da sentença, objeto direto (DObj), objeto indireto (IndObj) e dependente (Dep), função que abarca casos de adjunção, além de aposto (Aposto) e núcleo (Núcleo) funções utilizadas quando as palavras alvo são nomes ou adjetivos (TORRENT & ELLSWORTH, 2013).

Por fim, a camada de TS contempla a classificação do sintagma em que um dado constituinte se encontra, por exemplo: se verbal, é marcado como SV, se nominal, como SN.

Na Figura 8, temos um exemplo de descrição do *frame* Quebrar que serve de base para a anotação lexicográfica em camadas. Neste exemplo observamos a definição do *frame* e a divisão dos EFs em nucleares e não nucleares.

## Quebrar

Definição	
Uma única <b>Subparte</b> se quebra do <b>Todo</b> .	
Exemplo(s)	
Elementos de Frame Nucleares	
<b>Subparte</b> [subpart]	A parte do <b>Todo</b> que é concebida como estando quebrada ou tirada da parte remanescente do Todo.
<b>Todo</b> [whole]	O pedaço que é apresentado em uma posição menos proeminente e que é o ponto de referência ou o todo restante do qual a Subparte se quebrou.
Elementos de Frame Não-Nucleares	
<b>Critério</b> [criterion]	Este elemento de frame expressa a propriedade das Partes ou componentes do <b>Todo</b> que se correlaciona com as localizações (e.g.) de fragmentações do <b>Todo</b> .
<b>Lugar</b> [place]	A localização onde a ruptura ocorreu.
<b>Meios</b> [means]	Uma ação, evento ou estado que viabilizou o evento de ruptura.
<b>Modo</b> [manner]	Qualquer descrição do evento ou dos participantes não abarcados por Elementos de Frame mais específicos.
<b>Tempo</b> [time]	O tempo quando a ruptura da <b>Subparte</b> aconteceu.

Figura 8: Frame de Quebrar

A partir da atribuição de EFs aos sintagmas da sentença *O cabo da panela quebrou*, podemos fazer a identificação da realização sintática dos EFs em termos de seus Tipos Sintagmáticos e Funções Gramaticais, conforme a Figura 9.

Save Approve Disapprove Ignore Release Hide AS Add Cxn	
[13679] AST_MS_APP NI	0 c a b o d a p a n e l a q u e b r o u .
Quebrar.quebrar.v	q u e b r o u
FE	S u b p a r t e
FE	T o d o
GF	E x t
PT	N P
Other	
Verb	
Sent	

Figura 9: Exemplo de anotação lexicográfica

A Figura 9 nos mostra a sentença *O cabo da panela quebrou* anotada lexicograficamente. Primeiramente, note que as duas primeiras camadas correspondem ao EF do frame Quebrar evocado pela UL quebrou. Nesse caso, *o cabo da panela* é o único material linguístico a instanciar EFs na sentença por isso, se constitui, ao mesmo tempo como a Subparte e o Todo. A Subparte

aparece na primeira camada porque o *cabo* é o núcleo do Sintagma Nominal, já o *Todo* é representado pelo adjunto adnominal *da panela*. A anotação lexicográfica é realizada sintagmaticamente, por isso, *o cabo da panela*, que se constitui um sintagma, não pôde ser dividido em duas partes. Sendo assim, temos um sintagma nominal que se configura nos dois EFs evocados pelo frame. Em sequência, são anotadas a FG, no caso, a de Externo, e o TS, Sintagma Nominal.

A outra forma de anotação, de texto corrido, é realizada pela anotação de todas as ULs que evocam *frames* em um texto: cada UL encontrada é anotada. Diferente da anotação lexicográfica, a anotação de texto corrido não busca uma UL alvo para ser anotada, então, não há um frame único a ser anotado; como é a UL que evoca o frame e são distintas as ULs a se anotar, distintos serão, também, os frames a se anotar.

Observa-se, na Figura 10, um exemplo de anotação de texto corrido. A anotação decorre de uma monografia em que foram selecionadas redações de alunos do Ensino Médio para se realizar tal anotação com o objetivo de se verificar a sua adequação aos critérios propostos pelo ENEM (SILVA & ESTRADA, 2012).

Nessa anotação, a numeração indica cada sentença analisada. As letras indicam as sentenças anotadas mais de uma vez, por possuírem mais de uma UL evocadora de *frame*. Ao lado de cada UL, vem anotado o nome do *frame* por ela evocado. Os EFs são estabelecidos em função do *frame* evocado pela UL.

O objetivo de tais anotações é mostrar exemplos claros de possibilidades combinatórias básicas para uma UL alvo e a seleção de material para ser anotado deve cobrir os vários sentidos de padrões combinatórias encontrados no *corpus*.

Como as *framenets* são capazes de descrever as propriedades combinatórias de palavras individuais, então, veio a possibilidade de analisar todas as sentenças ou textos e fazer uma anotação. Contudo, observaram-se sentenças em que a organização sintática e semântica não podia ser totalmente explicada nos tipos de estruturas organizadas na base de anotações da FN. Nesse contexto, surge a necessidade de um novo recurso,

desenvolvido para atender justamente estruturas mais complexas e idiomáticas: o Constructicon.

### Horário Eleitoral

1a. [O horário eleitoral obrigatório<sub>Entidade</sub>] é<sub>cop</sub> **bom**<sub>Usefulness</sub> [para o povo e bom também para a democracia<sub>Finalidade</sub>].

2a. [Nele<sub>Meio</sub>], [a população<sub>Agente</sub>] tem<sub>Sup</sub> a [**oportunidade**<sub>Oportunidade</sub>]<sub>Opportunity</sub> [de conhecer melhor os candidatos e suas propostas<sub>Situação\_desejável</sub>].

2b. Nele, [a população<sub>Conhecedor</sub>] tem a oportunidade de **conhecer**<sub>Familiarity</sub> [melhor<sub>Grau</sub>] [os candidatos e suas propostas<sub>Entidade</sub>].

3a. [Pesquisa do Datafolha<sub>Entidade</sub>] **mostra**<sub>Cause\_to\_perceive</sub> [que 64% da população do Estado de São Paulo acha necessária a manutenção do horário eleitoral<sub>Fenômeno</sub>].

3b. Pesquisa do Datafolha mostra que [64% da população do Estado de São Paulo<sub>Pensador</sub>] **acha**<sub>Opinion</sub> [necessária a manutenção do horário eleitoral<sub>Parecer</sub>].

4a. Ao se **conceder**<sub>Giving</sub> [espaço nas emissoras de tv's<sub>Tema</sub>] [para os partidos políticos<sub>Recipiente</sub>], o povo tem a oportunidade de ver e analisar melhor as propostas de seus candidatos. [INC=Doador]

4b. [Ao se conceder espaço nas emissoras de tv's para os partidos políticos<sub>Condição</sub>], [o povo<sub>Agente</sub>] tem<sub>sup</sub> a [**oportunidade**<sub>Oportunidade</sub>]<sub>Opportunity</sub> [de ver e analisar melhor as propostas de seus candidatos<sub>Situação\_desejável</sub>].

4c. [Ao se conceder espaço nas emissoras de tv's para os partidos políticos<sub>Evidência</sub>], [o povo<sub>Pensador</sub>] tem a oportunidade de **ver**<sub>Grasp</sub> e analisar [melhor<sub>Maneira</sub>] [as propostas de seus candidatos<sub>Fenômeno</sub>].

4d. [Ao se conceder espaço nas emissoras de tv's para os partidos políticos<sub>Evidência</sub>], [o povo<sub>Avaliador</sub>] tem a oportunidade de ver e **analisar**<sub>Assessing</sub> [melhor<sub>Maneira</sub>] [as propostas de seus candidatos<sub>Fenômeno</sub>].

5a. No entanto, [algumas mudanças<sub>Ação</sub>] devem ser<sub>cop</sub> **feitas**<sub>Intentionally\_act</sub> [a fim de se obter um horário eleitoral mais justo<sub>Finalidade</sub>]. [INC=Agente]

5b. No entanto, algumas mudanças devem ser feitas a fim de se **obter**<sub>Getting</sub> [um horário eleitoral mais justo<sub>Tema</sub>]. [INC=Recipiente]

6a. [O tempo de visibilidade<sub>Todo</sub>], por exemplo, deve ser<sub>cop</sub> **distribuído**<sub>Separating</sub> [de forma igualitária<sub>Modo</sub>] [para todos os candidatos e coligações<sub>Recipientes</sub>]. [IND=Partes], [INC=Agente]

7a. Deve-se, ainda, **exigir**<sub>Being\_obligated</sub> [que os candidatos<sub>Parte\_responsável</sub>] [usem seu tempo para os debates e apresentação de propostas<sub>Dever</sub>], [evitando-se assim o uso de propagandas<sub>Consequência</sub>].

7c. Deve-se, ainda, exigir que [os candidatos<sub>Agente</sub>] **usem**<sub>Using</sub> [seu tempo<sub>Instrumento</sub>] [para os debates e apresentação de propostas<sub>Finalidade</sub>], [evitando-se assim o uso de propagandas<sub>Resultado</sub>].

7d. Deve-se, ainda, exigir que os candidatos usem seu tempo para os debates e apresentação de propostas, **evitando-se**<sub>Avoiding</sub> [assim<sub>Maneira</sub>] [o uso de propagandas<sub>Situação\_indesejável</sub>]. [INC=Agente]

8a. Outro ponto importante é<sub>cop</sub> a **proibição**<sub>Prohibiting</sub> [da divulgação de pesquisa eleitoral por parte dos candidatos<sub>Estado\_de\_coisas</sub>]. [IND=Princípio]

9a. [O eleitor<sub>Protagonista</sub>] deve **guiar-se**<sub>Compliance</sub> [somente<sub>Maneira</sub>] [pela análise de seu candidato<sub>Norma</sub>], não sendo influenciado por pesquisas, que em alguns casos são mal intencionadas.

9b. [O eleitor<sub>Entidade\_dependente</sub>] deve guiar-se somente pela análise de seu candidato, não sendo<sub>Cop</sub> **influenciado**<sub>Objective\_Influence</sub> [por pesquisa<sub>Entidade\_influenciadora</sub>], [que em alguns casos são mal intencionadas<sub>Descriptor</sub>].

10a. Portanto, [o horário eleitoral obrigatório<sub>Ator</sub>] **constitui**<sub>Performers\_and\_roles</sub> [um importante mecanismo para o sistema democrático<sub>Papel</sub>].

11a. [Ele<sub>Ajudador</sub>] **ajuda**<sub>Assistance</sub> [na divulgação de informação<sub>Meta</sub>] e auxilia o cidadão na escolha de seu.

11b. [Ele<sub>Auxiliador</sub>] ajuda na divulgação de informação e **auxilia**<sub>Assitence</sub> [o cidadão<sub>Beneficiário</sub>] [na escolha de seu candidato<sub>Meta</sub>].

12a. Dessa forma, **tem-se**<sub>Coming\_to\_believe</sub> [que o horário eleitoral é bom para o eleitor<sub>Conteúdo</sub>]. [INC=Pensador], [IND=Evidência]

12b. Dessa forma, tem-se que [o horário eleitoral<sub>Entidade</sub>] é<sub>cop</sub> **bom**<sub>Usefulness</sub> [para o eleitor<sub>Finalidade</sub>].

Figura 10: Anotação de texto corrido (SILVA & ESTRADA, 2012 apud LAGE, 2013, p. 24-25)

Fillmore, Kay e O'Connor (1988) apresentam construções como as regras que licenciam signos linguísticos baseados em outros signos linguísticos. As estruturas licenciadas por uma ou mais construções são chamadas de construtos. São eles que servirão de base para a anotação construcional. Cada anotação captura as propriedades particulares de um construto com respeito a uma construção que o licencia.



O propósito do Constructicon seria, então, o de 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, 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.

Isso não significa dizer, entretanto, que Lexicon e Constructicon sejam dois recursos totalmente separados. Ambos se conectam na medida em que: (a) itens lexicais são construções e são, portanto, licenciados por construções lexicais; (b) tanto itens lexicais quanto construções não lexicais evocam frames, que constituem o coração da FrameNet. A divisão que se propõe, portanto, é muito mais didática e metodológica do que teórica. Dessa forma, as duas frentes estão interligadas a fim de proporcionar, tanto no ambiente lexical como no construcional, uma descrição adequada dos fenômenos linguísticos.

Lage (2013) apresenta um conjunto de construções do PB que não são analisáveis por padrões de valência, ou seja, não podem ser anotadas lexicograficamente, merecendo, assim, uma análise construcional. Com a análise das construções da família *Para Infinitivo* (TORRENT, 2009), as quais serviram de objeto no processo de implantação do Constructicon do PB, encontramos construções como a de Dativo com Infinitivo<sup>2</sup>, exemplificada em (12).

(12) Então eles me davam aquele dinheiro para eu ir no bonde normal.

Analisando essa construção pelo viés lexicográfico, elegemos o verbo *dar*, presente na sentença, como a UL alvo. Assim, o frame evocado seria o de Dar. Através dessa anotação, é possível localizar o Tema que foi dado – o *dinheiro* –, seu Recipiente – o pronome de primeira pessoa –, o Doador – o pronome de terceira – e com que finalidade o Doador dá o Tema ao Recipiente. Contudo, não se consegue depreender de parte alguma dessa

---

<sup>2</sup> Todas as construções mencionadas neste trabalho podem ser consultadas em: <http://webtool.framenetbr.ufff.br/maestro20/index.php/fnbr20/report/cxn/main?datasource=fnbr&lang=1>

análise o fato de que a construção presente na sentença evoca o frame de Suficiência, segundo o qual, um determinado Item – nesse caso, o *dinheiro* – possibilita a ocorrência de uma Situação\_habilitada – no caso, *ir no bonde normal*. Isso ocorre porque não é o verbo “dar” que evoca o frame, mas a construção. Assim, a anotação das sentenças no Constructicon é realizada com intuito de trazer informações diferentes acerca das propriedades sintático-semânticas das sentenças.

Torrent et al. (2014), a partir das discussões feitas por Fillmore (2008) acerca das implicações da continuidade entre gramática e léxico, apresentam três princípios que orientam a anotação construcional. Primeiro, é necessário decidir se um item linguístico deve ser tratado como uma instância de uma construção no *Constructicon* do PB ou como um padrão de valência de um item lexical no banco de dados lexicográficos. Para isso, é necessário observar se existe ou não algum material linguístico lexicalmente especificado na construção.

Como nenhuma das construções estudadas neste trabalho possui material lexicalmente especificado, nenhuma delas satisfaz absolutamente esse critério, que pode ser formalizado da seguinte maneira:

Critério 1: Sendo X um material lexicalmente especificado, existe X na construção em potencial?

As construções Transitiva Direta Ativa, Ergativa e CAC podem licenciar construtos diversos, sem que seja obrigatória a presença de um item lexical específico. A resposta negativa nos direciona ao tratamento construcional, ao passo que a positiva nos revela que X poderia ser uma UL e, portanto, poderia ser tratada via padrão de valência. Somente esse critério poderia ser usado para tais construções receberem um tratamento construcional, porém discutiremos os outros critérios a fim de reforçar essa justificativa.

O segundo critério diz respeito à evocação ou não de algum frame pela construção em potencial e pode ser formalizado como:

Critério 2: Sendo F um frame e X um material lexicalmente especificado, X evoca F?

Como as construções aqui analisadas não possuem material lexicalmente especificado, também não se encaixam nesse critério. Contudo, tais construções apresentam um frame próprio que é evocado independente do verbo que as compõem.

Por fim, temos o terceiro critério que está relacionado com o fato de a UL em potencial evocar ou não o mesmo frame em padrões de valência distintos.

Critério 3: Sendo F um frame e X um material lexicalmente especificado, X evoca F em outro padrão de valência?

Como afirmado anteriormente, nem a Transitiva Direta Ativa, nem a Ergativa e nem a CAC possuem material lexicalmente e suas ULs em outro padrão de valência podem evocar frames distintos. Com isso temos mais uma afirmação da necessidade de modelagem dessas construções no Constructicon. Veja-se, por exemplo, no Quadro 1, tomando-se como base as sentenças utilizadas no experimento a ser descrito na seção 3, a diversidade de frames verbais que podem aparecer em construtos licenciados por essas três construções.

O Quadro 1 nos mostra realizações da Transitiva Ativa, da CAC e da Ergativa. Como podemos observar, tais construções licenciam construtos e cada uma possui um frame que abarca o significado geral da construção. A Transitiva Direta Ativa evoca o frame de Ação\_transitiva, a Ergativa o frame de Ser\_afetado e a CAC evoca os frames de Ser\_afetado e Parte\_todo. A variedade de verbos e suas realizações polissêmicas evidenciam a produtividade dessas construções na língua, reforçando o fato que um único lexema verbal pode aparecer em realizações distintas.

<b>Construção Transitiva Direta Ativa   Frame evocado: Ação_transitiva</b>	
<i>Sentença</i>	<i>Frame evocado pelo verbo</i>
A faxineira <b>torceu</b> o pano.	Dar_forma

A menina <b>arrebentou</b> o lacre.	Danificar
O cachorro <b>machucou</b> o menino.	Causar_ferimento
Maria <b>quebrou</b> a noz.	Causar_fragmentar
A criança <b>furou</b> os balões.	Danificar
A faxineira <b>trincou</b> o copo.	Danificar
O bandido <b>queimou</b> as provas.	Danificar
A vizinha <b>soltou</b> o cachorro.	Soltar
A professora <b>rasgou</b> o papel.	Danificar
A criança <b>descolou</b> o adesivo.	Descolar
<b>CAC   Frames evocados: Ser_afetado e Parte_todo</b>	
<i>Sentença</i>	<i>Frame evocado pelo verbo</i>
O atleta <b>torceu</b> o tornozelo.	Experenciar_ferimento_corporal
O menino <b>arrebentou</b> o nariz.	Experenciar_ferimento_corporal
O jogador <b>machucou</b> o joelho.	Experenciar_ferimento_corporal
O estudante <b>quebrou</b> o pé.	Experenciar_ferimento_corporal
O pedreiro <b>furou</b> o dedo.	Experenciar_ferimento_corporal
O celular <b>trincou</b> a tela.	Tornar-se_não-operacional
O computador <b>queimou</b> o HD.	Tornar-se_não-operacional
A panela <b>soltou</b> o cabo.	Quebrar
A calça <b>rasgou</b> o bolso.	Tornar-se_não-operacional
O sapato <b>descolou</b> a sola.	Quebrar
<b>Construção Ergativa   Frame evocado: Ser_afetado</b>	
<i>Sentença</i>	<i>Frame evocado pelo verbo</i>
O cabo <b>torceu</b> .	Tomar_forma
A corda <b>arrebentou</b> .	Quebrar
O menino <b>machucou</b> .	Experenciar_ferimento_corporal
O vaso <b>quebrou</b> .	Despedaçar
O balão <b>furou</b> .	Tornar-se_não-operacional
A janela <b>trincou</b> .	Despedaçar
O mato <b>queimou</b> .	Pegar_fogo
O botão <b>soltou</b> .	Quebrar
O lençol <b>rasgou</b> .	Despedaçar
O cartaz <b>descolou</b> .	Tornar-se_solto

Quadro 1: Exemplos de construtos licenciados pelas construções Transitiva Ativa, CAC e Ergativa

Ainda apresentando critérios para a inserção de construções no *Constructicon*, Lage (2013) afirma que tanto a FrameNet como o Constructicon foram desenvolvidos com a finalidade principal de fomentar pesquisas e tarefas em PLN. Sendo a FrameNet um recurso lexical, suas anotações concernem Unidades Lexicais e os frames por elas evocados e, sendo o *Constructicon* um recurso para atender uma série de estruturas linguísticas que não podiam ser incorporadas à anotação lexicográfica. No entanto, Lage (2013) pontua que alguns fenômenos, tipicamente analisados como construções na literatura em

Linguística Cognitiva, podem ser tratados no modo lexicográfico, visto que o frame evocado será o mesmo, independentemente de sua estrutura argumental. Isso indica que, se uma determinada estrutura já tivesse sido anotada na FrameNet, não seria necessário ser anotada outra vez no Constructicon, isto é, não seria preciso que uma mesma informação constasse mais de uma vez no banco de dados.

Na implantação do Constructicon, optou-se, portanto, pelo Princípio da Não-Redundância, ou seja, o não armazenamento redundante de dados, para evitar que construções tratáveis como padrões de valência fossem tratadas no Constructicon. Como ambos os projetos, o Lexicon e o Constructicon, se voltam a tarefas em PLN, a ideia era a de que seria prejudicial ter uma mesma informação anotada duas vezes, uma vez que isso aumentaria o custo de construção do sistema sem, por outro lado, representar ganho significativo.

De acordo com esses critérios, nem todas as construções deveriam ou poderiam ser modeladas no Constructicon, mas apenas aquelas construções em que o padrão de valência das ULs que nelas figuram não fosse suficiente para fornecer todas as informações necessárias para o significado da construção.

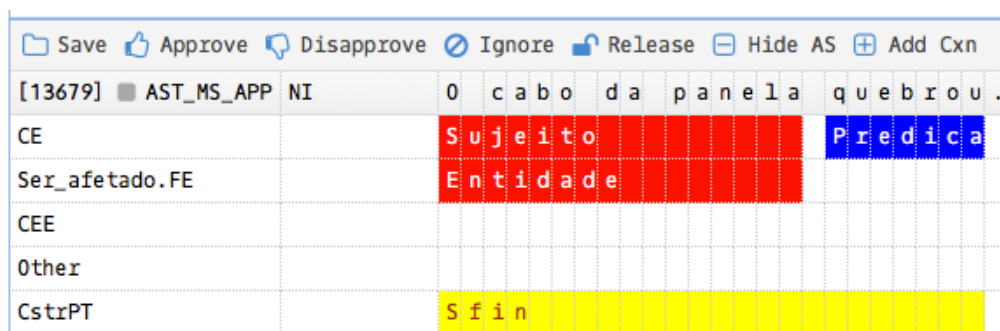
Se considerarmos apenas esses critérios, muitas das realizações das construções que serão descritas no próximo capítulo, as quais foram necessárias para a construção da rede mínima de construções apresentada, poderiam ser descritas apenas segundo o padrão de valência dos itens lexicais que funcionam como seus núcleos e, sendo assim, não seriam necessárias na base construcional da FN-Br. Porém, a nossa perspectiva atual difere do que foi proposto por Lage (2013), uma vez que um dos nossos objetivos com este trabalho é fomentar a criação de um *parser* que seja capaz de reconhecer características sintáticas e semânticas de construções.

Sendo assim, a nossa proposta visa abarcar tanto construções *core* (Transitiva Ativa Direta e Ergativa) como construções periféricas (CAC), com o intuito de constituir uma gramática do PB que inclua também padrões construcionais que se manifestam como padrões de valência das ULs. Tal decisão se ancora, ainda, no fato de que os frames tipicamente evocados por construções como as que estudamos são demasiadamente genéricos para estarem associados a uma UL. Tanto é assim que o frame Ação\_transitiva é

caracterizado na FrameNet como não-lexical, ou seja, não evocável por nenhuma UL em particular.

Diante desses critérios, a anotação construcional ocorre da mesma forma que a anotação lexicográfica, na medida em que busca um item alvo para ser anotado. Por outro lado, ela também se assemelha com a anotação de texto corrido, uma vez que não se prende apenas a unidade alvo, sendo assim, qualquer informação relevante para a construção será anotada.

A anotação construcional também é realizada em camadas. Nela, encontramos uma camada para identificar a expressão licenciada pela construção, outra camada para identificar os componentes constituintes da expressão. Juntas essas camadas identificam o construto descrito pela construção. Pode haver ainda outras camadas para identificar os elementos externos do construto, como, por exemplo, o contexto.



	O	c	a	b	o	d	a	p	a	n	e	l	a	q	u	e	b	r	o	u	.
CE																					
Ser_afetado.FE																					
CEE																					
Other																					
CstrPT																					

Figura 11: Anotação construcional

A Figura 11, apresenta a mesma sentença trazida na Figura 9, *O cabo da panela quebrou*, anotada construcionalmente. Nesse exemplo, observamos, na primeira camada, os elementos da construção (CEs), o *cabo da panela* e *quebrou* anotados respectivamente como Sujeito e Predicado. Na segunda camada, encontramos a relação entre a construção e o frame que ela evoca, *Ser\_afetado*, através da unificação entre o CE e o EF, e a última camada anotada corresponde ao Tipo Sintagmático do construto. As diferenças entre as formas de anotação ocorrem porque, enquanto o método lexicográfico representa apenas a UL e os seus elementos de acordo com seu padrão de valência, o método construcional mapeia uma construção formalmente, unificando-a com um frame específico que resulta no licenciamento de um

construto. Nesse exemplo, o construto *O cabo da panela quebrou* é licenciado pela construção Ergativa, unificada ao frame Ser\_afetado.

Portanto, a ancoragem do Constructicon na BCG se manifesta também devido à forma como essa ferramenta é capaz de representar a unificação, a qual é tratada anotando-se as instâncias da construção em camadas de modo a alinhar os CEs com os valentes do frame evocado pela construção.

Tal recurso destina-se a permitir uma integração mais profunda entre os dois tipos de anotação desenvolvidos na FrameNet Brasil, a lexical e a construcional. Assim sendo, as camadas usadas na anotação de sentenças representam a unificação de um conjunto de valores para os atributos, Elementos da Construção, Tipo Sintagmático, Função Gramatical, associados ao próprio construto. Através dessa tarefa, pode-se modelar adequadamente a continuidade entre o léxico e gramática do PB pelo desenvolvimento de um banco de dados em que o Lexicon e o Constructicon funcionem juntos.

Diante das informações acerca do enquadre computacional da BCG e do *Constructicon*, justificamos a inserção desse trabalho nos estudos sobre o Processamento de Língua Natural, a fim de fornecer bases para a pesquisa de reconhecimento de construções por máquina. Desse ponto trataremos na próxima seção.

## **1.3 Processamento de Língua Natural**

### *1.3.1 Bases do Processamento de Língua Natural*

O Processamento de Língua Natural (PLN) consiste no desenvolvimento de modelos computacionais para a realização de tarefas que dependem de informações expressas em alguma língua natural.

De acordo com Dias da Silva (1996), o estudo do PLN visa à implementação de sistemas computacionais em que a comunicação entre o homem e a máquina possa se dar através de uma língua natural, ou “pseudolíngua natural”, e não pelo uso de comandos de uma linguagem de programação qualquer. Sendo assim, o estudo do PLN deve fornecer a base

para a implementação de programas computacionais específicos que, de alguma forma, sejam capazes de lidar com objetos linguísticos.

Na sequência de enunciados *João bateu com o carro. Ele quebrou o braço.*, sabemos que, na segunda sentença, o pronome “ele” está relacionado ao referente “João”, porque carro não tem braço. Mas como um sistema computacional identificaria essa relação? Essas e outras questões inerentes à linguagem natural são tratadas no campo do PLN, que envolve tanto conhecimentos em computação como em linguística.

Algumas das primeiras investigações em PLN começaram no início da década de 50 com projetos sobre um novo campo de pesquisa denominado “Tradução Automática”, “Tradução Mecanizada” ou simplesmente MT (*Machine Translation*). Traduzir, então, era considerado um procedimento semelhante ao de decifrar códigos criptografados. O *Georgetown Experiment*, que propôs a tradução do Russo para o Inglês através desse método foi cercado de grande otimismo quanto a avanços eminentes, porém, alguns dos principais problemas desse tipo de tradução, especialmente os relacionados à desambiguação semântica, ainda eram difíceis de resolver por meios automáticos (KOEHN, 2010).

O experimento incluiu a tradução completamente automática de mais de 60 frases do russo para o inglês. A experiência foi vista, a princípio, como um grande sucesso e marcou o começo de uma era em que a investigação em tradução automática foi abundantemente financiada. Os autores consideravam que em três a cinco anos a tradução automática seria um problema trivial.

Originalmente concebido e desenvolvido para atrair o interesse do público e do governo e mostrar as possibilidades de tradução automática, o sistema dispunha de seis regras gramaticais e duzentos e cinquenta elementos no seu vocabulário, sendo especializado no campo da química orgânica.

Contudo, o progresso verificado foi muito mais lento do que o previsto. Os EUA, em 1966, financiaram o ALPAC (Automatic Language Processing Advisory Committee) para realizar um estudo sobre tradução por máquina. O estudo revelou que a tradução por máquina não era mais barata ou mais rápida que a tradução humana o que ocasionou uma interrupção de investimentos na área, mostrando os perigos de altas promessas para sistemas de tradução automática, uma vez que o problema da representação do significado de uma



maneira formal revelou-se como um grande desafio da inteligência artificial (KOEHN, 2010).

Devido a esse fato, em meados de 1970, os estudos para os trabalhos de tradução automática, bem como para as outras tarefas de PLN, ganharam uma atitude mais acadêmica e realista. Os pesquisadores passaram a ser mais cautelosos diante do complexo processo de tradução e da própria sofisticação do código linguístico. Pesquisas mostraram que a máquina poderia ser programada para processar objetos linguísticos, como interpretar perguntas, asserções e comandos; inferir, explicar ações e até mesmo “aprender” palavras novas.

Esse viés mais acadêmico dos estudos em PLN ganhou força, uma vez que se percebeu que o financiamento devia ir também para a pesquisa em linguística. A partir da década de 80, encontramos uma sofisticação dos sistemas de PLN através do desenvolvimento de teorias linguísticas que eram motivadas pelos estudos na área. De 1990 em diante, buscou-se o desenvolvimento de projetos de sistemas de PLN complexos com o intuito de integrar os vários tipos de conhecimentos linguísticos e extralinguísticos, bem como as estratégias de inferência envolvidas no processamento da linguagem por seres humanos.

Ainda segundo Koehn (2010), nas décadas de 80 e 90, o formalismo sintático tornou-se mais sofisticado com gramáticas reversíveis utilizadas tanto para geração como para análise de sentenças. A noção de representação do significado de maneira formal agrupou várias vertentes de investigação, como a inteligência artificial e a linguística computacional. O estudo do PLN, enquanto área multidisciplinar, hoje é uma realidade. Temos o desenvolvimento de trabalhos cooperativos em áreas da Linguística, Inteligência Artificial e Ciência da Computação, tornando evidente a necessidade de equacionar competências específicas de cada área para a realização desse tipo de pesquisa.

Nesse contexto, Dias da Silva (1996) propõe uma estratégia de pesquisa para o estudo do PLN que envolve a integração de questões em três domínios: **linguístico, representacional, e implementacional**. Segundo autor, as três fases podem ser entendidas da seguinte maneira:

- **Fase Linguística:** construção do corpo de conhecimentos sobre a própria linguagem, dissecando e compreendendo os fenômenos linguísticos necessários para o desenvolvimento do sistema. Nesta fase, a análise dos fenômenos linguísticos é elaborada em termos de modelos e formalismos desenvolvidos no âmbito da teoria linguística.
- **Fase Representacional:** construção conceitual do sistema, envolvendo a seleção e/ou proposição de sistemas formais de representação para os resultados propostos pela fase anterior. Nesta fase, projetam-se as representações linguísticas e extralinguísticas em sistemas formais computacionalmente tratáveis.
- **Fase Implementacional:** codificação das representações elaboradas durante a fase anterior em termos de linguagens de programação e planejamento global do sistema. Nesta fase, além de transformar as representações da fase anterior em programas computacionais, estudam-se as questões referentes à integração conceitual e física dos vários componentes envolvidos, bem como questões referentes ao ambiente computacional em que o sistema será desenvolvido e implementado (DIAS DA SILVA, 1996, p. 91 ).

Dotado de uma base gerativista, Dias da Silva (1996, p.85) afirma que, para a construção de sistemas computacionais que simulem a competência e o desempenho linguísticos dos seres humanos, devem ser definidos os conhecimentos e habilidades que os falantes possuem. Assim, com base nas considerações de Lasnik (1990 apud DIAS-DA-SILVA, 1996, p.87), o autor faz uma correlação de sua proposta de divisão em níveis com os estudos das ciências cognitivas. Dessa forma, esse equacionamento pode ser resumido da seguinte maneira:

A explicitação do conhecimento e uso linguísticos envolve questões do domínio linguístico, uma vez que é nessa fase que os fatos da língua e do seu uso são especificados. Conceitos, termos, regras, princípios, estratégias de resolução de problemas e formalismos linguísticos são os elementos trabalhados. No domínio da representação, questões referentes à escolha ou à proposição de sistemas de representação, que incluem, por exemplo, a lógica, redes semânticas, regras de reescrita e *frames*, bem como estratégias de codificação dos elementos trabalhados no domínio anterior, entram em foco. No domínio da implementação, além das questões que envolvem a implementação das representações por meio de programas, há questões que dizem respeito à montagem do próprio sistema computacional em que o programa será alojado. (DIAS DA SILVA, 1996, p.101)

DOMÍNIOS	PROBLEMAS	RECURSOS
Lingüístico	explicitar o conhecimento e o uso lingüístico	teorias da competência e do desempenho
↓↑	↓↑	↓↑
Representacional	representá-los	linguagens formais de representação
↓↑	↓↑	↓↑
Implementacional	coficar as representações	linguagens de programação e sistemas computacionais

Figura 12: Estratégia para o equacionamento entre domínios, problemas e recursos no PNL (DIAS-DA-SILVA, 1996, p.100)

O esquema da Figura 12 (DIAS DA SILVA, 1996, p.100) resume a estratégia para o equacionamento desses níveis no PLN.

Esta pesquisa, que busca modelar computacionalmente as Transitiva Direta Ativa, Ergativas e CACs, se enquadra dentro da fase representacional, uma vez que a fase linguística já foi proposta, conforme se mostrará no próximo capítulo. Esta última fase apresentou a descrição dos padrões licenciados pelas construções em estudo, levando em consideração aspectos sintáticos, semânticos e pragmáticos.

Partindo dessa descrição, apresentaremos, nos próximos capítulos, no domínio da representação, uma proposta de formalização de construções da Berkeley Construction Grammar, o Constructicon; e, no domínio da implementação, os resultados do emprego da técnica de Spreading Activation para a interpretação adequada dessas construções pela máquina.

Com base no esquema de equacionamento de Dias da Silva (1996), este trabalho pode ser concebido conforme a Figura 13.

De acordo com Dias da Silva (1996, p.94), a fase representacional está entre a representação fornecida pela teoria linguística e o algoritmo e é, além de necessária, “*estrategicamente positiva*”, pois, além de funcionar como um projeto arquitetônico que traz as informações necessárias para a construção do sistema, garante maior “*transportabilidade*” dos resultados já alcançados para

tipos de máquinas diferentes, que possam vir a empregar linguagens de programação também diferentes daquelas originalmente utilizadas nos primeiros testes implementacionais.

<b>DOMÍNIOS</b>	<b>PROBLEMAS</b>	<b>RECURSOS</b>
Linguístico	Explicitar o conhecimento e o uso linguístico	Teorias da Gramática das Construções e Semântica de Frames
↓↑	↓↑	↓↑
Representacional	Representá-los	Modelo de formalização da Berkeley Construction Grammar (Constructicon).
↓↑	↓↑	↓↑
Implementacional	Codificar as representações	<i>Parser</i> semântico CARMA

Figura 13: Proposta para o equacionamento dos três níveis no âmbito deste trabalho

No caso deste trabalho, o produto da fase representacional foi inicialmente aplicado a uma tarefa de *parsing*, área do PLN sobre a qual discorreremos a seguir.

### 1.3.2 *Parsing*

Dentro do domínio do PLN, um vasto conjunto de tarefas está envolvido, além da tradução automática, que vão desde tratamentos superficiais do texto até análises profundas do ponto de vista sintático e semântico. Muitas dessas tarefas necessitam de tratamentos que incluam análises precisas que permitam fornecer, para uma dada sentença de entrada, as suas possibilidades de análise sintática e semântica, através das possibilidades combinatórias das classes de palavras contidas na sentença. Esta transformação, conhecida como *parsing*, consiste em um processo de combinar os símbolos de um texto em um grupo que pode ser substituído por outro símbolo mais geral. Este novo símbolo pode, por sua vez, ser combinado em um outro grupo, e assim por diante, até que surja uma estrutura tratável computacionalmente.

De acordo com Othero e Menuzzi (2005), o *parsing* diz respeito à interpretação automática de sentenças de linguagem natural por meio de programas de computador conhecidos como *parsers*, que são capazes de classificar morfosintaticamente as palavras e expressões de sentenças em uma dada língua e, principalmente, de atribuir às sentenças a sua estrutura de constituintes, baseando-se em modelo de gramática. Atualmente, nos sistemas de *parsing*, a análise de um texto acontece de forma modular, ou seja, o texto é processado sequencialmente por um tokenizer (que divide o texto em tokens), por um tagger morfológico (que cria uma etiqueta para cada token), por um componente sintático e, finalmente, por um componente semântico. Cada componente usa a saída do componente prévio como entrada e gera sua própria saída, enriquecida com mais informação linguística. O *parser* semântico é geralmente o último estágio do processamento do texto. Como é possível perceber, o que se busca com todo este processamento, até chegar ao nível semântico, é uma interpretação (representada formalmente) do texto sob análise.

Para o PB há poucos *parsers*, sendo um deles o PALAVRAS (BICK, 2000) desenvolvido por Eckhard Bick na Universidade de Århus (Dinamarca). O PALAVRAS se apoia em um léxico de 50.000 lemas e milhares de regras gramaticais para fornecer uma análise morfológica e sintática de qualquer texto. O formalismo aplicado integra-se a tradição da Constraint Grammar.

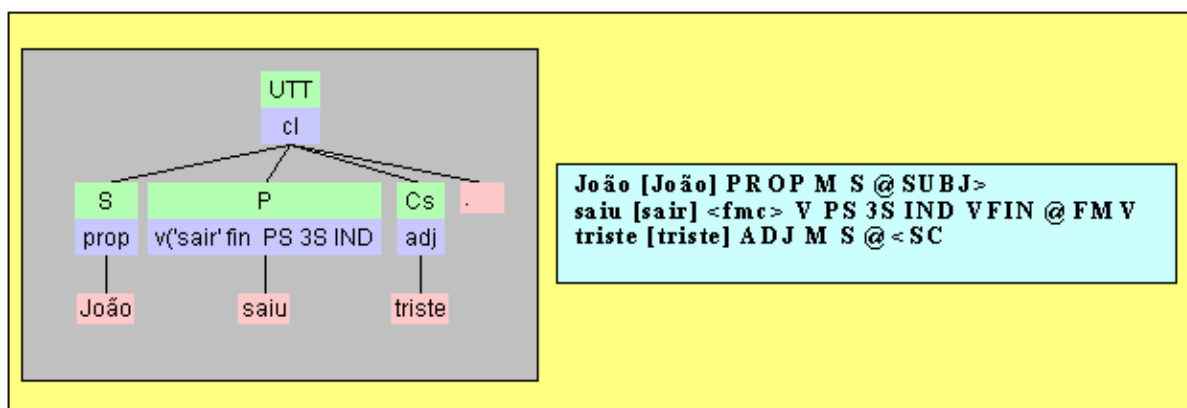


Figura 14: Teste da sentença João saiu triste no PALAVRAS (CONTERATTO, 2006, p. 6).

A Figura 14 mostra um teste realizado no *parser* PALAVRAS para a sentença *João saiu triste*. Nela, a etiqueta “PROP” indica que o sujeito *João* é

constituído de um nome próprio (masculino/singular). A flexão de *sair* — que está na terceira pessoa do singular do pretérito perfeito do modo indicativo — é indicada pelos caracteres V PS 3S IND, e o predicativo *triste* possui as etiquetas ADJ M S, indicando um adjetivo masculino singular.

De acordo com Conteratto (2006), a análise sintática no PALAVRAS é realizada com a ajuda de etiquetas morfológicas e sintáticas aplicadas ao texto de entrada pelo analisador e etiquetador morfológico. A cada palavra do texto em análise é associada uma descrição que inclui a classe morfológica (indica se é um substantivo, um advérbio, etc.) e uma forma sintática, representada essa por marcadores de dependência – @, > –, que indicam o núcleo da unidade.

Outro projeto, o CURUPIRA, (MARTINS ET AL, 2003), é um analisador de uso geral para o PB, que analisa sentenças através de uma gramática funcional livre de contexto desenhada para o padrão escrito da língua. O CURUPIRA tem como enfoque primário analisar sequências de palavras, independentemente da sua gramaticalidade. As relações de dependência não são verificadas, nem a desambiguação lexical é levada em consideração. A entrada do CURUPIRA pode ser uma frase isolada ou um texto e a saída constitui-se de uma notação própria, na qual procura-se etiquetar sintaticamente cada um dos itens lexicais da sentença, por menores e menos expressivos que sejam. O *parser* usa um conjunto razoavelmente extenso de etiquetas sintáticas, baseadas na Nomenclatura Gramatical Brasileira.

Segundo Martins et al (2002, p.17-18), o CURUPIRA processa a sentença em sentido descendente, recursivo, da esquerda para a direita, a partir da informação disponibilizada pelo dicionário e dá prioridade de aplicação das regras estipulada pela gramática. Para tanto, a sentença é processada seguindo os seguintes passos:

- *Tokenização* da sentença, em que a frase é quebrada em itens lexicais e cada um deles é classificado como um *token*;
- Recuperação das categorias e atributos dos itens lexicais no dicionário e verificação de clíticos pronominais e formas compostas;

- Desambiguação dos itens lexicais alocados em mais de uma categoria distinta, com base nas categorias do próprio item e em informações dos itens lexicais vizinhos à esquerda e à direita;
- Análise sintática dos itens lexicais através da aplicação das regras do *parser*, as quais são reaplicadas recursivamente enquanto houver sucesso nos mapeamentos entre as regras e os itens lexicais;
- Formatação e armazenamento dos resultados em um arquivo de saída, percorrendo-se as árvores sintáticas formadas na memória do programa;
- Carregamento do arquivo de saída pela interface para ser visualizado.

Martins et al (2002) apresentam – Figura 15 – o resultado do *parsing* da sentença *O português é a língua oficial do Brasil*, mostrando todos os atributos dos itens lexicais e as regras de reescrita categorial que foram usadas durante a análise sintática:

```
ART/PROINDVAR/PROPOS/PREP/CONTR/NUM + *ADJ/PART* + VERBO/~SUBST --> SUBST (português)

0) o --> ART
// Art(masculino, singular, artigo definido) Pro(masculino, singular, [terceira pessoa,
numero indefinido]) Sub(masculino, singular, grau nulo)

1) português --> SUBST
// Adj(masculino, singular, grau nulo) Nom(masculino, singular, grau nulo)

2) é --> VERBO[ser]-> Lig TransInd
// Ver(presente, terceira pessoa, singular, genero indefinido)

3) a --> ART
// Art(feminino, singular, artigo definido) Pre() Pro(feminino, singular, [terceira
pessoa, numero indefinido]) Abrev(masculino, singular) Sub(masculino, singular, grau
nulo)

4) língua --> SUBST
// Sub(masculino, singular, grau nulo) Sub(feminino, singular, grau nulo)

5) oficial --> ADJ
// Adj(masculino, singular, grau nulo) Adj(feminino, singular, grau nulo)
Sub(masculino, singular, grau nulo) Sub(feminino, singular, grau nulo)

6) do --> PREPOSICAO
// Pre()
```

```

7) do --> ART
// Art(masculino, singular, artigo definido)
8) Brasil --> NOMPRO
// Nom(masculino, singular, grau nulo)
9) #>>>>>>>>> DELIMITADOR

```

Figura 15: Parsing da sentença O português é a língua oficial do Brasil (MARTINS ET AL, 2002, p.11)

A Figura 15 mostra, em parte, como as categorias gramaticais que estão em azul foram analisadas sintaticamente pelo CURUPIRA. Para cada categoria, observamos características que as definem, tais como gênero, número, grau etc.

Por fim, cabe salientar que, entre os objetivos do CURUPIRA, não está o de desambiguar a estrutura sintática das sentenças da língua portuguesa. A ferramenta fornece todas as possibilidades de análise sintática, apresentadas segundo a prioridade de aplicação das regras da gramática que lhe serve de base, sem qualquer compromisso com a indicação da estrutura "correta", ou "mais adequada", ou mesmo "mais provável" para a sentença de entrada. Como o CURUPIRA não realiza nenhuma espécie de análise semântica, tudo o que consegue fazer é combinar palavras e classes de palavras de forma a oferecer, como saída, suas possibilidades de estruturação sintática, muitas das quais se revelarão, evidentemente, descabidas e impertinentes, por não admitirem nenhuma possibilidade de projeção semântica (MARTINS ET AL, 2002).

Um problema predominante para o *parsing* é a transformação de uma sentença de entrada, potencialmente ambígua, em uma forma não ambígua que possa ser manipulada internamente em um sistema computacional. Diante disso, algumas questões devem ser levantadas, a partir da análise destas ferramentas.

Primeiramente, observamos a centralidade da morfossintaxe. Em alguns casos, vemos a definição do processo de *parsing* com um processo de análise exclusivamente sintática, ignorando as questões associadas à semântica. Assim sendo, sentenças licenciadas pelas construções Transitiva Direta Ativa e



CAC, apesar de semanticamente distintas, teriam o mesmo tratamento em *parsers* como os apresentados.

Uma forma de aumentar a eficiência e a abrangência no processo de *parsing*, vencendo algumas das limitações descritas, é agregar fatores semânticos à descrição gramatical. Uma possibilidade é o uso de gramáticas baseadas em construções, uma vez que a análise a ser realizada pelo *parser* não envolve apenas a análise sintática, como também a análise semântica. Nesse contexto, o aspecto semântico não deve ser limitado exclusivamente à semântica formal. Além disso, a análise construcional deve ser capaz de possibilitar o teste de variadas gramáticas, a inclusão de aspectos contextuais e apresentar um modelo de interpretação que seja cognitivamente plausível.

A partir dessas questões, uma nova proposta para o *parsing*, que engloba fatores semânticos que estão vinculados às teorias como a Gramática das Construções e da Semântica de Frames, pode ser apresentada. Um sistema que difere dos *parsers* apresentados anteriormente é o Constructional Analyzer (BRYANT, 2008). Este sistema é usado no contexto da ECG (Embodied Construction Grammar) para a interpretação cognitivamente plausível de uma dada sentença. O *Analyzer* emprega a ECG como espinha dorsal: construções e estruturas linguísticas cognitivas, tais como esquemas imagéticos e frames, formam a base de cada interpretação desenvolvida pelo analisador. O *Analyzer* utiliza uma unidade computacional chamada de reconhecedor de construção, que é projetado para verificar a forma e o significado das restrições das construções.

Em vez de um *parser* monolítico que usa regras gramaticais como entrada, o próprio *Analyzer* é um repertório de reconhecedores de construções que trabalham em conjunto para gerar a análise construcional. Devido ao emparelhamento inicial de forma e significado proporcionado por construções, esse analisador com acesso às informações semânticas é capaz de usar como *input* soluções heurísticas para os problemas comuns associados a análise da linguagem real.

Bryant (2008) afirma que a chave dessa abordagem e o que torna o sistema mais robusto é a ideia que cada enunciado está tentando comunicar algum tipo de cena e, nesta cena, os participantes são parametrizados em

fames e esquemas imagéticos na ECG, assumindo que uma melhor análise é aquela que melhor descreve a cena.

Como exemplo, considere-se a especificação semântica gerada pelo *Analyzer* para a sentença *She gave him a cookie*, na Figura 16.

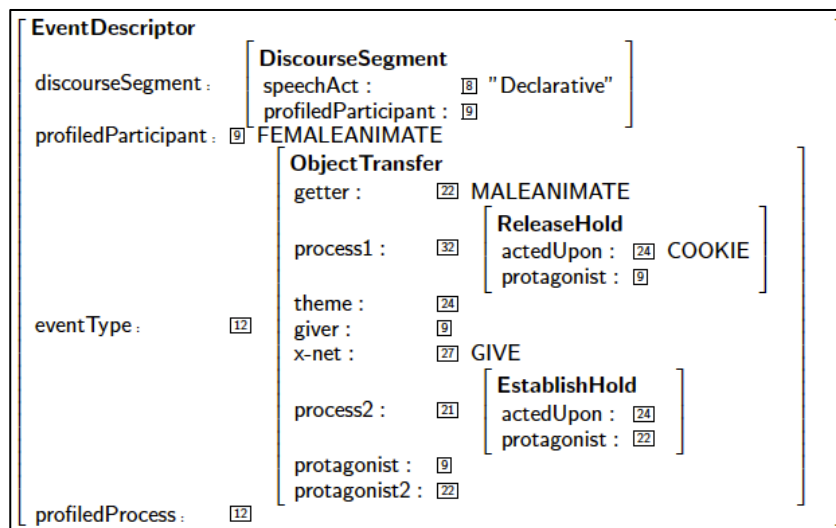


Figura 16: Especificação semântica gerada pelo *Analyzer* (BRYANT, 2008, p.145)

A Figura 16 mostra a especificação semântica para a sentença, *She gave him a cookie*. Nela, o FEMALEANIMATE é o perfil de identificação do participante e está substituindo o pronome *She*. O tipo de evento é perfilado como uma transferência de objeto. Esse tem o receptor definido como MALEANIMATE (*him*), o doador está definido como FEMALEANIMATE, e o tema definido como o *cookie*. Notamos que os números representam a unificação dos participantes de cada cena.

Um *parsing* que abarca aspectos semânticos através de cenas e esquemas consegue tratar de forma mais satisfatória os fenômenos na língua, por isso, um *parser* com subsídios construcionais pode contribuir para os estudos no PLN.

Nesse sentido, na busca de contribuir para o desenvolvimento de um *parser* construcional apresentaremos na próxima seção a ferramenta utilizada para o reconhecimento das construções em estudo.

## 1.4 CARMA

O reconhecimento das construções Transitiva Direta Ativa, Ergativa e Argumento Cindido ocorreu pelo CARMA (Constructional Analyzer using Relations among Multiple AVMs), *parser* em desenvolvimento pela FrameNet Brasil. Sendo assim, através das duas frentes da FN-Br, o Lexicon e o Constructicon, o parser utiliza a modelagem realizada por ambos recursos para gerar especificações sintático-semânticas de sentenças do PB.

O CARMA é um framework para a análise de construções gramaticais. A partir de um construto, ou seja, uma sentença em língua natural, a sua tarefa é encontrar as construções associadas a esse construto. Com o uso do Lexicon e do Cosntructicon, o CARMA se diferencia de outros *parsers* por utilizar, durante o processamento, informações semânticas relacionadas aos frames verbais e construcionais dos construtos em análise.

Para o processo de análise, no CARMA, é criada uma rede semântica cujos nós são as Palavras, Lexemas, Lemas, Unidades Lexicais, Frames, Elementos de Frame, Construções e Elementos de Construção que estão disponíveis na base de dados da FrameNet Brasil. Os links entre os nós são baseados nas relações existentes entre os diversos elementos: relações entre Frames, entre Elementos de Frames, entre Construções e Frames, entre Unidades Lexicais e Frames etc., e um outro link específico entre as palavras é capaz de representar a ordem em que elas aparecem na sentença.

A estrutura do CARMA apresenta a relação de Evocação entre uma construção e um frame de maneira invertida na rede, pois a ocorrência de um dado frame que é evocado por uma construção é utilizada como evidência para a ocorrência da construção, ou seja, é o frame que ativa a construção e não a construção que ativa o frame. Por outro lado, a relação entre um frame e seus EFs ocorre de duas maneiras, tanto a ocorrência de um frame indica a ocorrência de seus elementos de frame, quanto a ocorrência de um dado elemento de frame indica a ocorrência do frame.

Uma relação que não está explícita no banco de dados é a relação entre um lexema e a construção lexical correspondente. No CARMA foram criadas construções lexicais representando as classes de palavras (Nome, Verbo,

Adjetivo, Artigo, etc.). Com base na classe do lexema, ele é associado automaticamente à construção lexical correspondente

Como já mencionado anteriormente, o processamento no CARMA utiliza a técnica da Ativação Propagada, seguindo um procedimento semelhante ao usado para o framework LUDI (MATOS, 2014). O processo é iniciado com a construção da Rede Completa. Esta rede é composta por todos os nós que podem potencialmente ser ativados, a partir das palavras existentes na sentença sob análise. O processo de Ativação Propagada é então aplicado sobre a Rede Completa, gerando a Rede Ativada. Quando o processo de ativação termina, é gerada uma Rede Alvo, a partir das construções e frames que foram ativados por todas as palavras constantes na sentença.

A estratégia de ativação consiste em iterações, ou seja, quando um nó é ativado, ocorre a análise dos nós ativados anteriormente. Os nós relativos às palavras têm um nível de ativação ( $A$ ) igual a 1, enquanto os demais nós têm o nível de ativação inicial igual a zero. A primeira palavra da sentença é então ativada, então cada iteração, cada nó, propaga sua ativação para os nós vizinhos, com uma função do valor atual do seu nível de ativação e dos pesos dos links que o conectam com cada nó vizinho. A cada iteração  $p$ , cada nó  $j$  tem um nível de ativação igual a  $A_j(p)$  e uma saída (output) igual a  $O_j(p)$ , que é uma função do seu nível de ativação, conforme a equação:

$$O_j(p) = f(A_j(p))$$

O processo de um nó que espalha sua saída para outros nós é denominado *firing*. A saída de cada nó afeta o nível de ativação da próxima iteração de qualquer nó  $k$  ligado a  $j$  por um link direto. Assim, o nível de ativação de cada nó  $k$  da rede é função da saída (na iteração anterior ( $p - 1$ )) de cada nó vizinho  $j$  que tenha um link direto  $e_{jk}$ , bem como uma função do peso do link  $W_{jk}$ , como mostrado na equação:

$$A_k(p) = \sum_j O_j(p - 1)W_{jk}$$

A função de saída  $O$  (uma variação da função logística) foi escolhida cuidadosamente, para evitar um excesso de ativações nos nós da rede e é apresentada na equação:

$$O_j(p) = \begin{cases} 0, & \text{se } A_j(p) < \tau \\ \frac{1 - \exp(5 * (-A_j(p)))}{1 + \exp(-A_j(p))}, & \text{caso contrário} \end{cases}$$

O valor  $\tau$  é o threshold, ou seja, o valor limite, abaixo do qual a saída não é mais gerada, impedindo que nós com níveis de ativação muito baixos influenciem seus nós vizinhos. Em termos de implementação computacional, o framework CARMA trabalha com um valor para o limite  $\tau = 0.001$  e, como condição de parada, com a inatividade dos nós (ou seja, quando não há mais nós a serem ativados).

Quanto à estrutura interna do nó, cada nó na rede – Lemas, Lexemas, Frames, ULs, EFs, CEs – possui uma estrutura interna, que, por sua vez, é formatada em Matriz de Valores de Atributos, representada via uma Estrutura de Traços Tipados, o que permite uma implementação computacional, usando um modelo conexionista estruturado, facilitando a integração linguístico-computacional. Dessa forma, o Quadro 2 traz os principais elementos desta estrutura.

Elemento	Valor	Descrição
Id	String	Identificação do nó na rede.
A	Real	Nível de ativação do nó, computado com base nas entradas.
O	Real	Valor de output, calculado a cada ativação.
energy	Inteiro [0..10]	Um valor inteiro, com o objetivo de evitar que ocorram laços infinitos durante a ativação dos nós, em caso de ativação mútua (quando o nó A ativa o nó B e vice-versa). O valor de energy é diminuído em 1 a cada ativação e quando chega a 0 o nó não é mais ativado.
Input	Array	Representa os vários links de entrada no nó (vindos de outros nós da rede). Um link pode ser opcional ou

		mandatário. Um nó só dispara ( <i>firing</i> ) quando todas as entradas obrigatórias estão ativas.
Output	array	Representa os vários links de saída do nós (em direção a outros nós).
Status	String	'free': O nó está liberado para <i>firing</i>  'blocked': O nó pode receber mais ativação mas não pode <i>firing</i>  'waiting': Quando o nós está esperando que as entradas obrigatórias sejam ativadas, para que possa <i>firing</i>
Before	Inteiro	Indica a identificação de um nó que deve estar ativo antes que o nó atual possa <i>firing</i> .

Quadro 2: Elementos estruturais do CARMA

Devido ao fato de o CARMA estar em fase de desenvolvimento, o sistema apresenta algumas limitações, como, por exemplo, a ausência de Tipos Ontológicos, que podem acrescentar mais evidências da ocorrência de um Frame/Elemento de Frame, como também o fato de que a ocorrência de uma relação específica entre nós pode evidenciar a ocorrência de outro nó. Da mesma forma, a evocação de um frame por uma UL está sendo representada diretamente com um link entre a UL e o frame, já que uma abordagem mais cognitivamente plausível deveria considerar as valências da UL em relação ao frame, que podem ser obtidas através das anotações disponíveis na FN-Br. Por fim, não estão sendo tratadas as possíveis ocorrências de Instanciação Nula.

Diante dos recursos existentes na fase inicial dessa ferramenta, apresentaremos na próxima seção os resultados obtidos no processo de análise que é denominado Especificação Semântica (SemSpec – Semantic Specification), para as construções Transitiva Direta Ativa, Ergativa e CAC.A partir da apresentação do Analyzer e do CARMA, pontuamos que um *parser* capaz de abarcar aspectos semânticos em suas análises possui a capacidade de tratar de forma mais satisfatória dos fenômenos linguísticos.

O Analyzer através de esquemas, que representam nossas experiências corporais, sejam sensoriais ou perceptuais em nossa interação com o mundo, apresenta a semântica das construções como conceitos abstratos que

emergem da nossas experiências. Nesse *parser* a análise, ou especificação semântica de uma sentença ocorre por meio de AVMs, em que são calculadas a pontuação de probabilidade para uma interpretação. Para isso, o sistema considera informações sintáticas e semânticas em um modelo probabilístico conjunto, que alia essas informações.

O CARMA apresenta em sua base frames e construções conforme as relações existentes no Lexicon e no Constructicon. Diferente do Analyzer, a análise no CARMA ocorre através de uma rede de ativação, em que o reconhecimento de sentenças é realizado de acordo com o seu nível de ativação, para tanto utiliza da técnica da Ativação Propagada.

Cabe ressaltar ainda, que o CARMA funciona como um reconhecedor de construções que já estão modeladas no Constructicon, ou seja, seu papel é mostrar como a modelagem construcional contribui para que os padrões construcionais podem ser reconhecidos pela máquina. O Analyzer, apresenta a relação dos constituintes das sentenças, mostrando como eles estão relacionados semanticamente aos esquemas e frames.

## **2 CONSTRUÇÕES DE ESTRUTURA ARGUMENTAL DO PORTUGUÊS DO BRASIL**

Neste capítulo, apresentamos as construções analisadas nesse trabalho. Primeiramente, reportamos descrições linguísticas já existentes na literatura e, depois, trazemos a modelagem construcional pela FN-Br.

Em uma descrição de construções nucleadas do inglês, Kay e Fillmore (1999) propõem a existência de três principais tipos de construções: (i) aquelas herdeiras da construção Núcleo\_Complemento (HC), que consistem de um núcleo lexical seguido por um ou mais complementos; (ii) as herdeiras da construção X\_Núcleo (XH), compostas de um núcleo não necessariamente lexical precedidas por um sintagma máximo, um especificador; e (iii) construções de modificação, em que uma palavra ou sintagma nuclear é precedido ou antecedido por um modificador.

As construções analisadas nesse trabalho são herdeiras da construção X\_Núcleo, composta por um núcleo que é especificado por um elemento à esquerda, o que significa, de acordo com Kay e Fillmore (1999), que a construção herdeira contém toda a informação da construção herdada e outras informações acrescidas, referentes aos elementos que preenchem a construção.

Através das relações de herança, apresentaremos como as construções nucleadas propostas por Kay e Fillmore (1999) foram necessárias para o mapeamento das construções descritas nesse trabalho. Mas, antes, passaremos à apresentação de cada uma das três construções-alvo.

### **2.1 A Construção Transitiva Direta Ativa**

O conceito de transitividade na língua portuguesa está geralmente relacionado com a exigência ou não de objeto por parte do verbo. Perini (2009) afirma que, tradicionalmente, um verbo transitivo exige a presença de objeto direto em sua oração, ou seja, em orações com verbos transitivos o objeto direto é obrigatório. Não vamos discutir aqui os problemas relacionados à definição tradicional, mas usaremos essa definição para compreender as descrições linguísticas propostas para a construção Transitiva Direta Ativa.



Castilho (2010) engloba essas construções em grupo maior denominado sentenças biargumentais. Segundo o autor, essas sentenças exibem um argumento externo sujeito e um argumento interno objeto direto proporcional a um pronome acusativo como em (13) e (14).

(13) João comprou uma bola

(14) Maria fez um bolo

A estrutura sintagmática dessas sentenças é  $[SN^1_{\text{sujeito}} + V + SN^2_{\text{objeto direto}}]$ , na qual o  $SN^1$  é geralmente agentivo e o  $SN^2$ , paciente. O sujeito e o objeto direto podem ser preenchidos por um sintagma de núcleo nominal ou um sintagma nominal de núcleo pronominal. O objeto pode ainda se instanciar por uma sentença finita ou por uma sentença infinitiva.

Perini (2010) apresenta a mesma configuração para as construções transitivas, destacando que o primeiro SN é sempre sujeito, e que o segundo nem sempre precisa ser especialmente marcado, pois trata-se de um SN que não é o sujeito da oração

Em um tratamento construcional e sociocognitivo, Ferreira (2010) apresenta uma rede de construções motivada por relações de herança (Figura 16), afirmando que um simples esquema sintático pode deflagrar conceptualmente uma sequência intrincada de ações, que pode ser focalizável na sequência da interação comunicativa.

Conforme foi explicado anteriormente, um modelo construcional é capaz de abranger generalizações sintáticas abstratas que tratam das correspondências entre os papéis temáticos e as funções gramaticais. Assim, na rede construcional do PB proposta por Ferreira (2010), encontramos as relações de herança que expressam as propriedades de algumas construções específicas, como a construção transitiva agentiva, e, a partir delas, aquelas propriedades correspondentes a esquemas mais abstratos dos quais ela é herdeira.

Nessa rede, a construção transitiva agentiva aparece como uma instanciação das construções transitivas da qual herda importantes características semânticas e sintáticas. A construção é caracterizada por sua predicação agentiva, expressando sujeito e objeto que são, respectivamente, agente e paciente.

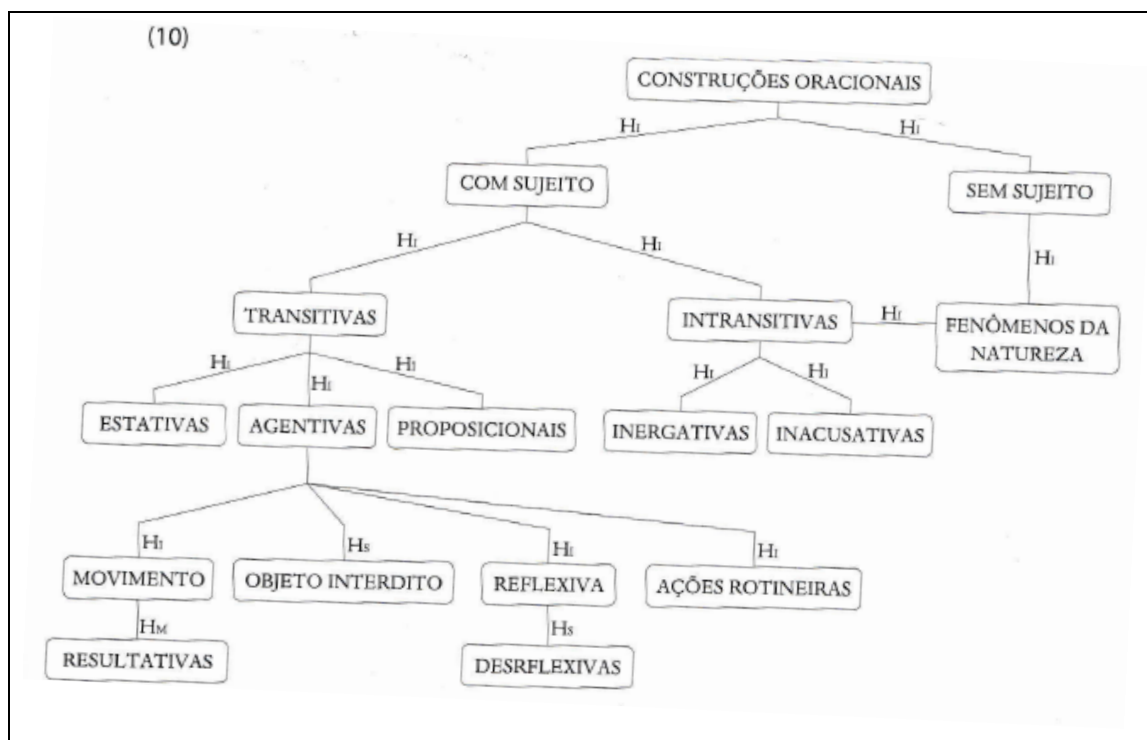


Figura 16: Rede de Construções (FERREIRA, 2010, p.111)

## 2.2 A Construção Ergativa

O termo ergativo é usado para definir verbos intransitivos que têm um correspondente transitivo tal que o complemento objeto do verbo transitivo corresponde ao sujeito do verbo intransitivo. O argumento único dos verbos ergativos exibe propriedades típicas do objeto direto dos verbos transitivos.

Castilho (2010) apresenta como características das sentenças ergativas: o sujeito como um argumento não controlador do estado de coisas descrito pelo verbo e a não explicitação do causador da ação, conforme vemos em (15) e (16).

(15) A cadeira quebrou

(16) A porta bateu

A construção ergativa tem a configuração [SN +V] em que o SN é o sujeito paciente, uma vez que não pratica nenhuma ação, sofrendo o efeito do evento expresso pelo verbo. Tem como característica um argumento interno afetado por uma causa potencialmente externa. No exemplo *A porta bateu* não se está atribuindo a mudança de estado da porta a nada, isto é, a porta pode

ter batido por si só ou por outro motivo, por isso se diz que é um evento de causa potencialmente externa.

Sentenças ergativas expressam eventos, têm tempo marcado e possuem uma interpretação episódica, indicando um acontecimento. Para Perini (2009), essas sentenças são construções em que há a demissão do argumento externo e a promoção do argumento interno para a posição de sujeito. Por isso, o autor afirma que deve-se considerar essas sentenças sob a ótica desse fenômeno sintático. Isso explicaria porque existem verbos que formam sentenças ergativas e transitivas, como as sentenças usadas em nosso teste de julgamento semântico.

Na rede proposta por Ferreira (2010), as construções ergativas (chamadas de inacusativas) são instanciações das construções intransitivas, herdando sua estrutura monoargumental.

### **2.3 Construções de Argumento Cindido (CACs)**

Através de uma abordagem construcionista, Sampaio (2010) apresenta a identificação de uma construção de estrutura argumental do PB, que é denominada Construção de Argumento Cindido (CAC). Com a observação de inúmeras ocorrências da CAC, fundamentou-se a postulação de uma Família de Construções de Argumento Cindido, que foram rotuladas como: CAC-artefato; CAC-parte do corpo; CAC-posse alienável; CAC-entidade atributo. Para a análise dessas construções, Sampaio (2010) buscou integrar suas características semânticas, sintáticas e discursivas.

Como nas CACs observou-se um fenômeno conhecido como *mismatch* (desencontro), a autora afirma que “*o termo tem sido usado para descrever um conjunto de fenômenos linguísticos que envolvem um mapeamento entre elementos ou estruturas incongruentes, sendo esta incongruência definida em relação a uma condição típica ou default*” (SAMPAIO, 2010, p.48), ou seja, não há uma correspondência direta entre os argumentos sintáticos e semânticos em uma estrutura.

A partir dessas concepções, Sampaio (2010) observa que, no caso da CAC, há uma possível paráfrase com a Ergativa, porém, com diferenças na estrutura informacional, mais especificamente, na estrutura de tópico. Em sua

análise, Sampaio (2010) defende que essa função corresponde à motivação da CAC. Dessa forma, assume-se que a CAC é uma construção ergativa com um único argumento semântico que é sintaticamente cindido.

Assim, a cena conceptual evocada pela CAC é a mesma evocada pela Ergativa. As duas são sintaticamente diferentes, mas semanticamente sinônimas, devendo, portanto, segundo o Princípio da Não-Sinonímia (GOLDBERG, 1995), ser pragmaticamente distintas. Sampaio (2010) explica que:

Assumir simplesmente que a estrutura da CAC é resultado do alçamento do SN-possuidor do sujeito da Ergativa Canônica ignora o fato de que a estrutura da CAC apresenta características intrínsecas, independentes de sua possível relação com a Ergativa Canônica. As restrições de uso, aliadas às especificidades da CAC, demonstram que não se trata de uma simples derivação, mas do estabelecimento de um padrão sintático-semântico-pragmático produtivo e muito frequente. (SAMPAIO, 2010, p. 92).

A postulação de uma família de CACs por Sampaio levou em consideração, principalmente, as classes dos verbos licenciados, o tipo semântico da entidade afetada e os padrões de uso das diferentes versões da CAC. Os dados obtidos permitiram à autora a identificação e caracterização de quatro membros dessa família de construções: a CAC-artefato, em (17); a CAC-parte do corpo, em (18); a CAC-posse alienável, em (19), e a CAC-entidade/atributo, em (20)<sup>3</sup>.

(17) OLÁ , **meu celular trincou o vidro**, mas ele funcional tudo só ã dá p ver nada.... gostaria de saber se o q devo trocar é o vidro mesmo

(18) No treino de ontem **Ranielli torceu o tornozelo**

(19) **Rubinho quebrou o carro..**

(20) E para completar **a câmara que comprei aumentou o preço**, já estou prevendo o que vem pela frente, é um descaso total, uma falta de respeito.

A CAC-artefato ocorre sempre que o dano é conceptualizado pelo usuário como uma espécie de evento espontâneo e não como resultado da ação de um agente. Em relação aos verbos licenciados pela CAC-artefato, constituem um grupo que evoca frames relacionados ao domínio de Dano.

<sup>3</sup> Os exemplos (17-20) foram coletados por Sampaio (2010).

Além disso, nas CAC-artefato, o sujeito é um artefato e seu objeto uma parte específica desse artefato, configurando uma relação parte-todo.

Já na CAC-parte do corpo, o verbo que evoca o frame de *Experenciar\_ferimento\_corporal*. Seu sujeito é um ser vivo e seu objeto uma parte específica do corpo desse ser, configurando uma relação de posse inalienável.

Por sua vez, como afirma Sampaio (2010, p. 119) a CAC-posse alienável pode ser interpretada em condições bastante específicas. Na verdade, o que temos na CAC-posse alienável é um tipo especial de conhecimento por parte do falante/ouvinte (por exemplo, saber que, em condições normais, o Rubinho não quebraria seu próprio carro). Não havendo algum conhecimento prévio que imponha a interpretação ergativa, a interpretação desse tipo de estrutura tende a ser causativa, atribuindo ao sujeito o papel de agente e ao objeto o papel de paciente.

Na CAC-entidade/atributo, são evocados frames do domínio da mudança escalar; ou seja, seus enunciados expressam uma mudança no posicionamento de um dos atributos de uma dada entidade em uma escala virtual.

A partir da descrição da Família CAC, Sampaio (2010) pontua que a diferença entre a CAC e uma Ergativa reside no fato de que o sujeito da CAC corresponde ao elemento que, na Ergativa, é expresso como um adjunto adnominal; enquanto o que seria o núcleo do sujeito da Ergativa aparece, na CAC, em posição de objeto direto. Do ponto de vista semântico, o sujeito da CAC representa o TODO da entidade afetada e seu objeto corresponde a uma PARTE específica dessa entidade.

Neste trabalho, optou-se por tratar apenas duas CACs da família proposta por Sampaio, a CAC-artefato e a CAC-parte do corpo. Essa escolha é justificada pelo fato de que construções com essas características são mais produtivas na língua, ocorrem com maior número de verbos e, por isso, são mais facilmente identificadas por falantes de PB. Outro fator de influência foi que, através desse recorte, os experimentos computacionais puderam ser realizados em tempo hábil.

## 2.4 Rede Mínima de Construções Necessária ao Modelo

A rede construções descritas nesta seção foi elaborada com base nas relações de herança completa propostas por Kay e Fillmore (1999). Diante disso, descrevemos todas as construções que foram necessárias para a modelagem de uma rede que possibilitasse o reconhecimento das construções Transitiva Direta Ativa, Ergativa e CAC.

Nossa intenção ao formular essa rede é a de que seja possível, pelo acompanhamento das relações de herança, predizer as características de uma construção específica.

Como afirma Goldberg (1995), as construções estão normalmente relacionadas entre si, integrando redes construcionais que podem apresentar diferentes tipos de relações. A ideia é que cada construção tem um significado próprio e uma forma esquemática e convencional que é capaz de fornecer as mais abstratas generalizações para o licenciamento de outras construções.

Por exemplo, uma Construção Transitiva, que licencia sentenças como *A professora rasgou o papel*, definida formalmente por [SN [V SN]], fornece um padrão formal para uma Construção de Argumento Cindido, que licencia construtos como *A calça rasgou o bolso*, ao mesmo tempo em que compartilha a informação semântica de uma Construção Ergativa, que, por sua vez, licencia *O lençol rasgou*. A Figura 17 mostra as relações entre as três construções, todas herdeiras da Construção Sujeito\_Predicado, que, por sua vez, é herdeira da Construção X\_Núcleo. As relações de herança são representadas por linhas vermelhas, conectando círculos verdes, que representam as construções. O compartilhamento de estrutura semântica entre a CAC e a Ergativa é representado pela relação de Evocação – seta azul – entre essas construções e o frame Ser\_afetado, representado por um quadrado vermelho.

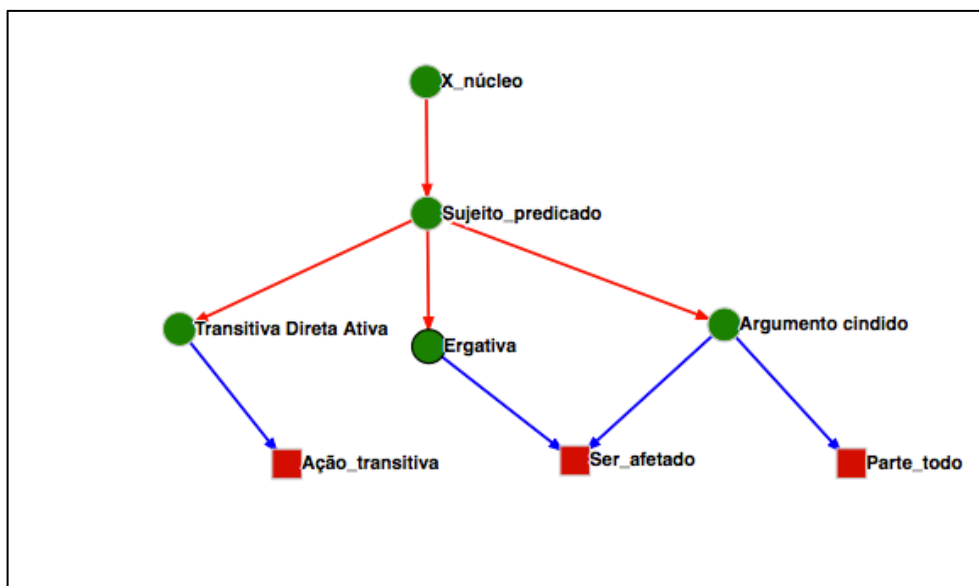


Figura 17: Relações entre as construções e frames

A postulação de uma rede de construções fornece meios para capturar generalizações gramaticais e apresentá-las de uma forma acessível para o linguista. Segundo Croft (2001), os modelos de rede são capazes de capturar padrões gramaticais que são difíceis ou impossíveis de capturar utilizando taxonomias ou construções de categorias independentes, por isso, esses modelos esquemáticos conseguem representar o conhecimento gramatical dos falantes de forma compatível com um modelo computacional da mente humana.

Para a construção da nossa rede, foi necessária a modelagem de 17 construções do PB que, de alguma forma, estão relacionadas com as três construções analisadas neste trabalho. Como já mencionado anteriormente, nosso intuito com essa modelagem é fornecer subsídios para a construção de um *parser* sintático-semântico, sendo assim, todas as construções descritas aqui foram indispensáveis para a identificação da Transitiva Direta Ativa, da Ergativa e da Argumento Cindido pela máquina.

As construções que compõem a rede são todas herdeiras das construções nucleadas propostas por Kay e Fillmore (1999), são elas:

1. N\_Cxn (Construção Nome): essa construção licencia todos os núcleos nominais do PB e apresenta um Nome como único elemento da construção.

2. NP\_Cxn (Construção Sintagma Nominal): construção de um Sintagma Nominal genérico, formado, obrigatoriamente por um Nome na posição de núcleo.
3. Mod\_NP (Construção SN\_Determinado): uma estrutura nominal não máxima (o Nome) se combina a um Determinante, criando um sintagma nominal.
4. Mod\_Cxn (Construção Modificação): construção genérica de modificação, formada por um Modificador que tem escopo sobre o núcleo.
5. Det\_Cxn (Construção Determinante): construção formada por um Determinante que tem escopo sobre o núcleo.
6. Mod\_Head\_Cxn (Construção Modificador Núcleo): construção em que o Modificador modifica o Núcleo.
7. Head\_Cxn (Construção Núcleo): construção genérica que congrega todos os tipos de núcleos do PB.
8. V\_Cxn (Construção Verbo): essa construção licencia todos os núcleos verbais do PB.
9. VP\_Cxn (Construção Sintagma Verbal): construção de sintagma verbal genérico, formado, obrigatoriamente por um Verbo na posição de núcleo.
10. Comp\_VP\_Cxn (Construção Sintagma Verbal com Complemento): tipo de construção Núcleo\_Complemento em que o Núcleo é um Verbo. Esta construção de SV deve apresentar ao menos um Complemento solicitado pela valência do Verbo.
11. Comp\_W\_VP\_Cxn (Sintagma verbal sem Complemento): construção Núcleo em que o Núcleo é um Verbo. Esta construção de SV não apresenta nenhum Complemento.
12. Head\_Comp\_Cxn (Construção Núcleo Complemento): construção genérica que congrega todas as construções do PB em que um Núcleo seleciona um ou mais Complementos
13. X\_Head\_Cxn (Construção X\_Núcleo): Nessa construção o Núcleo é especificado por um elemento X à esquerda.



14. Subj\_Pred\_Cxn (Construção Sujeito Predicado): a construção de Sujeito\_predicado fornece um argumento externo, o Sujeito a uma sentença que não o tem, o Predicado.
15. Act\_Trans\_Dir\_Cxn (Construção Transitiva Ativa Direta): tipo de construção SN\_V\_SN. Essa construção exibe um argumento externo Sujeito e um Predicado. O argumento interno é um objeto direto.
16. Split\_Obj\_Cxn (Construção Argumento Cindido): tipo de construção SN\_V\_SN. Essa construção exibe um argumento externo Sujeito e um Predicado. O argumento interno é um objeto direto. A construção apresenta uma relação PARTE/TODO entre os argumentos.
17. Ergative\_Cxn (Construção Ergativa): tipo de construção SN\_V. Essa construção exibe um argumento interno Sujeito e um Predicado.

Nas seções anteriores foram mostradas as propriedades sintáticas e semânticas que definem as três construções em estudo. A intenção ao formular essa rede é apresentar como essas construções são licenciadas a partir da interação com outras construções. Ainda que grande parte das propriedades das construções Transitiva Direta Ativa, Ergativa e CAC, que foram expostas, sejam atinentes a essas construções em si, outras construções contribuem para sua definição, seja através de herança, ou porque compõem, como signos-filhos, as construções em estudo.

Por exemplo, a construção Transitiva Direta Ativa é uma construção da família Sujeito\_predicado, que herda a composição binominal abstrata da construção X\_Núcleo. Nessa medida, a construção Transitiva Direta Ativa traz as informações adicionais de que o elemento X não só é um sujeito, mas que este é agente. Por outro lado, os CEs Sujeito e Predicado da construção Transitiva Direta Ativa são definidos como instâncias licenciadas pelas construções de Sintagma Nominal e Sintagma Verbal com Complemento, respectivamente.

A representação de sintagmas e categorias gramaticais em construções é uma forma de mostrar as generalizações gramaticais, como ocorre na

construção Sintagma Verbal, que captura generalizações gramaticais de um grande número de construções da língua, e também captura uma generalização mais ampla da construção HC (Núcleo com Complemento).

De acordo com Croft (2001), construções altamente esquemáticas, como é o caso das construções HC\_Cxn, Mod\_Cxn e XH\_Cxn, são baseadas em padrões morfossintáticos que são postulados para justificar um tipo elevado de semelhança morfossintática e semântica. Dessa forma, essas construções são usadas para definir outras construções através das relações de herança e fornecer padrões para a descrição linguística.

Cada construção da rede apresenta uma relação com as outras construções, seja por relações de herança, ou por restrições sintáticas e semânticas. As Figuras 18, 19 e 20 mostram como essas construções estão relacionadas entre si de acordo com sua configuração sintática e semântica.

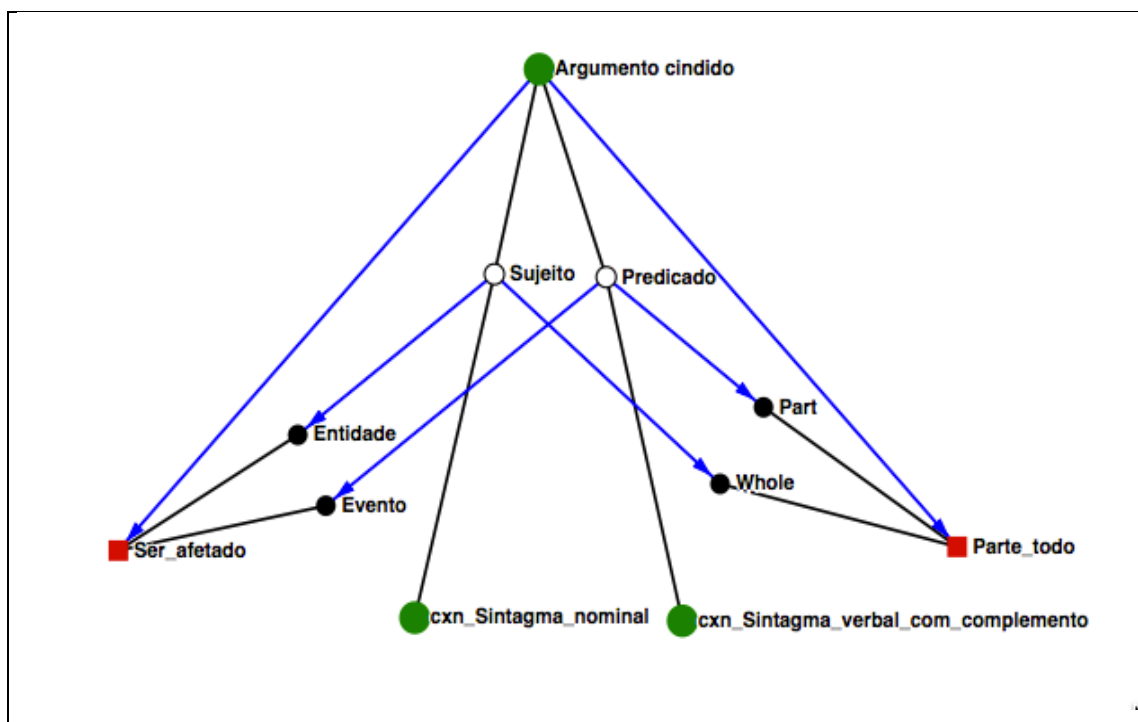


Figura: 18: Modelagem da Construção de Argumento Cindido

A Figura 18 apresenta as relações da CAC com as outras construções da rede, como também sua configuração sintático-semântica. A Construção de Argumento Cindido é herdeira da Construção Sujeito\_Predicado, que, por sua vez, é herdeira da Construção X\_Núcleo. Os CEs, Sujeito e Predicado, que estão representados por círculos brancos, constituem-se pelas construções

Sintagma Nominal e Sintagma Verbal com Complemento representadas pelos círculos verdes. Os frames Ser\_afetado e Parte\_Todo, que estão representados pelos quadrados vermelhos, estão relacionados à construção pela relação de Evocação, seus Efs, círculos pretos, que estão unificados aos CEs, constituem a Entidade e o Evento no frame Ser\_afetado, enquanto no frame Parte\_Todo, constituem o Todo e a Parte.

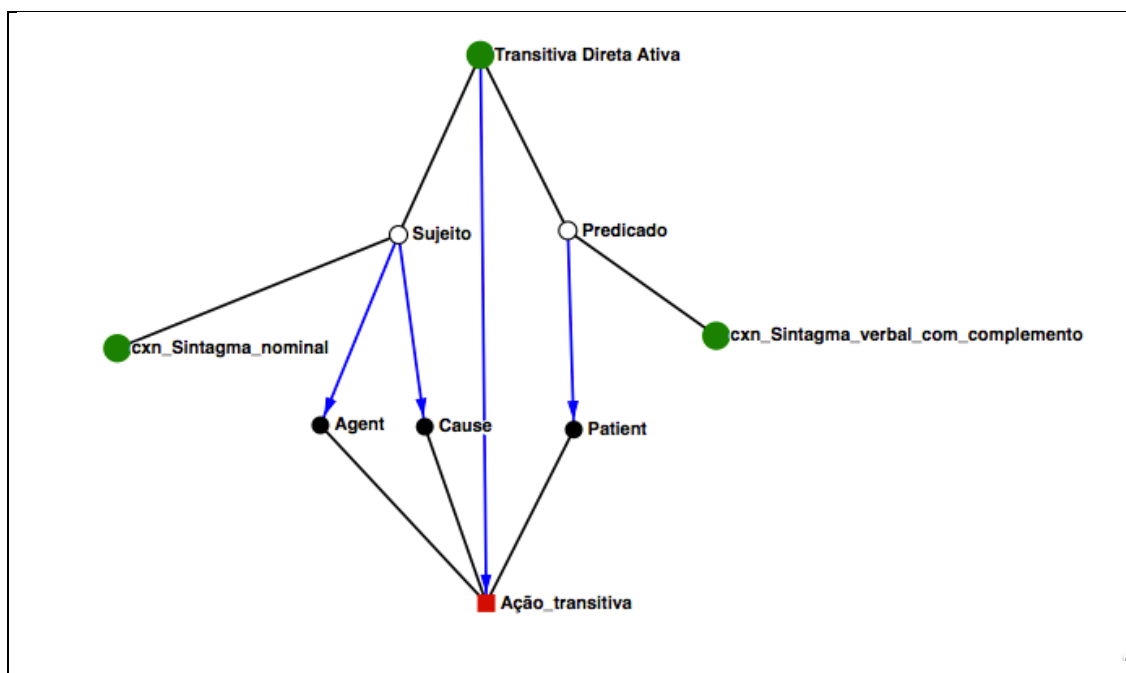


Figura 19: Modelagem da Construção Transitiva Direta Ativa

Na Figura 19, encontramos a modelagem da Construção Transitiva Direta Ativa, que também é herdeira da Construção Sujeito\_Predicado e possui como CEs as mesmas construções da Construção de Argumento Cindido. Apesar da contraparte sintática da CAC ser idêntica à da Transitiva Direta Ativa, notamos pela relação de Evocação que a distinção entre elas se encontra na contraparte semântica, que se dá pela relação todo-parte ou possuidor-posse. O frame aqui evocado é de Ação\_Transitiva em que um Agente ou uma Causa afeta um Paciente. Esses Efs, que estão representados pelos círculos pretos, estão unificados aos CEs, círculos brancos, que se constituem pelas construções de Sintagma Nominal e Sintagma Verbal com Complemento, representadas pelos círculos verdes.

Por fim, a Construção Ergativa, representada pela Figura 20, também é herdeira da construção Sujeito\_Predicado porém, a Construção Ergativa é

constituída pelas construções, círculos verdes, de Sintagma Nominal e Sintagma Verbal sem Complemento, já que a construção Ergativa possui uma estrutura monoargumental. Na contraparte semântica, evoca, assim como a CAC, o frame *Ser\_afetado*, que possui apenas um EF, Entidade, que está unificado ao Sujeito da construção. O Predicado não está unificado a nenhum EF porque, apesar de a construção apresentar dois CEs, a Construção de Sintagma Verbal sem Complemento indica que esse CE é apenas um núcleo verbal.

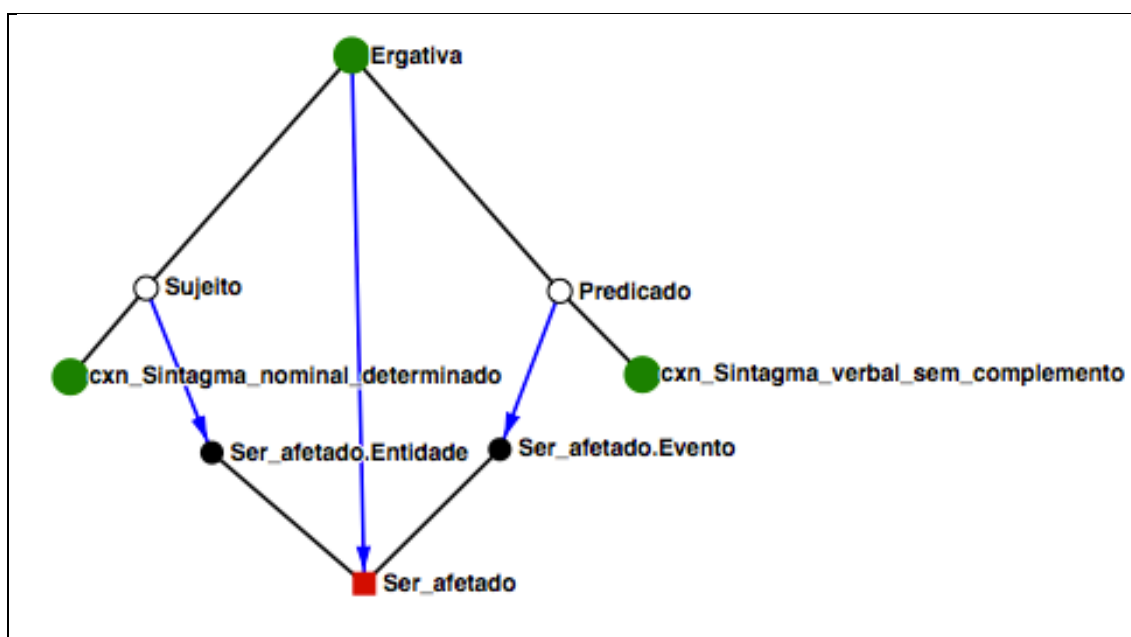


Figura 20: Modelagem da Construção Ergativa

A modelagem dessas construções no Constructicon consegue abarcar aspectos peculiares que vão além da estrutura sintática, e um tratamento construcional baseado em unificação e em frames fornece para os estudos de PNL uma visão holística de fenômenos linguísticos, uma vez que forma e significado recebem o mesmo tratamento.

No próximo capítulo, apresentaremos a metodologia experimental que usamos neste trabalho. Descreveremos, a partir da modelagem construcional realizada nesta seção, como os falantes do PB reconheceram o padrão das construções em estudo e apresentaremos o desenho experimental para o reconhecimento de construções pela máquina.

### 3 METODOLOGIA EXPERIMENTAL

#### 3.1 Aplicação de Teste de Julgamento Semântico para Estabelecimento de Gold Standard

Com a finalidade de verificar se as construções em estudo possuem um padrão semântico que pode ser facilmente identificado por falantes do PB, e fornecer evidência quantitativa para sustentar essa hipótese, foram desenvolvidos testes de julgamento de similaridade semântica por meio de uma pesquisa virtual, criada e aplicada através da ferramenta web *Survey Monkey*.

Esse é um software que disponibiliza ferramentas de desenvolvimento de pesquisas, coleta e análise de dados online. Nesse sentido, esses testes foram propostos com o objetivo de verificar se as várias sentenças inicialmente produzidas como exemplares das construções Transitiva Direta Ativa, Ergativa e Argumento Cindido compartilham propriedades reconhecidas por qualquer falante nativo do PB.

Para esta pesquisa, foi elaborado um teste contendo 40 sentenças, sendo 10 licenciadas pela construção Transitiva Direta Ativa, 10 pela Ergativa, 10 pela CAC e 10 sentenças distratoras. Para cada sentença, foram apresentadas três opções de resposta: (i) alguém realizou uma ação por vontade própria / um objeto foi utilizado por alguém para realizar uma ação, (ii) algo aconteceu com alguém independente da sua vontade / algo aconteceu com um objeto, causando uma mudança em seu estado, (iii) nenhuma das duas. Cada participante julgou qual das opções representava melhor o sentido de cada construção como mostra a Figura 21.

As sentenças alvo que continham o padrão das três construções estudadas foram adaptadas conforme as descrições linguísticas propostas por Ferreira (2009), Castilho (2010), Perini (2010) e Sampaio (2010), de modo que cada tipo de construção pudesse ter aproximadamente o mesmo tamanho, a fim de que a estrutura da sentença não interferisse no julgamento dos falantes. A utilização de sentenças criadas ocorreu para que fosse possível um mapeamento satisfatório da diversidade de frames e verbos encontrados nas três construções. Ademais, a criação das sentenças para os testes teve o

intuito de mostrar de forma padronizada como realizações verbais distintas podem estar relacionadas a significado mais genérico e como os frames dos verbos participam na elaboração desse significado geral. A Figura 21 mostra exemplos de cada construção de referência que foram utilizadas no teste, e como cada participante devia proceder para o julgamento semântico de cada uma delas.

Para cada uma das sentenças abaixo, você deve escolher uma opção que melhor apresenta, de maneira genérica, o sentido da sentença sob julgamento. Você pode, ainda, escolher a opção "Nenhuma das duas", para os casos em que o sentido da sentença sendo julgada não puder ser parafraseado pelas opções fornecidas.

**\* 1. O atleta torceu o tornozelo.**

- Alguém realizou uma ação por vontade própria.
- Algo aconteceu a alguém independentemente de sua vontade.
- Nenhuma das duas

**\* 2. O celular trincou a tela.**

- Um objeto foi utilizado por alguém para realizar uma ação.
- Algo aconteceu com um objeto, causando uma mudança em seu estado.
- Nenhuma das duas.

**\* 6. O bandido queimou as provas.**

- Alguém realizou uma ação por vontade própria.
- Algo aconteceu a alguém independentemente de sua vontade.
- Nenhuma das duas

**\* 7. O balão furou.**

- Um objeto foi utilizado por alguém para realizar uma ação.
- Algo aconteceu com um objeto, causando uma mudança em seu estado.
- Nenhuma das duas.

Figura 21: Recorte do teste aplicado

A Tabela 1, que mostrou a diversidade de verbos que ocorrem nessas construções, juntamente com os frames genéricos que elas evocam, nos traz também todas as sentenças que foram utilizadas em nosso teste de julgamento semântico.

O objetivo desse teste foi fazer com que cada participante avaliasse em que medida essas sentenças podem ser consideradas construtos que são licenciados pelas construções em estudo.

Desse modo, as sentenças licenciadas pela construção Transitiva Direta Ativa, que evoca o frame de Ação\_Transitiva, indicando um Agente que realiza uma ação sobre um Paciente, deveriam ser marcadas como *Alguém realizou uma ação por vontade própria*. As sentenças licenciadas pela construção Ergativa, que aponta para uma entidade que sofreu a ação de evento, evocando o frame Ser\_Afetado, deveriam ser marcadas como *Algo aconteceu com um objeto causando a mudança em seu estado*. Já as sentenças licenciadas pela construção de Argumento Cindido, que evoca o mesmo frame que as Ergativa, deveriam ser marcadas como *Algo aconteceu com um objeto causando a mudança em seu estado* ou *Algo aconteceu a alguém independente da sua vontade*.

Os testes buscaram confirmar a hipótese de que cada uma das três construções, apesar de licenciarem construtos diversos, evocam frames que abarcam o significado geral da construção. Os diferentes verbos e suas realizações polissêmicas indicariam que essas construções são produtivas na língua, o que é reforçado pelo fato de que um único verbo pode aparecer em realizações distintas, e que, por isso, uma análise exclusivamente lexical não abarcaria todas as propriedades dessas construções.

A partir das condições já citadas, o teste foi realizado contendo algumas sentenças distratoras, com a intenção de minimizar a possibilidade de o respondente compreender o objetivo do teste. Outra estratégia usada foi apresentar as alternativas contendo as sentenças de forma aleatória. Com isso, evitamos a repetição de uma mesma estrutura da sentença, o que, por sua vez, não permite que o participante crie um padrão de escalonamento das alternativas. Tivemos o cuidado, ainda, de adaptar as sentenças seguindo um mesmo tamanho, para evitar que sentenças de tamanhos diferentes pudessem influenciar nas respostas dos participantes.

A pesquisa contou com 76 participantes. No entanto, apenas 59 voluntários responderam ao questionário por completo, o que fez com que considerássemos apenas as respostas destes e descartássemos as dos demais participantes. Para critério de inclusão, todos os participantes deveriam

estar cursando algum curso de graduação e possuir ainda, o PB como língua materna.

Os resultados dos testes constituem o *gold standard* em relação ao qual discutiremos a capacidade do algoritmo modelado em reconhecer padrões construcionais. Tais resultados serão apresentados e discutidos na seção 4.1.

### 3.2 Desenho Experimental

Para a tarefa de reconhecimento de construções por máquina, utilizamos uma técnica de travessia em redes chamada Ativação Propagada ou Spreading Activation. Essa técnica é aplicada em Redes de Ativação Propagada (RAP), que são geralmente usadas em processos de desambiguação lexical.

A Ativação Propagada é um método para pesquisa de redes associativas, redes neurais ou redes semânticas. Nessa técnica, o processo é iniciado pela marcação de um conjunto de nós de origem, por exemplo, os conceitos de uma rede semântica, que possuem um peso e recebem um nível de ativação, em seguida, após a ativação do nó, há a propagação ou espalhamento para outros nós ligados a esse nó de origem.

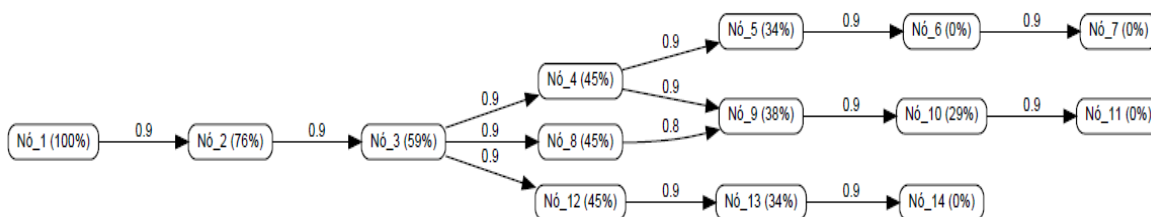


Figura 22: Exemplo de Ativação Propagada

A Figura 22 ilustra o método: a cada passagem por um nó, o nível de ativação é diminuído. Quando o nível de ativação fica abaixo de um certo limiar previamente definido (threshold), ou quando uma condição de parada previamente estabelecida é alcançada, a ativação não é mais realizada. A propagação da ativação é iniciada no nó 1 (100%). Cada link tem o mesmo valor para o peso (0.9), exceto o link entre os nós 8 e 9 (0.8). O fator de decaimento é de 0.85 e o threshold é de 35%. A Figura 22 mostra a ocorrência



de 5 ciclos de propagação da ativação. A ativação também pode ocorrer por caminhos alternativos, identificados por meio de marcadores distintos, e só termina quando dois caminhos alternativos alcançam o mesmo nó (MATOS, 2014).

A Ativação Propagada busca modelar a forma como o cérebro funciona, na medida em que impulsos se movem através de toda uma rede de neurônios para recuperar informações específicas. Essa técnica apresenta uma matriz de conceitos dentro da nossa memória como unidades cognitivas, cada uma consistindo de um nó, e os seus elementos ou características associadas, todas ligados entre si e relacionados a um conceito original.

Matos (2014), usando a Ativação Propagada em uma proposta de desambiguação lexical baseada na Semântica de Frames, afirma que essa técnica possui um processamento massivo paralelo que ajuda a diferenciá-la de outras técnicas no âmbito da Inteligência Artificial, como também apresenta restrições na comunicação entre as unidades de processamento e mensagens não-simbólicas, tornando o processo independente de uma interpretação específica. Ainda, o resultado obtido no processo pode ser avaliado por um componente externo, um avaliador de caminhos que compara e analisa os caminhos na rede que tenham unidades altamente ativadas.

A Figura 23 apresenta a desambiguação relativa a alternância causativo-incoativo do verbo *abrir*, estudado por Matos (2014), usando a Ativação Propagada. Nesse exemplo, vemos como entidades físicas são conceptualizadas como contêineres, e como a FN-Br, aliada a uma ontologia, contribuiu para que o processo de desambiguação fosse possível. Portanto, temos uma UL *abrir* que evoca um frame *Abrir\_físico\_incoativo*, este que possui um EF, Container que está relacionado a um tipo ontológico com função gramatical de Externo. A Figura 23 mostra o caminho percorrido após a ativação de cada elemento na rede, ou seja, na medida em que há uma ativação de um elemento, outro elemento é ativado liberando um outro nível de ativação.

Dessa forma, o uso da técnica da Ativação Propagada para o PLN pode ser considerado cognitivamente plausível se os algoritmos e estruturas de dados empregados puderem ser reduzidos a redes conexionistas. O modelo de gramática usado nesse trabalho, a BCG, que é um modelo implementado

computacionalmente por estrutura de traços, permite tal redução na medida em que as propriedades das construções são definidas em termos das relações representadas nas Figuras 18, 19 e 20.

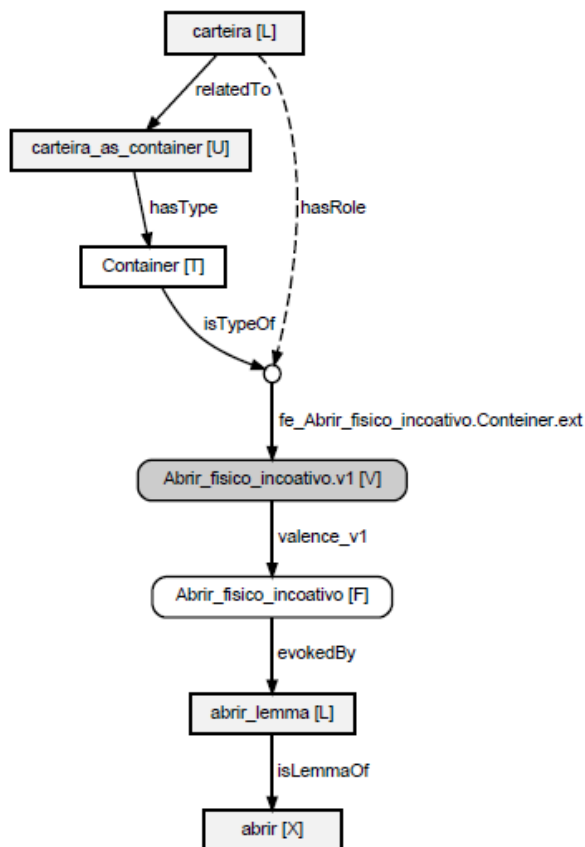


Figura 23: Desambiguação Abrir\_carteira (MATOS, 2014, p. 148)

## 4 TAREFA DE DESCOBERTA CONSTRUCIONAL POR MÁQUINA

A identificação das construções em estudo pelo CARMA consistiu, primeiramente, na realização de um teste de julgamento semântico por falantes do PB, das sentenças apresentadas na Tabela 1, com a finalidade de estabelecer um *gold standard*. Posteriormente, essas mesmas sentenças foram submetidas a análise pela máquina, para que pudéssemos estabelecer uma comparação entre o julgamento semântico feito por falantes e o resultado das análises realizadas pelo CARMA, verificando possíveis equivalências entre ambas. Assim, tendo como objetivo maior a implementação de um *parser* sintático-semântico, a tarefa de descoberta construcional é justificada pelo fato de um grande número de análises em PLN estar baseado em aspectos exclusivamente formais e, com isso, aspectos funcionais deixam de ser considerados. Nas próximas seções, apresentaremos os resultados das análises realizadas, como também a discussão sobre esses resultados.

### 4.1 Resultados

Primeiramente apresentaremos os resultados no nosso teste de julgamento semântico, cuja finalidade foi testar o reconhecimento dos padrões construcionais por falantes do PB para as construções Transitiva Direta Ativa, Ergativa e Argumento Cindido. Após a aplicação do teste, encontramos os resultados apresentados na Tabela 3. A coluna *score* indica o percentual dos falantes consultados que atribuiu às sentenças sendo julgadas uma semântica compatível com a da construção que as licencia.

A Tabela 3 mostra que os 59 falantes que responderam ao teste, em sua maioria, identificaram os 30 construtos utilizados no experimento como sendo licenciados pelas construções objeto deste trabalho. Indicam ainda que as construções Transitiva Direta Ativa, Ergativa e CAC apresentam um significado geral e esquemático, que vai além do significado verbal de cada sentença e que foi parafraseado nas opções de resposta apresentadas aos falantes.

Os scores apresentados na Tabela 1 indicam a porcentagem de falantes consultados que julgou cada sentença de acordo com o significado geral esperado para a construção. Assim, por exemplo, a sentença *A faxineira*

*trincou o copo*, que é licenciada pela construção Transitiva Direta Ativa, foi identificada pelo significado de um Agente que realiza uma ação sobre um Paciente por 94,92% dos participantes.

<b>Construção Transitiva Direta Ativa</b>		<b>Score</b>
<b>Frame evocado: Ação_transitiva</b>		
<i>Sentença</i>	<i>Frame evocado pelo verbo</i>	
A faxineira torceu o pano.	Dar_forma	94,92
A menina arrebentou o lacre.	Danificar	81,36
O cachorro machucou o menino.	Causar_ferimento	72,88
O menino quebrou a noz.	Causar_fragmentar	86,44
A criança furou os balões.	Danificar	91,53
A faxineira trincou o copo.	Danificar	93,22
O bandido queimou as provas.	Danificar	96,61
A vizinha soltou o cachorro.	Soltar	93,22
A professora rasgou o papel.	Danificar	93,22
A criança descolou o adesivo.	Descolar	89,83
<b>CAC</b>		<b>Score</b>
<b>Frames evocados: Ser_afetado e Parte_todo</b>		
<i>Sentença</i>	<i>Frame evocado pelo verbo</i>	
O atleta torceu o tornozelo.	Experenciar_ferimento_corporal	94,92
O menino arrebentou o nariz.	Experenciar_ferimento_corporal	93,22
O jogador machucou o joelho.	Experenciar_ferimento_corporal	94,92
O estudante quebrou o pé.	Experenciar_ferimento_corporal	94,92
O pedreiro furou o dedo.	Experenciar_ferimento_corporal	98,31
O celular trincou a tela.	Tornar-se_não-operacional	96,61
O computador queimou o HD.	Tornar-se_não-operacional	88,14
A panela soltou o cabo.	Quebrar	86,44
A calça rasgou o bolso.	Tornar-se_não-operacional	86,44
O sapato descolou a sola.	Quebrar	93,22
<b>Construção Ergativa</b>		<b>Score</b>
<b>Frame evocado: Ser_afetado</b>		
<i>Sentença</i>	<i>Frame evocado pelo verbo</i>	
O cabo torceu.	Tomar_forma	89,83
A corda arrebentou.	Quebrar	96,61
O menino machucou.	Experenciar_ferimento_corporal	98,31
O vaso quebrou.	Despedaçar	94,92
O balão furou.	Tornar-se_não-operacional	96,61
A janela trincou.	Despedaçar	96,61
O mato queimou.	Pegar_fogo	81,36
O botão soltou.	Quebrar	77,97
O lençol rasgou.	Despedaçar	93,22
O cartaz descolou.	Tornar-se_solto	72,88

Tabela 1: Score das sentenças no teste de julgamento semântico

Na próxima seção discutiremos esses resultados encontrados, comparando-os com as análises obtidas pelo CARMA.

A modelagem realizada no Constructicon, que considerou as construções Transitiva Direta Ativa, Ergativa e CAC, de acordo com suas propriedades sintáticas, por meio da configuração dos CEs, e suas propriedades semânticas, pelos frames evocados por cada construção, como também as relações de herança entre as três construções, serviram de subsídio para que fosse realizada a tarefa de reconhecimento construcional por máquina.

As análises no CARMA, assim como no teste de julgamento semântico, foram realizadas com todas as sentenças da Tabela 1, que, após serem modeladas, foram submetidas à ferramenta com o intuito de verificar em que medida a máquina, da mesma forma que os falantes, é capaz de reconhecer o padrão de cada construção. Apresentaremos agora uma amostra de como foi esse reconhecimento.

A Figura 24 apresenta a análise do construto *O vaso quebrou*, que é licenciado pela Construção Ergativa. Na Figura 24 vemos como ocorreu a ativação da Rede Alvo em que as palavras do construto estão representadas por círculos vermelhos, que também representam os frames. Os círculos brancos representam os lemas, as unidades lexicais, as construções lexicais, os elementos de frame e os elementos da construção, já os círculos verdes representam as construções.

A rede ativada nos mostra o caminho percorrido para identificação da construção Ergativa. Vemos como um nó vai ativando outro de acordo com as relações existentes na FN-Br (herança, evocação, restrições). Nesse exemplo, cada palavra ativou um lema, ou uma construção lexical, que, por sua vez, ativaram uma construção ou uma UL. Com a ativação das ULs, houve a ativação de frames, cujos EFs podem ativar uma construção, através da ativação dos CEs que estão relacionados a eles por evocação.

No caso da construção Ergativa, o que permitiu sua identificação foi a ativação do EF Entidade, no frame Ser\_afetado, que, por sua vez, ativou o CE Sujeito da construção Ergativa, da mesma forma que o CE Predicado foi ativado pela construção Sintagma Verbal sem Complemento que possui como CE um verbo.

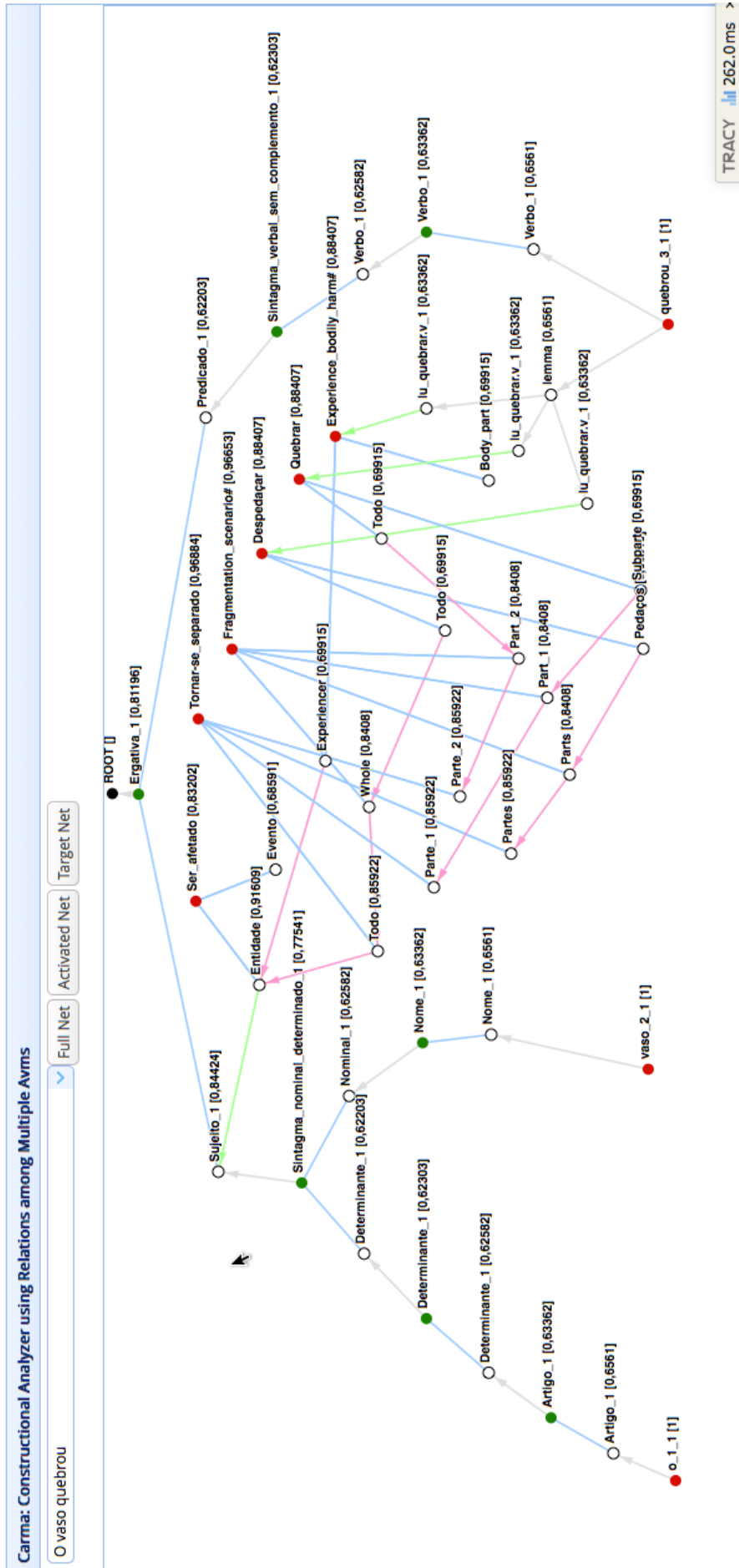


Figura 24: Rede ativada na análise do construto O vaso quebrou

O processo de ativação propagada para as demais sentenças pode ser consultado em:

<http://webtool.framenetbr.ufjf.br/maestro20/index.php/fnbr20/carma/carma>.

Os resultados das análises das 30 sentenças submetidas ao CARMA estão apresentados na Tabela 2 que mostra o nível de ativação de cada sentença conforme a construção reconhecida. Como o nível de ativação dos nós iniciais é igual a 1, os resultados apresentados na Tabela 2 mostram os níveis finais em cada sentença após a ativação da rede. Assim, na sentença *O cachorro machucou o menino* reconhecida como Transitiva Direta Ativa apresentou nível de ativação de 0,85082, mostrando como os pesos e os links entre os nós influenciaram nesse resultado.

Construção Transitiva Direta Ativa Frame evocado: Ação transitiva		ERG	TDA	CAC
<i>Sentença</i>	<i>Frame evocado pelo verbo</i>			
A faxineira <b>torceu</b> o pano.	Dar forma	-	0,85308	0,83934
A menina <b>arrebentou</b> o lacre.	Danificar	-	0,84366	0,83957
O cachorro <b>machucou</b> o menino.	Causar ferimento	-	0,85082	0,83771
O menino <b>quebrou</b> a noz.	Causar fragmentar	-	0,84067	0,83959
A criança <b>furou</b> os balões.	Danificar	-	0,84059	0,84026
A faxineira <b>trincou</b> o copo.	Danificar	-	0,84059	0,83945
O bandido <b>queimou</b> as provas.	Danificar	-	0,84067	0,84300
A vizinha <b>soltou</b> o cachorro.	Soltar	-	0,84128	0,83813
A professora <b>rasgou</b> o papel.	Danificar	-	0,84059	0,83945
A criança <b>descolou</b> o adesivo.	Descolar	-	0,84129	0,83969
<b>CAC</b>				
<b>Frames evocados: Ser afetado e Parte todo</b>				
<i>Sentença</i>	<i>Frame evocado pelo verbo</i>			
O atleta <b>torceu</b> o tornozelo.	Experenciar ferimento corporal	-	0,85308	0,86228
O menino <b>arrebentou</b> o nariz.	Experenciar ferimento corporal	-	0,84366	0,86238
O jogador <b>machucou</b> o joelho.	Experenciar ferimento corporal	-	0,85082	0,86153
O estudante <b>quebrou</b> o pé.	Experenciar ferimento corporal	-	0,84067	0,86239
O pedreiro <b>furou</b> o dedo.	Experenciar ferimento corporal	-	0,84059	0,86270
O celular <b>trincou</b> a tela.	Tornar-se não-operacional	-	0,84059	0,86200
O computador <b>queimou</b> o HD.	Tornar-se não-operacional	-	0,84214	0,86367
A panela <b>soltou</b> o cabo.	Quebrar	-	0,84128	0,86140
A calça <b>rasgou</b> o bolso.	Tornar-se não-operacional	-	0,84059	0,86083
O sapato <b>descolou</b> a sola.	Quebrar	-	0,84129	0,86076
<b>Construção Ergativa</b>				
<b>Frame evocado: Ser afetado</b>				
<i>Sentença</i>	<i>Frame evocado pelo verbo</i>			
O cabo <b>torceu</b> .	Tomar forma	0,81168	-	-
A corda <b>arrebentou</b> .	Quebrar	0,81193	-	-
O menino <b>machucou</b> .	Experenciar ferimento corporal	0,80990	-	-
O vaso <b>quebrou</b> .	Despedaçar	0,81196	-	-
O balão <b>furou</b> .	Tornar-se não-operacional	0,81270	-	-
A janela <b>trincou</b> .	Despedaçar	0,81180	-	-
O mato <b>queimou</b> .	Pegar fogo	0,81271	-	-
O botão <b>soltou</b> .	Quebrar	0,81036	-	-
O lençol <b>rasgou</b> .	Despedaçar	0,81180	-	-
O cartaz <b>descolou</b> .	Tornar-se solto	0,81208	-	-

Tabela 2: Resultados das análises no CARMA

## 4.2 Discussão

De acordo com as análises desenvolvidas, observamos que, quanto ao teste para o estabelecimento do *gold standard* – Tabela 1 – as sentenças das construções em estudo obtiveram *scores* que indicam o reconhecimento dos padrões construcionais dessas construções por parte dos falantes do PB.

Os construtos licenciados pela construção Transitiva Direta Ativa, que evoca o frame de Ação\_Transitiva, indicando um Agente que realiza uma ação sobre um Paciente, apresentaram *scores* maiores que 72,88, sugerindo um reconhecimento desse padrão pelos participantes. Já os construtos da construção de Argumento Cindido mostraram *scores* maiores que 86,44, evidenciando que os participantes reconheceram que essa construção aponta para uma Entidade que sofreu a ação de um Evento, evocando o frame Ser\_Afetado. Por fim, a construção Ergativa que evoca o mesmo frame da CAC, foi reconhecida com *scores* acima de 72,88.

Esses resultados evidenciam que cada uma das três construções, apesar de licenciarem construtos diversos, abarcam significados gerais. O uso de diferentes verbos com realizações distintas para as mesmas construções reforçou a ideia que uma análise exclusivamente lexical não daria conta de todas as propriedades dessas construções. Por isso, esse teste colaborou para que confirmássemos o padrão sintático-semântico de cada construção, uma vez que observamos que há frames tipicamente evocados por construções que são demasiadamente genéricos para estarem associados a uma UL.

O CARMA, por sua vez, conseguiu reconhecer as construções. Notamos que 29 das 30 sentenças apresentaram resultado coerente com o *gold standard*. Isso indica que, da mesma maneira que um padrão construcional foi reconhecido por um falante de língua natural, o sistema também reconheceu o padrão das três construções. A identificação através dos construtos da construção Transitiva Direta Ativa ocorreu por níveis de ativação superiores a 0,84059. Já a construção de Argumento Cindido foi identificada com níveis de ativação que ultrapassam 0,86076. E os construtos reconhecidos como instanciações da construção Ergativa apresentaram níveis de ativação maior que 0,80990.



Como a Tabela 4 indica, os construtos licenciados pela construção Ergativa receberam *scores* apenas para esta construção. Isso se deveu ao fato de que a estrutura sintática dos construtos, [SN [V]], só está associada no Constructicon da FN-Br à construção Ergativa. Em outras palavras, apesar de os frames ativados serem pertinentes, foi a sintaxe a principal responsável pelo reconhecimento incontestado desse padrão construcional.

Por outro lado, os construtos licenciados pelas construções Transitiva Direta Ativa e CAC ativaram as duas construções durante o processamento, porém aquela que apresentou maior nível de ativação que foi reconhecida pelo *parser* como licenciadora da sentença em questão. A diferença nos *scores* de ativação, nesses casos, é devida à semântica das construções em questão. Isso porque as duas opções só foram ativadas porque tanto a Transitiva Direta Ativa quanto a CAC apresentam o mesmo padrão sintático [SN[V SN]]. Em outras palavras, um *parser* apenas sintático não seria capaz de diferenciar essas duas construções.

Quando analisamos o processamento das 30 sentenças, observamos de que maneira a modelagem de construções na FN-Br contribuiu para os resultados obtidos pelo CARMA. Através das duas frentes, Lexicon e Constructicon, integradas, a FN-Br fornece uma base de dados em que os significados são modelados via frames, permitindo maior abrangência dos fenômenos semânticos. Por outro lado, com uma base construcional que conta com um modelo de gramática implementado via sistema de traços por AVMs, é capaz de proporcionar aos fenômenos linguísticos um enfoque computacional mais eficiente, tanto do ponto de vista sintático, quanto de sua ligação com aspectos semânticos. Dessa forma, os resultados obtidos pelo CARMA se devem ao tipo de tratamento linguístico-computacional fornecido pela FN-Br.

Através das análises realizadas pelo CARMA, notamos que os frames responsáveis pelo significado geral das construções não são evocados por ULs específicas, uma vez que a ativação desses frames não ocorreu via Unidade Lexical.

Apesar do bom desempenho do CARMA, encontramos dois problemas em suas análises. Primeiramente, a sentença *O bandido queimou as provas*, que deveria ser licenciada pela construção Transitiva Ativa Direta, o foi pela construção de Argumento Cindido. Essa sentença apresentou os seguintes

níveis de ativação 0,84067 para a construção Transitiva e 0,84300 para a CAC. Isso pode ter ocorrido pelo fato de o frame *Atear\_fogo* (*Setting\_fire*), que é tipicamente agentivo e, portanto, responsável pela ativação do EF Agente no frame *Ação\_transitiva*, apesar de ter um nível de ativação maior, não ter sido capaz de superar a conjunção de outros dois frames não-agentivos – *Pegar\_fogo* e *Tornar-se\_não\_operacional* – evocados por *queimar.v*. Isso fez com que a CAC tivesse um valor final de ativação maior. O que percebemos com isso é que, quando há ativação com mais potência de um frame, não há um bloqueio na propagação dos outros frames com ativação menor.

O mesmo problema é encontrado na especificação semântica gerada para alguns construtos licenciados pela construção Ergativa. Temos como exemplo o fato de que, no reconhecimento da construção Ergativa no construto *O vaso quebrou*, Figura 24, o frame *Ser\_afetado* foi ativado tanto pela relação de herança entre o EF Todo do frame *Despedaçar* e o EF Entidade de *Ser\_afetado*, quanto pela relação de herança entre o mesmo EF Entidade e o EF Experienciador do frame *Experenciar\_ferimento\_corporal*. Notamos, ainda, que o nível de ativação do frame *Ser\_afetado* é menor que o dos frames ativados via UL, mas nenhum desses frames foi responsável pela ativação da construção Ergativa.

Outra limitação encontrada no CARMA é a impossibilidade de um tratamento adequado das relações Parte-Todo. Como já sabemos, uma das características da CAC é a relação parte-todo entre os sintagmas objeto e sujeito, respectivamente. Em *O celular trincou a tela* sabemos que *tela* se constitui uma parte do artefato *celular*. Por isso além do frame *Ser\_afetado*, essa construção evoca o frame *Parte\_todo*. A FN-Br tem dois frames que, em tese, seriam capazes de modelar essa relação: *Artefato* e *Artefato\_subparte*. Veja a Figura 25.

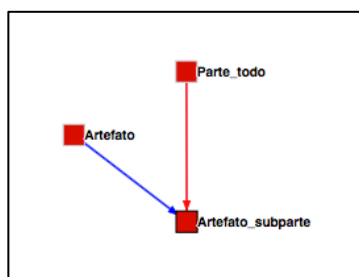


Figura 25: Relações de Herança e Subframe entre os frames *Parte\_todo*, *Artefato* e *Artefato\_subparte*

Na Figura 25, vemos que o frame *Artefato\_subparte* herda (seta vermelha) de *Parte\_todo*, o que indica que tal frame modela um tipo específico de relação parte-todo. Por outro lado, é subframe de *Artefato*, o que indica que modela subpartes dos elementos que evocam o frame de *Artefato*.

Porém, como as relações se estabelecem entre frames e não entre as palavras que os evocam, a base de dados da FN-Br não é capaz ainda de mapear esse tipo de relação adequadamente. Prova disso se manifesta se submetemos a sentença *O pano trincou a tela* para análise no CARMA. obteremos o seguinte resultado apresentado na Figura 26. Notamos que não há relação parte-todo entre *pano* e *tela*, mas como *pano* evoca o frame *Artefato* e *tela* evoca *Artefato\_subparte*, a ativação propagada reconhece esse construto como uma instanciação da CAC.

Uma possível solução para esse problema seria a ontologização da base de dados da FN-Br que ocorreria através de uma associação entre Tipos Ontológicos, classes de uma Ontologia, e Elementos de Frame.

Matos (2014) propôs um alinhamento entre uma Ontologia e a FN-Br como opção metodológica para tratar de recursos lexicais. Segundo ele, o uso de uma Ontologia permite a representação de alguns elementos da FN-Br como classes ou propriedades. Na proposta de Matos (2014), os frames de entidade, que representam essas classes e propriedades, seriam substituídos por Classes de uma Ontologia. Assim, essas Classes que representam um conjunto fechado de valores seriam tratadas em termos de estruturas *qualia* (PUSTEJOVSKY, 1995), que são relações semânticas que capturam como os seres humanos entendem os objetos e as relações destes no mundo e provê uma explicação mínima para o comportamento linguístico dos itens lexicais.

Pustejovsky (1995) define quatro papéis *qualia*, o Formal, que é tipicamente a descrição de sua categoria básica; Constitutivo, que expressa a relação entre um dado objeto e seus constituintes ou suas partes, tais como o material, o peso ou partes características; o Télico, que traz informação sobre a função ou o propósito do objeto, tais como a intenção de um agente executando uma dada ação ou a função intrínseca do objeto, e o Agentivo, que determina a origem do objeto, tais como seu criador, tipo de origem (natural ou artificial) ou sua causa inicial.

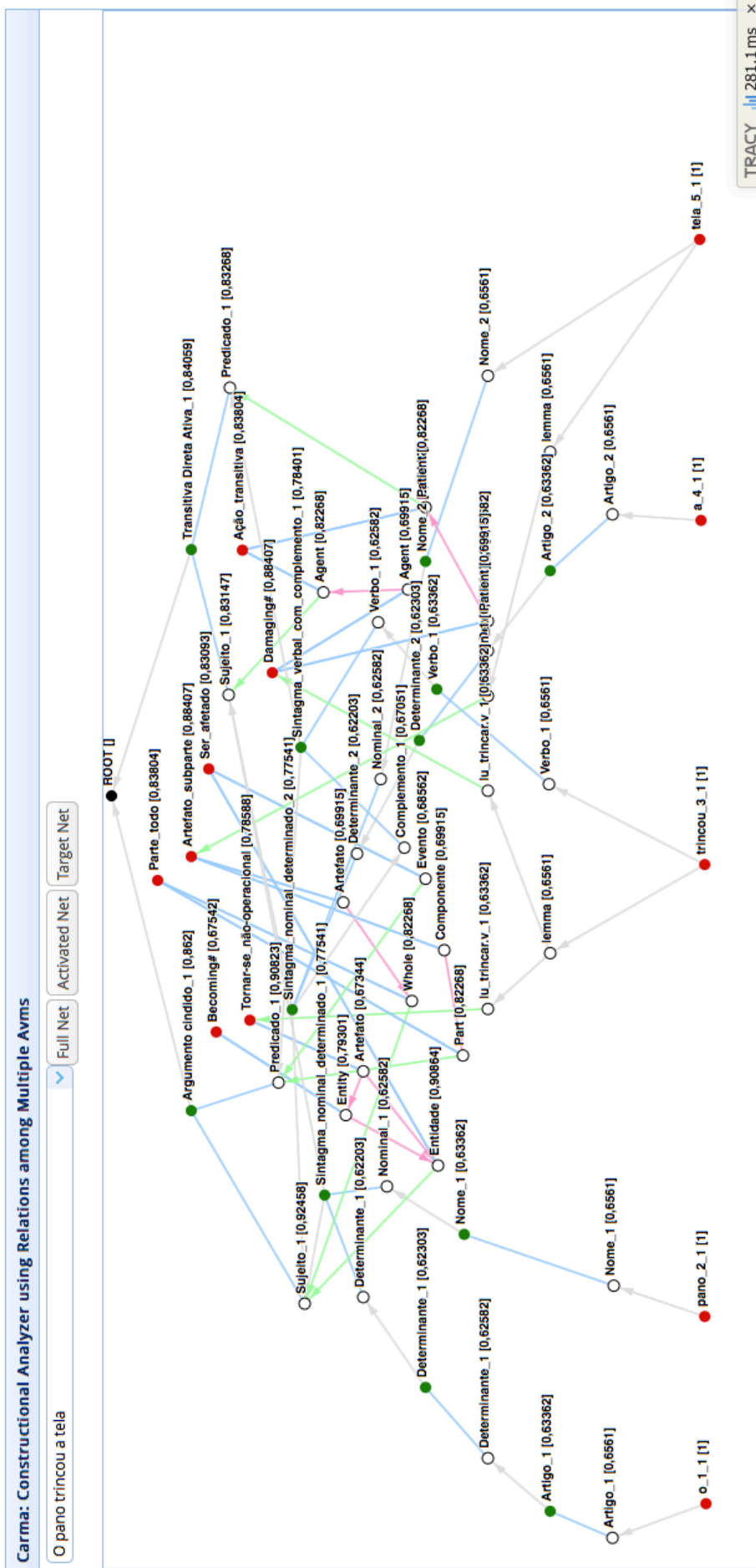


Figura 26: Rede ativada na análise do construto *O pano trincou a tela*

Se a FN-Br estivesse ontologizada, as relações parte-todo poderiam estar alinhadas a um *qualia* Constitutivo, sendo assim, a base teria a informação que *tela* é Constitutivo de *celular*, mas não é Constitutivo de *pano*, ou seja, uma sentença como *O pano trincou a tela* não seria reconhecida pelo parser, uma vez que não é possível estabelecer uma relação meronímica entre *tela* e *pano*, o que torna a sentença improvável no PB.

De acordo com Matos (2014), a ideia é que toda a base da FN tenha uma ontologia de apoio, em que Classes sejam, no mínimo, as entidades não-relacionais. Assim, essa ontologia estabeleceria as relações conceptuais entre estas entidades, independentemente de sua realização linguística, com a função de permitir uma distinção motivada entre os diversos tipos de EFs.

Pelo fato de a FN-Br ser um banco de dados que não só apresenta relações entre ULs e os frames evocados por elas, mas também conta com um repertório de construções do PB, temos um banco ainda limitado com poucas construções. Para análises mais completas, seria necessário um banco de dados com o maior número de construções existentes na língua.

Na próxima seção discutiremos as implicações do experimento para o desenvolvimento de Constructicons, apresentando como este trabalho pode contribuir para o reconhecimento de construções pela máquina.

## 5 IMPLICAÇÕES DO EXPERIMENTO PARA O DESENVOLVIMENTO DE CONSTRUCTICON

Uma vez que o desenvolvimento dessa pesquisa modelou as construções Transitiva Direta Ativa, Ergativa e CAC no Constructicon do PB com base na BCG e, posteriormente, submeteu construtos dessas construções a uma análise no *parser* CARMA, é necessário que façamos algumas considerações no que tange às implicações da análise empreendida para esse estudo.

Primeiramente, este trabalho focou-se na interação entre a Linguística e a Computação a fim de propor a modelagem de construções através do Constructicon do PB. Dessa forma, essa modelagem, que se encaixa dentro da fase representacional (DIAS DA SILVA, 1996), buscou representar formalmente as construções Transitiva Direta Ativa, Ergativa e Argumento Cindido, tornando-as computacionalmente tratáveis.

Os resultados obtidos apontam para a importância da interação entre os recursos disponíveis na FN-Br. Conforme apresentado, o Lexicon e o Constructicon precisam funcionar juntos para que a continuidade entre o léxico e gramática (FILLMORE, 2008) seja mapeada de forma satisfatória. A perspectiva de tratar os fenômenos semânticos via frames e formalizar o significado das sentenças juntamente com seus aspectos sintáticos fornece à modelagem de construções apresentada nesse trabalho um alinhamento entre o modelo de descrição linguística da Gramática das Construções de Berkeley e a implementação computacional da FN-Br.

Entretanto, ambos os recursos estão ainda em desenvolvimento, por isso, não abrangem um grande número de ocorrências construcionais e lexicais do Português do Brasil. Apesar da limitação de extensão, a FN-Br tem buscado a ampliação do seu banco de dados, não só desenvolvendo pesquisas de cunho lexicográfico, que contemplam a criação de dicionários e tradução por máquina, como também investindo na catalogação de construções nucleares e periféricas do PB.

Outra consideração a ser pontuada é a de que, mesmo a que modelagem aqui realizada tenham proporcionado uma abordagem holística dos fenômenos linguísticos, contemplando aspectos da sintaxe e da semântica

das sentenças submetidas ao CARMA, os testes revelaram a necessidade da ontologização da FN-Br para que se possam estabelecer restrições combinatórias entre ULs.

Assim como a FN-Br, as ontologias também são uma forma computacional de expressar a semântica associada aos frames, porém através de uma semântica mais formal. Uma ontologia define uma especificação formal e explícita dos termos de um domínio e das relações entre eles. Dessa forma, as ontologias são úteis do ponto de vista computacional, pois muitos recursos computacionais foram desenvolvidos para lidar com ontologias e associá-las à FrameNet Brasil facilitaria o tratamento de aspectos por ela ainda não contemplados.

Com os resultados dos testes realizados pelo CARMA, percebemos a importância de tratar os frames relacionados a entidades e qualidades de maneira diferenciada. A não identificação de relações restrições às relações parte-todo pelo CARMA, que identificou como CAC a sentença *O pano quebrou a tela*, nos alerta para como a representação dos frames e seus EFs apresenta lacunas.

Entidades como *tela*, *pano*, *celular* representam uma classe, ou Tipo Ontológico e cada traço correspondente de uma classe pode ser representado ontologicamente por uma relação *qualia*, já que essa relação foca nos objetos representados por nomes da mesma forma que eventos como *trincar* são representados na FN-Br por uma estrutura de traços, em que cada traço corresponde a um EF.

Assim, para um processamento mais completo por parte do CARMA, é fundamental que as relações semânticas entre construções, frames, EFs e também ULs estejam bem estabelecidas. As relações entre elementos de frame e os tipos ontológicos seriam definidas quando uma UL representar uma entidade da ontologia, ou quando um EF puder ser associado com um tipo ontológico que represente um conjunto de entidades. Assim, EFs como Artefato (celular, pano) e Artefato\_subparte (tela) estariam relacionados via estrutura *qualia*, com a informação de *tela* ser constitutivo de celular, mas não de *pano*.

Essa falha pode ser justificada pelo fato de que estamos usando a estrutura da FN-Br para finalidades que não estavam contempladas em seu

projeto inicial, tais como a desambiguação de frames e construções, a tradução, a conexão com outras bases, a resolução de metonímias etc.

Por fim, outro fator que merece ser considerado é a necessidade do tratamento de Entidades Nomeadas. Por exemplo, em nosso teste de julgamento semântico, tínhamos a sentença *Maria quebrou a noz* como instanciação da construção Transitiva Direta Ativa, mas, como a FN-Br não possui nenhum frame para nomes próprios e nem os relaciona a nenhum EF, para o processamento dessa sentença no CARMA, foi necessário fazer uma alteração. Assim a sentença processada foi *A menina quebrou a noz*, pois é possível relacionar *menina* a um frame.

Tratar das Entidades Nomeadas não é um problema fácil de ser resolvido. No caso de *Maria.n*, essa UL evocaria um frame específico? Se sim, então, quais seriam seus EFs? Ou teríamos um caso de um único EF incorporado? *Maria* poderia estar também relacionado a algum frame mais genérico? A FN-Br vem buscando responder essas questões.

Diante dos apontamentos realizados, é necessário pensar que o maior detalhamento semântico a ser alcançado pelo Lexicon e pelo Constructicon é imprescindível para que a FN-Br não sirva apenas como recurso lexical, uma vez que seu banco possui o potencial de fornecer dados que podem colaborar com diversas pesquisas no campo do Processamento de Língua Natural.

Portanto, em desdobramentos futuros da tarefa de reconhecimento de construções por máquina, é necessário (i) ampliar o banco de dados do Constructicon, de modo abarcar mais construções produtivas na língua que, por mais nucleares e frequentes que sejam, ainda não receberam um análise computacional, como é o caso das construções passivas, construções com verbo suporte etc.; (ii) continuar uma análise comparativa entre construções de estrutura argumental semântica ou sintaticamente semelhantes, a fim de encontrar as motivações dessas semelhanças; (iii) descrever políticas de alinhamento entre FrameNets e Ontologias, a fim de complementar as limitações de cada recurso.



## 6 CONCLUSÕES

Neste trabalho, buscou-se apresentar a modelagem de construções do PB de modo a aprimorar o Constructicon da FN-Br e possibilitar através dessa modelagem o reconhecimento de construções por máquina. Para tanto, as construções Transitiva Direta Ativa, Ergativa e Argumento Cindido foram modeladas de acordo com suas características sintático-semânticas.

Uma vez que o Constructicon descreve construções em termos de suas propriedades gramaticais e seu potencial semântico, realizando descrições de modo compatível com os pressupostos teóricos da Berkeley Construction Grammar, a formalização encontrada nele se dá pelo processo de unificação. Nesse modelo, as construções e seus signos filhos combinam-se, projetando uma nova AVM, a qual contém os valores e atributos das AVMs que se uniram. 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.

As restrições construcionais foram aqui pontuadas, considerando tanto os aspectos semânticos quanto sintáticos dos signos filhos – ou CEs. Assim, as restrições sintáticas foram tratadas em termo dos CEs que são representados em sintagmas e categorias gramaticais como uma forma de mostrar as generalizações gramaticais de cada construção. Por outro lado, CEs também foram unificados a um determinado componente semântico, ou seja, um EF do frame evocado pela construção.

Diante disso, foi possível explicitar as relações fundamentais entre as construções em uma rede integrada em que as relações de Herança – entre construções – e de Evocação – entre construções e frames, juntamente com as restrições, serviram para modelar as construções aqui estudadas.

Com a modelagem de 16 construções – Construção Nome, Construção Sintagma Nominal, Construção SN\_Determinado, Construção Modificação, Construção Determinante, Construção Modificador Núcleo, Construção Verbo, Construção Sintagma Verbal, Construção Sintagma Verbal com Complemento, Construção Núcleo Complemento, Construção X\_Núcleo, Construção Sujeito Predicado, Construção Transitiva Ativa Direta, Construção Argumento Cindido e Construção Ergativa –, todas herdeiras das construções nucleadas propostas

por Kay e Fillmore (1999), apresentamos de que forma as propriedades das construções Transitiva Direta Ativa, Ergativa e CAC, que foram aqui expostas em diálogo com as descrições linguísticas de Ferreira (2009), Castilho (2010), Perini (2010) e Sampaio (2010), estão relacionadas com outras construções que contribuem para sua definição, seja através de herança, ou composição.

Com a elaboração da rede, percebemos que as construções em estudo apresentaram uma relação de herança entre a construção mais genérica Sujeito\_Predicado, que, por sua vez, é herdeira da Construção X\_Núcleo. Porém uma distinção entre os CEs das construções Transitiva Direta Ativa, CAC e Ergativa foi corresponsável pela modelagem dessas construções no *Constructicon*. Enquanto as construções Transitiva Direta Ativa e CAC apresentam como CEs as construções Sintagma Nominal e Sintagma Verbal com Complemento, a construção Ergativa é configurada pelas construções Sintagma Nominal e Sintagma Verbal sem Complemento.

A unificação entre a contraparte semântica e sintática dessas construções foi representada através do mapeamento dos CEs para os EFs dos frames evocados por elas. A modelagem das construções conseguiu mapear que, de um lado, a CAC e a Ergativa são semanticamente sinônimas, mas com uma diferença entre elas, pois, enquanto a Ergativa apresenta uma estrutura monoargumental, em que uma entidade é afetada por um evento, a CAC apresenta uma estrutura de dois argumentos com a mesma informação semântica da CAC, acrescida por uma relação parte-todo entre eles.

Por outro lado, foi possível demonstrar como a CAC é semelhante à Transitiva Direta Ativa sintaticamente, devido a uma motivação que se manifesta pelo fato de ambas as construções herdarem de um tipo específico da construção Sujeito-Predicado, que especifica a transitividade do núcleo do Predicado como direta.

Diante das informações acerca do enquadre computacional dado às construções pela BCG e pelo *Constructicon*, justificamos a necessidade de incluir neste trabalho discussões acerca dos estudos sobre o Processamento de Língua Natural, a fim de fornecer bases para a pesquisa de reconhecimento de construções por máquina.

Esta pesquisa se enquadrou na fase representacional, uma vez que a fase linguística já havia sido proposta por Sampaio (2010), Ferreira (2009),

Castilho (2010) e Perini (2010), que descreveram os padrões licenciados por essas construções, levando em consideração aspectos sintáticos, semânticos e pragmáticos.

Partindo dessa descrição, apresentamos, no domínio da representação, através do modelo de formalização de construções da Berkeley Construction Grammar, o Constructicon, uma proposta capaz de tratar de forma amigável ao usuário a interpretação adequada dessas construções pela máquina.

A fase implementacional não foi desenvolvida nesta dissertação, mas adaptada de Matos (2014). Seus resultados, porém, foram aqui discutidos e utilizados em uma tarefa de descoberta construcional por máquina através do *parser* em desenvolvimento pela FN-Br, o CARMA.

Para critério de comparação com os resultados obtidos pelo CARMA, foi realizado um teste de julgamento semântico com a finalidade de estabelecimento de um *gold standard*. O teste de revelou o reconhecimento, por parte dos falantes do PB, dos padrões das construções em estudo, servindo ainda para avaliarmos em que medida esses resultados seriam equivalentes aos resultados do processamento no CARMA.

A partir da comparação dos resultados, constatamos que o *parser* reconheceu as construções de maneira satisfatória. Entretanto, percebemos que os recursos disponíveis na FN-Br ainda não são suficientes para tratar de relações semânticas do tipo parte-todo.

Esta dissertação, portanto, a partir das análises empreendidas, propõe os seguintes avanços:

- (i) a partir das discussões teórico-metodológicas empreendidas neste trabalho, propõem-se a modelagem de 16 construções para um empreendimento de processamento sintático-semântico de construções no PB;
- (ii) modeladas as construções, apresentaram-se os resultados do primeiro *parser* sintático-semântico do PB que utiliza a técnica de Ativação Propagada;
- (iii) dadas as limitações do Constructicon, deixa-se abertura para investigações futuras, as quais, partindo do alinhamento entre a FN-Br e ontologias, visem a analisar as propriedades semânticas dos itens lexicais em termos de relações *qualia*.

## REFERÊNCIAS BIBLIOGRÁFICAS

BICK, E. **The Parsing System PALAVRAS: Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework**, PH. D. thesis, Arhus University, 2010.

BRYANT, John. **Best-Fit Constructional Analysis**, PhD. dissertation, UC Berkeley, 2008.

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

CONTERATTO, Gabriela Betania Hinrichs. Semântica e Computação: uma interação necessária o aperfeiçoamento de sistemas PLN, Letras de Hoje. Porto Alegre. v. 41, nº 2, p. 353-367, junho, 2006.

CROFT, W. **Radical construction grammar: syntactic theory in typological perspective**. Oxford: Oxford University Press, 2001.

DIAS-DA-SILVA, B. C. **A face tecnológica dos estudos da linguagem: O processamento automático das línguas naturais**. Araraquara, 1996. Tese de Doutorado em Letras - Faculdade de Ciências e Letras, Universidade Estadual Paulista, Araraquara, 1996.

FERREIRA, M. A Construção de Ação Rotineira no Português do Brasil. In\_\_\_MIRANDA, N. S.; SALOMÃO, M. M. M. **Construções do Português do Brasil: da gramática ao discurso**. Belo Horizonte: UFMG, 2009.

FILLMORE, C. J. The case for case. In: BACH, E. & HARMS, R. (eds) **Universals in Linguistic Theory**. New York: Holt, Rinehart and Winston, 1968.

\_\_\_\_\_. The case for case reopened. In: COLE, P.; SADDOCK, J. (eds) **Grammatical Relations**. New York: Academic Press, 1977. p.59-81.

\_\_\_\_\_. Frame Semantics. In: THE LINGUISTIC SOCIETY OF KOREA (org.). **Linguistics in the morning calm**. Seoul: Hanshin, 1982.

\_\_\_\_\_. Border Conflicts: FrameNet Meets Construction Grammar. In: **EURALEX,13, 2008, Barcelona. Anais**. Barcelona: Universitat Barcelona Fabra, 2008a.

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

FILLMORE, C. J., KAY, P. and O'CONNOR, M. C. Regularity and Idiomaticity in Grammatical Constructions: the case of *Let Alone*. In: **Language**, Vol. 64, No. 3, 1988. p. 501-538.

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.

KAY, P. & FILLMORE, C. J. Grammatical Constructions and Linguistic Generalizations: the What's X Doing Y Construction. **Language**, vol. 75, nº 1, 1999.

KOEHN, P. **Statistical machine translation**. Cambridge. Cambridge University Press, 2009.

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.

LASNIK, H. (1990) "The Study of Cognition." *In*: D.N. Osherson & H. Lasnik, (eds.). **Language: an Invitation to cognitive science**. Cambridge, Mass., The MIT Press. p.11-15.

MARTINS, R. T.; HASEGAWA, R.; NUNES, M. G. V. **Curupira: um parser funcional para o português**. NILC-TR-02-26, 2002.

MARTINS, R.; NUNES, G.; HASEGAWA, R. Curupira: A Functional Parser for Brazilian Portuguese. In: NUNO J. M.; TRANCOSO, I.; BAPTISTA, J.; NUNES, M. G. V. (orgs.). **Computational Processing of the Portuguese Language**, v. 2721, p. 179-183. Berlin: Springer-Verlage, 2003.

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.

OTHERO, G. A.; MENUZZI, S. M. **Linguística Computacional**: teoria & pratica. São Paulo: Parábola Editorial, 2005.

PERINI, Mário A. **Gramática do português brasileiro**. São Paulo: Parábola Editorial, 2010.

- PERINI, Mário A. **Gramática Descritiva do Português**. São Paulo: Editora Ática, 2009.
- PUSTEJOVSKY, J. **The Generative Lexicon**. Cambridge, MA: MIT Press, 1995.
- SALOMÃO, M. M. M. FrameNet Brasil: um trabalho em progresso. **Calidoscópico**, São Leopoldo: UNISINOS, vol. 7 n. 3, p. 171-182, set/dez 2009a.
- SALOMÃO, M.M.M.; SAMPAIO, T.F. ;TORRENT, T.T. A linguística cognitiva encontra a linguística computacional: notícias do projeto FrameNet Brasil. **Cadernos Estudos Linguísticos**. Universidade Estadual de Campinas. Campinas. nº 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.
- SILVA, A. B. C.; ESTRADA, C. C. D. **Coesão e coerência em redações do Enem sob a perspectiva da FrameNet Brasil**. Monografia de pós-graduação em Ensino de Língua Portuguesa. UFJF, Juiz de Fora, 2012.
- TORRENT, T. T. **A Rede de Construções em Para (SN) Infinitivo**: uma abordagem centrada no uso para as relações de herança e mudança construcionais. 2009. 166 f. Tese de Doutorado em Linguística. Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2009.
- TORRENT, T.T.; LAGE, L.M.; SAMPAIO, T.F.; TAVARES, T.S.; MATOS, E.E.S. Revisiting border conflicts between FrameNet and Construction Grammar. In: BORIN L.; MELO G.; HEPPIN, K.F.; TORRENT, T.T. (Eds.), **Frames, constructions, and computation**. Special issue of *Contructions and Frames* 6:1 (2014). 2014. v, 135 pp. (pp. 34–51).