

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

Adrieli Bonjour Laviola da Silva Darowisch

**Constructicografia multilíngue em ação: diretrizes linguístico-computacionais para
o alinhamento de constructicons**

Juiz de Fora
2019

Adrieli Bonjour Laviola da Silva Darowisch

Constructicografia multilíngue em ação: diretrizes linguístico-computacionais para o alinhamento de constructicons

Tese de Doutorado 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 Doutora em Linguística.

Orientador: Prof. Dr. Tiago Timponi Torrent

Juiz de Fora
Outubro de 2019

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

Darowisch, Adrieli Bonjour Laviola da Silva.

Constructicografia multilíngue em ação: : diretrizes linguístico computacionais para a alinhamento de constructicons / Adrieli Bonjour Laviola da Silva Darowisch. -- 2019.

145 f.

Orientador: Tiago Timponi Torrent

Tese (doutorado) - Universidade Federal de Juiz de Fora, Faculdade de Letras. Programa de Pós-Graduação em Linguística, 2019.

1. Gramática das Construções. 2. Semântica de Frames. 3. Constructicografia Multilíngue. 4. Universal Dependecies. 5. Constructicon. I. Torrent, Tiago Timponi, orient. II. Título.

Adrieli Bonjour Laviola da Silva Darowisch

**Constructicografia Multilíngue em Ação: diretrizes linguístico-computacionais
para o alinhamento de constructicons**

Tese de Doutorado 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 Doutora em Linguística.

Aprovada em 25 de outubro de 2019

BANCA EXAMINADORA

Prof. Dr. Tiago Timponi Torrent – Orientador
Universidade Federal de Juiz de Fora

Prof. Dr. Ely Matos
Universidade Federal de Juiz de Fora

Prof. Dra. Patrícia Fabiane Amaral da Cunha Lacerda
Universidade Federal de Juiz de Fora

Prof. Dra. Flavia Hirata-Vale
Universidade Federal de São Carlos

Prof. Dr. Sandro Rigo
Universidade do Vale do Rio dos Sinos

Aos meus amores – Otávio, Miria,
Adenilson e Matheus

AGRADECIMENTOS

Alcançar objetivos, finalizar uma jornada, obter êxito em um empreendimento são motivos que nos inspiram a ser gratos. Entretanto, acredito que a gratidão é o que nos move e o que nos permite chegar à realização e concretização de sonhos. Por isso, tenho certeza que a finalização desta tese de doutorado só foi possível devido à gratidão infinita que existe em mim a todas as pessoas e instituições envolvidas nessa trajetória.

Dessa maneira, inicio, agradecendo à Universidade Federal de Juiz de Fora (UFJF), onde cheguei há dez anos e, desde então, encontrei minha profissão, descobri-me como pesquisadora e me desenvolvi em tantos aspectos. Por extensão, agradeço ao Programa de Pós-Graduação em Linguística da Faculdade de Letras da UFJF, por sua excelência e qualidade, e, especialmente, por todo apoio durante minha trajetória como aluna de pós-graduação.

Agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo suporte em minha formação, bem como ao programa CAPES/STINT pela bolsa de doutorado sanduíche na Universidade de Gotemburgo, na Suécia, (processo número 88887.130226/2016-00) que me permitiu realizar trocas com outros pesquisadores, o que foi fundamental para o desenvolvimento desta pesquisa. Incluo, portanto, aqui um agradecimento ao meu orientador na Suécia, professor Benjamin Lyngfelt, pelas discussões enriquecedores e pelos trabalhos desenvolvidos em parceria, mas, principalmente, pela acolhida e pela amizade.

À FrameNet Brasil e seu incrível time, agradeço pelas oportunidades, pelas parcerias, por todos os trabalhos compartilhados, pelos cansaços e pelas conquistas. Foram oito anos que não só solidificaram aprendizados, mas, principalmente, amizades. Dentre os colegas queridos, agradeço em especial a Ludmila que se tornou família, companheira de sonhos e vitórias. Agradeço pela companhia em viagens a congressos, pelas apresentações compartilhadas, pelos estudos realizados em parceria, e, principalmente, pelo tempo que compartilhamos na Suécia.

Em especial, agradeço e dedico mais essa conquista à pessoa que não só orientou este trabalho e esta pesquisa durante os últimos quatro anos, mas que me orienta profissionalmente há oito anos, meu orientador, Tiago Torrent. Muito obrigada pelos incentivos, pela motivação, pela amizade e por todo o apoio desde o início da minha jornada acadêmica. Obrigada por me direcionar e me inspirar a buscar sempre

mais. Obrigada pelo trabalho impecável que realiza como professor, pesquisador e orientador.

Agradeço, aos amores minha vida, que suportaram minhas ausências e que se encheram de orgulho a cada retorno meu e comemoraram cada pequena conquista. Ao meu marido, companheiro de todos os momentos e parceiro de aventuras, Otávio, muito obrigada por ser você. Obrigada por se aventurar comigo e por apoiar meus sonhos, que se fizeram seus. Agradeço aos meus pais, Adenilson e Miria, por me mostrarem que sonhos são possíveis, que a gratidão transforma a vida, que tudo está em nossa mente e que somos seres infinitos. Ao meu irmão, Matheus, agradeço por sempre me lembrar o quão maravilhoso é estudar e fazer o que gostamos. Sua coragem, sua motivação e sua força de vontade se parecem tanto com as minhas, e isso me dá tanto orgulho e me enche de gratidão.

Agradeço, ainda, a Elenise e Douglas, meus tios e meus pais do coração, que desde o início da minha formação foram meu lar em Juiz de Fora, quando eu precisei de cuidado, acolhida e carinho de casa. Aos meus sogros, Lydia e Rubens, agradeço pela compreensão, pelo carinho, e por torcerem sempre pela minha vitória e felicidade. A meus avós e demais antepassados, muito obrigada! Sem vocês, nada disso seria possível. Finalmente, a Deus, agradeço por ser essa força infinita que habita em mim e guia cada decisão da minha vida. Muito obrigada!

RESUMO

Esta tese propõe diretrizes linguístico-computacionais para o alinhamento de constructicons, tendo como ponto de partida o Constructicon do português do Brasil (PB), desenvolvido juntamente à FrameNet Brasil (FN-Br) na Universidade Federal de Juiz de Fora. Um constructicon é um recurso sintático, um repositório de construções, conforme descrevem Fillmore et al. (2012). Após a fase inicial de implementação do Constructicon do PB, o atual momento de desenvolvimento de tal recurso envolve esforços no sentido de ampliar relações entre construções, frames e seus elementos, aumentando a generalidade e a coerência interna das análises (ALMEIDA, 2016; LAGE, 2018). Ademais, considerando a existência de constructicons em desenvolvimento para outras línguas – inglês, sueco, alemão, russo e japonês –, uma das intenções do projeto brasileiro é conectar os demais recursos em uma base multilíngue. Entretanto, ainda que todos tenham como base teórica e como referência a Gramática das Construções e o Constructicon do inglês, foram adotadas diferentes metodologias e abordagens para a implementação desses recursos. Devido a essas particularidades, ao se pensar na proposição de um alinhamento entre constructicons, é fundamental buscar uma representação que permita alinhá-los apesar das suas especificidades e enclivamentos analíticos. Nesse sentido, esta tese discute quais procedimentos metodológicos poderiam ser aplicados aos esforços de alinhamento entre constructicons, empregando uma metodologia de trabalho que se divide em três etapas: 1) análise contrastiva de construções do PB, do sueco e do inglês; 2) levantamento de possibilidades de representação do alinhamento desses recursos; e 3) implementação da proposta de alinhamento. Como resultado das três etapas, foram elaborados critérios de comparação de construções que podem ser utilizados por pesquisas futuras de contraste entre construções de línguas distintas e foram utilizadas UDs e framenets como representações intermediárias para o alinhamento de construções e constructicons, permitindo que outros projetos de línguas distintas realizem alinhamentos construcionais em direção a um constructicon multilíngue. Dessa maneira, este trabalho contribui para o alinhamento entre constructicons de línguas distintas no âmbito linguístico e computacional.

Palavras-chave: 1. Gramática das Construções. 2. Semântica de Frames. 3. Constructicografia Multilíngue 4. Universal Dependencies. 5. Constructicon

ABSTRACT

This dissertation proposes linguistic-computational guidelines to the alignment of constructicons, having as the starting point the Brazilian Portuguese (BP) Constructicon, developed within FrameNet Brasil (FN-Br) at the Federal University of Juiz de Fora. A constructicon is a syntactic resource, a repository of constructions, as defined by Fillmore et al. (2012). After the initial phase of the BP Constructicon implementation, the current moment of the resource development includes efforts in order to improve the relations among constructions, frames and their elements, increasing the generality and intern coherence of the analysis (ALMEIDA, 2016; LAGE, 2018). Besides that, considering the existence of constructicons for other languages currently under development – English, Swedish, German, Russian and Japanese –, one of the goals of the Brazilian project is to connect those resources in a single multilingual database. However, even with all the projects having the same theoretical basis and the same reference – namely the Construction Grammar and the English Constructicon – they have used different methodologies and approaches for the implementation of their resources. Because of those particularities, when thinking of a proposal of an alignment among constructicons, it's important to seek a representation that allows them to be aligned despite their specificities and analytical biases. In this matter, this dissertation discusses which methodological procedures could be applied to the efforts of alignment employing a work methodology that is divided into three stages: 1) contrastive analysis of BP, Swedish and English constructions; 2) survey of possibilities of representation of the alignment of these resources; and 3) implementation of the alignment proposal. As a result of the three steps, constructions comparison criteria were elaborated that they can be used by future contrast researches between constructions of different languages, and UDs and framenets were used as intermediate representations for the alignment of constructions and constructicons, allowing other language projects perform constructional alignments towards a multilingual constructicon. Therefore, this work aims contributes to the alignment of constructicons of different languages in the computational and linguistic scope.

Keywords: 1. Construction Grammar; 2. Frame Semantics; 3. Multilingual Constructicography; 4. Universal Dependencies; 5. Constructicon

SUMÁRIO

INTRODUÇÃO	12
1 MODELOS COMPUTACIONALMENTE IMPLEMENTÁVEIS DE ANÁLISE LINGÜÍSTICA	19
1.1 GRAMÁTICA DAS CONSTRUÇÕES	19
1.2 UNIVERSAL DEPENDENCIES	36
2 LINGUISTICA COMPUTACIONAL	44
2.1 PROCESSAMENTO DE LÍNGUA NATURAL - NLP	44
2.2 COMPREENSÃO DE LÍNGUA NATURAL – NLU	46
3 CONSTRUCTICOGRAFIA MULTILÍNGUE	50
3.1 CONSTRUCTICOGRAFIA	50
3.1.1 Constructicon do inglês	52
3.1.2 Constructicon do sueco	59
3.1.3 Constructicon do alemão	64
3.1.4 Constructicon do russo	67
3.1.5 Constructicon do japonês	70
3.1.6 Constructicon do português do Brasil	72
3.1.7 Contraste entre Constructicons	81
3.2 CONSTRUCTICOGRAFIA MULTILÍNGUE X GRAMÁTICA DAS CONSTRUÇÕES CONTRASTIVA.....	84
4 METODOLOGIA	89
5 PROPOSTA DE ALINHAMENTO DE CONSTRUCTICONS	94
5.1 COMPARAÇÃO TRILÍNGUE ENTRE OS CONSTRUCTICONS DO INGLÊS, DO SUECO E DO PORTUGUÊS DO BRASIL.....	95
5.2 MAPEAMENTO DE RESTRIÇÕES ÀS CONSTRUÇÕES DO PB.....	102
5.3 MODELAGEM DE UDS NO CONSTRUCTICON DO PB	110
5.4 COMPARAÇÃO PB, INGLÊS E SUECO: ALINHAMENTO DAS CONSTRUÇÕES DE DCI	114

6 CONCLUSÕES	132
REFERÊNCIAS BIBLIOGRÁFICAS	137

INTRODUÇÃO

O objetivo principal desta pesquisa é a proposição de diretrizes linguístico-computacionais para o alinhamento de constructicons, tendo como ponto de partida o Constructicon¹ do português do Brasil (PB), desenvolvido juntamente à FrameNet² Brasil (FN-Br) na Universidade Federal de Juiz de Fora. O alinhamento de constructicons pode ser definido como qualquer esforço que vise a implementar computacionalmente ligações entre as bases de dados de recursos computacionais cujo objetivo seja o de serem repositórios estruturados de construções de uma língua. Convém ressaltar que o alinhamento não necessita ser sempre total, sendo possíveis e até mesmo esperadas correspondências parciais. Sendo esta pesquisa voltada para discussões e proposições linguístico-computacionais, e uma vez que buscamos propor um modelo representacional que relacione construções de línguas distintas por meio dos seus aspectos formais e funcionais, esta pesquisa se enquadra como uma aplicação em NLU (Compreensão de Língua Natural), um dos conceitos fundamentais da Linguística Computacional juntamente ao NLP (Processamento de Língua Natural). Este, em geral, desenvolve sistemas que não tratam o significado, o que, para esta pesquisa é de extrema relevância. Isso porque o alinhamento entre constructicons pode servir para várias aplicações tecnológicas, desde sistemas de compreensão de língua a sistemas de geração linguística, bem como tradução por máquina, ensino de segunda língua, entre outros. Com isso em mente, faz-se fundamental entender, inicialmente, o que vem a ser um constructicon e o contexto em que ele se insere.

Fillmore et al. (2012) definem Constructicon como um recurso sintático, um repositório de construções, como um recurso análogo a um Lexicon. Nesse sentido, um constructicon pode ser desenvolvido autonomamente, ou a partir de um outro recurso linguístico-computacional já existente. No escopo desta tese, trataremos de constructicons desenvolvidos a partir de alguma relação com uma framenet. A

¹ Neste trabalho, optou-se por utilizar Constructicon e Lexicon com inicial maiúscula ao se referir a um dos projetos existentes em desenvolvimento, por exemplo, Constructicon do PB ou Constructicon do sueco. No entanto, quando utilizado para abordar os recursos sintáticos de forma geral, lexicon e constructicon serão escritos com inicial minúscula.

² Da mesma maneira, FrameNet com inicial maiúscula será utilizada para se referir aos projetos específicos de cada língua e, com inicial minúscula, ao se referir aos recursos lexicográficos computacionais baseados na Semântica de Frames de forma genérica.

framenet é um projeto lexicográfico de agrupamento e descrição de palavras de acordo com as cenas ou situações em comum que elas descrevem (PETRUCK, 2018) e a primeira iniciativa nesse sentido foi desenvolvida para o inglês – a Berkeley FrameNet (FN), a qual está localizada no International Computer Sciences Institute, em Berkeley. O objetivo central do projeto é desenvolver um Lexicon, o qual, dito de outro modo, é um recurso no qual sob uma entrada de um frame são agrupadas unidades lexicais (LUs – Lexical Units, em inglês) que se relacionam semanticamente. Tal recurso baseia-se na Semântica de Frames (FILLMORE, 1982) e é sustentado por evidências em corpora, ou seja, para cada LU ou frame em análise, são buscadas sentenças-exemplo em corpora diversos que evidenciem a cena conceptual esperada. Sendo consideradas bons exemplos, essas sentenças são anotadas com base na política de anotação de cada uma das framenets.

Um frame, como compreendido nas framenets, é uma representação esquemática de uma situação envolvendo participantes e outros papéis conceptuais, denominados elementos de frame (FEs). Uma LU, por sua vez, é um pareamento da forma de uma palavra a seu significado (FILLMORE ET AL., 2012), o qual é expresso pela relação entre um lema e o frame evocado por ele (LEE-GOLDMAN & PETRUCK, 2018). Assim, nas framenets, estão representadas tanto relações entre uma LU e um frame – uma vez que uma LU evoca um determinado frame – quanto relações entre frames, por exemplo, a relação de herança. A definição dessas relações é importante, pois é isso o que permite estruturar uma framenet como uma rede e, segundo argumentam Ruppenhofer et al. (2016), o estabelecimento de relações permite compreender mais facilmente um frame muito complexo ao relacioná-lo a um frame mais concreto, ou ainda é possível explicar um frame mais específico, tomando como referência FEs mais genéricos.

O Constructicon, por sua vez, conforme proposto por Fillmore et al. (2012), daria conta de documentar computacionalmente amplas possibilidades sintáticas, semânticas e pragmáticas, isto é, aspectos da língua que não seriam possíveis de ser tratados unicamente no âmbito do Lexicon, por exemplo, expressões idiomáticas, padrões sintáticos esquemáticos e estruturas complexas. A ideia inicial era construir os dois recursos de forma interrelacionada e complementar, em que a base de dados da FN seria aumentada ao incluir informação construcional (LEE-GOLDMAN & PETRUCK, 2018). Além disso, ao assumir a definição clássica de construção como um pareamento de forma e sentido (GOLDBERG, 1995; 2006), a contraparte

semântica de uma construção, quando possível, seria especificada através de um frame descrito na FN.

A partir da iniciativa de Berkeley para o inglês, foram desenvolvidos projetos para outras línguas, seguindo aproximadamente a mesma proposta e os conceitos da Gramática das Construções de Berkeley (BCG³). No Brasil, Lexicon e Constructicon são desenvolvidos no laboratório da FrameNet Brasil, na Universidade Federal de Juiz de Fora, compartilhando a mesma base de dados, com relações mapeadas entre frames e construções e entre seus elementos.

A FN-Br teve início em 2008, com o desenvolvimento do Lexicon. Anos mais tarde, em 2010, iniciou-se a pesquisa voltada para a implementação do Constructicon do PB, que, em sua primeira fase, contou com a descrição de construções da família Para Infinitivo⁴ (TORRENT, 2009; LAGE, 2013) e de Quantificação Indefinida (BRODBECK, 2010; TAVARES, 2013). Além disso, foram definidas as políticas de anotação de construções e os critérios de análise de uma determinada estrutura via Constructicon ou via Lexicon (LAGE, 2013). Ainda nesse primeiro momento, inserida no contexto do projeto Frames e Construções em Contraste (SALOMÃO, SAMPAIO & TORRENT, 2013), uma pesquisa de perspectiva contrastiva foi realizada, em que 69 construções das 72 descritas no Constructicon de Berkeley foram comparadas a possibilidades em PB (LAVIOLA, 2015). Essa pesquisa julgou cada construção do Constructicon do inglês como sendo equivalente, parcialmente equivalente ou não existente em relação a possibilidades em PB, a partir de suas especificações formais e funcionais e de exemplos anotados. Essas análises, entretanto, basearam-se em uma perspectiva subjetiva, dependente da noção de correspondência empregada pela pesquisadora. Contudo, para que se possa estabelecer uma relação entre constructicons, a análise contrastiva deve ser baseada em critérios mais concretos e deve extrapolar a simples comparação entre sentenças-exemplo⁵. No atual momento

³ BCG é a sigla para Berkeley Construction Grammar, a versão em inglês de Gramática das Construções de Berkeley.

⁴ As construções da família Para Infinitivo descritas no Constructicon do PB são: i. Construção de Dativo com Infinitivo; ii. Construção Modal Epistêmica; iii. Construção Modal Deontica; iv. Construção de Propósito Qualificado; v. Construção de Gap de Sujeito; vi. Construção de Gap de Objeto; vii. Construção de Gap de Adjunto; viii. Construção Volitiva; ix. Construção Manipulativa; x. Construção de Tempo Iminente; e xi. Construção de Aspecto Iterativo.

⁵ Essa discussão foi feita tanto em Laviola (2015) quanto em Bäckström et al. (2014), os quais desenvolveram análises bilíngues de comparação entre as construções do Constructicon de

do Constructicon do PB, esforços têm sido empreendidos no intuito de ampliar relações entre construções, frames e seus elementos, aumentando a generalidade e a coerência interna das análises (ALMEIDA, 2016; LAGE, 2018).

Além do inglês, do PB e do sueco, o alemão, o russo e o japonês também possuem Constructicons em diferentes níveis de desenvolvimento. Ainda que todos tenham como base teórica e como referência a BCG e o Constructicon de Berkeley, foram adotadas diferentes metodologias e abordagens para a implementação desses recursos, por exemplo, no que tange à seleção e inclusão de construções no constructicon e à relação deste com as framenets de cada língua. Devido a essas particularidades, as quais serão melhor exploradas na seção 3.2, ao se pensar na proposição de um alinhamento entre constructicons, é fundamental analisar não somente a estrutura sintático-semântica de cada construção e como elas se aproximam ou se diferenciam de uma outra língua em comparação, mas, inclusive, a metodologia utilizada por cada projeto. Com isso, emerge o problema norteador desta pesquisa:

Consideradas as especificidades e enviezamentos analíticos inerentes a cada constructicon, que metodologia poderia ser aplicada aos esforços de alinhamento entre recursos construcionais de modo a produzir análises objetivas e coerentes?

Dias-da-Silva et al. (2007), no que tange a pesquisas em PLN (Processamento de Linguagem Natural), sugerem que um trabalho de base linguística e aplicação computacional – características desta pesquisa – passa por três fases: a fase linguística, a fase representacional e a fase implementacional. Esta pesquisa, ao propor diretrizes linguístico-computacionais para o alinhamento de constructicons, aborda essas três fases.

Berkeley e possibilidades em PB e em sueco, respectivamente. Essas análises levaram ao desenvolvimento do trabalho comparativo trilingue, de cooperação entre pesquisadores do PB e do sueco, que teve como ponto de partida as análises empreendidas durante o desenvolvimento desta pesquisa, o qual resultou na proposição de quatro critérios para a comparação entre construções de línguas distintas (LYNGFELT ET AL., 2018a).

Na fase linguística, realizamos uma análise que buscou comparar as construções já descritas no Constructicon do inglês com possibilidades construcionais no português do Brasil e no sueco. Este trabalho, conduzido pela autora desta tese em colaboração com a equipe sueca, buscou definir critérios de comparação entre construções de línguas distintas com o intuito de desenvolver uma metodologia aplicável para a comparação de construções de várias línguas (LYNGFELT ET AL., 2018a⁶).

Como ação da fase representacional, foram realizadas discussões acerca de quais medidas são necessárias para que se possa alinhar computacionalmente constructicons desenvolvidos para línguas distintas. Segundo Lyngfelt et al. (2018a), o alinhamento computacional de constructicons requer a discussão das seguintes questões: como as entidades (neste caso, as entradas construcionais) seriam representadas na base de dados, quais propriedades das construções precisariam ser representadas e como estas seriam formalizadas. Surge, portanto, a hipótese deste trabalho para solucionar o problema de pesquisa proposto, segundo a qual:

Representações intermediárias e interlinguais, como as framenets e as Universal Dependencies, podem ser utilizadas para o alinhamento de constructicons, uma vez que podem ser aplicadas independentemente de escolhas metodológicas de cada time responsável por um constructicon.

Lyngfelt et al. (2018a) apresentam brevemente a proposta de Vossen (1998) e Lönneker-Rodman (2007) de duas metodologias para o alinhamento de recursos lexicais. A primeira delas é uma abordagem de expansão (*expand approach*), na qual o esqueleto estrutural do recurso fonte é transferido para a língua alvo e preenchido com os dados específicos dessa língua. A segunda é uma abordagem de fusão (*merge approach*), em que dois recursos já existentes são mapeados entre si com

⁶ Este trabalho cooperativo surgiu a partir da análise comparativa das entradas construcionais do inglês, descritas no Constructicon de Berkeley, com possibilidades em português do Brasil, realizada pela autora desta tese (LAVIOLA, 2015). A partir desses resultados, o próximo passo na tentativa de se propor um constructicon multilíngue foi sugerir a realização de uma comparação trilingue que abarcasse também os resultados da comparação das construções do inglês com o sueco e estabelecesse uma metodologia de comparação de construções aplicável a várias línguas. Tal comparação também foi conduzida pela autora desta tese, em colaboração com a equipe sueca.

base nas correspondências que existem em seus elementos constituintes (LYNGFELT ET AL., 2018a). De acordo com os autores, a abordagem que melhor se adequaria ao propósito de alinhamento entre constructicons já existentes seria a de fusão, ainda que esta garanta um alinhamento apenas parcial, visto que os recursos foram construídos de forma independente.

Embora Lyngfelt et al. (2018a) apresentem os resultados de um trabalho de comparação trilingue entre construções dos constructicons do inglês, do PB e do sueco, nenhuma tentativa de alinhamento computacional de recursos construcionais foi feita até o momento. Esta pesquisa, portanto, apresenta, como ação da fase implementacional, propostas de se representar computacionalmente e, conseqüentemente, implementar esse alinhamento entre constructicons.

Uma primeira proposta, descrita por Lyngfelt et al. (2018a), envolve a utilização de uma base de dados relacional, em que construções, frames, unidades lexicais, elementos de frames, elementos de construções (e outros) são todos definidos como entidades as quais são caracterizadas por meio de relações e restrições estabelecidas entre si. A FN-Br, ao propor uma base de dados relacional para o Lexicon, desenvolveu também um conjunto de relações e restrições para o Constructicon do PB, as quais inicialmente foram criadas para melhor representar as propriedades de cada construção. Apesar disso, tais relações e restrições podem ser bastante relevantes para o alinhamento de construções de línguas distintas, uma vez que duas construções de línguas diferentes que apresentam similaridade semântica e diferenças formais, por exemplo, podem ser modeladas como tendo uma relação de evocação de um mesmo frame, mas apresentando diferentes configurações de restrições.

Além das relações entre entidades, relações de dependência entre palavras em uma determinada sentença podem ser úteis por meio de uma comparação entre sentenças anotadas para cada língua. Isso pode ser obtido através do uso das Universal Dependencies (UDs), as quais objetivam fornecer uma anotação gramatical consistente para várias línguas, com o objetivo de serem úteis para diversos projetos de perspectiva multilíngue e aplicáveis translinguisticamente (UNIVERSAL DEPENDENCIES PROJECT, 2019). As relações de dependência, portanto, objetivam maximizar o paralelismo, permitindo que uma mesma relação gramatical seja anotada de maneira semelhante em línguas distintas. Este trabalho propõe, dessa forma, mostrar como as UD's podem ser incluídas na base de dados do Constructicon e como

é possível mapear as semelhanças entre construções de recursos de línguas diferentes, objetivando o alinhamento computacional destes – embora as UDs sejam um projeto ainda em desenvolvimento.

Dado o panorama posto para a tese, o capítulo 1 dedica-se a apresentar os modelos computacionalmente implementáveis de análise linguística, a saber a abordagem da Gramática das Construções de Berkeley em que se baseia o desenvolvimento do Constructicon do PB, e as Universal Dependencies. O capítulo 2 apresenta os conceitos de Processamento de Língua Natural (em inglês, Natural Language Processing – NLP) e Compreensão de Língua Natural (em inglês, Natural Language Understanding (NLU)), ambos centrais para a Linguística Computacional. Por sua vez, o capítulo 3 destina-se a definir Constructicografia Multilíngue e a demonstrar como ela se diferencia da abordagem adotada pela Gramática das Construções Contrastiva. O capítulo 3 também discute o que vem a ser Constructicografia e como esse conceito se articula em relação ao conceito de Lexicografia. O capítulo 4 descreve a metodologia aplicada no desenvolvimento da pesquisa, explicando os experimentos realizados para se proporem diretrizes para o alinhamento de Constructicons, a que se dedica o capítulo 5. Ao final desta tese, são apresentadas as conclusões que se pôde alcançar e quais as contribuições se depreendem desta pesquisa.

1 MODELOS COMPUTACIONALMENTE IMPLEMENTÁVEIS DE ANÁLISE LINGUÍSTICA

Este capítulo apresenta os modelos de análise linguística computacionalmente implementáveis em que se baseia a proposta desta tese. O primeiro deles refere-se à Gramática das Construções de Berkeley, a qual é uma das abordagens da Gramática das Construções. Esta se desenvolveu a partir da Gramática de Casos, proposta por Fillmore (1968, 1977) e das primeiras versões da Semântica de Frames (FILLMORE, 1982; 1984). A partir disso, várias abordagens da Gramática das Construções foram desenvolvidas, o que, por esse motivo, torna conveniente chamá-la de “gramáticas das construções” e, também por isso, nos referimos, neste trabalho, a “abordagens construcionais”. Destacam-se a Gramática Cognitivista das Construções (GOLDBERG, 1995; 2006), Gramática das Construções de Berkeley (KAY & FILLMORE, 1999; FILLMORE, 2013), Gramática das Construções Baseada em Signos (BOAS & SAG, 2012), Gramática das Construções Corporificada (BERGEN & CHANG, 2013), por exemplo. Nesta pesquisa, enfocamos a perspectiva da Gramática das Construções de Berkeley, uma vez que o estudo e o desenvolvimento de constructicons – como o fazemos – fundamentam-se nessa abordagem.

O segundo modelo apresentado é o projeto Universal Dependencies, o qual é uma iniciativa multilíngue de anotação gramatical baseada nas especificações da Gramática de Dependência.

1.1 GRAMÁTICA DAS CONSTRUÇÕES

A Gramática das Construções tem como objeto central a noção de construção gramatical. Esta, embora sempre tenha sido muito utilizada por várias áreas da linguística, era considerada periférica até a década de 70, uma vez que os estudos gramaticais até então eram voltados para uma análise composicional e, portanto, consideravam o item lexical, a palavra, os morfemas como itens fundamentais da análise gramatical, preterindo as construções. Com o surgimento da Linguística Cognitiva⁷ e o rompimento desta com certos paradigmas da Linguística Gerativa, o

⁷ Salomão (2007 e 2009) descreve a trajetória da Linguística, com foco para o surgimento da Linguística Cognitiva. Segundo ela, esta tem suas bases na proposição chomskiana de que

estudo da significação passou a ter mais relevância e, paralelamente, o estudo das construções passou a ser fundamental para as abordagens construcionistas^{8,9}, as quais consideram construções como unidades básicas da língua e, conseqüentemente da gramática, ou ainda, segundo Boas (2013), blocos fundamentais da linguagem.

Segundo Perini (2010), construção é uma fórmula que combina uma descrição formal e uma definição semântica ou funcional, a partir da qual sentenças variadas são produzidas. Goldberg (2013, p. 17), por sua vez, define construção como “pareamentos convencionais e aprendidos de forma e função em níveis variados de abstração e complexidade”¹⁰. Essa ideia está relacionada à definição saussuriana de signo linguístico, constituído por um significante (a forma ou imagem acústica, nos termos de Saussure) e um significado (o conceito ou a representação mental). Assim como os signos de Saussure, uma construção, no âmbito da Gramática das Construções, pode ser um padrão sintático, lexical ou morfológico, ou seja, qualquer estrutura linguística em que um padrão formal esteja associado a um determinado significado ou função.

Conforme Fried & Ostman (2004), tanto palavras quanto sentenças possuem propriedades construcionais e são, por esse motivo, representadas como construções

a teoria linguística é mentalista, momento em que Chomsky faz o primeiro rompimento com a tradição estruturalista. Apesar disso, o modelo gerativista – liderado por Chomsky – apresenta insuficiência no tratamento semântico e ausência de descrições relacionadas à idiomatidade das línguas. Na década de 70, portanto, surgem os primeiros trabalhos em linguística cognitiva, como os estudos de George Lakoff, acerca da categorização; de Eleanor Rosch, Brent Berlin, Paul Kay e colaboradores, os quais identificaram os efeitos de prototipia nos processos de categorização, com base no léxico das cores; e, dentre outros, de Charles Fillmore em relação à Semântica de Frames.

⁸ Todas as traduções contidas nesta tese são de responsabilidade da autora. Os trechos originais estão inseridos nas notas de rodapé.

⁹ Fala-se em abordagens construcionais, uma vez que a Gramática das Construções é um dos empreendimentos da Linguística Cognitiva, a qual tem um escopo analítico diversificado e dividido entre aqueles que desenvolvem pesquisa linguística voltada para a significação do discurso – focadas na situação comunicativa – e aqueles que se voltam para a significação da sentença – relacionada às descrições do léxico ou das construções gramaticais (SALOMÃO, 2007). Além disso, várias linhas de pesquisa nesse campo foram desenvolvidas, existindo, pois, várias abordagens da Gramática das Construções.

¹⁰ ...conventional, learned form-function pairings at varying levels of complexity and abstraction.

(p.12). Além disso, segundo Lee-Goldman & Petruck (2018), uma construção pode ser esquemática, como a construção de *Sujeito_Predicado*, ou pode ser parcialmente especificada lexicalmente, como a construção de quantificação indefinida *Uma_chuva_de_N*, descrita no Constructicon do Português do Brasil. Entendemos, dessa forma, que uma língua é um sistema construcional.

Isso nos leva a concordar com Perini (2010), quando este, retomando os estudos construcionistas fundadores, propõe que um usuário da língua tem conhecimento acerca das construções de sua língua e é a partir desse conhecimento que realiza a comunicação. Fried & Ostman (2004) confirmam essa ideia ao sugerir que “conhecer uma língua significa conhecer suas construções”¹¹. Nesse conceito, centra-se a ideia da Gramática das Construções de que “a língua é um inventário de construções”¹² (FRIED & OSTMAN, 2004, p. 13).

Nesse momento, convém diferenciar construções de construtos. Como visto, as primeiras podem ser definidas como generalizações de padrões gramaticais, ou seja, como as regras que licenciam uma determinada expressão linguística. Por outro lado, os construtos representam as realizações concretas de tais construções, em outras palavras, construtos são as instanciações das construções em situações reais.

A partir desse conceito e tendo em vista que há muitas unidades linguísticas com níveis diferentes de complexidade – ou seja, nas quais a relação da forma com o significado não pode ser definida por meio de cálculos combinatórios entre suas partes componentes (SALOMÃO, 2009) –, muitos elementos da língua não poderiam ser analisados apenas seguindo uma abordagem puramente lexical. Centra-se nisso a ideia constitutiva da emersão das abordagens construcionais. Corroborando essa ideia, Goldberg (1995) define construções como sendo um pareamento de forma e significado, e apenas são consideradas como tal

se alguns aspectos da sua forma ou algum aspecto do seu significado não são estritamente previsíveis a partir das partes componentes ou de construções já estabelecidas como existentes na língua. (GOLDBERG, 1995, p.4)¹³

¹¹ ...'knowing a language' means knowing its constructions.

¹² ...language *is* the inventory of its constructions.

¹³ ...if something about their form or meaning is not strictly predictable from the properties of their component parts or from other constructions.

Nesse sentido, ao se adotar uma abordagem construcional, conforme reforçam Kay e Fillmore (1999), um gramático das construções analisa tanto padrões gramaticais gerais, quanto padrões idiomáticos, a fim de “desenvolver um sistema de representações capaz de uma codificação econômica e sem perda de generalização de todas as construções (ou padrões) da língua”¹⁴ (KAY & FILLMORE, 1999, p. 2). Uma construção é idiomática quando seu sentido especifica relações semânticas e pragmáticas distintas daquelas obtidas a partir da soma do significado do grupo de construções menores que a compõe, o que torna insuficiente um modelo semântico essencialmente composicional.

Lyngfelt (2018, p. 5) enumera as cinco características da Gramática das Construções, conforme são apresentadas por Goldberg (2013):

- (i) **Construções gramaticais** são as unidades básicas, incluindo todo o léxico e padrões gramaticais abstratos, bem como as relações e os princípios de combinação entre eles;
- (ii) Construções são **generalizações de superfície**, ou seja, o significado está associado diretamente à forma, em oposição a abordagens mais derivacionais, como o Programa Minimalista (CHOMSKY, 1995);
- (iii) As construções de uma língua são organizadas como uma **rede** – isto é, um constructicon;
- (iv) Construções são específicas de cada língua, uma vez que são convenções, e, por isso, não é ideal tentar criar uma categoria universal de uma determinada construção. Ao contrário, a abordagem da Gramática das Construções seria comparar uma dada construção em diferentes línguas para discernir quais semelhanças e quais diferenças ela apresenta em cada língua – enfatizando a **variabilidade e a generalização translinguística**. Do ponto de vista da constructicografia, diferentemente, o principal objetivo é estabelecer construções (ou entradas construcionais) correspondentes mais próximas para línguas diferentes, muito semelhante ao que é feito na

¹⁴ develop an explicit system of representation, capable of encoding economically and without loss of generalization, all the constructions (or patterns) of the language.

lexicografia multilíngue.

- (v) A Gramática das Construções é **baseada no uso**, ou seja, fundamentada empiricamente.

Além dessas características, Fried e Ostman (2004), definem Gramática das Construções como sendo uma abordagem

- (i) **não-modular** – isto é, abarca informações fonológicas, sintáticas, semânticas, tratadas simultaneamente, garantindo, com isso, descrições de signos complexos;
- (ii) **gerativa** – por tratar de todas as construções gramaticais da língua, com destaque para as que requerem certo grau de formalização;
- (iii) **não-derivacional** – significando que não há regras de derivação de um nível a outro, diferentemente da abordagem gerativista; e
- (iv) **baseada em unificação** – ou seja, as construções e especificações que formam as construções são combinadas de forma não-derivacional. Isso significa que, ao se tentar combinar peças linguísticas incompatíveis umas com as outras, não é possível licenciar uma construção.

A Gramática das Construções de Berkeley (BCG – Berkeley Construction Grammar), nas palavras de Fillmore (2013), compreende a gramática de uma língua como “o conjunto de suas construções gramaticais, as regras que unificam informação semântica e formal em vários tipos de objetos linguísticos”¹⁵. Formalmente, a BCG é uma gramática de estrutura sintagmática – portanto, baseada na noção de constituinte – cujos nós são traços complexos. A principal operação deste modelo é a unificação. Essa operação é caracterizada pela combinação de duas ou mais Matrizes de Atributo e Valor (AVMs¹⁶) para a formação de uma nova AVM, a qual contém os mesmos valores e atributos das AVMs originais combinadas.

¹⁵ ...the set of its **grammatical constructions**, the rules that unite formal and semantic information into various kinds of linguistic objects...

¹⁶ AVM é a sigla para Attribute-Value Matrix, em inglês, a qual é utilizada dessa maneira por convenção.

Uma AVM é um conjunto de traços que especifica as características de cada elemento, e, assim, também define as possibilidades ou restrições de combinações de dois ou mais elementos em uma construção. A unificação, nesse sentido, tem como tarefa assegurar que os atributos com valores incompatíveis falhem ao se combinar e não cheguem a licenciar uma construção (FRIED & OSTMAN, 2004). Entende-se, pois, esse processo como responsável por receber ou rejeitar constituintes candidatos a preencher posições sintagmáticas particulares, em que a aceitação ou rejeição será determinada a partir da análise de compatibilidade entre os valores atribuídos aos elementos linguísticos associados e os que são requeridos pelas posições da construção.

Para melhor visualizar esse procedimento de análise da BCG, partiremos do exemplo do PB apresentado em (1).

(1) [Tem_V] [louça_{SN2}] para [mim_{SN3} lavar_{VINF}].

O exemplo (1) é uma instância da construção de *Dativo_com_Infinitivo_Obrigaçã*¹⁷. Esta herda da construção *Cláusula_Relativa_Infinitiva_com_Para*, assim como a construção de *Dativo_com_Infinitivo_Suficiência*, exemplificada em (2). As duas construções são diferenciadas, principalmente, pela contraparte semântica, visto que a *DCI_Obrigaçã* evoca o frame de *Ser_Obrigado*, enquanto a *DCI_Suficiência* evoca o frame de *Suficiência* – com as respectivas descrições apresentadas nas Figuras 1 e 2.

¹⁷ Para fins notacionais, nomes de construções serão representados em fonte *Courrier New Itálico*, e os nomes de seus elementos constituintes em *VERSALETE ITÁLICO*. De modo semelhante, nomes de frames serão grafados em fonte *Courrier New* e seus elementos, em fonte *VERSALETE*.

Suficiência

Definição

Um **Item** está localizado em uma **Escala** relativamente a um valor crítico que é determinado por alguma **Situação_habilitada**. Geralmente, quando a **Situação_habilitada** não é expressa, ela é interpretada como o desejo de que o **Item** esteja localizado em um ponto particular da **Escala**. Com muitos dos adjetivos nesse frame, a **Escala** não é mencionada explicitamente. Nesses casos, a **Escala** é geralmente a quantidade incorporada no significado dos adjetivos.

Exemplo(s)

Elementos de Frame Nucleares

FE Core:

Escala [Scale]	A Escala é o atributo do Item que está sendo descrito.
Item [Item]	A entidade cuja propriedade escalar é especificada.
Situação_habilitada [Enabled_situation]	Este EF identifica a Situação_habilitada , em relação à qual um valor crítico é julgado.

Elementos de Frame Não-Nucleares

Relações

Unidades Lexicais

Figura 1: Frame de Suficiência conforme descrito na base da FrameNet Brasil

Ser_obrigado

Definição

Sob alguma **Condição**, geralmente implícita, uma **Parte_responsável** deve realizar alguma **Obrigaçao**. Se a **Obrigaçao** não se cumpre, pode haver Consequências indesejáveis, que podem ou não ser ditas.

Exemplo(s)

Elementos de Frame Nucleares

FE Core:

Obrigaçao [duty]	A ação que a Parte_responsável é obrigada a realizar. Eu tenho que tirar um visto para o Canadá.
Parte_responsável [responsible_party]	A pessoa que precisa realizar a Obrigaçao . Todos devem votar.

Elementos de Frame Não-Nucleares

Relações

Unidades Lexicais

Figura 2: Frame de Ser_obrigado conforme descrito na base da FrameNet Brasil

Formalmente, ambas as construções apresentam a mesma estrutura, tendo como principais elementos um sintagma nominal nuclear e uma oração infinitiva

introduzida pela preposição para. Dessa maneira, tanto no exemplo (1), quanto no exemplo (2), característico da *DCI_Suficiência*, é possível verificar a seguinte estrutura: [(SN1) V SN2 para (SN3) VINF].

(2) [Ela_{SN1}] [deu_V] [dinheiro_{SN2}] para [mim_{SN3}] [viajar_{VINF}].

Nota-se que tanto o SN1 quanto o SN3 podem estar presentes ou não nas sentenças sem prejuízo para a semântica da construção. No primeiro caso, SN1 representa o sujeito da oração finita, conforme visível em (2). O SN3, por sua vez, representa o sujeito do verbo infinitivo, podendo ou não aparecer em exemplos que instanciam ambas as construções, uma vez que, ainda que o verbo infinitivo tenha sempre um sujeito, este poderá estar explícito na frase ou ser inferido pelo contexto.

Semanticamente, a construção de *DCI_Obrigação* descreve uma situação obrigatória cuja realização envolve um item codificado no *SINTAGMA_NOMINAL_NÚCLEO* (*NP_NÚCLEO*). Assim, em (1), “louça” corresponde ao objeto da ação obrigatória codificada na *SENTENÇA_INFINITIVA_PARA* (*PARA_SINF*), em “para lavar”. Como a *DCI_Obrigação* é uma cláusula relativa, tal objeto é mapeado ao antecedente da relativa.

Por sua vez, (2) representa um propósito cuja realização depende da existência de um recurso codificado no *NP_NÚCLEO*. Em (2), portanto, “dinheiro” corresponde ao ITEM e “para mim viajar”, à *SITUAÇÃO_HABILITADA*, de modo que tal item habilita a concretização da ação presente na sentença infinitiva.

Realizamos, a seguir, uma análise, através do uso de AVMs, da sentença (1), instância da construção de *DCI_Obrigação*, mais especificamente o trecho “louça para lavar”, a fim de exemplificar o processo de unificação tão relevante para a BCG. A descrição é feita por meio da inclusão de AVMs para cada elemento da construção a ser analisada, sendo essas AVMs representadas em caixas dispostas em ordem linear em conformidade com a ordem em que aparecem os elementos da sentença. Começaremos, entretanto, pelo último termo da sentença, o verbo no infinitivo “lavar”. A representação da descrição desse elemento está na Figura 3.

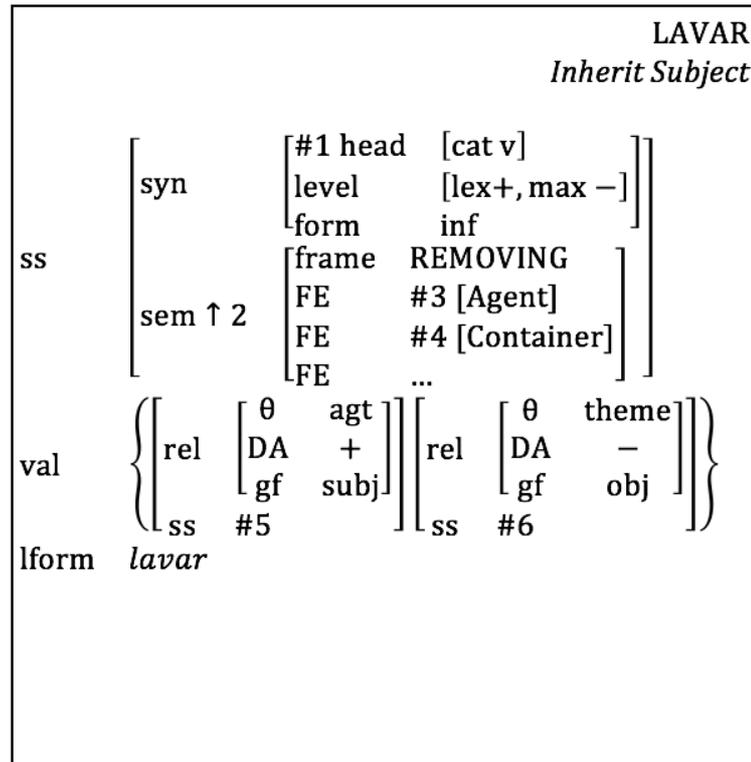


Figura 3: Representação do termo “lavar”.

Na representação de uma AVM, primeiramente, tem-se a descrição sintático-semântica (ss) do elemento. Assim, sintaticamente, “lavar” é núcleo do sintagma (head) e classificado como verbo (cat v). Além disso, o termo é marcado em “level” como lex+ e max-, o que significa, respectivamente, que se trata de uma representação lexical e que não é uma projeção máxima de “lavar”. Há ainda a informação quanto à forma do termo, que o especifica como um infinitivo (inf). Na Figura 3, há alguns números usados para indicar o processo de unificação entre os termos do construto. Eles serão melhor explicados, porém, quando outras caixas forem apresentadas, completando a AVM.

Na sequência, há a descrição semântica. No caso de “lavar”, este evoca o frame de *Removing*, o qual possui, pelo menos, dois FEs: AGENTE (AGENT) E TEMA (THEME). Outra informação relevante é a valência do termo, que indica que o verbo “lavar” requer um argumento com papel semântico de agente e função gramatical de sujeito (gf subj), além de outro argumento com função gramatical de objeto (gf obj). Este, contudo, é mapeado ao antecedente da relativa. Por fim, tem-se a representação da forma de manifestação lexical desse construto (lform): “lavar”. Há, ainda, a

informação de que esse verbo herda propriedades da construção de *Sujeito* (*Inherit Subject*).

Adicionando mais um elemento da instância em análise, a Figura 4 demonstra a representação da sentença infinitiva (Sinf) “mim lavar”.

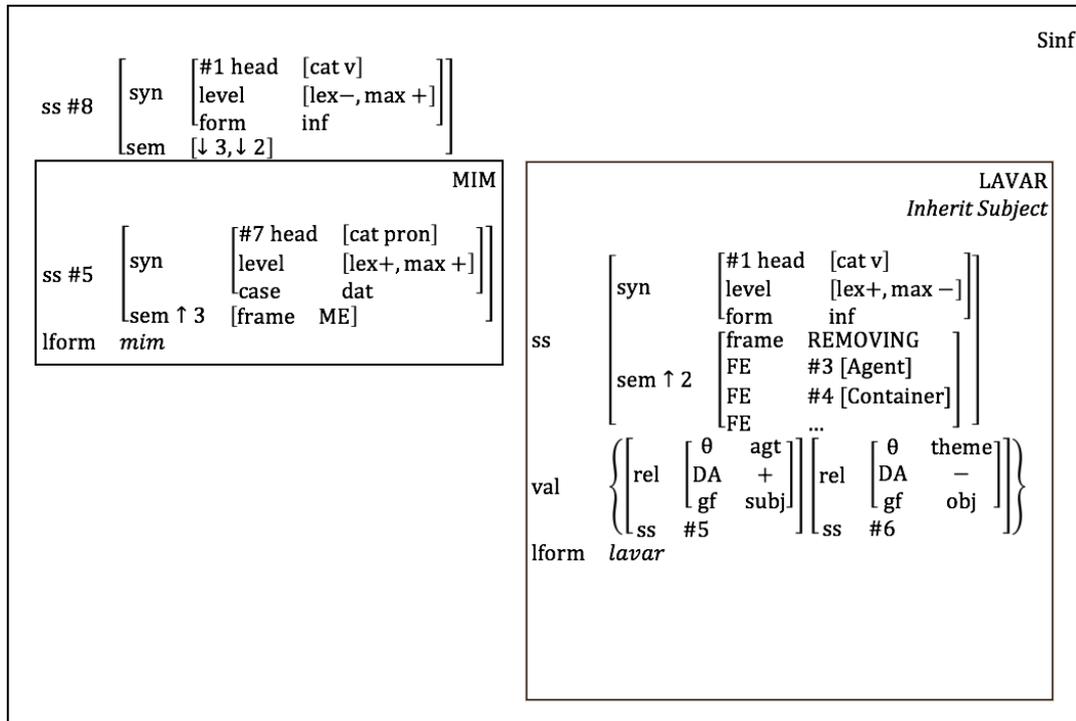


Figura 4: Representação sentença infinitiva “mim lavar”.

Na Figura 4, há duas caixas menores inseridas em uma caixa maior. A caixa maior inclui, portanto, as descrições dos dois elementos que compõem a sentença infinitiva: o verbo “lavar”, representado na Figura 3, analisada acima, e o termo “mim”. Este tem sua especificação sintático-semântica unificada àquela prevista pela valência de “lavar” e indexada com o número 3 (#3), o que indica que “mim” é sujeito agente do verbo “lavar”. Sintaticamente, “mim” está marcado como núcleo, e é da categoria pronome (“cat pron”). Além disso, é um pronome do caso dativo (dat) e é uma representação lexical em sua projeção máxima (max +). Semanticamente, está coindexado ao número 4, visto que “mim” é o elemento que cumpre o papel semântico de Agente em relação ao verbo “lavar”.

Na caixa maior, há a descrição da sentença infinitiva completa. A descrição sintática (syn) apresenta os valores de núcleo (head) coindexado em #1, uma vez que é representado pelo elemento “lavar”, de categoria verbal (cat v). Em relação ao

“level”, esse sintagma não é uma representação lexical, por isso, “lex-“ para esse traço. Por outro lado, essa é a sua representação máxima, sendo, dessa forma, o traço representado por “max +”. No que tange à semântica (sem), há nesse traço os números 4 e 2 com setas apontando para baixo, o que significa que a semântica do sintagma corresponde à semântica dos dois termos que o compõem. Assim, o número 4 está coindexado à semântica de “mim”, enquanto o número 2, à semântica de “lavar”.

A Figura 5 apresenta a descrição da sentença infinitiva regida pela preposição “para” e inclui, além das duas caixas dos elementos “mim” e “lavar” já descritos, a caixa que apresenta os traços e valores da preposição “para”. Sintaticamente, esse elemento é de categoria preposição (cat p) e é caracterizado como núcleo. Ademais, constitui-se em uma representação lexical (lex+), mas não está em sua projeção máxima (max-).

Semanticamente, ele evoca o frame de *Finalidade* (Purpose), o qual tem como principais elementos o AGENTE e o OBJETIVO. No que diz respeito à valência, esta requer um elemento que tenha papel de OBJETIVO (GOAL) e função gramatical de dependente (dep), enquanto a especificação sintático-semântica é unificada à da sentença infinitiva (Sinf) através da coindexação com #6.

Ainda sobre a Figura 5, no topo, à esquerda, há a descrição sintático-semântica da sentença infinitiva regida por “para” (*PARA_SINF*) como um todo. Em relação à sintaxe, o núcleo de *PARA_SINF* é a preposição “para”, identificado pela coindexação com #8; tal sintagma não é uma representação lexical (lex-) e está em sua projeção máxima (max+). A semântica do sintagma é especificada pelos números 7, 4 e 2 com setas direcionadas para baixo, significando que o sentido de *PARA_SINF* é uma composição do sentido das demais partes que formam o sintagma. Desse modo, 7 é coindexado ao frame evocado pela preposição “para”; 4 ao frame evocado por “mim” e 2, ao frame evocado por “lavar”.

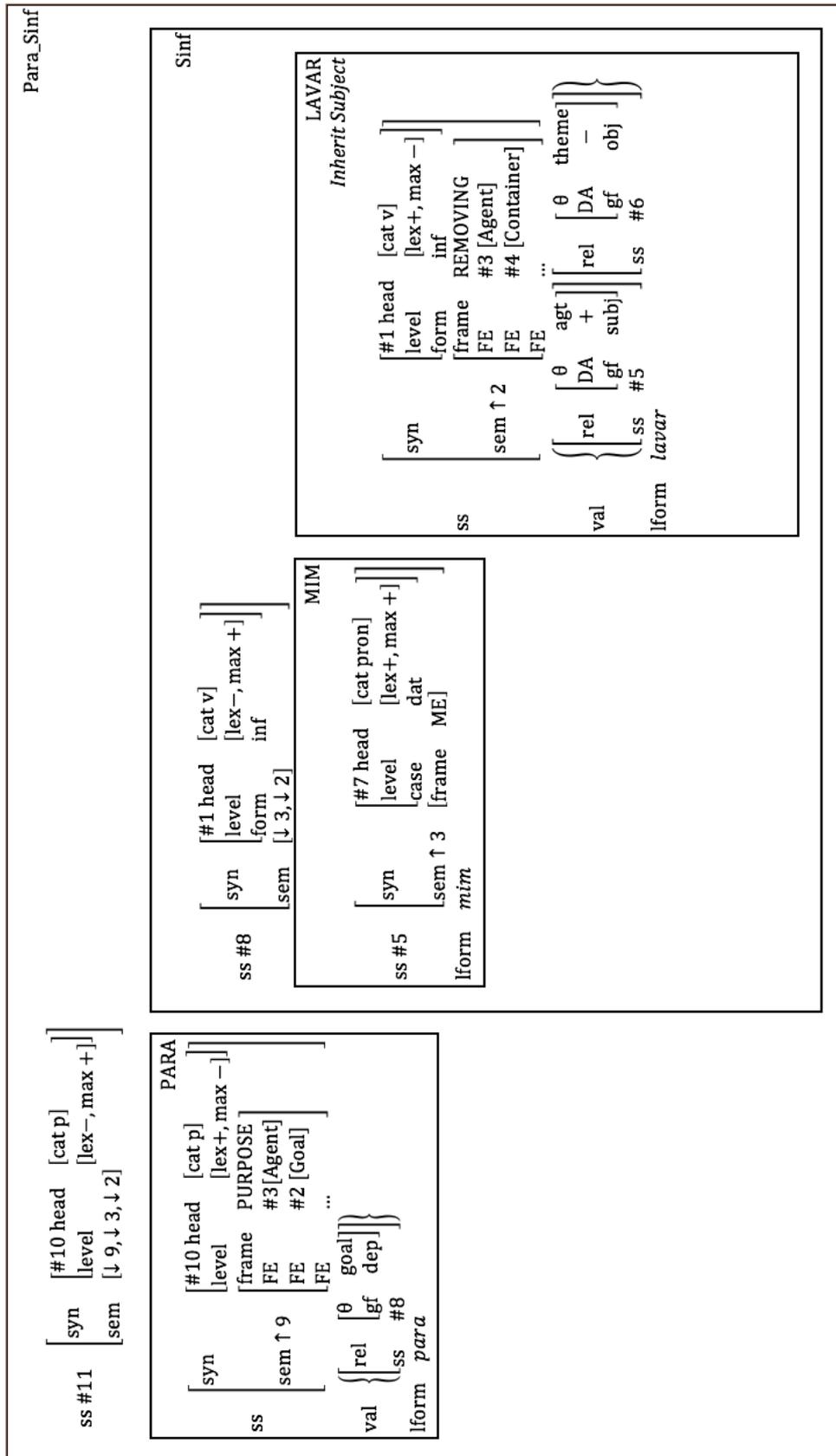


Figura 5: Representação de “para mim lavar”

Por fim, a Figura 6 demonstra a descrição de “louça para mim lavar”, a qual abarca a caixa que contém a descrição de *PARA_SINF*, vista na Figura 5, e inclui a descrição do sintagma nominal “louça”. Há, portanto, uma caixa maior (*HEAD_NP*) que envolve duas outras caixas. A menor delas contém o termo “louça”, o qual é, sintaticamente, um elemento nuclear, de categoria nominal (cat n), de representação lexical (lex+), mas que não está em sua projeção máxima (max -). No tangente à semântica do termo, “louça” evoca o frame de *Artefato* (*Artifact*), que possui o FE (*ARTIFACT*). A segunda caixa engloba o termo “louça” e o apresenta como um sintagma nominal (NP). As únicas diferenças concentram-se no fato de que esse elemento deixa de ser uma representação lexical (lex-) e está em sua projeção máxima (max+).

Há, na sequência, a caixa enquadrada como *SINTAGMA_NOMINAL_NÚCLEO* (*HEAD_NP*). Nesta, a diferença com as duas outras caixas menores centra-se no traço max-, ou seja, não é mais uma projeção máxima, visto que apresenta requisitos de valência, quais sejam um elemento de papel *OBJETIVO* (*GOAL*) com função gramatical de dependente (dep). Note-se que esta valência não é prevista prototipicamente por *louça.n*, mas pela construção de *Dativo_com_Infinitivo_Obrigaçao*. Além disso, o frame evocado é o de *Ser_Obrigado* (*Being_Obligated*), o qual tem o FE *OBRIGAÇÃO* (*OBLIGATION*), coindexado ao #9.

Por fim, na caixa maior que representa toda a estrutura de uma instância da construção de *Dativo_com_Infinitivo_Obrigaçao* (“louça para mim lavar”), apresenta-se como núcleo do construto o SN “louça”, devido à unificação em #11. Além disso, a estrutura do construto apresenta os traços lex- e max+, significando que não se trata de uma representação lexical e que está em sua projeção máxima, respectivamente. Em relação à contraparte semântica, esta é representada pelos números 10, 13 e 7, com setas direcionadas para baixo, o que indica que a semântica do construto remete aos frames de *Artefato* (#10), *Ser_Obrigado* (#13) e *Finalidade* (#7).

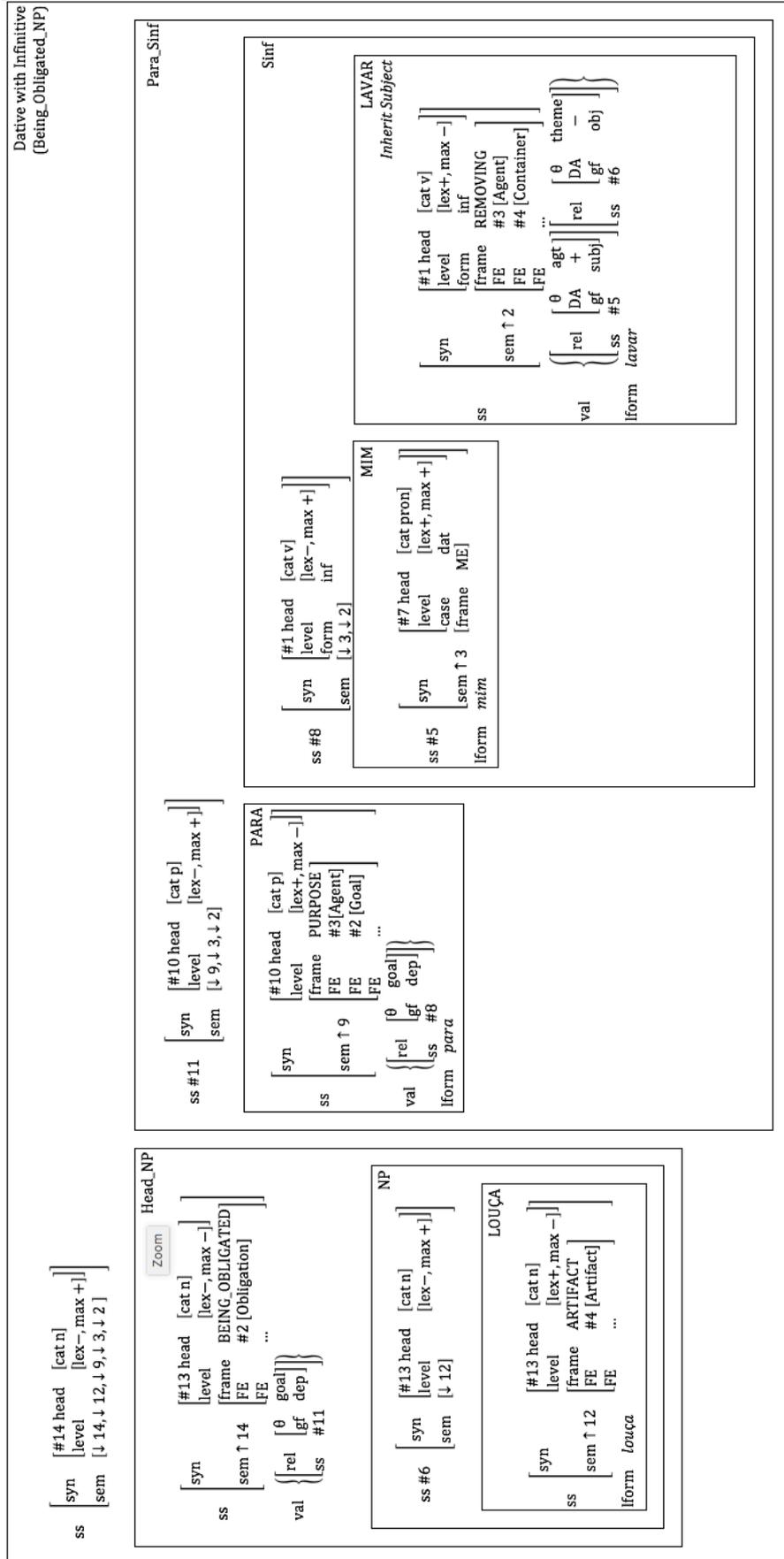


Figura 6: Representação de “louça para mim lavar”

Lage (2018) apresenta a descrição das AVMs do construto “dinheiro para mim viajar”, instância da construção de *Dativo_com_Infinitivo_Suficiência*, presente em Torrent et al. (2018). Como se pode notar na Figura 7, comparada à Figura 6, a diferença fundamental entre as construções de *DCI_Obrigaçao* e *DCI_Suficiência* está no frame evocado e na valência do verbo infinitivo. A primeira evoca o frame de *Ser_Obrigado*, enquanto a segunda, o frame de *Suficiência*. Assim, a primeira tem um item que é parte da OBRIGAÇÃO, e, portanto, o SN “louça” complementa a valência do verbo no infinitivo “lavar”. No que diz respeito à *DCI_Suficiência*, por outro lado, o SN é um ITEM que se apresenta como um recurso para a execução de uma determinada atividade, mas não exerce função de complemento do sentido do verbo no infinitivo, mas de adjunto instrumental.

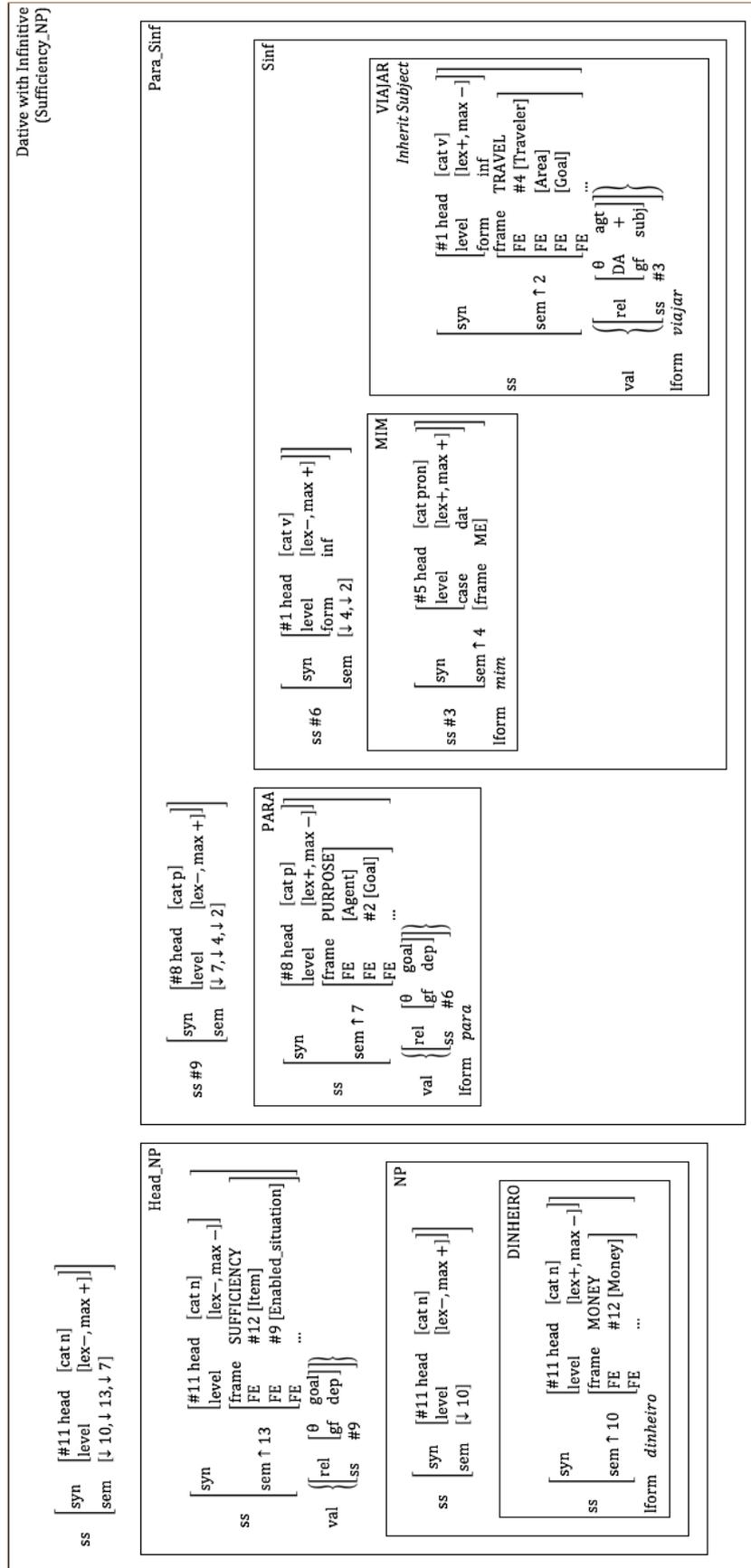


Figura 7: Representação de “dinheiro para mim viajar”. (LAGE, 2018; TORRENT ET AL., 2018)

Outro conceito de grande relevância para a BCG é o de herança construcional. Conforme mencionado anteriormente, a gramática é uma rede de construções que se organiza a partir da conexão das construções entre si por meio de links de herança (FILLMORE, 1999). O conceito de herança está ligado à noção de família, em que duas construções podem pertencer a uma mesma família que apresente função semelhante. Assim, uma delas pode ser a construção filha, a qual herda suas características da construção mãe. Essa relação de herança pode ser de dois tipos: herança única ou herança múltipla. Na primeira, a construção filha pode herdar apenas de uma construção mãe, enquanto na segunda, a construção filha pode ter mais de uma construção mãe, ou seja, herda de mais de uma construção dominante.

A BCG – e, por consequência, o constructicon – assume a noção de herança múltipla. As construções de *DCI_Suficiência* e *DCI_Obrigação*, por exemplo, são herdeiras da construção *Cláusula_relativa_infinitiva_para*, o que, na Figura 8, é representado pelas setas vermelhas. Na origem da seta, está a construção mãe, e no seu final, as construções filhas. Ambas apresentam os CEs *PARA_SINF* e *NP_NÚCLEO* que também se fazem presentes na *Cláusula_relativa_infinitiva_para*. Esta, por sua vez, é herdeira de duas outras construções: a *Cláusula_relativa_infinitiva* e a *Adjuntiva_final_infinitiva* conforme se observa na Figura 8. Da *Cláusula_relativa_infinitiva*, a *Cláusula_relativa_infinitiva_para* herda a propriedade de ser uma cláusula infinitiva, enquanto que da *Adjuntiva_final_infinitiva*, ela herda a propriedade de a cláusula relativa ser um sintagma verbal infinitivo introduzido por uma preposição para, em que esta indica, semanticamente, uma finalidade.

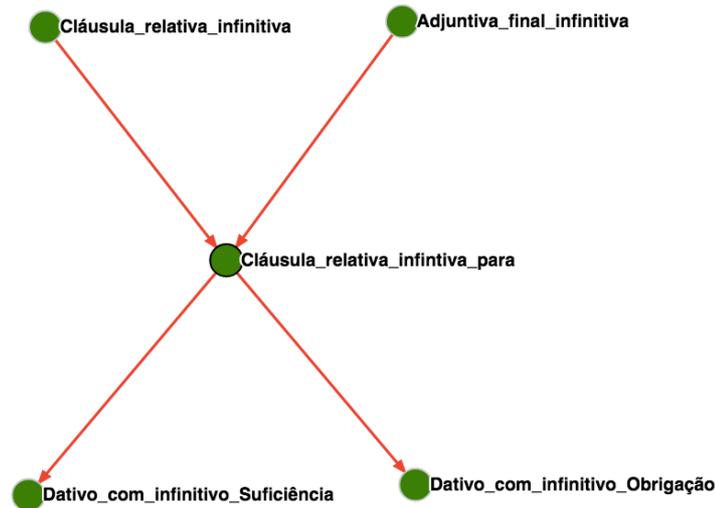


Figura 8: Grafo da relação de Herança entre as construções gerado a na base de dados da FrameNet Brasil.

A BCG é a abordagem da Gramática das Construções que embasou o surgimento dos constructicons. Além das especificações da BCG e da Semântica de Frames, utilizamos para a proposta de alinhamento entre constructicons o projeto Universal Dependencies. Sobre as características das UDs, dedica-se a próxima seção.

1.2 UNIVERSAL DEPENDENCIES

A gramática de dependência, de acordo com Nivre (2005), desenvolveu-se como uma forma de representação sintática e compreende uma ampla diversidade de teorias gramaticais e formalismos que compartilham certas premissas básicas acerca de estrutura sintática, em particular a de que a estrutura sintática consiste de elementos lexicais ligados por relações binárias assimétricas chamadas dependências. Assim, a propriedade formal comum das estruturas de dependência, como comparada a representações baseadas em constituência, é a ausência de nós sintagmáticos (NIVRE, 2005, p.2).

Hudson (1980) define constituência como uma relação parte-todo, enquanto dependência é caracterizada como uma relação parte-parte. De modo semelhante, nas palavras de Langacker (2018), constituência é uma “questão de relação parte-todo, enquanto dependência consiste em relações entre uma parte ‘principal’ e as partes ‘auxiliares’” (LANGACKER, 2018, p. 36). Assim, nas relações de constituência,

uma palavra núcleo de um constituinte é central na organização de um constituinte maior (um nome em um sintagma nominal, ou um verbo em um sintagma verbal, por exemplo). As palavras restantes são direta ou indiretamente dependentes do núcleo. Por sua vez, nas relações de dependência, a relação núcleo-dependente é feita de forma explícita através de uma ligação direta entre os núcleos e as palavras que são imediatamente dependentes deles (JURAFSKY & MARTIN, 2009).

Embora consistência e dependência sejam assumidos como modelos de gramática diferentes, Langacker (2018) argumenta como os dois tipos de relações podem ser representados juntamente – algo que foi alcançado pela Gramática Cognitiva e, a partir da proposta desta tese, também no Constructicon da FrameNet Brasil. Segundo ele, desde o seu início, a abordagem construcional da Gramática Cognitiva inclui ambas as relações, utilizando para a representação das construções gramaticais diagramas que tipicamente mostram relações de consistência, mas que também incorporam relações representadas em árvores de dependência. Assim, em um diagrama da Gramática Cognitiva, apresenta-se a estrutura componente cujo perfil é herdado pelo todo composto, o que equivale ao núcleo em representações de consistência. Além disso, de acordo com Langacker (2018), os diagramas da Gramática Cognitiva também indicam correspondências que abarcam relações de dependência, por exemplo, como as estruturas componentes são conectadas umas as outras e à estrutura composta. Isso vai ao encontro da visão adotada nesta tese de que relações de consistência e dependência não são excludentes.

Com base nisso, entende-se a noção de dependência como baseada na ideia de que a estrutura sintática de uma sentença constitui-se de relações assimétricas entre as palavras da sentença. Uma relação de dependência se mantém entre um núcleo e um dependente (NIVRE, 2005, p.4). Para estabelecer as relações de dependência e para distinguir o núcleo do dependente nestas relações, foi necessário estabelecer alguns critérios. Nivre (2005) cita alguns dos critérios usados para identificar uma relação sintática entre um núcleo H e um dependente D em uma construção C:

1. H determina a categoria sintática de C e pode substituir C.
2. H determina a categoria semântica de C; D fornece a especificação semântica de C.
3. H é obrigatório; D pode ser opcional.
4. H seleciona D e determina se D é obrigatório ou opcional.
5. A forma de D depende de H (concordância ou regência).

6. A posição linear de D é especificada com referência a H. (NIVRE, 2005, p.4)¹⁸

Nivre (2005, p. 23) ainda expõe quais são as vantagens e benefícios de se usar representações baseadas em dependência em analisadores sintáticos em oposição a representações mais tradicionais baseadas em constituência. A primeira delas relaciona-se ao fato de ligações de dependência estarem mais próximas às relações semânticas necessárias para o próximo passo de interpretação. A segunda vantagem é que as árvores de dependência contêm um nó por palavra – uma vez que o trabalho do analisador é apenas conectar nós existentes, e não criar novos nós, a tarefa de parsing é, de certo modo, direta. Por último, Nivre (2005) afirma que um parser de dependência realiza uma operação de uma palavra por vez ao invés de esperar por sintagmas completos.

Nesse contexto, destaca-se o projeto Universal Dependencies (UDs), o qual “é um projeto que desenvolve anotação treebank¹⁹ consistente para várias línguas, com o objetivo de facilitar o desenvolvimento de parsers multilíngues, aprendizado translinguístico e pesquisas sobre parsing a partir de uma perspectiva de tipologia linguística”²⁰ (UNIVERSAL DEPENDENCIES PROJECT, 2019).

Embora apresente vários argumentos favoráveis à utilização das UDs como esquema de representação, Nivre (2015) tem consciência de que as UDs podem não ser um esquema de anotação ideal para todos os projetos. Nas palavras do autor,

o objetivo principal das UDs é fornecer um tipo de *língua franca* para a anotação gramatical, a qual pode ser usada para intercâmbio de dados e desenvolvimento de sistemas multilíngues, mas não possui a ambição de abarcar todas as

¹⁸ 1. H determines the syntactic category of C and can often replace C. 2. H determines the semantic category of C; D gives semantic specification. 3. H is obligatory; D may be optional. 4. H selects D and determines whether D is obligatory or optional. 5. The form of D depends on H (agreement or government). 6. The linear position of D is specified with reference to H.

¹⁹ Treebank envolve corpus de dados linguísticos transcritos, enriquecidos com anotação de informações sintáticas e/ou semânticas na forma de representações arbóreas em que as relações entre os elementos são dadas no interior de sentenças ou nos fragmentos de sentenças.

²⁰ ... is a project that is developing cross-linguistically consistent treebank annotation for many languages with the goal of facilitating multilingual parser development, cross-lingual learning, and parsing research from a language typology perspective.

informações que são codificadas em um esquema específico de anotação treebank. (NIVRE, 2015, p.3,4)²¹

Apesar disso, ele acredita que “as UD’s poderiam ser uma escolha conveniente para iniciar novos projetos, dada a disponibilidade de orientações disponíveis para várias línguas” (NIVRE, 2015, p.4).²² Isso também pelo fato de o objetivo das UD’s ser “prover um inventário universal de categorias e orientações para facilitar anotação de construções similares entre línguas”²³ (UNIVERSAL DEPENDENCIES PROJECT, 2019). Como o uso das UD’s nesta tese encontra-se mais focado na promoção do alinhamento de constructicons multilíngues do que na anotação treebank, as preocupações externadas por Nivre não representam um impeditivo à escolha das UD’s como representação interlingual para os fins desta pesquisa.

A estrutura básica das UD’s envolve sentenças segmentadas em palavras e tais palavras descritas por propriedades morfológicas e ligadas por relações sintáticas. A decisão de tratar palavras como as unidades básicas de análise constitui um compromisso com a hipótese lexicalista na sintaxe, mas é também consistente com práticas de processamento de língua natural.

A descrição morfológica de uma palavra nas UD’s consiste de três níveis de representação: um lema (ou forma base da palavra), uma etiqueta universal de part-of-speech (POS) – representando a categoria lexical abstrata associada à palavra –, e um conjunto de traços representando propriedades lexicais e gramaticais que são associadas com a forma da palavra em particular. O esquema de anotação das UD’s é uma versão revisada e combinada da Universal Stanford Dependencies, do conjunto de etiquetas do Google Universal Part-of-Speech Tagset e do Interset interlingua para

²¹ The main goal is to provide a kind of lingua franca for grammatical annotation, which can be used for data interchange and development of multilingual systems, but we do not have the ambition to capture all the information that is encoded in specific treebank annotation schemes.

²² we think that UD could be a convenient choice for quick-starting new annotation projects given the availability of consistent guidelines for many languages.

²³ The general philosophy is to provide a universal inventory of categories and guidelines to facilitate consistent annotation of similar constructions across languages, while allowing language-specific extensions when necessary.

conjuntos de etiquetas morfossintáticas (UNIVERSAL DEPENDENCIES PROJECT, 2019).

Os lemas são geralmente determinados por cada língua de forma específica, enquanto as etiquetas de POS e os traços morfológicos são universais. A lista de etiquetas de POS é uma lista fixa que contém 17 etiquetas, como mostradas na Figura 9. Algumas etiquetas universais podem não ser aplicáveis a todas as línguas da mesma forma, apesar disso, para as etiquetas de POS, não é possível estender a lista com categorias específicas de cada língua. É possível, entretanto, obter classificações mais refinadas através do uso de traços (UNIVERSAL DEPENDENCIES PROJECT, 2019). Estes são informação adicional acerca da palavra, sua categoria gramatical e propriedades morfossintáticas. “Cada traço tem a forma `Name=Value` e cada palavra pode ter qualquer número de traços, separados por uma barra vertical, como em `Gender=Masc | Number=Sing`”.²⁴ (UNIVERSAL DEPENDENCIES PROJECT, 2019). A Figura 10 apresenta os traços conforme especificados no projeto Universal Dependencies (2019).

Open class words	Closed class words	Other
<u>ADJ</u>	<u>ADP</u>	<u>PUNCT</u>
<u>ADV</u>	<u>AUX</u>	<u>SYM</u>
<u>INTJ</u>	<u>CCONJ</u>	<u>X</u>
<u>NOUN</u>	<u>DET</u>	
<u>PROPN</u>	<u>NUM</u>	
<u>VERB</u>	<u>PART</u>	
	<u>PRON</u>	
	<u>SCONJ</u>	

Figura 9: Tabela contendo etiquetas de POS (UNIVERSAL DEPENDENCIES PROJECT, 2019)

²⁴ Every feature has the form `Name=Value` and every word can have any number of features, separated by the vertical bar, as in `Gender=Masc | Number=Sing`.

Lexical features	Inflectional features	
	<i>Nominal*</i>	<i>Verbal*</i>
<u>PronType</u>	<u>Gender</u>	<u>VerbForm</u>
<u>NumType</u>	<u>Animacy</u>	<u>Mood</u>
<u>Poss</u>	<u>NounClass</u>	<u>Tense</u>
<u>Reflex</u>	<u>Number</u>	<u>Aspect</u>
<u>Foreign</u>	<u>Case</u>	<u>Voice</u>
<u>Abbr</u>	<u>Definite</u>	<u>Evident</u>
	<u>Degree</u>	<u>Polarity</u>
		<u>Person</u>
		<u>Polite</u>
		<u>Clusivity</u>

Figura 10: Tabela contendo traços de POS (UNIVERSAL DEPENDENCIES PROJECT, 2019)

A anotação sintática, por sua vez, é composta por relações de dependência entre palavras, em que a representação de dependência básica forma uma árvore enraizada, contendo uma relação *root* para palavras que não dependem de nenhuma outra, mas apresentam um dependente, que, por sua vez, possui outro dependente e assim por diante. As relações de dependências universais foram criadas para capturar o conjunto de funções gramaticais que funcionam em diferentes línguas. Mais precisamente, o objetivo das UD's é maximizar o paralelismo ao permitir que a mesma relação gramatical seja anotada de forma semelhante em línguas diferentes, e permitindo fazer distinções que sejam suficientes para separar construções que não ocorrem da mesma maneira em todas as línguas (UNIVERSAL DEPENDENCIES PROJECT, 2019).

As UD's usam uma versão revisada das relações originalmente definidas pelas Universal Stanford Dependencies, e especificam 37 relações sintáticas, listadas na Figura 11. Na parte de cima da tabela, as linhas indicam categorias funcionais em relação ao núcleo e as colunas correspondem a categorias estruturais dos dependentes. A parte de baixo da tabela lista relações que não são relações de dependência no sentido estrito.

	Nominals	Clauses	Modifier words	Function Words
Core arguments	nsubj obj iobj	csubj ccomp xcomp		
Non-core dependents	obl vocative expl dislocated	advcl	advmod* discourse	aux cop mark
Nominal dependents	nmod appos nummod	acl	amod	det clf case
Coordination	MWE	Loose	Special	Other
conj cc	fixed flat compound	list parataxis	orphan goeswith reparandum	punct root dep

Figura 11: Tabela contendo relações sintáticas das UDs (UNIVERSAL DEPENDENCIES, 2019)

Também são distinguidos elementos nucleares (core arguments), como sujeitos, objetos e complementos oracionais. A primeira linha da Figura 11 mostra relações para argumentos nucleares de predicados com uma coluna para argumentos nominais ([nsubj](#), [obj](#), [iobj](#)) e uma para argumentos oracionais ([csubj](#), [ccomp](#), [xcomp](#)). Para complementos oracionais, são diferenciadas orações com controle obrigatório ([xcomp](#)) e orações com outros tipos de licenciamento de sujeito ([ccomp](#)). A segunda linha da Figura 11 mostra as relações para dependentes não-nucleares de predicados, as quais diferenciam dependentes nominais, oracionais e outros. A terceira linha apresenta dependentes de núcleos nominais.

A parte inferior da tabela mostra relações que podem ocorrer com praticamente todo tipo de núcleo e que não necessariamente correspondem a relações gramaticais tradicionais. A primeira coluna cobre relações de coordenação, a qual é analisada como uma relação assimétrica, em que o primeiro conjunto é o núcleo dos quais os outros conjuntos dependem via relação [conj](#). Conjunções coordenadas e pontuação delimitando os conjuntos são anexadas usando as relações [cc](#). A coluna a seguir apresenta relações de expressões de multpalavras fixas ([mwe](#)). As demais são

categorias especiais para casos específicos e, por último, há uma coluna Outros, em que está incluída a relação *root*.

O projeto de Universal Dependencies, como se pôde notar, é uma iniciativa focada em criar orientações para anotação gramatical consistente em várias línguas (NIVRE, 2015). Isso torna, portanto, crescente o interesse de projetos com objetivo multilíngue em utilizar as etiquetas das UD's nos processos de anotação linguística, como exemplo, pode-se citar o Constructicon do russo (JANDA ET AL., 2018) e a recém inaugurada Latvian FrameNet (GRUZITIS ET AL., 2018). Tendo isso em vista, e porque há orientações disponíveis para várias línguas, esta pesquisa propõe uma metodologia de anotação e postulação de restrições sobre as construções que combine a anotação construcional do Constructicon da FN-Br associada à anotação das relações de dependência proposta pelas Universal Dependencies, o que será detalhado no capítulo 4 desta tese.

2 LINGUISTICA COMPUTACIONAL

Este capítulo tem como objetivo situar a pesquisa desenvolvida no domínio da Linguística Computacional, uma vez que são propostas neste trabalho análises de âmbito linguístico que sejam implementáveis computacionalmente. Para tanto, discutiremos os conceitos de Processamento de Língua Natural e de Compreensão de Língua Natural, fundamentais para a Linguística Computacional, a fim de demonstrar de que maneira esta pesquisa se relaciona a eles.

2.1 PROCESSAMENTO DE LÍNGUA NATURAL – NLP

Esta seção apresenta uma das áreas centrais para a Linguística Computacional. Esta é a responsável pela investigação do tratamento computacional das línguas naturais e pelo desenvolvimento de sistemas que abarcam aspectos linguísticos.

Os trabalhos em Linguística Computacional tiveram início nos anos 1950 e, nessa época, centravam-se em estudos de tradução por máquina. Segundo Spärck-Jones (2007), os primeiros esforços nesse sentido envolveram o desenvolvimento de algoritmos de *parsing* sintático, bem como a formulação de estrutura de dados apropriados para gramáticas e dicionários. Às primeiras iniciativas computacionais, juntaram-se várias abordagens teóricas, especialmente as ideias centrais das pesquisas de Noam Chomsky (1956; 1957), bem como propostas da gramática de dependência, além de estudos em semântica – desenvolvidos tanto para a seleção dos sentidos, quanto para as formas lógicas, a fim de apoiar a representação do e o raciocínio sobre o conhecimento. Em resumo, uma vez que trabalhos computacionalmente motivados cobriam uma ampla gama de abordagens, funções e tarefas, linguistas computacionais e linguistas chomskyanos compartilhavam uma crença geral na importância de um aparato formal que pareasse cadeias de palavras com descrições de estrutura gramatical (SPÄRCK-JONES, 2007, p.439).

Nos anos 70, começou a haver divergências entre a linguística tradicional e a sua vertente computacional, o que teve continuidade nas décadas que se seguiram, tornando-se ainda mais evidente na década de 90. Isso fez com que os trabalhos em linguística computacional se voltassem para as pesquisas baseadas em corpus – as quais incorporavam ideias probabilísticas, fornecendo sua própria contribuição para

os modelos de linguagem (SPÄRCK-JONES, 2007, p. 439) – não sendo inexplicável, portanto, por que muitos trabalhos em Linguística Computacional atuais se baseiam em modelos puramente matemáticos. A Linguística de Corpus, por exemplo, estuda determinados fenômenos linguísticos e sua ocorrência em grandes amostras de uma determinada língua por meio de análises estatísticas.

Com o rompimento entre linguística computacional e linguística tradicional, a primeira continuou a se desenvolver sem a relação de dependência que havia previamente com a segunda. Segundo Spärck-Jones (2007), isso não significa que a Linguística Computacional não possa se valer de aspectos da linguística tradicional. Contudo, o foco não é mais a importância das teorias linguísticas para o desenvolvimento das pesquisas em linguística computacional, mas a importância que os computadores podem ter para as pesquisas em linguística. Nesse sentido, Spärck-Jones (2007, p. 440) reforça o quanto a Linguística Computacional tem crescido e associa esse crescimento ao desenvolvimento das pesquisas em Processamento de Língua Natural (NLP – *Natural Language Processing*), em tecnologia de linguagem humana, entre outros, as quais são úteis para testar as ideias da Linguística.

Entretanto, embora tais abordagens predominantemente estatísticas tenham obtido sucesso até o momento, conforme destaca Spärck-Jones (2007), elas têm suas próprias limitações e, por esse motivo, não são as únicas perspectivas utilizadas em NLP. Segundo Dias-da-Silva et al. (2007), o NLP apresenta-se, como “um campo de estudos bastante heterogêneo e fragmentado” (DIAS-DA-SILVA ET AL., 2007). Em contraposição aos modelos de predominância estatística, Dias-da-Silva et al. (2007) adotam a perspectiva de Winograd (1972), segundo o qual é indispensável um embasamento linguístico nos empreendimentos em NLP, ou seja, para simular uma língua natural de modo satisfatório em um sistema de NLP, este precisa conter vários sistemas de conhecimento linguístico.

Nesse contexto, Dias-da-Silva et al. (2007) explicam que a montagem de um sistema de NLP exige o desenvolvimento de, no mínimo, três etapas. A primeira delas especifica os conhecimentos sobre a língua e seus usos – conceitos, regras e princípios – e, por esse motivo, é denominada fase linguística. A segunda etapa dedica-se a tratar de “questões referentes à escolha ou à proposição de sistemas de representação” (p.13), por exemplo, a definição de estratégias de codificação dos elementos especificados na etapa linguística. A terceira fase, denominada fase implementacional, discute a programação das representações definidas na segunda

fase, além da programação do próprio sistema computacional da tarefa em questão. Nas palavras de Dias-da-Silva et al. (2007), a três fases são definidas como:

- **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 ET AL., 2007, p.13).

Dessa maneira, conforme explicam Dias-da-Silva et al. (2007), projetar um sistema de NLP envolve principalmente a especificação, representação e codificação sistemática de um amplo volume de informações, a fim de projetar e implementar sistemas computacionais avançados que permitam que a comunicação entre o homem e o computador ocorra através de códigos linguísticos, e não por meio de instruções e comandos codificados em uma linguagem artificial.

Apesar da relevância do desenvolvimento de trabalhos em NLP para a Linguística Computacional e, especialmente, da importância, para esta pesquisa, da metodologia de montagem de sistemas de NLP, descrita por Dias-da-Silva et al. (2007), convém tratar ainda do que vem a ser NLU e da inserção deste trabalho nessa vertente da Linguística Computacional.

2.2 COMPREENSÃO DE LÍNGUA NATURAL – NLU

A Compreensão de Língua Natural (ou *Natural Language Understanding* – NLU) é entendida, em alguns contextos, como um subconjunto do NLP. Entretanto, os programas em NLP e NLU, conforme explica McShane (2017), possuem objetivos distintos e atuam de formas diferentes, embora apresentem a mesma intenção de

contribuir para o desenvolvimento de tecnologias que irão facilitar a experiência humana e sejam, portanto, ambos de grande relevância.

De acordo com McShane (2017), durante os últimos 25 anos de pesquisa em Linguística Computacional, trabalhos de NLU ficaram em segundo plano, enquanto eram desenvolvidos sistemas baseados em aplicações de técnicas estatísticas na manipulação de corpora, que, muitas vezes, utilizavam anotação manual para alimentar o processo de aprendizagem da máquina. Tais trabalhos vêm produzindo uma série de aplicações bem-sucedidas, no que tange, por exemplo, à extração de conhecimento, geração automática de sentenças, tradução por máquina, entre outros. Contudo, conforme argumenta a autora, os sistemas de NLP tendem a apresentar uma “inabilidade de computar e registrar o significado (a base da compreensão de língua natural ou NLU)” (MCSHANE, 2017, p. 44).

McShane (2017) discute cinco crenças relacionadas a NLU e NLP, que as colocam em oposição e competição, e apresenta argumentos que refutam tais crenças:

- (i) Sistemas de NLP não necessitam de conhecimento para resolver problemas, enquanto sistemas de NLU sim. Dessa forma, sistemas de NLP seriam mais econômicos. Contudo, segundo McShane (2017), sistemas de NLP necessitam de conhecimento na forma de anotações: quanto melhor o sistema, mais e melhores anotações em corpora são necessárias. O que eles fazem, muitas vezes, é evitar fenômenos e aplicações que poderiam requerer um suporte de conhecimento indisponível nesses sistemas.
- (ii) Sistemas de NLU, baseados no conhecimento, foram testados e falharam. Segundo a autora, sistemas de NLU, na verdade, nunca foram testados, pois sistemas baseados em análises estatísticas prevaleceram nos últimos anos.
- (iii) NLU é uma extensão de NLP. Essa crença reside no fato de se aplicar o termo NLU a todo sistema que incorpora características semânticas e pragmáticas, ainda que superficialmente. Análises estatísticas de dados de corpora seriam NLP, enquanto qualquer tratamento desses dados no que tange à semântica e à pragmática seria NLU. McShane (2017), por outro lado, propõe que o termo NLU seja aplicado apenas a sistemas de compreensão profunda.

- (iv) NLP e NLU são programas opostos e em competição. McShane (2017) refuta essa crença ao afirmar que, na verdade, NLU e NLP são complementares. Ambos têm ambições e métodos diferentes, mas podem trabalhar em complementação para criar sistemas mais complexos e que deem conta de um maior número de fenômenos.
- (v) Sistemas de NLP são realistas e sistemas de NLU são irrealistas. Essa crença errônea, segundo McShane (2017), decorre de uma ênfase indevida em compartimentalização, ou seja, em segmentar a cognição e isolar a NLU. Não se pode considerar retirar NLU da cognição e isolá-la e, ainda assim, esperar que o significado seja analisado perfeitamente.

Com essas cinco crenças refutadas, McShane (2017) reforça a tese de que sistemas de NLP e NLU, embora apresentem diferentes objetivos e utilizem métodos distintos, intencionam contribuir para o desenvolvimento de tecnologias que visem a ampliar a experiência humana. Contudo, se a própria linguagem humana tem sido frequentemente caracterizada como um sistema que relaciona forma e significado, é de se esperar que sistemas baseados em perspectivas puramente estatísticas falhem ou sejam incompletos, já que, no processamento de linguagem natural, negligenciam o aspecto do significado ou tratam tal aspecto de forma superficial.

Ramalho et al. (2018), Bonial et al. (2018) e Bender et al. (2015) demonstram em seus trabalhos que projetos em NLP e NLU se beneficiariam se utilizassem métodos que envolvessem representação semântica. Bender et al. (2015) explicam que a máquina não é capaz de captar o significado pragmático de certas expressões, visto que este é obtido a partir da análise de outros fatores, por exemplo, a intenção comunicativa do interlocutor. Bonial et al. (2018) propõem um método de anotação semântica que intenciona “capturar as facetas principais do significado, abstraindo-se dos fatos sintáticos idiossincráticos” (BONIAL ET AL., 2018). Por sua vez, Ramalho et al. (2018) apresentam um método de incorporação do significado de língua natural capaz de entender a descrição de uma cena, isto é, que seja capaz de capturar as variações que a representação do significado humano implica. Como exemplo, Ramalho et al. (2018) exploram a linguagem natural espacial que é notoriamente ambígua e difícil de processar computacionalmente. Por esse motivo, eles apresentam uma arquitetura inovadora que permite representar cenas a partir da descrição linguística. Como consequência de trabalhos que dão relevância à

semântica e não sejam apenas baseados em perspectivas estatísticas, pode-se obter melhores resultados na representação sintática, confirmando a ideia de que, sendo a linguagem humana por si própria um sistema que relaciona forma e significado, um bom sistema de NLU ou NLP deverá ser baseado na relação forma-significado, dando relevância a ambos.

Para esta tese, investimos na constituição de um sistema de representação de construções que, naturalmente, unem aspectos de forma e sentido dos fenômenos linguísticos. Assim, buscamos propor um modelo representacional capaz de relacionar construções de línguas distintas através dos seus aspectos formais e funcionais e implementá-lo, demonstrando, de fato, como o alinhamento entre construções de línguas distintas é possível. Diante disso, esse trabalho se encaixa como aplicação de NLU, visto que, segundo MacShane (2017), sistemas de NLP em geral não são capazes de tratar o significado.

Ainda assim, como o intuito desta tese é propor diretrizes linguístico-computacionais para o alinhamento dos constructicons existentes, que possam ser úteis na produção de um constructicon multilíngue, metodologicamente, utilizamos as três fases descritas por Dias-da-Silva et al. (2007), mesmo que estas tenham sido elaboradas para o desenvolvimento de sistemas de NLP. Nesse sentido, ainda que, metodologicamente, a proposta das fases de Dias-da-Silva et al. (2007) se faça bastante útil para a estruturação deste trabalho, é essencial reforçar nosso posicionamento de que sistemas de NLU – que abarcam a representação semântica – tendem a ser mais bem sucedidos no tratamento de linguagem natural.

3 CONSTRUCTICOGRAFIA MULTILÍNGUE

Neste capítulo, apresentamos a proposta de constructicografia multilíngue, a qual envolve, de um lado, segundo Lyngfelt et al. (2018a), análises contrastivas e, de outro, o desenvolvimento de ferramentas para a ligação entre os recursos e de um formato de representação útil para línguas diversas. Antes disso, porém, introduzimos o conceito de constructicografia como a área de estudos voltada para o desenvolvimento de constructicons e, a partir dela, apresentamos os seis constructicons que estão em desenvolvimento atualmente – o constructicon do inglês, do sueco, do alemão, do russo, do japonês e do português do Brasil.

Em seguida, discutimos o que vem a ser constructicografia multilíngue e qual a sua relação com a teoria da Gramática das Construções Contrastiva. Uma vez que constructicografia é muitas vezes assumida como um cruzamento entre Gramática das Construções e lexicografia, a constructicografia multilíngue pode ser considerada uma combinação entre a Gramática das Construções contrastiva e a lexicografia multilíngue (LYNGFELT ET AL., 2018a, p.256) e é o que embasa a proposta desta tese, que consiste na elaboração de um sistema que alinhe constructicons de línguas distintas.

3.1 CONSTRUCTICOGRAFIA

Na Gramática das Construções, uma das ideias centrais é a concepção de língua como uma rede de construções – uma das definições de um constructicon. Outra definição é a adotada neste trabalho e primeiramente introduzida por Fillmore no desenvolvimento do constructicon do inglês, que caracteriza constructicon como sendo um repositório de descrições construcionais, isto é, um dicionário de construções (LYNGFELT, 2018).

Lyngfelt (2018) argumenta que o desenvolvimento prático de um constructicon pode ser caracterizado como uma mescla entre Gramática das Construções e Lexicografia, surgindo dessa mescla o termo constructicografia, a qual pode ser definida como a área que estuda o desenvolvimento de constructicons. Essas duas tradições – Gramática das Construções e Lexicografia – sobrepõem-se em alguns aspectos, por exemplo, no que diz respeito à morfologia, mas são bastante diferentes em outros. Um exemplo é a forma com que a lexicografia geralmente apresenta a

língua: por meio de listas de entradas lexicais. Essa escolha reflete o foco principal da lexicografia, que é o de dar conta de sentidos convencionais das palavras, ao invés de tentar cobrir todos os possíveis usos. Assim, as definições encontradas em dicionários são generalizações idealizadas, que equivalem a descrições prototípicas. A sintaxe, por outro lado, preocupa-se com outros tipos de generalizações idealizadas, sendo seu principal objetivo estabelecer um conjunto preferencialmente pequeno e finito de regras genéricas para dar conta de um conjunto infinito de expressões gramaticais (LYNGFELT, 2018).

Apesar disso, como mencionado anteriormente, Fried e Ostman (2004) propõem a Gramática das Construções como uma abordagem gramatical não modular, o que é motivado pela existência de muitos padrões que combinam propriedades lexicais e gramaticais. A partir dessa concepção, rejeita-se a distinção tradicional entre gramática e léxico e, por outro lado, propõe-se um *continuum* de construções. Isso significa que, segundo Lyngfelt (2018), a maioria das generalizações linguísticas não são nem totalmente gramaticais, nem completamente lexicais, mas localizadas em algum ponto entre esses dois aspectos. Dessa forma, seria possível afirmar o lexicon como sendo um subconjunto do constructicon. No entanto, os constructicons existentes por motivos de aplicação prática – uma vez que são todos recursos linguístico-computacionais –, fazem distinção entre lexicon (FN) e constructicon.

No que tange especificamente a um constructicon, duas características relevantes devem ser destacadas: (i) um constructicon é uma rede de construções, ou um conjunto de redes, que podem interagir umas com as outras, e (ii) construções (ou construtos) são passíveis de combinação, o que, em geral, é proposto a partir do processo de unificação, mas que é insuficiente para tratar fenômenos de coerção e outras complexidades (LYNGFELT, 2018).

Assim, um constructicon pode ser definido como um dicionário de construções gramaticais, uma vez que seus propósitos, condições de uso e as exigências do formato e descrição apresentam uma forte correlação com aqueles da lexicografia. Entradas lexicais em dicionários, em geral, buscam dar conta do uso, são dispostas em listas ou algum outro formato de busca, preferem descrições simples e de fácil compreensão pelo usuário. As entradas construcionais devem ser tanto legíveis por usuários humanos – e, portanto, seguem a mesma lógica das entradas lexicais em dicionários – quanto por máquinas, visto que um constructicon é também um recurso

computacional. Nesse contexto, aplicações computacionais – e tentativas de se conectar recursos de línguas diferentes, como a que é objeto desta tese – requerem descrições formalizadas e estratégias relacionais, enquanto usuários humanos preferem descrições mais fáceis de se ler e compreender.

As seções a seguir dedicam-se a apresentar os constructicons que vêm sendo desenvolvidos para diferentes línguas – a saber, o do inglês, do sueco, do alemão, do russo, do japonês e do português do Brasil – e o estágio de desenvolvimento em que cada um se encontra. Alguns deles são projetos bastante recentes – alemão, russo e japonês –, enquanto outros encontram-se em estágios mais avançados, com uma quantidade maior de construções descritas, bem como aplicações em desenvolvimento – como o constructicon do inglês, do sueco e, especialmente, do PB. Além disso, intenciona-se mostrar de que maneira cada recurso se relaciona com o desenvolvimento da FrameNet de cada língua. Para tanto, é necessário antes verificar se há uma framenet disponível na língua em questão – não há, por exemplo, uma framenet em desenvolvimento para o russo. Nos casos em que há uma FrameNet disponível na língua, apresentaremos quais os constructicons que mantém relação com a FrameNet da respectiva língua e quais são recursos independentes. A conexão entre os dois recursos se justifica por uma perspectiva que entende que, sendo uma construção um par forma e função, toda construção evocaria um frame. Já a independência entre constructicon e framenet pode ser justificada pela existência de construções que são padrões sintáticos abstratos e que não evocariam frame. Entende-se como evocação de um frame por uma construção a relação entre informações da construção e especificações de um dado frame descrito na língua. Sabendo disso, é possível haver construções que não evocam frames porque não há um frame ainda descrito na framenet da língua que se relacione à função característica da construção, ainda que essa função seja abstrata. Dessa maneira, com a necessária expansão da framenet da respectiva língua adicionando novos frames de acordo com as especificidades das construções descritas no constructicon da língua, seria possível relacioná-los através da relação de evocação e, assim, conectar os dois recursos.

3.1.1 Constructicon do inglês

O Constructicon do Inglês, desenvolvido em Berkeley, é a primeira iniciativa de constructicografia como descrevemos neste trabalho. Ele foi inicialmente desenvolvido como um apêndice à FrameNet com o intuito de tratar construções periféricas que não eram suficientemente tratáveis por um viés lexicográfico. Sendo assim, FrameNet e Constructicon, desde o nascimento deste, têm sido desenvolvidos como complementares e interrelacionáveis e assim segue sendo feito por alguns dos projetos para as demais línguas – como no que tange à FrameNet Brasil e ao Constructicon do Português do Brasil.

A FrameNet surgiu em Berkeley, em 1997, sob coordenação do professor Charles Fillmore, como uma iniciativa lexicográfica computacional que intencionava “descobrir e descrever as propriedades semânticas e distribucionais das palavras do vocabulário genérico do inglês” (FILLMORE ET AL., 2012). Portanto, desde a época de seu nascimento, a FN objetiva descrever as propriedades combinatórias de cada palavra, tanto semântica quanto sintaticamente. Para tanto, utiliza a Semântica de Frames para fornecer as bases da descrição semântica das unidades lexicais e faz uso de sentenças extraídas de corpora para sustentar as análises através da anotação lexicográfica (FILLMORE ET AL., 2012).

A Semântica de Frames, segundo Petruck (1996), é “um programa de pesquisa em semântica empírica que enfatiza a continuidade entre linguagem e experiência” (p.1). Essa afirmação vai ao encontro do que propôs Fillmore (1977) ao afirmar que “os significados são relativizados às cenas”, o que aponta para o fato de que os significados apresentam uma estrutura interna organizada a partir de sua inserção em cenas estruturadas culturalmente, os frames. Estes são representações esquemáticas de situações envolvendo participantes e outros papéis conceptuais (LEE-GOLDMAN & PETRUCK, 2018). Uma vez que frames são compostos por elementos (FEs) – definidos como participantes nas cenas conceptuais –, para compreender um dado conceito, é necessário entender os participantes envolvidos na cena em questão. Frames, nesse sentido, podem ser definidos como uma idealização individualizada da percepção, da memória, da experiência, da ação ou de um objeto por parte do usuário da língua. Assim, o falante toma esses conceitos relativizados às cenas para compreender sentenças, elementos do léxico ou mesmo as próprias situações cotidianas das quais participa.

Um indivíduo, por exemplo, se move a partir de um determinado local em direção a um destino, passando por um caminho específico. Essa descrição simples

envolve três elementos que compõem o frame de *Motion*: SOURCE (local fonte), GOAL (local alvo) e PATH (caminho do deslocamento). Além disso, o verbo mover-se é uma Unidade Lexical (LU) evocadora desse frame, uma vez que LU é definida como um sentido de uma palavra, expressa pela relação entre um lema e o frame que ela evoca. A Figura 12 exemplifica a descrição do frame de *Motion*, como apresentado na base de dados da FrameNet de Berkeley.

Motion

[Lexical Unit Index](#)

Definition:

Some entity (**Theme**) starts out in one place (**Source**) and ends up in some other place (**Goal**), having covered some space between the two (**Path**). Alternatively, the **Area** or **Direction** in which the **Theme** moves or the **Distance** of the movement may be mentioned.

That kite you see just to the right of his head was **MOVING** **around** **pretty fast** but the camera seemed to catch it ok.

There are several accounts of **the stench** **DRIFTING** **to shore** **from the ships in the middle of the river**

Dust particles **FLOATING** **about** made him sneeze uncontrollably.

The grill, unsecured, **ROLLED** **a few feet** **across the yard**.

Figura 12: Entrada do frame de *Motion* na base de dados da FrameNet

Atualmente, a base de dados da FN de Berkeley possui cerca de 1.200 frames, aproximadamente 13.000 LUs e quase 200.000 sentenças anotadas manualmente (LEE-GOLDMAN & PETRUCK, 2018).

Lee-Goldman e Petruck (2018) explicam ainda que, em um determinado momento analítico, a FN encontrou elementos da língua que carregam significado ou que evocam frames, mas que não são necessariamente puramente lexicais, por exemplo, expressões idiomáticas, expressões lexicais complexas ou mesmo padrões sintáticos esquemáticos. Além disso, notou-se que certas sentenças poderiam apresentar características particulares dependendo da relação entre suas partes componentes e, assim, evocariam frames diferentes daqueles evocados pela LU separadamente. Isso fez com que os pesquisadores da FN percebessem que o Lexicon sozinho não daria conta de capturar todas as unidades significativas da língua (LEE-GOLDMAN & PETRUCK, 2018). Surgia, assim, em 2008, o Constructicon do

inglês a partir do projeto Beyond The Core (FILLMORE ET AL., 2012), o qual tinha como objetivo ampliar a base de dados da FN, incluindo informação construcional que abarcasse estruturas que apresentam organização sintática e semântica que não pudessem ser totalmente explicadas nos mesmos moldes das estruturas já descritas na base de dados da FN.

De acordo com Lee-Goldman e Petruck (2018), através da teoria da Semântica de Frames, a FN já fornecia informação sobre o mapeamento de forma e sentido – referenciando a definição clássica de construção como um pareamento entre forma e significado (GOLDBERG, 1995; 2006) –, primeiramente enfocando construções lexicais. Nas palavras dos autores, a FrameNet “inclui muita informação construcional, incluindo o entendimento fundamental (para a Gramática das Construções) de que frames constituem o lado semântico das construções²⁵” (LEE-GOLDMAN & PETRUCK, 2018, p.12).

O Constructicon, portanto, foi desenvolvido com o intuito de documentar as possibilidades sintáticas, semânticas e pragmáticas do inglês através de uma ferramenta computacional online. Nesse sentido, busca-se descrever as características semânticas e sintáticas das construções e anotar sentenças exemplo de cada uma delas a fim de demonstrar essas características, aplicando basicamente as mesmas ferramentas usadas para o trabalho com o Lexicon (FILLMORE ET AL., 2012. p. 309).

Antes de inserir uma construção na base de dados, porém, é feita uma análise a fim de verificar se determinado padrão deve ser tratado como uma construção. Segundo Lee-Goldman e Petruck (2018), essa análise é guiada pela seguinte pergunta: “este padrão de uso da língua, incluindo sua forma e seu significado, é previsível a partir de outros princípios da língua, ou ele possui características que demandam uma análise específica?²⁶” (p. 19). Se for necessária uma análise específica, então o padrão em questão é considerado uma construção que pode ser tratada via Constructicon.

²⁵ includes much information about constructions, including the fundamental (to Construction Grammar) understanding that frames constitute the semantic side of constructions.

²⁶ is this pattern of language use, including both form and meaning, predictable on the basis of other (perhaps more general) principles of the language, or does it have features that demand specific analysis?

Atualmente, no Constructicon de Berkeley, há 74 padrões construcionais descritos. Na primeira linha de uma entrada construcional, está o nome da construção, seguido, na segunda linha, pelas relações que ela apresenta. Uma relação construcional é a relação de Herança Construcional, em que uma determinada construção, a construção filha, apresenta as mesmas restrições semânticas, pragmáticas e formais de uma outra construção, denominada construção mãe, o que significa que tal construção filha herda essas características da construção mãe. Relacionada a essa relação, há a relação de Herança de Elemento da Construção (CE), a qual envolve um CE que é herdado de outro CE, quando todas as restrições formais, semânticas e pragmáticas do CE pai também se aplicam ao CE filho. Lee-Goldman e Petrucci (2018) ainda apresentam outras relações: a ligação de um CE a uma construção, em que um CE em uma construção tem como especificidade a instanciação de outra construção; e a ligação de uma construção a um frame, em que o significado desta é o significado de um frame em particular.

Na Figura 13, por exemplo, a construção *Way_manner* mantém uma relação com o frame de *Motion*, o que indica que a contraparte semântica dessa construção é definida pelo frame que ela evoca. Além disso, ela possui uma relação de Herança, em que herda atributos da construção *Way_neutral*. Nas linhas seguintes, é apresentada a descrição sintática e semântica da construção, a qual pode ser seguida por exemplos e, quando necessário, por referências bibliográficas. Há, ainda, uma segunda parte da entrada construcional em que são definidos os elementos da construção, conforme se pode ver na Figura 14.

Way_manner

Evokes the Motion frame.
Inherits Way_neutral.

- A verb exceptionally takes *one's way* (the CEE) as a direct object, where *one's* is a possessive pronoun coindexed with the external argument of the verb. Together, they indicate that some entity moves while performing the action indicated by the manner verb. The manner verb is either transitive or intransitive, and thus labeled either *Transitive_manner_verb* or *Intransitive_manner_verb*. Following *one's way* is an obligatory frame element indicating some core aspect of motion (Source, Path, Goal, Direction).
- The semantics of this construction is identical (or at least very close to) that of the frame Motion: A **Theme** moves under its own power from a Source, in a Direction, along a Path, to a Goal, by a particular means. In many cases the path traversed by the Self_mover is also created by them as they go, in a particular manner (i.e., while performing some temporally coextensive action) (as in *he whistled his way through the plaza*).

• [_{the} **She**] [_{t_man} whistled] [_{cee} **her way**] [_{Path} **down the lane**] [_{goal} **to the silo**].

• References:

- Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: Chicago University Press.
- Kuno, Susumu and Takami Ken-ichi. 2004. *Functional Constraints in Grammar: On the Unergative-Unaccusative Distinction*. Amsterdam: John Benjamins Publishing Company.

Figura 13: A construção *Way_manner* do Constructicon do inglês

- **CEE(cee)**: The construction-evoking element is the noun phrase *one's way*, where *one's* is coindexed to the **Theme**.
ex.: She whistled [_{cee} her way] down the lane to the silo. TRANSLATIONS 1 2
- **Direction(dir)**: The direction that the **Theme** heads in during the motion.
ex.: She whistled her way down the lane [_{dir} towards the silo]. TRANSLATIONS 1 2
- **Goal(goa)**: **Goal** is used for any expression which tells where the **Theme** ends up as a result of the motion.
ex.: She whistled her way down the lane [_{goa} to the silo]. TRANSLATIONS 1 2
- **Intransitive_manner_verb(i_man)**: The **Intransitive_manner_verb** takes the CEE as its object, and indicates the action performed by the **Theme** while it moves.
ex.: She [_{i_man} whistled] her way down the lane to the silo. TRANSLATIONS 1 2
- **Manner(man)**: Any expression which describes a property of motion which is not directly related to the trajectory or rate of motion expresses the frame element Manner. Descriptions of steadiness, grace, means of motion, and other things count as Manner expressions.
ex.: She [_{man} gracefully] whistled her way down the lane to the silo. TRANSLATIONS 1 2
- **Means(mea)**: An action which enables the **Theme** to move.
ex.: She laughed her way home [_{mea} by not thinking about all the horrible things that had happened]. TRANSLATIONS 1 2
- **Modifier(mod)**: The *way* in the **CEE** may be modified by an adjective. It often depicts a state of the **Theme** related to or resulting from their motion, but in some cases it modifies the path taken, motion event itself, or (rarely) a trait of the **Theme** unrelated to the fact that it is in motion. The **Modifier** is always indicated on the second layer.
ex.: She whistled her [_{mod} cheerful] way down the lane to the silo. TRANSLATIONS 1 2
- **Path(pat)**: The space traversed by the **Theme** between the **Source** and **Goal**.
ex.: She whistled her way [_{pat} down the lane] to the silo. TRANSLATIONS 1 2
- **Source(sou)**: **Source** is used for any expression which implies a definite starting-point of motion. In prepositional phrases, the prepositional object expresses the starting point of motion. With particles, the starting point of motion is understood from context.
ex.: She whistled her way [_{sou} from the schoolhouse] down the lane. TRANSLATIONS 1 2
- **Theme(the)**: The **Theme** moves, often under its own power. Normally it is expressed as an external argument.
ex.: [_{the} She] whistled her way down the lane to the silo. TRANSLATIONS 1 2
ex.: She whistled [_{the} her] way down the lane to the silo. TRANSLATIONS 1 2
- **Transitive_manner_verb(t_man)**: The **Transitive_manner_verb** takes the CEE as its object, and indicates the action performed by the **Self_mover** as it moves. It normally takes a post-verbal (i.e., non-subject) argument, but in this construction that FE is unexpressed and specified to be interpreted indefinitely (i.e., INI interpretation). Note that here "transitive" is meant to cover not just NP-taking verbs but also verbs that take PP, AP, AVP, etc. complements.

Figura 14: Elementos da Construção *Way_manner*

Contudo, convém ressaltar que, apesar de serem definidas na entrada construcional, nenhuma das relações previstas – herança, evocação ou CE-construção – foi implementada no projeto Beyond The Core. Isso decorre principalmente do fato de este ter sido um projeto piloto de dois anos de duração, que, por não ter sido renovado, não teve continuidade e, com isso, muitas das ideias discutidas e propostas não puderam ser implementadas.

No que tange ao procedimento de anotação, Fillmore et al. (2012, p.310) sugerem que ele tem como base o método de identificação e classificação de sintagmas – da mesma forma como é proposto na anotação lexicográfica. Além disso, intenciona indicar: (i) as porções de língua que licenciam a construção em questão – no exemplo em (3), que instancia a construção *Rate.cost_time*, isso é representado pelo sintagma entre chaves { } em (3a); (ii) os segmentos dos trechos que são determinados pelas exigências da construção – representados pelos sintagmas entre colchetes [] em (3b); e (iii) os elementos do contexto cujas propriedades requerem uma instância da construção dada ou são selecionados por ela – evidenciados pela preposição disposta fora das chaves em (3c).

- (3) a. Ela foi contratada por {quatro dólares a hora}
 b. Ela foi contratada por {[quatro dólares] [a hora]}
 c. Ela foi contratada [por] {[quarto dólares] [a hora]}

Rate.cost_time NoColor NoTag ColorTag summary

Inherits from Rate.

- The **Rate.cost_time** construction indicates rate of transfer of assets in terms of **Cost** for each period of (iterated) **Time**.
ex.: I am now going to have to pay [_{Cost} **five pounds**] [_{Time} **per day**] for these children. TRANSLATIONS 1 2

- Cost(cos)**: **Cost** expresses amount of assets transferred for each period of (iterated) Time.
ex.: I am now going to have to pay [_{Cost} **five pounds**] per day for these children. TRANSLATIONS 1 2
- Time(tim)**: **Time** expresses the period of (iterated) time over which the amount of assets indicated by Cost are transferred.
ex.: I am now going to have to pay five pounds [_{Time} **per day**] for these children. TRANSLATIONS 1 2

Figura 15: Construção *Rate.cost_time* conforme constante da base de dados do Constructicon da FN

A informação entre chaves indica o signo mãe do construto e é a primeira informação a ser identificada por meio da anotação. As porções entre colchetes separados são os denominados Elementos da Construção (CEs), ou signos filhos, conforme se pode ver na Figura 15 a qual especifica a descrição da construção *Rate.cost_time*. Com base nisso, os elementos em colchetes da sentença (3) [*quatro dólares*] e [*a hora*] equivalem, respectivamente, aos CEs *COST* (CUSTO) e *TIME* (TEMPO).

Na base de dados do Constructicon da FN, a anotação é feita em camadas, semelhante à anotação lexicográfica. Há, pois, camadas em que se anotam os Elementos da Construção (CEs), as Funções Gramaticais (GFs) e Tipos Sintagmáticos (PT) dos CEs, além de uma camada em que se marca o Elemento Evocador da Construção (CEE) e outra em que é anotado o Tipo Sintagmático do Construto (CstrPT). A anotação construcional ilustra a gama de realizações construcionais de uma maneira simples, ou seja, o processo de anotação permite visualizar os parâmetros das construções, que são ilustrados através dos CEEs possíveis, da presença ou ausência de elementos opcionais e de elementos externos que representam os FEs, quando a construção evoca frame. Além disso, Lee-Goldman e Petruck (2018), mencionam que a anotação ajuda no processo de desenvolvimento e refinamento das definições das construções, visto que, no decorrer do processo de anotação, é possível perceber que as propriedades previamente especificadas para a construção estudada nem sempre são representadas pelos dados encontrados nas sentenças retiradas de corpora, e isso requer adaptações na definição da construção.

A partir do Constructicon do inglês, outras iniciativas têm sido desenvolvidas, como o Constructicon do sueco (LYNGFELT ET AL., 2012), do japonês (OHARA ET

AL., 2013), do alemão (ZIEM & BOAS, 2017), do russo (JANDA ET AL., 2018), além do Constructicon do português do Brasil. As seções seguintes dedicam-se a apresentar cada um desses projetos.

3.1.2 Constructicon do sueco

O Constructicon do Sueco (SweCcn) desenvolve-se no Departamento de Sueco da Universidade de Gotemburgo, tem como base, assim como a grande maioria dos demais Constructicons em desenvolvimento, conceitos da Gramática das Construções de Berkeley (FILLMORE, 2013) e compartilha algumas das motivações do Constructicon do Inglês. Além disso, o SweCcn é produto de uma ação cooperativa de três unidades de pesquisa situadas no mesmo departamento: Sprakbanken (“the Swedish language bank”), Lexikaliska institutet (‘Instituto Lexical’ ou ‘Centro de lexicologia e lexicografia’) e Institutet för svenska som andrasprak (‘Instituto de sueco como segunda língua’), o que não só o torna uma ferramenta linguística e um recurso para estudos na área de tecnologia da linguagem, mas também o faz útil para aplicações pedagógicas, por exemplo, no ensino de sueco como língua estrangeira. Desse modo, conforme explicado por Lyngfelt et al. (2012), o SweCcn é um projeto colaborativo que envolve pesquisadores em Gramática das Construções, Tecnologia da Linguagem, Lexicografia, Fraseologia, Ensino de Sueco como Segunda Língua e Semântica.

Antes de tratar especificamente do SweCcn, faz-se relevante descrever brevemente o trabalho desenvolvido pelo Sprakbanken, uma vez que o SweCcn é desenvolvido como uma adição ao projeto de nome Swedish FrameNet ++ (SweFN++) e está relacionado a outros recursos também desenvolvidos no mesmo departamento. O Sprakbanken foi criado em 1975, com financiamento governamental, como um centro nacional para pesquisa em recursos linguísticos em sueco, especialmente recursos lexicais e de criação de corpora. Desde então, o Sprakbanken tem trabalhado no desenvolvimento e no refinamento de recursos linguísticos e ferramentas de tecnologia da linguagem, bem como as aplicações desses recursos como ferramentas de pesquisa e de ensino nas mais variadas áreas da linguística (LYNGFELT ET AL., 2012).

O Sprakbanken desenvolveu um macrorecurso integrado, a SweFN++, o qual tem como principais objetivos o desenvolvimento de uma FrameNet para o Sueco, a integração de vários recursos lexicais existentes (desenvolvidos no Sprakbanken ou não), o desenvolvimento de uma metodologia e um modo de trabalho que explore ao máximo ferramentas de tecnologia da linguagem e outras com o intuito de minimizar os esforços humanos necessários para desenvolver a SweFN++, e o uso da SweFN++ em aplicações concretas de tecnologia da linguagem (LYNGFELT ET AL., 2018b).

Quatro componentes relevantes para o desenvolvimento do SweCcn foram produzidos a partir dos esforços empregados no desenvolvimento da SweFN++, a saber: o macrorecurso lexical, composto por quatorze projetos em desenvolvimento, incluindo o Constructicon do Sueco; a infraestrutura lexical Karp, que também é composta por outros 23 recursos lexicais; a FrameNet do Sueco; e a infraestrutura de corpora denominada Korp, a qual permite acesso a *corpora* do sueco de aproximadamente 10 bilhões de palavras. Um importante recurso que ocupa a posição central no macrorecurso lexical é o SALDO, um léxico morfológico e semântico do sueco moderno (BORIN, FORSBERG & LÖNNINGREN, 2013), que permite a conexão dos demais recursos, seja pelo conteúdo, através dos identificadores de sentido do SALDO, seja pela forma, por meio dos seus identificadores de *lemgram* (isto é, um lema representado junto a seu comportamento gramatical).

A FrameNet do sueco, por sua vez, contém cerca de 34 mil Unidades Lexicais distribuídas entre aproximadamente 1.200 frames, dos quais 50 não estão descritos em outras framenets, ou seja, são exclusivos da SweFN++. Apesar do intuito de conectar recursos lexicais do projeto SweFN++ e do intuito de tornar a FN do sueco e o SweCcn compatíveis tanto entre si quanto com as suas contrapartes para outras línguas, através do estabelecimento de links entre as entradas construcionais no Constructicon sueco e os frames descritos na base de dados da SweFN++ (LYNGFELT ET AL., 2018b), ambos continuam sendo recursos independentes e são bem menos integrados do que constructicons e framenets de outras línguas, como a FN-Br e o Ccn do PB.

No entanto, Lyngfelt et al. (2018b) apontam como uma das vantagens de se fazer tal ligação a possibilidade de se compararem construções de diferentes línguas através dos frames evocados por elas, desde que elas sejam construções que evocam frames, isto é, construções que podem ser ligadas a um frame através de uma relação – como a relação de *Evokes* (Evocação), em que determinada construção se relaciona

a um frame porque ela evoca tal frame. Até o momento, apenas metade das construções descritas no SweCcn tem uma ligação a frames na SweFN++, e, segundo eles mesmos afirmam, não é um objetivo do projeto ter todas as suas entradas construcionais ligadas a um frame do outro recurso, mas, sim, distinguir construções evocadoras de frame de construções que não o sejam. Lyngfelt et al. (2018b) apontam como um exemplo para o primeiro tipo construções de estrutura argumental, com núcleos verbais, com um potencial de alterar a valência do verbo, uma vez que essas construções em geral se referem a tipos de eventos e podem, conseqüentemente, evocar um frame que se relacione ao seu sentido central. As segundas, por sua vez, são construções que descrevem um padrão abstrato e que, em geral, têm uma função mais gramatical a nível da sintaxe. Além disso, construções que não evocam frames podem ser também aquelas que têm um significado que não se relaciona a nenhum frame até então descrito pela FN. Isso ocorre com construções que possuem um sentido pragmático ou uma função que não é tratada satisfatoriamente pela FrameNet, e referem-se menos a cenas e entidades.

Lyngfelt et al. (2018b) propõem, ainda, que a relação entre um frame da FN e uma construção deve ser baseada na relevância dessa ligação para futuros trabalhos. Pode ser relevante, por exemplo, para aprimorar as interconexões entre Constructicon e Framenet, ou então para estabelecer relações e conexões entre construções de línguas diferentes. Apesar disso, os linguistas reforçam não ser uma ambição do SweCcn estabelecer ligações entre todas as construções descritas nesse recurso e os frames da FrameNet.

Um benefício da abordagem sueca, de não estabelecer uma relação tão direta entre a FN do sueco e o SweCcn, é, segundo Lyngfelt et al. (2018b), a possibilidade de se ter uma análise construcional menos enviesada em direção às análises da FrameNet do inglês, e conseguir, assim, uma análise mais apurada das construções do sueco. Contudo, os autores também realçam um ponto negativo nessa abordagem: uma compatibilidade reduzida com os Constructicons de outras línguas desenvolvidos de forma mais aproximada com as suas respectivas Framenets.

O Constructicon do Sueco possui, atualmente, 400 entradas construcionais, sendo a maioria construções esquemáticas, situadas no limite entre léxico e gramática. O projeto tem como objetivo principal dar conta do maior número de construções existentes no continuum léxico-gramática, o que justifica a existência de uma considerável quantidade de entradas construcionais que dão conta de padrões

parcialmente esquemáticos que combinam propriedades lexicais e gramaticais. Construções lexicais apenas, contudo, não são incluídas no Constructicon, uma vez que já são tratadas pelos demais recursos lexicais, como o SALDO e a FN do Sueco, por exemplo, com os quais o SwCcn está interconectado.

O processo de seleção e análise de construções candidatas a serem tratadas no âmbito do Constructicon segue os seguintes passos:

1. Coletar as construções candidatas, o que pode ser feito manualmente, a partir de textos de uso autêntico da língua, ou seja, produzidos em situações reais de uso; ou a partir de fontes secundárias, como dicionários, gramáticas, estudos construcionais anteriores, dentre outros.
2. Buscar construtos em corpora adequados, sendo os mais utilizados textos de jornais e blogs, os quais apresentam maiores números de tokens. Quando os padrões são muito infrequentes nos corpora buscados, pode-se fazer uma nova busca em sites de busca da internet.
3. Definir se a construção candidata, após os resultados da busca em corpora, será incorporada ao Constructicon, reformulada (no caso de uma construção que está sendo revisada) ou rejeitada. Após essa etapa, se a construção for escolhida para ser incluída no Constructicon, segue-se o processo de definição da entrada construcional.

Cada entrada construcional tem, em geral, até quinze campos de informação em sua descrição, destes, três campos (*definition*, *structure sketch* e *examples*) são centrais para a caracterização da construção, uma vez que fornecem as informações mais básicas da construção no que diz respeito à estrutura, à semântica, à pragmática e à distribuição. No campo de definição, é fornecida uma explicação do sentido do padrão construcional e, em alguns casos, de como ele é utilizado. O segundo campo especifica a forma da construção e seus elementos componentes. Por último, o campo de exemplos traz pelo menos um exemplo anotado, com o intuito de ilustrar como a construção é usada no que diz respeito à estrutura sintática, à informação estrutural, à variabilidade e à distribuição (LYNGFELT ET AL, 2018b), ou seja, o exemplo é importante na medida em que torna visíveis as especificações da construção de uma

maneira prática. A Figura 16 exemplifica como é apresentada a construção de *Aposto* (*Apposition*) no SweCcn.

NAME **apposition** – *Ulf Olsson, kommunstyrelsens ordförande*
DEFINITION Två nominalfraser som följer på varandra, där den ena utgör [huvud]Entity och den andra utgör [attribut]Property. Attributet kan stå före eller efter huvudet.
STRUCTURE [NP₁ , NP₂]
EXAMPLES Hon bor i [[Stockholm]Entity , [Sveriges huvudstad.]Property]apposition

Figura 16: Entrada construcional da construção *Apposition* no Constructicon do Sueco

Na primeira linha, consta o nome da construção, *Apposition*, enquanto que, na segunda linha, está a definição do padrão, o qual é formado por duas frases nominais que se seguem uma a outra, a primeira marcada como o Elemento da Construção *ENTITY* (*ENTIDADE*), e a segunda, um atributo da primeira, identificado como o Elemento de Construção *PROPERTY* (*PROPRIEDADE*). Uma *PROPRIEDADE*, portanto, é atribuída a uma *ENTIDADE*, que se torna, dessa maneira, especificada por tal propriedade. Na terceira linha, o padrão é descrito formalmente como composto por dois Sintagmas Nominais (NP1 e NP2). Por último, tem-se um exemplo, o qual pode ser lido em português em (4).

(4) Ela mora em [[Estocolmo]Entidade, [capital da Suécia.]Propriedade]apposition

É relevante, ainda, atentar-se para o fato de que, no SweCcn, os exemplos encontrados na entrada construcional são anotados com base nos CEs. Etiquetas com o nome do elemento são adicionadas ao fragmento da sentença que faz o papel desse elemento. Na sentença (4), Estocolmo recebe a etiqueta Entidade, referente ao CE de mesmo nome, enquanto capital da Suécia, característica de Estocolmo, recebe a etiqueta Propriedade. Além disso, como é notado no exemplo (4), a construção como um todo é também anotada no SweCcn, portanto, uma etiqueta com o nome da construção é aplicada ao conjunto dos dois signos que compõem a construção de *Apposition*.

Além desses três campos centrais, também é possível haver um campo que indique Herança, ou seja, a construção em questão é uma variante mais específica

de outra mais genérica e herda desta algumas de suas propriedades. É possível também que a construção herde propriedades de várias construções, e, assim, o campo de Herança será preenchido por uma rede de construções. Outra informação que também pode estar disponível na entrada construcional é o frame evocado pela construção, quando há uma ligação possível, que fica indicada no campo Frame. Além disso, se a construção apresenta um correspondente no Constructicon de Berkeley, o nome de tal construção é adicionado à entrada construcional no campo Berkeley ID.

3.1.3 Constructicon do alemão

O Constructicon do alemão vem sendo desenvolvido em colaboração entre a Universidade de Austin, nos Estados Unidos, e a Universidade de Düsseldorf, na Alemanha (ZIEM & BOAS, 2017). A FrameNet do alemão desenvolvida pelo projeto SALSA é, até o momento, o seu mais bem elaborado repositório ancorado em uma base construcional. Entretanto, assim como é feito na FrameNet de Berkeley, os esforços se concentram em tratar expressões linguísticas evocadoras de frames, ou seja, o foco está em estudar expressões lexicais, ignorando, de certa maneira, informações construcionais que ultrapassem o nível da palavra (BOAS & ZIEM, 2018). Por isso, surge a necessidade de um recurso que dê conta especificamente das construções da língua.

Os principais objetivos relativos à proposição de um constructicon para o alemão envolvem (a) criar um Constructicon para o alemão em paralelo ao Constructicon do inglês; (b) obter um recurso compatível com outros recursos sintáticos desenvolvidos para outras línguas, como o inglês, o sueco, o PB, por exemplo, e, por conseguinte, (c) fazer parte dos esforços da FrameNet e do Constructicon multilíngues. Nesse sentido, segundo Ziem & Boas (2017), o projeto pretende utilizar duas abordagens metodológicas para inserção de construções no Constructicon do alemão. A primeira delas é a investigação de quais entradas construcionais do Constructicon do inglês (FILLMORE ET AL. 2012) podem ser reutilizadas para criar entradas paralelas no recurso do alemão.

Essa abordagem metodológica sugere um trabalho contrastivo como o ponto de partida para a criação de entradas construcionais no recurso. Nesse sentido, segundo os linguistas, há um interesse em explorar que tipos de informações contidas

nas construções cadastradas na base de dados do Constructicon do inglês (FILLMORE ET AL. 2012) podem ser reutilizadas para criar entradas construcionais comparáveis no Constructicon do alemão. Essa metodologia é semelhante a que foi adotada por Boas (2002) para reutilização de frames da FrameNet do inglês em outras línguas e, também, vai ao encontro do trabalho de comparação entre as construções descritas no Constructicon do inglês e possíveis contrapartes em português do Brasil e em sueco apresentado nesta tese e também em Lyngfelt et al. (2018a), a partir das reflexões iniciais contidas em Laviola (2015) e Bäckström et al. (2014).

A segunda abordagem proposta relaciona-se ao estudo de fenômenos gramaticais específicos da língua que podem influenciar a arquitetura do Constructicon, bem como podem ser importantes para a discussão de um possível contraste entre constructicons de outras línguas, uma vez que, sendo específicas da língua, podem não apresentar relação com nenhuma outra entrada em outro constructicon (BOAS & ZIEM, 2018-07-30).

Em resumo, Boas & Ziem (2018) intencionam propor as bases para uma metodologia contrastiva centrada, inicialmente, em uma extensão baseada em comparações das entradas construcionais do inglês para o alemão e, sequencialmente, em análises de fenômenos internos da língua e na inserção dos padrões específicos da língua que não são tratáveis a partir da abordagem contrastiva.

Os linguistas afirmam que reutilizar entradas construcionais do Constructicon do inglês nem sempre é uma abordagem metodológica eficaz, devido às particularidades das línguas. É por esse motivo que é sugerida uma metodologia baseada em corpus que esteja focada na criação de entradas construcionais específicas do alemão, garantindo que o Constructicon evolua de forma semelhante ao Constructicon do inglês, ainda que permanecendo conceptualmente independente deste (ZIEM & BOAS, 2017).

Um exemplo de construção descrita na base de dados do Constructicon do alemão é a construção *Exklamativ_Was*. Essa construção evoca os frames de *Stimulate_emotion* e *Stimulus_focus*. Portanto, semanticamente, essa construção indica uma emoção de surpresa do falante em relação a algum estímulo. Assim, os CEs são *ESTÍMULO (STIMULUS)* e o Elemento Evocador da Construção (CEE) expresso pelo termo *Was* (em português, “Que”, e em inglês “How”). Em alguns

casos, pode aparecer também o elemento *GRAU (DEGREE)* usado como um intensificador do estímulo. Vejamos o exemplo (5).

- (5) [_{CEE} Was] [_{STIMULUS} war sie hübsch geworden!]
 “How beautiful she became!”
 “Que linda ela se tornou!”

A descrição dos elementos da construção é apresentada na Figura 17. Nota-se, primeiramente, a descrição dos elementos internos (Interne Konstruktionselemente), que, na construção *Exklamativ_Was*, é o *ESTÍMULO (STIMULUS)*. Em seguida, tem-se o elemento evocador da construção (Konstruktionsevozierende Elemente) e, por fim, os elementos de correlação (Korrelierende Elemente), que podem ser associados a elementos não-nucleares, uma vez que não são obrigatórios e não aparecem em todos os exemplos da construção.

Exklamativ_Was	
Elemente der Konstruktion	
Interne Konstruktionselemente (KE)	
STIMULUS	Beispielbelegstelle: Auch bei Tisch, angesichts ihres Gatten, der stillschweigend' sein Essen hinunterschläng, blieb Frau Landaul nachdenklich gestimmt. (Was-Exklamativ[KEE]Was) (STIMULUS war aus dem trefflichen Plauderer, dem beredten und galanten Mann von damals geworden)! Heute war Herr Landaul so stumm wie einst Herr Aubevoile, aber sein mürrisches und feindseliges Schweigen hatte nichts mit der Schweigsamkeit seines Vorgängers gemein, die vielleicht der Schüchternheit entsprang und voll Ehrerbietung war. (Die Zeit, 12.02.1948, Nr. 07)
Konstruktionsevozierende Elemente (KEE)	
WAS	Beispielbelegstelle: Und da, hinter einer Bühne, sah ich auf einmal, Maria. (Was-Exklamativ[KEE]Was) (STIMULUS war sie hübsch geworden)! Gewachsen in all den Jahren. (Die Zeit, 17.04.1947, Nr. 16)
Korrelierende Elemente (KorE) exemplarisch	
DOCH	Beispielbelegstelle: Im Radio ist es gekommen. (Was-Exklamativ[KEE]Was) (STIMULUS KorE doch das Radio KorE alles bringt)! Ich denke mir, der Rundfunk ist schneller, die Zeitungen werden es auch noch
ALLES	veröffentlichen. (Die Zeit, 23.10.1947, Nr. 43)

Figura 17: Descrição dos elementos da construção *Exklamativ_Was* do alemão como consta no Constructicon do alemão

Os procedimentos de análise no constructicon do alemão envolvem anotação sintática, através de um parser sintático, o qual provê anotação automática de Tipos Sintagmáticos e Funções Gramaticais; e anotação semântica, realizada a partir da ferramenta WebAnno.

No que diz respeito às relações entre Constructicon e FrameNet, não há uma relação tão direta quanto há no Constructicon do inglês e do PB (como veremos na seção 3.1.6).

3.1.4 Constructicon do russo

Assim como os demais projetos, o constructicon do russo surgiu tendo como base teórica as noções da Gramática das Construções, da semântica lexical e da análise quantitativa de dados linguísticos e o intuito de desenvolver tanto materiais pedagógicos para estudantes de russo como segunda língua, quanto recursos tecnológicos para usuários do russo, visto que, segundo Janda et al. (2018), não havia, até então, um vasto inventário de construções do russo que permitisse tais aplicações. Ainda que alguns dicionários e outros trabalhos com construções tivessem descrito determinadas construções russas, nenhum deles chegou a ser exaustivo – especialmente da forma como é necessário que um recurso para estudantes de russo como segunda língua seja –, ou ainda intencionavam tratar construções em um nível mais genérico e abstrato, focando na sintaxe básica da língua (JANDA ET AL., 2018, p. 166).

De acordo com Janda et al. (2018), apesar de o russo ser a sexta língua mais falada no mundo – em relação ao número total de falantes – e a oitava no que se refere ao número de falantes de russo como primeira língua, ainda não há muitos empreendimentos no desenvolvimento de recursos tecnológicos voltados para o russo. O desenvolvimento do constructicon do russo vem ao encontro dessa necessidade, abordando tanto a perspectiva do falante nativo quanto do usuário não-nativo. Dessa maneira, conforme Janda et al. (2018, p. 168) explicam, o constructicon do russo evoluiu de trabalhos de referência tradicionais e de recursos de corpora eletrônicos, já que pesquisas baseadas em estudos de corpus criaram condições para estudos sistemáticos de construções linguísticas. Embora muito úteis, as pesquisas empreendidas nesse contexto ainda deixam lacunas no que tange às particularidades semântico-sintáticas do russo, as quais pretendem ser preenchidas pelo Constructicon do russo (JANDA ET AL., 2018, p. 169).

O projeto é empreendido a partir da colaboração entre pesquisadores russos – os linguistas da Higher School of Economics, em Moscou, na Rússia – e estrangeiros, a saber: pesquisadores do grupo de pesquisa CLEAR (Cognitive Linguistics: Empirical Approaches to Russian) localizado na Arctic University of Norway em Tromsø, Noruega; e os linguistas que atuam no Constructicon do sueco na Universidade de

Gotemburgo, na Suécia (JANDA ET AL., 2018). Por ter uma equipe composta por falantes nativos e não-nativos, os pesquisadores do Constructicon do russo adotam uma estratégia em que os pesquisadores de outras nacionalidades são encarregados de identificar as construções que são desafiadoras para os estudantes de russo, enquanto apenas os pesquisadores russos podem interpretar e anotar as construções (JANDA ET AL., 2018, p.172).

Embora o projeto tenha como objetivo final produzir um Constructicon que dê conta de descrever um amplo número de construções da língua, no atual momento de desenvolvimento, foram feitas escolhas quanto a quais tipos de construções deveriam ser tratadas primeiramente. Assim, tendo em vista que o projeto intenciona ser útil tanto como um recurso tecnológico quanto como um recurso pedagógico e preencher as lacunas de ambos no tangente a construções russas, foi decidido que o foco recairia não sobre as construções no nível da palavra – visto que estas são descritas em dicionários – mas, sim, sobre as construções polillexêmicas (*multi-word units*) (JANDA ET AL., 2018, p. 172).

Até o momento, há cerca de 600 construções descritas na base de dados do Constructicon do russo, o qual utiliza a mesma estrutura do Constructicon do sueco e, portanto, apresenta as mesmas características de busca e de apresentação das construções e, ainda, é desenhado para que seja comparável entre línguas. Uma entrada construcional no Constructicon russo pode incluir até cinco aspectos: nome da construção, que pode ser um exemplo ou pode ser mais esquemático, dependendo da construção; definição, a qual define a semântica da construção, com etiquetas para os elementos; estrutura; exemplos e comentários, quando necessário acrescentar alguma informação relevante sobre a construção.

A estrutura da construção é especificada com base nas Universal Dependencies, por exemplo, ao caracterizar a construção apresentada na Figura 18, o equivalente em russo à expressão “X quilômetros por hora” do PB. A informação contida em Structure – [root NP [nummod Num] [nmod [case v] NP-Acc] – significa que há dois sintagmas nominais (NP) e a preposição v ‘em’ na construção. O primeiro NP pode conter um numeral e um NP quantificado por tal numeral – por exemplo, “60 quilômetros”. O segundo NP – algo equivalente a “por hora” –, por sua vez, é governado pela preposição no caso acusativo (JANDA ET AL., 2018, p.174).

NAME	60_километров_в_час – <i>два раза в день, 60 километров в час</i>
DEFINITION	Употребляется для обозначения скорости или частоты. Обозначает [расстояние]Distance, [количество квантов]Quantity [повторяющегося действия]Event, [объем вещества]Quantity или [расход денег]Cost, приходящихся на [период времени]Time.
STRUCTURE	[root NP [nummod Num] [nmod [case в] NP-Acc]]
EXAMPLES	Разрешенная [скорость]Measure [движения]Event на этом крайне опасном участке дороги — не более [[40]Quantity [км]Unit [в]cee [час]Time]60_километров_в_час. [Поезд будет курсировать между Миланом и Неаполем]Event со [скоростью]Measure [[300]Quantity [километров]Unit [в]cee [час]Time]60_километров_в_час. [Выпивай по стакану чудесного напитка]Event [[два]Quantity [раза]Unit [в]cee [день]Time]60_километров_в_час. А хотите, я скажу вам, как [продавать]Event [[300]Quantity [литров]Unit [в]cee [день]Time]60_километров_в_час? Значит, [прибавка к пенсии]Event не превысит [[88]Quantity [рублей]Unit [в]cee [месяц]Time]60_километров_в_час. Такие красавцы, я думаю, [рождаются]Event [[раз]Quantity [в]cee [сто лет]Time]60_километров_в_час.
COMMENT	В письменной речи обозначается также как 60 км/ч. Обратите внимание, что расход топлива обозначается с помощью другой конструкции (ср. семь литров на 100 километров).
SWECCN	AVBCLFmN9pg0dX7Prwnn

Figura 18: Exemplo de entrada construcional no Constructicon russo (JANDA ET AL., 2018, p. 174)

Diferentemente do alemão e do sueco – que possuem uma *framenet* e um *constructicon* que se estruturam separadamente, mas que podem ser relacionáveis – e do inglês, do PB e do japonês – os quais desenvolveram *constructicons* para suas línguas como projetos anexos às respectivas *framenets* –, o russo, por sua vez, não possui uma *framenet*. Apesar disso, há um projeto correlacionado, o *FrameBank*, que objetiva descrever e analisar características morfossintáticas e semânticas de lexemas individuais ao invés de tratar de frames generalizados, como o faz a FN. De acordo com Lyashevskaya & Kashikin (2015), o *FrameBank* é um dicionário de construções lexicais do russo e um corpus anotado de usos de tais construções com esquema de anotação semelhante ao da FN.

Ainda que tenha como base teórica os princípios da Gramática das Construções, o *FrameBank* é um recurso construcional que dá conta apenas de construções lexicais, mais especificamente de padrões morfossintáticos. Por esse motivo, o *constructicon* do russo – que objetiva tratar, a princípio, de construções polilexêmicas – vai além dos casos descritos no *FrameBank*. No entanto, uma vez que tem por objetivo maior modelar toda a língua russa em termos de construções, para consegui-lo, pretende, como passos futuros do projeto, integrar à sua base os demais dicionários e recursos desenvolvidos para o russo, como *FrameBank*, a fim de incluir

tanto unidades menores que uma palavra, quanto concatenações de construções em unidades do discurso maiores (JANDA ET AL., 2018, p.176).

3.1.5 Constructicon do japonês

O Constructicon do japonês é um projeto em desenvolvimento, que teve início como um anexo à FrameNet japonesa, da mesma maneira como foi feito para o inglês e para o PB, por exemplo, em relação às framenets de suas respectivas línguas, descritos neste trabalho nas seções 3.1.1 e 3.1.6 respectivamente. A FrameNet do japonês (JFN), por sua vez, teve início em 2003, em colaboração com a FN de Berkeley, como um projeto de descrição lexical de aplicação da Semântica de Frames, e tendo como objetivo a descrição da valência de palavras japoneses, baseada na Semântica de Frames e em dados de corpus. Assim, a JFN intenciona descrever o sentido de cada unidade lexical no que tange ao frame evocado por elas e anotar exemplos encontrados em corpus.

Por ser um projeto em desenvolvimento em colaboração com a FrameNet de Berkeley, a JFN tem sua base de dados e sua ferramenta de anotação formulada nos mesmos padrões da FN do inglês, seguindo os mesmos métodos no processo de anotação. Ademais, os frames da JFN foram importados da FN, a partir da metodologia de expansão.

Apesar disso, a JFN, de acordo com Ohara (2013), também intenciona analisar os contrastes existentes entre o japonês e o inglês, ao questionar de que forma a abordagem da Semântica de Frames é suficiente para analisar o léxico do japonês. Outro questionamento está relacionado a de que maneira os frames existentes na FN do inglês são totalmente aplicáveis para caracterizar as unidades lexicais do japonês.

Como mencionado acima, diferentemente do Constructicon do sueco e do russo, o Constructicon do japonês é um projeto anexo à JFN e foi proposto com o intuito de incluir elementos gramaticais que não podem ser tratados pelo viés lexicográfico da JFN, de forma muito semelhante ao Constructicon de Berkeley e ao Constructicon do PB, do qual trataremos na próxima seção. Segundo Ohara (2013), ao realizar anotações de texto corrido na JFN, os pesquisadores perceberam que as práticas de anotação existentes no projeto lexical não eram suficientes para tratar alguns tipos de expressões, como partículas complexas, auxiliares complexos, e

construções de nível oracional (OHARA, 2013, p. 22). Assim, o projeto é uma proposta de aplicação prática das teorias da Semântica de Frames e da Gramática das Construções que seja capaz de tratar expressões que não podem ser anotadas na JFN (OHARA, 2018).

Ohara (2018) descreve cinco maneiras de classificação de construções gramaticais baseadas em frames: (1) construções que evocam frames semânticos, como a construção japonesa *Comparative-inequality*; (2) construções que evocam frames interacionais – por exemplo, a construção *Suspended-clause*; (3) construções que não evocam frames, mas que são interpretáveis composicionalmente, exemplificadas pelas construções *Head-complement*, *Subject-predicate* e *Modifier-head*; (4) construções que, como um todo, não evocam frames, mas que possuem interpretações individuais sob diferentes condições – por exemplo, construções idiomáticas, específicas da língua; e (5) construções que não evocam frames e que omitem constituintes repetitivos específicos de uma dada posição, como a construção *Shared-completion*. Segundo a linguista, adotar a classificação acima é útil e eficiente para separar construções gramaticais, uma vez que tais tipos são mutualmente exclusivos.

Assim, para descrever construções gramaticais em japonês, Ohara (2018) menciona que são utilizados exemplos anotados de corpus, além das cinco maneiras de classificação de construções baseadas em frames apresentadas acima. Isso significa que, paralelamente ao que é feito no Constructicon de Berkeley, o procedimento de análise e descrição de construções na JFN segue o seguinte caminho: primeiramente, identifica-se o conjunto de elementos evocadores da construção (CEEs); em seguida, identificam-se os elementos da construção (CEs); e, finalmente, a interpretação é fornecida.

Tomemos como exemplo a construção *Comparative-inequality*, mencionada acima e exemplificada pela sentença em (6) (OHARA, 2018). Essa construção evoca o frame de *Comparison_inequality* e, por esse motivo, é classificada no primeiro grupo de construções do japonês, isto é, o grupo de construções que evocam frames. Semanticamente, essa construção indica desigualdade entre duas Entidades, as quais são argumentos de um adjetivo pleno. Assim, os CEs são *ENTIDADE1 (ENTITY1)*, *ENTIDADE2 (ENTITY2)* e *CARACTERÍSTICA (FEATURE)*.

(6) [_{Entity1} *kore (no hoo) ga*] [_{Entity2} *are*] [_{CEE} *yori*] [_{Feature} *nagai*]
 This GEN side NOM that than longer
 “*This is longer than that*”.
 “Isso é mais comprido do que aquilo”.

Há, no exemplo (6), a presença dos CEs (*ENTITY1*, *ENTITY2* e *FEATURE*) e também está marcado o CEE (*yori* “do que”), o qual é o elemento que permite perceber que está sendo realizada uma comparação entre duas entidades. A interpretação semântica, por fim, é dada pelo frame, segundo o qual uma *ENTIDADE* (*ENTITY1*) é comparada a um determinado *PADRÃO* (*ENTITY2*) em relação a seus valores para uma dada *CARACTERÍSTICA* (*FEATURE*).

3.1.6 Constructicon do português do Brasil

O Constructicon do português do Brasil teve início em 2010 a partir da análise das construções da família Para Infinitivo (TORRENT, 2009; LAGE, 2013; LAVIOLA, 2015). Estas já haviam sido analisadas e descritas com base nos pressupostos da Gramática das Construções. Assim, quando de sua implementação, o Constructicon do PB seguia as mesmas diretrizes adotadas pelo Constructicon do inglês, ao propor representações computacionais de construções que já haviam sido estudadas anteriormente.

No momento de implementação do Constructicon, Lage (2013) propôs o estabelecimento de critérios para que existisse um princípio claro que definisse quais construções seriam tratadas via Constructicon e quais seriam tratadas via padrão de valência no Lexicon. Os critérios são perguntas de polaridade – com respostas sim ou não – e a proposição da pergunta seguinte é dependente de uma resposta afirmativa ao critério anterior.

O primeiro critério é referente à existência de algum material linguístico lexicalmente especificado na construção em potencial, e, para que esta seja de princípio analisada via Constructicon, a resposta a essa pergunta deve ser negativa. O segundo avalia se o material linguístico lexicalmente especificado na construção em potencial evoca um frame, se não o faz, a construção candidata deve ser analisada

via Constructicon, se sim, passa-se ao terceiro critério. Este último verifica a possibilidade de o material linguístico lexicalmente especificado evocar o mesmo frame em diferentes padrões de valência. Se a resposta a isso for positiva, o frame é evocado pela LU alvo e receberá, portanto, um tratamento lexicográfico, somente se a resposta for negativa, o tratamento da construção em potencial será feito via Constructicon (LAGE, 2018). É importante notar que os critérios propostos por Lage (2013; 2018) são mais refinados e objetivos do que aquele reportado por Lee-Goldman e Petruck (2018), o qual se apresenta muito mais intuitivo e impreciso.

Definido o material linguístico como uma construção, passa-se ao processo de criação da entrada construcional na base de dados. O primeiro passo é a definição da construção e de seus elementos (CEs) com suas especificidades, de uma forma que seja tanto legível por usuários humanos, quanto por máquina, visto que uma das finalidades do Constructicon do PB é alimentar um parser sintático-semântico, o CARMA (MATOS ET AL., 2017), que vem sendo desenvolvido pela FN-Br.

As informações necessárias na descrição de uma construção na base de dados do Constructicon do PB são o nome da construção; uma definição; o nome da entrada construcional – que será o nome interlingual da construção, sempre iniciado por “cxn”, uma abreviação para “construção” –; uma especificação se a construção é abstrata – se apenas estrutura a rede hierárquica de construções – ou não – se licencia construtos na língua –; e uma informação se essa construção está ativa na base de dados que estrutura o CARMA, significando que ela foi revisada e liberada para uso no parser. Ao se inserir os CEs, estes também receberão um nome, uma definição, um nome para a entrada do elemento, sempre iniciado por “ec”, além de informações sobre se o elemento é nuclear ou periférico, ou ainda se pode ser instanciado mais de uma vez.

Diferentemente da maioria dos constructicons em desenvolvimento para outras línguas, o Constructicon do PB compartilha a mesma base de dados com a FN-Br e se estrutura em uma rede de heranças (LAGE, 2018). Isso significa que há uma série de relações e restrições mapeadas entre construções, frames e LUs, “o que interliga de forma sistemática as bases de dados lexical e construcional, permitindo descrições mais abrangentes e completas das construções” (LAGE, 2018, p. 120).

As relações entre construções são estruturadas por meio da relação de Herança. O Constructicon da FN-Br adota tanto a herança múltipla – em que as construções interagem de modo que herdaram de mais de uma construção, somando

as propriedades de cada uma das construções que estão em sua linhagem – e o modo completo de herança, em que a construção filha herda toda a informação constante da construção mãe, sem exceções, sendo, portanto, a construção filha um modo mais específico da construção mãe (LAGE, 2018). Para exemplificar como se processa a relação de herança no Constructicon da FN-Br, Lage (2018) utiliza como exemplo a construção de *Dativo_com_Infinitivo*, amplamente estudada no âmbito do Constructicon da FN-Br (TORRENT, 2009; LAGE, 2013; LAVIOLA & TORRENT, 2012; LAVIOLA, 2015). Essa construção pertencente à família Para Infinitivo apresenta uma dupla interpretação, decorrente de uma diferença de evocação do frame – uma evoca o frame de *Suficiência* e a outra o frame de *Ser_obrigado*. Devido a isso, ela foi dividida em duas construções diferentes, diferenciadas apenas pela contraparte semântica: *Dativo_com_infinitivo_Suficiência* e *Dativo_com_Infinitivo_Obrigaçã*o, conforme demonstram as Figuras 19 e 20, respectivamente.

Dativo_com_infinitivo_Suficiência [Sufficiency_NP]

Definição	
Uma Para_Sinf indica um propósito cuja realização depende de um recurso codificado em um NP_Núcleo . A valência do NP_Núcleo é aumentada de modo a requerer a Para_Sinf em uma estrutura semelhante à de relativização.	
Exemplo(s)	
Elementos da Construção	
NP_Núcleo [Head_NP]	Nome que codifica um recurso empregado para a realização do propósito indicado pela Para_Sinf .
Para_Sinf [Para_Sinf]	Oração que indica o propósito cuja realização é habilitada pelo recurso codificado pelo NP_Núcleo .
Relações	
Evoca	Suficiência
Herda de	Cláusula_relativa_infinitiva_para

Figura 19: Construção de *Dativo_com_Infinitivo_Suficiência* na base de dados do Constructicon da FN-Br

Dativo_com_infinitivo_Obrigaç o [Obligation_NP]

Defini�o
Uma Para_Sinf indica uma obriga�o cuja realiza�o envolve um item codificado em um NP_N�cleo . O item indicado no NP_N�cleo deve, necessariamente, instanciar um EF nuclear que se mapeie a uma fun�o tem�tica n�o-agentiva no frame evocado pelo verbo infinitivo.
Exemplo(s)
Elementos da Constru�o
NP_N�cleo [Head_NP] Nome que codifica um item sobre o qual a a�o codificada pelo verbo da Para_Sinf deve ser realizada.
Para_Sinf [Para_Sinf] Ora�o que indica a obriga�o que deve ser realizada sobre o NP_N�cleo .
Rela�es
Evoca Ser_obrigado
Herda de Cl�usula_relativa_infinitiva_para

Figura 20: Constru o de *Dativo_com_Infinitivo_Obriga o* na base de dados do Constructicon da FN-Br

Embora se diferenciem pela contraparte sem ntica e pela proemin ncia do *NP_N CLEO* na grade argumental do verbo no infinitivo na constru o de *DCI_Obriga o*, as duas vers es da constru o de *Dativo_com_Infinitivo* herdam da mesma constru o – a *Cl usula_relativa_infinitiva_para* – definida na Figura 21. Trata-se de uma constru o abstrata, proposta unicamente para capturar as propriedades compartilhadas pelas duas constru es de DCI e realizar uma conex o entre elas.

Cl usula_relativa_infinitiva_para [Para_infinitival_relative]

Defini�o
Um NP_N�cleo � relativizado por uma Para_Sinf , uma senten�a infinitiva regida pela preposi�o "para".
Exemplo(s)
Elementos da Constru�o
NP_N�cleo [Head_NP] O NP_N�cleo � o antecedente da Para_Sinf .
Para_Sinf [Para_Sinf] A Para_Sinf relativiza o NP_N�cleo .
Rela�es
Herdado por Dativo_com_infinitivo_Obriga�o, Dativo_com_infinitivo_Sufici�ncia
Herda de Adjuntiva_final_infinitiva, Cl�usula_relativa_infinitiva

Figura 21: Constru o *Cl usula_relativa_infinitiva_para* na base de dados do Constructicon da FN-Br

A Figura 22 mostra como a relação de herança entre a construção de *Dativo_com_Infinitivo_Suficiência* e a construção de *Cláusula_relativa_infinitiva_para* é mapeada no *Grapher* da FN-Br. Na figura, tem-se a exemplificação de como a herança se dá entre os CEs das duas construções.

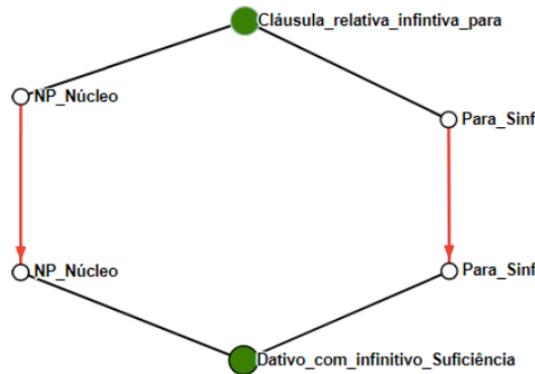


Figura 22: Grafo expandido da relação de Herança entre as construções *Dativo_com_infinitivo_Suficiência* e *Cláusula_relativa_infinitiva_para*

A conexão entre a rede de frames e o Constructicon, por sua vez, é feita através da relação entre construções e frames, marcada pela relação de Evocação, bem como através de restrições – as quais delimitam a presença de uma LU, um frame ou família de frames como preenchedores de um CE (LAGE, 2018). No que tange à relação de Evocação, esta foi definida em Lage (2018) e buscou verificar se o modelo semântico da FrameNet seria adequado à modelagem da contraparte semântica das construções descritas no Constructicon. Segundo Lage (2018), um dos benefícios da proposição de uma relação entre os frames da FN e as construções descritas no Constructicon é o fato de que, pela primeira vez, frames e construções são modelados descritiva e computacionalmente para atuar juntos. Assim, tem-se que uma construção pode evocar um determinado frame, de modo que a contraparte semântica da construção está relacionada a um determinado frame. Lage (2018) explica que, diferentemente da relação de herança entre construções, em que a construção filha recebe todas as informações da construção mãe, a relação de Evocação não implica que a construção deva conter toda a estrutura do frame. A Figuras 23 e 24 exemplificam as construções de *Dativo_com_infinitivo_Suficiência* e *Dativo_com_infinitivo_Obrigaçao*, respectivamente, as quais evocam respectivamente os frames de *Suficiência* e *Obrigaçao* e nelas, pode-se ver,

inclusive, a correspondência entre os CEs e os FEs, a qual é mapeada quando possível.

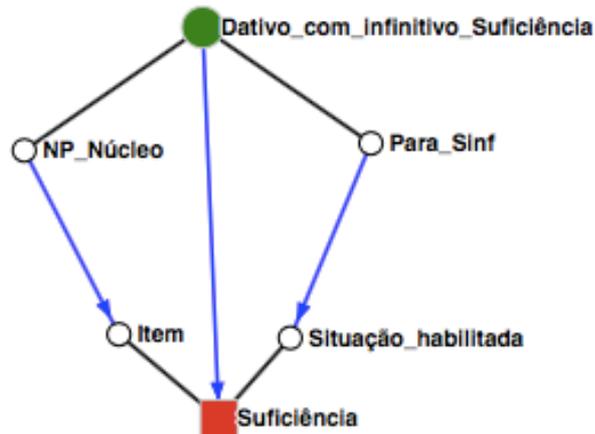


Figura 23: Grafo da relação de Evocação entre a construção de *Dativo_com_infinitivo_Suficiência* e o frame de *Suficiência*

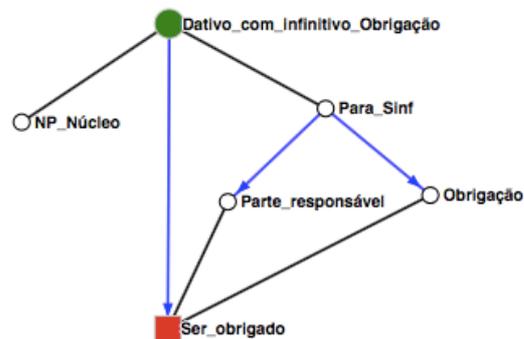


Figura 24: Grafo da relação de Evocação entre a construção de *Dativo_com_infinitivo_Obrigaçao* e o frame de *Ser_obrigado*

Além dessas informações, para que o constructicon seja útil para as tarefas de NLU a que se propõe, as informações precisam ser formalizadas no intuito de tornar possível a leitura delas por máquinas. Com esse objetivo, foi adicionado à base da FN um Editor de Restrições, conforme mostrado na Figura 25, composto pelas seguintes restrições: CE_Construction; CE_Frame; CE_Frame_Family; CE_LU; CE_Lemma; CE_Lexeme; CE_UDRelation; CE_before; CE_meets; CE_Number; CE_SemanticTypeofLU.

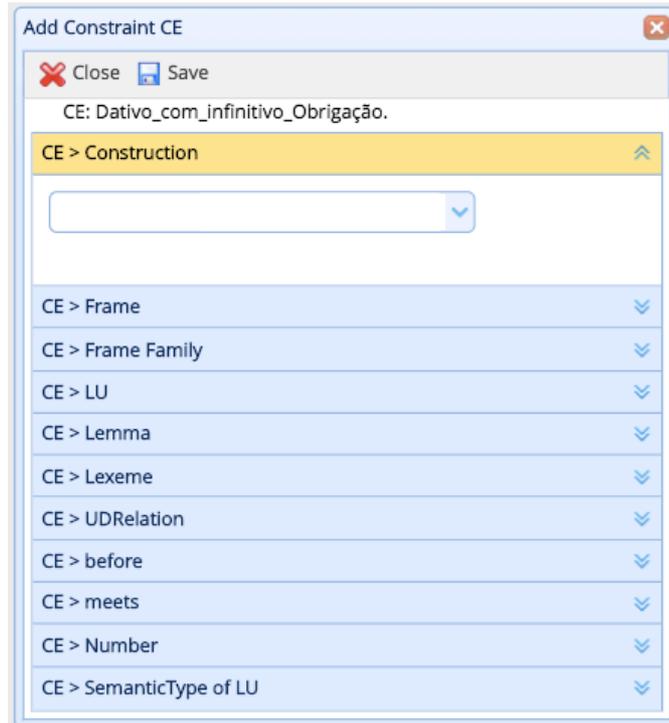


Figura 25: Editor de Construções do Constructicon da FN-Br.

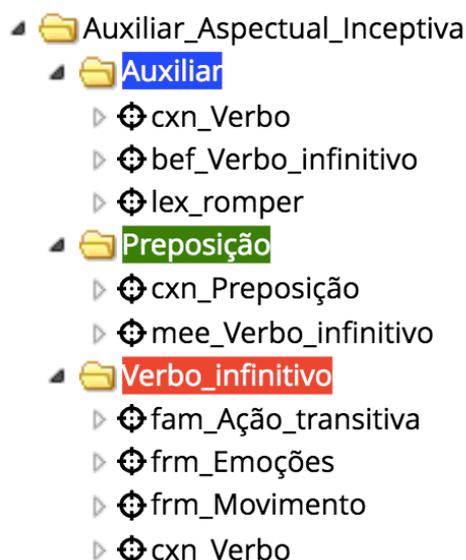
A restrição *CE_Construction* determina se o signo filho é licenciado por outra construção do Constructicon. Em outras palavras, se um CE for composto por uma construção já definida no Constructicon, essa informação poderá ser marcada. Isso possibilita economia ao sistema, na medida em que todas as restrições da construção mãe são automaticamente incorporadas ao CE. Na construção *Auxiliar_aspectual_inceptiva*, instanciada, por exemplo, em casos como *A menina rompeu a chorar*, os CEs *AUXILIAR*, *PREPOSIÇÃO* e *VERBO_INFINITIVO* são, respectivamente, licenciados pelas construções *Verbo*, *Preposição* e *Verbo*, que também se encontram descritas na base de dados da FN. Essa restrição modela aquilo que foi previsto para o Constructicon de Berkeley, mas não foi implementado (conforme seção 3.1.1), ou seja, que um dado CE seja licenciado por uma outra construção já descrita no recurso.

A restrição *CE_before* especifica se um dado CE deve vir antes de outro na sentença. Na construção *Auxiliar_aspectual_inceptiva*, necessariamente, o CE *AUXILIAR* aparece em uma posição anterior ao CE *VERBO_INFINITIVO*, o que torna necessário que se aplique essa restrição a esse elemento. A *CE_meets* indica que um CE deve vir antes de outro e que ambos devem ser adjacentes, ou seja, não pode haver nenhum material interveniente entre eles. Na construção em questão, essa

restrição é aplicada ao CE *PREPOSIÇÃO*, o qual se posiciona anteriormente ao CE *VERBO_INFINITIVO*, sem que possa haver entre eles qualquer outra palavra.

Lage (2018) incorporou outras restrições mais especificamente ligadas a restrições semânticas e de preenchimento das construções: *CE_frame*, *CE_frame_family*, *CE_lexeme*. A primeira indica que o material lexical que instancia um dado CE evoca por si próprio um frame, como o CE *VERBO_INFINITIVO* da construção *Auxiliar_aspectual_inceptiva*, cujos preenchedores evocam os frames de *Emoções* ou *Movimento*. Esse mesmo CE pode exemplificar a restrição *CE_frame_family*, uma vez que verbos que preenchem esse CE também podem evocar uma família de frames – *Ação_transitiva*. Por fim, há a restrição *CE_lexeme*, a qual pode ser exemplificada pelo CE *AUXILIAR*, da construção *Auxiliar_aspectual_inceptiva*, que deve ser preenchido pelo verbo “romper”, visto que essa restrição indica que uma determinada posição na construção deve ser ocupada por um lexema específico, pertencente, em geral, a um grupo semântico definido.

Após serem aplicadas, as restrições são registradas na base de dados e, com isso, permitem tanto que o analista veja as propriedades das construções cadastradas no Constructicon quanto que uma máquina possa processar as correlações expressas por elas. A Figura 26 mostra a construção *Auxiliar_aspectual_inceptiva* com seus CEs e as restrições aplicadas a eles como ela se encontra na base de dados da FN-Br.



de AVMs, como a mostrada na Figura 4, na seção 1.1, é feita de modo mais intuitivo e, com isso, é possível que seja apresentada de uma maneira mais amigável ao usuário, facilitando sua compreensão acerca da construção e das especificidades desta.

3.1.7 Contraste entre Constructicons

Nas seções anteriores, foram apresentados os constructicons existentes que estão em desenvolvimento. Embora todos eles tenham como ponto de partida a iniciativa de Berkeley e se baseiem nos pressupostos da Gramática das Construções, tais recursos apresentam algumas diferenças bastante importantes. Uma vez que este trabalho se propõe a discutir diretrizes para o alinhamento de constructicons, faz-se relevante resumir e discutir as distinções e as aproximações verificadas entre os constructicons acima apresentadas como uma contribuição nesse sentido.

A análise contrastiva será feita tomando o Constructicon do PB como referência – já que este é o recurso em que se insere esta pesquisa. Na Tabela 1, estão dispostas as características do Constructicon do PB – relevantes para a comparação – e nas demais colunas são apresentados os demais constructicons e como se comportam no que diz respeito às características do Constructicon do PB.

Critérios	PB	Inglês	Sueco	Alemão	Russo	Japonês
Desenvolve-se junto a uma FN da língua?	Sim.	Sim.	Não. FN e Constructicon são recursos independentes	Não. FN e Constructicon são recursos independentes	Não há FN para a língua.	Sim.
Há uma conexão entre a base de dados de frames e o Constructicon?	Sim.	Não. As relações são indicadas em prosa, mas não há conexão de banco de dados.	Há links entre algumas das entradas construcionais e os frames descritos na base de dados da FN do Sueco.	Há links entre algumas das entradas construcionais e os frames descritos na base de dados da FN.	Não. A contraparte semântica é dada apenas pela definição da construção.	Há a possibilidade de relacionar frames e construções, quando a construção evoca um determinado frame.
Há mapeamento de relações entre construções?	Sim.	Há a proposta, mas não há a aplicação na base de dados.	Sim.	Não é informado.	Não é informado.	Não é informado.
Apresenta um editor de restrições na base de dados?	Sim.	Não	Não	Não	Não	Não
Quantas entradas construcionais há?	143	74	400	Não é informado.	600	Não é informado.
A estrutura formal é especificada	Sim.	Sim.	Sim.	A representação sintática é	A contraparte sintática é especificada a	Sim.

através de Tipos Sintagmáticos que compõem cada cxn?				feita a partir do parser sintático de forma automática, com informação sobre Tipos Sintagmáticos.	partir das etiquetas das Universal Dependencies .	
Inseriu as construções do Ccn de Berkeley que apresentam contraparte na língua alvo?	Sim.	-	Sim, em sua maioria.	Intenciona inserir as construções do Ccn de Berkeley que apresentam contraparte em alemão	Não.	Não.
Inserir padrões específicos da língua a partir de busca em corpora?	Sim.	Sim.	Sim.	Sim.	Sim.	Sim.

Tabela 1: Contraste entre constructicons

Somente os recursos do PB, do inglês e do japonês são desenvolvidos a partir de uma framenet. Enquanto o sueco e o alemão possuem recursos independentes, produzidos por projetos independentes, o russo não possui uma framenet em desenvolvimento. Esse aspecto influencia o processo de criação de um constructicon multilíngue, uma vez que a hipótese desta tese envolve o uso de representações intermediárias e interlinguais, como as framenets e as Universal Dependencies, para o alinhamento de constructicons. Isso também ocorre ao verificar que nos demais constructicons não há uma representação da relação entre frames e construções, quando uma construção evoca um frame – ainda que nos constructicons do inglês, do sueco e do japonês seja possível estabelecer a relação, esta não é representada na base de dados como é feito no PB.

Outros dois aspectos fundamentais para o constructicon do PB e bastante relevantes para a proposta de alinhamento entre constructicons é a representação de relações e restrições na base de dados. O constructicon do sueco é o único dos recursos, além do constructicon do PB, que representa relações entre construções na sua base de dados. O constructicon do inglês, por sua vez, apresenta a proposta teórica, mas, devido ao encerramento do projeto, esta nunca foi implementada de fato. No que tange ao mapeamento de restrições, essa característica é exclusiva ao constructicon do PB e pode ser uma contribuição para o constructicon multilíngue.

A contraparte formal das construções é feita, pela maioria dos recursos, de forma manual e é especificada através dos tipos sintagmáticos de cada construção. As exceções são os constructicons do alemão e do russo: no alemão, a contraparte

sintática é gerada automaticamente pelo parser sintático contendo informações sobre os Tipos Sintagmáticos; já no constructicon russo, a contraparte sintática é especificada com base nas etiquetas das Universal Dependencies. Essa característica do constructicon russo vai ao encontro do que é proposto neste trabalho. Uma das propostas desta pesquisa envolve aplicar as etiquetas das Universal Dependencies de forma semiautomática às construções registradas no constructicon, visando uma uniformidade na representação sintática de construções por constructicons de línguas diferentes, contribuindo, assim, para o alinhamento de constructicons e o estabelecimento de relações entre construções de línguas distintas.

Outra característica relevante no contraste entre constructicons para a proposta de alinhamento entre eles é a metodologia adotada para a implementação de padrões construcionais em cada recurso. No Constructicon do PB, inicialmente se buscou inserir padrões específicos da língua (encontrados a partir de busca em corpora) que não fossem tratáveis via Lexicon. Em Laviola (2015), ao se realizar a análise comparativa entre padrões registrados no constructicon do inglês com possibilidades em PB, foram adicionadas as construções do constructicon de Berkeley que continham uma contraparte em PB. Além do Constructicon do PB, o único recurso que seguiu a mesma estratégia, pelo menos em certa medida, foi o Constructicon do sueco, o qual, através de um trabalho contrastivo, semelhante ao realizado em Laviola (2015), analisou quais construções do inglês possuem contraparte em sueco (BÄCKSTRÖM ET AL, 2014). O Constructicon do alemão também intenciona realizar uma análise comparativa a partir dos padrões do inglês e inserir aquelas que tiverem contraparte no alemão. Os demais recursos apenas usam a metodologia de realização de busca em corpora de padrões específicos de cada língua, estudo desses padrões e posterior inserção no constructicon.

Nota-se, dessa forma, que há distinções entre as características dos constructicons existentes que são bastante relevantes no tocante à metodologia de representação do Constructicon Multilíngue. Devido a isso, a proposta de utilização das representações intermediárias – framenets e UDs – pode ser uma estratégia de viabilizar o alinhamento desses recursos, já que, mesmo não sendo utilizadas por todos eles, são interlinguais e podem ser viabilizadas, ainda que seja necessário fazer pequenas adaptações nas metodologias próprias de cada constructicon para permitir a sua inclusão.

3.2 CONSTRUCTICOGRAFIA MULTILÍNGUE X GRAMÁTICA DAS CONSTRUÇÕES CONTRASTIVA

Lyngfelt et al. (2018a) explicam que a maioria dos trabalhos em Gramática das Construções é baseada na noção de que construções são específicas de cada língua. Essa ideia pode ser decorrente, em parte, da definição de construções como sendo pareamentos convencionalizados de forma e função e da abordagem orientada por dados. Algumas visões construcionistas são bastante categóricas ao afirmar, por exemplo como o faz Croft (2001), que “construções como configurações de propriedades morfossintáticas válidas entre línguas diversas não existem” (CROFT, 2001, p.283)²⁷. Existem abordagens menos categóricas, como a que é defendida por Goldberg (2013), em que não se elimina a possibilidade de haver pelo menos algumas construções válidas para línguas distintas, mas estas só são assumidas após uma análise detalhada de evidências empíricas sólidas (LYNGFELT ET AL, 2018a, p. 256).

Segundo Croft (2013), na Gramática das Construções Radical (RCG – Radical Construction Grammar), categorias e construções são específicas de cada língua e, por esse motivo, não podem ser usadas para comparar línguas distintas. Com isso, a RCG adota o método distribucional, ao reconhecer que este é “baseado na identificação de construções que funcionam como o contexto para a análise distribucional” (CROFT, 2013, p.)²⁸. Com base nisso, o linguista propõe a tipologia como uma extensão do método distribucional de comparação entre línguas. Assim, o princípio para a comparação de estruturas gramaticais nesse modelo é, portanto, a função de tais estruturas, uma vez que as línguas possuem uma grande variedade estrutural. Nas palavras de Croft²⁹ (2013),

²⁷ constructions as cross-linguistically valid configurations of morphosyntactic properties do not exist

²⁸ is based on the identification of constructions which serve as the contexts for distributional analysis.

²⁹ Croft (2001) apresentava um posicionamento categórico em relação à impossibilidade de comparação de construções de línguas distintas. Em 2013, Croft propõe a tipologia como uma metodologia para a comparação de construções. Em trabalho mais recente, Croft (no prelo) propõe os Comparative Concepts, uma outra proposta de representação intermediária para a comparação de construções de línguas distintas.

tipologistas comparam como as línguas codificam função na forma morfossintática. Assim, a tipologia morfossintática consiste no estudo de pareamentos de forma e função, que são diretamente comparáveis à Gramática das Construções (CROFT, 2013, p.)³⁰

De acordo com Boas (2010), estudos baseados na comparação entre línguas tiveram início na década de 1960, com foco em ensino de língua, mas expandiram para áreas de teoria da tradução, tipologia da linguagem e universais linguísticos. Nas últimas décadas, têm se tornado cada vez mais populares entre semanticistas e lexicógrafos computacionais. Apesar disso, a grande maioria dos trabalhos em Gramática das Construções consiste, ainda, de estudos de caso de construções ou grupos de construções particulares em uma única língua. Lyngfelt et al. (2018a) propõem que isso, no entanto, não elimina a possibilidade de um estudo contrastivo, visto que diferenças entre línguas são um bom motivo para o desenvolvimento de estudos comparativos de construções. Assim, a Gramática das Construções Contrastiva não necessariamente consiste de comparações entre construções correspondentes, podendo o ponto de partida ser um espaço funcional, ou um domínio, ou ainda um frame semântico, que pode ser realizado por diferentes conjuntos de construções em línguas distintas.

Boas (2010) apresenta um estudo realizado em 2002 em que comparou verbos do inglês e do alemão que evocam o frame de *Communication-Statement*. Nesse estudo, o linguista percebeu que o verbo *announce* do inglês permite que diferentes perspectivas de um evento de comunicação sejam expressas, o que é refletido pelas diferentes realizações sintáticas dos FEs. Em alemão, cada uma das realizações sintáticas de *announce* pode ser traduzida por diferentes verbos, cada um com um frame sintático próprio. Resumidamente, com esse estudo comparativo, Boas notou que a escolha de uma construção gramatical específica com o verbo *announce* depende diretamente do significado que se pretende expressar. Assim, usar frames semânticos para descrever diferenças sintáticas permite que se perceba essas especificidades tanto em um estudo monolíngue quanto em um trabalho comparativo

³⁰ typologists compare how languages encode function in morphosyntactic form. Hence morphosyntactic typology is the crosslinguistic study of form-function pairings, which is directly comparable to construction grammar.

entre línguas distintas, o que mostra, de acordo com Boas (2010), a relevância da Semântica de Frames para análises construcionais contrastivas.

Hilpert (2010), por outro lado, contrastou as construções comparativas do inglês com as do sueco nos níveis fonológico, morfológico, sintático, semântico e pragmático e concluiu que, embora haja uma analogia estrutural, as formas em cada uma das construções apresentam uma série de características imprevisíveis em cada um dos níveis gramaticais mencionados. Como exemplo, tem-se que falantes de inglês preferem pronomes de caso objetivo (*taller than me*), enquanto os falantes de sueco optam por usar pronomes nominativos no mesmo contexto sintático. Esse exemplo demonstra que o fato de uma escolha padrão em uma língua ser periférica em outra sugere que essa variável é potencialmente útil como um parâmetro para investigações tipológicas em outras línguas. Assim, ele advoga a favor de estudos contrastivos, sugerindo que análises entre línguas distintas são uma estratégia útil uma vez que podem direcionar atenção a parâmetros que não seriam detectados em uma análise monolíngue (HILPERT, 2010).

Nesse mesmo âmbito, Lyngfelt et al (2018a) afirmam que mesmo estudos primeiramente monolíngues são frequentemente contrastivos em vários níveis, ao contrastar propriedades de uma construção em relação a construções similares em outras línguas e que isso é pouco explorado. Outro campo pouco explorado é o que concerne a estudos de línguas em contato em configurações multilíngues, por exemplo, variedades em L2, traduções, entre outros. Nesses contextos, segundo Lyngfelt et al. (2018a), surgem questionamentos quanto a que medida se está lidando com variedades distintas ou com sistemas linguísticos mais ou menos integrados.

Ainda que, nos estudos contrastivos, raramente se assumam explicitamente uma correspondência construcional entre línguas, alguma noção de correspondência é pressuposta. A pergunta “Quais são as propriedades da construção C nas línguas X, Y e Z?” não presume que a construção C nas línguas X, Y e Z seja idêntica, mas garante uma percepção de C como um fenômeno aplicável em línguas distintas. Em uma afirmativa como “Função F é expressa por uma construção R em línguas X e Y”, também se tem uma noção de que R nas línguas X e Y são construções correspondentes em algum sentido (LYNGFELT ET AL., 2018a).

Apesar dessas pressuposições, nos estudos contrastivos, a noção de correspondência construcional não possui um papel central, visto que o seu propósito em tais estudos é servir como uma base para comparação e estabelecer relações de

correspondência. Na constructicografia multilíngue, por outro lado, assim como na lexicografia, estabelecer relações de correspondência é um dos objetivos centrais de comparações entre línguas. Segundo Lyngfelt et al. (2018a), a postura em relação à correspondência é uma característica chave de qualquer abordagem adotada.

Em lexicografia, a abordagem da FrameNet, que tem sua metodologia aplicada e desenvolvida para diferentes línguas, é bastante relevante ao se tratar de comparações lexicais entre línguas. Uma aplicação multilíngue da FN é geralmente baseada na ideia de que pelo menos alguns frames se aplicam a línguas distintas e que diferenças entre línguas podem ser atribuídas tanto a Unidades Lexicais (LUs) que instanciam os frames, quanto às partes que compõem a estrutura do frame – por exemplo, os FEs.

O desenvolvimento do Dicionário FrameNet Brasil da Copa do Mundo (TORRENT, SALOMÃO ET AL., 2014) – um dicionário eletrônico trilingue inglês, espanhol, português – nos domínios do futebol e do turismo é um exemplo de iniciativa no âmbito da lexicografia multilíngue que utiliza a FN como base. Esse dicionário depende da estrutura da FN para sugerir automaticamente correspondências entre LUs nas três línguas. Nele, os padrões de valência derivados da anotação de sentenças contendo verbos e nomes eventivos são comparados entre línguas como uma forma de prover traduções que melhor se encaixem na LU a ser procurada pelo usuário.

Nota-se, com os trabalhos em lexicografia multilíngue baseados em frames, conforme reforçam Lyngfelt et al. (2018a), que estes são bastante úteis para se comparar lexicas de línguas diversas. Isso ocorre porque alguns frames podem ser muito semelhantes em línguas diferentes, embora isso não se aplique a todos os frames em todas as línguas. Mesmo frames que ocorrem em várias línguas podem apresentar diferenças nas LUs que os evocam no que diz respeito a propriedades morfossintáticas, implicações situacionais, distribuição e frequência.

Lyngfelt et al. (2018a) sugerem que há uma aproximação entre lexicografia multilíngue baseada em frames e a constructicografia multilíngue. O primeiro motivo para tal afirmação relaciona-se ao fato de construções gramaticais poderem evocar frames assim como as palavras o fazem. Assim, é possível que relações entre línguas distintas e entre construções possam ser estabelecidas via frames da mesma forma como ocorre na lexicografia baseada em frames. Essa abordagem é discutida por Bäckström, Lyngfelt & Sköldbberg (2014) e Laviola (2015). Contudo, como,

aparentemente não são todas as construções que evocam frames semânticos, aplicar essa ideia a todas as construções da mesma maneira não é possível.

O segundo motivo que propõem os autores é o fato de as ferramentas adotarem um processo de anotação inspirada construcionalmente, em que

(i) análises de multicamadas mapeiam informação semântica (os elementos de frame) à realização linguística da LU que está sendo analisada (a qual pode incluir Funções Gramaticais e Tipos Sintagmáticos associados aos FEs), e (ii) os padrões de correspondência de forma-sentido derivados das análises representam a valência mínima da construção lexical que está sendo analisada. (LYNGFELT ET AL., 2018a, p.311)³¹

Lyngfelt et al. (2018a) concluem, portanto, que as análises da FN incluem informação construcional que pode ser útil para o desenvolvimento do construcion de uma única língua e, possivelmente, de um construcion multilíngue. Como uma abordagem baseada em frames funcionaria no desenvolvimento de constructicografia multilíngue de fato é o que, em certa medida, este trabalho busca, como um de seus objetivos, averiguar. Acerca disso, mais será tratado nas seções seguintes, antes, porém, apresentaremos a metodologia empreendida nesse trabalho.

³¹ (i) multilayer analyses map semantic information (the Frame Elements – FEs) to the linguistic realization of the Lexical Unit being analyzed (which may include Grammatical Functions and Phrase Types associated to the FEs), and (ii) the meaning-form correspondence patterns derived from the analyses represent the minimal valence of the lexical construction being analyzed (see Fillmore 2013 for a discussion of minimal valences as properties of lexical constructions).

4 METODOLOGIA

Este trabalho centra-se na discussão de uma proposta de alinhamento de constructicons e, mais especificamente, de representação computacional desse alinhamento. Devido a isso, a metodologia empregada nesta pesquisa envolve três etapas: 1) análise contrastiva de construções descritas em recursos existentes; 2) levantamento de possibilidades de alinhamento desses recursos; e 3) implementação da proposta de alinhamento. Tais etapas seguem, de modo geral, as três fases do desenvolvimento de um sistema de NLP apresentadas por Dias-da-Silva et al. (2007) – linguística, representacional e implementacional – e as identificaremos dessa maneira.

A fase linguística, nesta pesquisa, teve como principal função avaliar a viabilidade da proposta de alinhamento sugerida neste trabalho. Para tanto, tomamos como ponto de partida os resultados da pesquisa de mestrado (LAVIOLA, 2015) – que analisou as construções do Constructicon do inglês na busca por correspondentes em PB e verificou que tal comparação era possível – e os resultados semelhantes obtidos com a comparação das mesmas construções do inglês em relação ao sueco (BÄCKSTRÖM ET AL., 2014). A partir disso, propusemos uma quantificação do grau de correspondência de cada construção por meio da definição de critérios de comparação que fossem aplicáveis ao PB, ao sueco e ao inglês, e possíveis de serem replicados para a comparação de outras línguas. Essa primeira etapa, portanto, foi um trabalho cooperativo, que intencionou realizar uma comparação trilingue entre o PB, o sueco e o inglês – esta usada como língua fonte e mediadora entre o PB e o sueco – no âmbito de um projeto de cooperação bilateral em pós-graduação e pesquisa financiado pela CAPES e pelo STINT, no Brasil e na Suécia.

Tal empreendimento foi realizado no decorrer dos anos desta pesquisa de doutorado e teve a colaboração de pesquisadores envolvidos no desenvolvimento do Constructicon do PB e do Constructicon do Sueco. A equipe que se formou para o desenvolvimento dessa fase contou com a doutoranda brasileira, autora desta tese, e com uma doutoranda sueca da Universidade de Gotemburgo – Linnéa Bäckström – e a supervisão de seus respectivos orientadores, Professor Doutor Tiago Torrent e

Professor Doutor Benjamin Lyngfelt³². A equipe contou ainda com o pesquisador da FN-Br Ely Matos e com a lexicógrafa sueca Anna Hannesdóttir. A decisão de empreender um trabalho cooperativo entre os projetos brasileiro e sueco se deveu ao fato de a proposta de um alinhamento entre constructicons de línguas distintas requerer discussões que concernem a aspectos específicos de cada língua.

A comparação envolveu quatro critérios em formas de perguntas de polaridade, que fornecem uma pontuação à construção analisada, que varia de 0 a 3 pontos – em que 3 significa que a construção da língua fonte tem alta correspondência na língua-alvo, e 0 indica não haver nenhuma correspondência. O primeiro critério analisa se existe na língua-alvo uma construção como a que está em análise. Sendo a resposta afirmativa, passa-se para o segundo critério, o qual verifica se a construção comparável da língua-alvo possui a mesma função da construção da língua fonte. Em seguida, o terceiro critério dá conta da contraparte formal. Se a resposta ao critério for afirmativa – significando que a construção da língua-alvo tem forma equivalente à da língua fonte – a análise é encerrada. Caso a resposta seja negativa, passa-se para o quarto critério, o qual verifica se as diferenças formais existentes entre a construção fonte e a construção alvo são decorrentes de aspectos externos, ou seja, de propriedades gramaticais mais genéricas da língua alvo que não encontram correspondência na língua fonte.

Esses critérios foram aplicados a 66 das 73 entradas construcionais cadastradas no Constructicon do inglês, comparando-as a possibilidades em cada uma das línguas-alvo. Cada construção pôde, então, ser classificada com base no nível de correspondência apresentado. Assim, a **fase Linguística**³³ envolveu a definição de critérios para comparação de construções de línguas distintas, de modo a confirmar se tais construções são comparáveis e se, a partir disso, são possíveis de serem alinhadas em um recurso construcional multilíngue. Essa fase, portanto, é de extrema importância para trabalhos em NLU, visto que, nesse viés da Linguística Computacional, é necessário o estudo de um dado fenômeno sob a perspectiva dos

³² Os resultados dessa etapa foram publicados no capítulo 9 do livro *Constructicography: Constructicon Development across Languages* (2018) e aqui está referenciado como Lyngfelt et al. (2018a).

³³ Usaremos a denominação metodológica de Dias-da-Silva et al. (2007), embora o enfoque dado por ele é em tarefas de NLP, enquanto esta pesquisa se aproxima de sistemas de NLU.

linguistas para que se possa seguir para as demais etapas de caráter mais computacional.

Nesse sentido, segue-se para a **fase Representacional**. Nessa etapa, o objetivo desta pesquisa centrou-se em buscar métodos para representar o alinhamento multilíngue, de modo a estabelecer como as informações referentes a cada construção das línguas em comparação seriam mapeadas no recurso multilíngue. Com esse objetivo em mente, a primeira tarefa foi enquadrar na base de dados do Constructicon do PB as construções do Constructicon do inglês identificadas na fase anterior como tendo correspondência em PB. Assim, as tarefas nesta fase envolveram 1) adequação das construções expandidas do inglês às especificidades do Constructicon do PB, no que tange às definições e às representações das construções – visto que o Constructicon do PB focaliza a contraparte formal das construções ao defini-las na base de dados, deixando o aspecto semântico-funcional a cargo da rede de frames, o que é representado por meio de uma relação de evocação de frames; 2) mapeamento de restrições às construções expandidas do inglês.

No que tange às restrições, estas especificam quais elementos podem preencher cada um dos slots de um construto para instanciar uma dada construção. Dessa maneira, tendo restrições mapeadas para todas as construções que se deseja comparar entre línguas distintas, é possível visualizar, de maneira mais automatizada e mais amigável para o usuário, quais construções apresentam correspondência e como essa correspondência está mapeada. Além disso, tornam-se essas especificações legíveis também por máquina. Nesse processo, portanto, o primeiro passo, foi mapear restrições para cada uma das construções cadastradas na base de dados da FN-Br que foram identificadas como sendo comparáveis a partir do estudo empreendido na fase **Linguística** desta pesquisa. Para isso, foram adicionadas à base de dados cinco sentenças-exemplo de cada construção, as quais foram devidamente anotadas para nortear o trabalho e testar sua viabilidade.

Entretanto, a maior parte das restrições incluídas na base de dados tendem a ser dependentes das especificidades de cada língua e, por essa razão, podem não ser confiáveis na estruturação do alinhamento de Constructicons. Portanto, buscou-se uma representação intermediária que fosse aplicável a línguas distintas, e, com isso, foi proposta a inserção das etiquetas das Universal Dependencies na base de dados do Constructicon do PB.

Nesse sentido, adicionamos duas novas camadas à anotação construcional do Constructicon do PB, uma delas contendo as etiquetas de POS (Part-of-Speech) e a outra, as relações sintáticas das UD's. A anotação construcional do PB, portanto, passou a ter, além das camadas para anotar os CEs, os FEs – quando a construção evoca um frame –, o PT e a FG de cada elemento, camadas que apresentam as relações de dependência de cada elemento da construção, conforme se pode ver na Figura 28. Essa tarefa de anotação manual foi realizada apenas com o intuito de verificar a viabilidade de utilização das UD's na comparação de construções de línguas distintas. Para tanto, foram anotadas as cinco sentenças-exemplo que já haviam sido adicionadas à base de dados do Constructicon do PB com as duas novas camadas de anotação.

Layer	Annotation
CE	
CEE	
CstrPT	
UDPos	D ADJ
UDRelation	d root

Figura 28: Anotação construcional na base de dados da FN-Br contendo as camadas das UD's.

Verificada a viabilidade, foi proposta a modelagem de uma nova restrição no Constructicon do PB, com a qual fosse possível mapear restrições de etiquetas das UD's aos CEs. Uma vez que o intuito do projeto das Universal Dependencies é prover etiquetas que sejam universais – ou seja, aplicáveis a várias línguas –, acreditamos que, ao aplicar aos ECs uma restrição que indique que aquele slot do construto deve ser preenchido com um elemento que apresente uma relação sintática identificável também em línguas distintas, será possível indicar o alinhamento entre constructicons.

Finalmente, segue-se para a fase **Implementacional**, a qual intencionou demonstrar a implementação do alinhamento através de um estudo de caso, utilizando as construções de *Dativo_com_infinitivo* do PB e suas contrapartes no inglês e no sueco, aplicando a essas construções nas três línguas as restrições modeladas na base de dados do Constructicon do PB. De maneira esquemática, a Figura 29 apresenta os procedimentos metodológicos empreendidos em cada etapa.

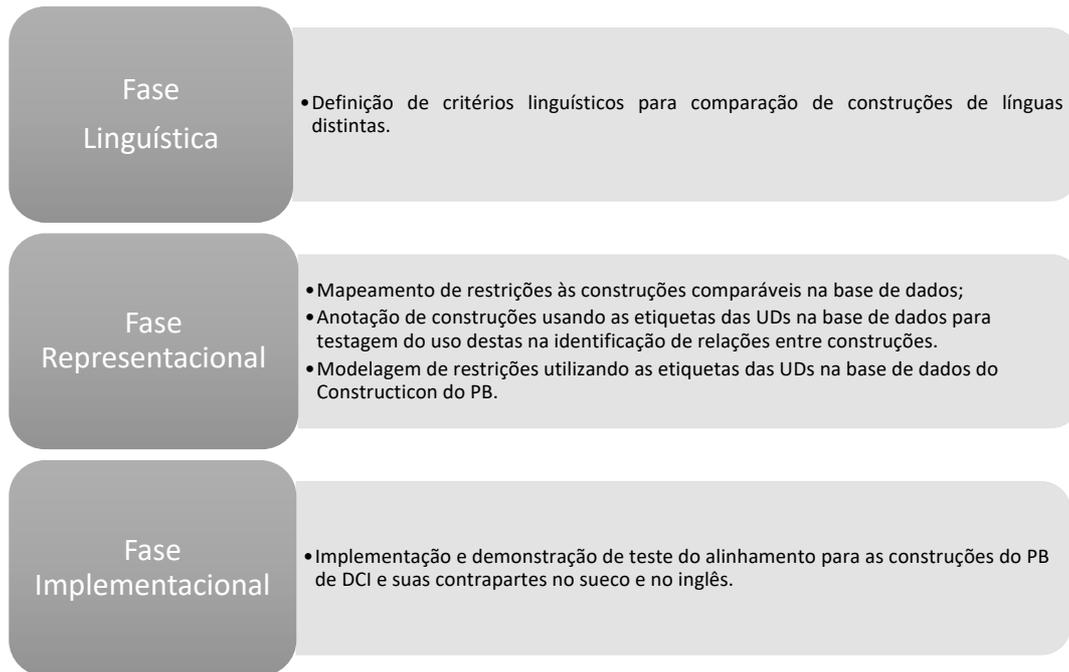


Figura 29: Resumo das etapas de desenvolvimento da pesquisa.

O capítulo 5 apresenta as análises empreendidas em cada etapa metodológica desta pesquisa no tangente à proposta do alinhamento computacional de constructicons.

5 PROPOSTA DE ALINHAMENTO DE CONSTRUCTICONS

Este capítulo apresenta a proposta para o alinhamento de constructicons, partindo do que foi feito em cada uma das etapas descritas no capítulo 4 deste trabalho. Inicialmente, apresentam-se os resultados do trabalho de comparação trilingue, em que as entradas construcionais do constructicon do inglês foram comparadas a possibilidades em sueco e em português do Brasil.

Essa comparação foi proposta a partir de análises bilíngues entre o mesmo conjunto de construções do inglês comparado ao sueco e ao português do Brasil separadamente e com abordagens metodológicas distintas. No trabalho descrito aqui, foi usada uma metodologia única para ambas as línguas, baseada em critérios de comparação, que podem ser aplicáveis a outras línguas. Além disso, conforme ressaltam Lyngfelt et al. (2018a, p. 255), o conjunto de análises desenvolvidas intenciona ser um material inicial para um constructicon multilíngue, bem como um campo de testes para problemas práticos e teóricos envolvidos em tal desenvolvimento.

Em um segundo momento, o trabalho concentrou-se na modelagem de restrições a construções do português do Brasil que, dada a investigação realizada na fase inicial, apresentam correspondência com aquelas definidas pelo Constructicon do Inglês. Em seguida, sentenças-exemplo das mesmas construções foram anotadas com as etiquetas das Universal Dependencies, as quais foram inseridas no processo de anotação do Constructicon do PB a fim de testar se seriam úteis para a representação das especificidades de cada construção. Tendo confirmado a viabilidade do uso das etiquetas das UD's, foram modeladas restrições de UD's na base de dados. Na seção 5.2, será demonstrado o mapeamento de restrições para a construção de *Dativo_com_Infinitivo* do PB. Por fim, será apresentado um estudo de caso feito comparando as versões de *Dativo_com_infinitivo* do PB e do sueco no que tange ao mapeamento de restrições, a fim de demonstrar a proposta do alinhamento de constructicons e suas construções.

5.1 COMPARAÇÃO TRILÍNGUE ENTRE OS CONSTRUCTICONS DO INGLÊS, DO SUECO E DO PORTUGUÊS DO BRASIL

Com o intuito de conectar constructicons de diferentes línguas, o primeiro passo foi o desenvolvimento de uma comparação bilíngue entre as entradas construcionais do constructicon de Berkeley com possibilidades em português do Brasil (LAVIOLA, 2015). Trabalho semelhante foi desenvolvido por Bäckström, Lyngfelt & Sköldberg (2014), comparando as construções do inglês ao sueco. A partir dessas iniciativas, propusemos um trabalho de parceria entre os pesquisadores suecos e brasileiros, o qual empreendeu uma comparação trilingue entre inglês, sueco e português do Brasil, com o intuito de harmonizar as metodologias utilizadas previamente e produzir critérios de análise que fossem aplicáveis para outras línguas (LYNGFELT ET AL, 2018a).

Tanto nos dois trabalhos bilíngues, quanto no de comparação trilingue, o ponto de partida é sempre o constructicon de Berkeley e o objetivo central é analisar se as construções do inglês também ocorrem em sueco e em português e, quando sim, qual o grau de correspondência existente entre elas. Dessa maneira, a comparação tem um caráter unidirecional, por partir sempre do inglês, como língua fonte, em relação às demais línguas, havendo, por esse motivo, uma comparação apenas indireta entre sueco e português.

Para tal comparação, foram consideradas 66 construções dentre as 73 descritas no constructicon de Berkeley. As sete construções não utilizadas na análise foram desconsideradas por apresentarem uma descrição muito incompleta ou, ainda, por serem muito abstratas – ou seja, construções que são definidas independentemente das propriedades específicas da língua – e, portanto, insuficientes como uma base para comparação entre línguas.

Considerando as construções do inglês, o objetivo da análise comparativa reportada em Lyngfelt et al. (2018a) envolveu procurar entre as entradas do constructicon do inglês aquelas que seriam mais ou menos equivalentes nas línguas alvo (sueco e PB) para, a partir disso, discernir as relações entre as entidades da língua fonte e das línguas alvo em termo de significado/função, forma e organização. Para operacionalizar a análise comparativa, foram elaboradas quatro perguntas com respostas sim ou não, as quais seguem uma sequência ordenada, em que a primeira

pergunta bloqueia a segunda em caso de uma resposta negativa, e a quarta só é respondida se a terceira pergunta obtiver resposta negativa.

- 1) Há uma construção, ou um conjunto de construções, correspondente na língua alvo?
- 2) Há uma construção na língua alvo que cobre toda a gama funcional da construção fonte e não é uma construção superordenada?
- 3) A construção fonte e a construção alvo mais próxima são semelhantes formalmente, exceto por diferenças lexicais?
- 4) Todas as diferenças formais são devidas a outras construções que não são do mesmo tipo que o par de construções da língua alvo e da língua fonte? (LYNGFELT ET AL, 2018a, p. 267)³⁴

Por serem perguntas de polaridade, elas permitem um score numérico para cada par de construção comparado. Respostas positivas equivalem a um score 1, enquanto respostas negativas, a um score 0, com exceção da quarta pergunta que, quando respondida afirmativamente, tem um score de 0.5. Assim, o score máximo para cada par de construção comparado é 3, o que indica que tal par tem alto grau de correspondência tanto formal quanto funcional.

A primeira pergunta “**Há uma construção, ou um conjunto de construções, correspondente na língua alvo?**”, em um sentido prático, questiona se há uma construção equivalente na língua alvo em um constructicon multilíngue. Essa pergunta, entretanto, não pressupõe correspondência formal ou funcional completa, mas requer que as construções sejam semelhantes o suficiente para serem consideradas relacionáveis. Uma resposta negativa a essa pergunta torna as demais perguntas irrelevantes e a pontuação final é 0. Um exemplo de construção que não apresenta correspondência em PB é a construção *Subject_auxiliary_inversion*, uma vez que ela não ocorre em PB.

A segunda pergunta “**Há uma construção na língua alvo que cobre toda a gama funcional da construção fonte e não é uma construção superordenada?**” trata de correspondência funcional, mas apenas de forma unidirecional, ou seja, há

³⁴ 1) Is there a corresponding construction, or set of constructions, in the target language? 2) Is there one construction in the target language, that covers the full functional range of the source construction and is not a superordinate construction? 3) Are the source construction and the closest target construction formally similar, except for lexical differences? 4) Do all formal differences follow other constructions not of the same type as the source-target constructional pair?

uma construção na língua alvo que apresenta função igual à construção da língua fonte, mas isso não necessariamente ocorre na direção contrária. Apesar disso, o que se espera com esse critério não é verificar se a construção na língua alvo tem exatamente a mesma função da construção na língua fonte, mas encontrar uma construção na língua alvo que não tenha distinções funcionais claras em relação à construção da língua fonte.

A pergunta três – **“A construção fonte e a construção alvo mais próxima são semelhantes formalmente, exceto por diferenças lexicais?”** – volta-se para a similaridade sintática, mais especificamente nesse contexto, similaridade na estrutura morfossintática, enquanto diferenças puramente lexicais são desconsideradas. A contraparte sueca da construção *Comparison_inequality* é considerada formalmente semelhante, visto que a única diferença é em relação ao conector *than* (em inglês) e *än* (em sueco). Já em PB, essa construção é considerada formalmente diferente, pois difere não somente em relação ao uso do conector *que*, mas também no que tange ao sintagma adjetival comparativo, que em inglês e em sueco pode ser tanto um comparativo morfológico (composto por um sufixo ou uma alteração na forma do adjetivo) e em PB só ocorre com quatro adjetivos, quanto um comparativo perifrástico – mais comum em PB (LYNGFELT ET AL, 2018a, p. 268).

Por fim, a quarta pergunta – **“Todas as diferenças formais são devidas a outras construções que não são do mesmo tipo que o par de construções da língua alvo e da língua fonte?”** – analisa se as diferenças são específicas da construção ou ocorrem a partir de outras construções, ou seja, são diferenças externas. Um exemplo é o fato de o sueco e o PB apresentarem concordância de gênero, enquanto o inglês basicamente não possui gênero gramatical. Tais diferenças afetam um número grande de construções sem necessariamente distinguir propriedades de cada uma delas – isso porque essas características estão associadas a construções muito genéricas e são passadas para construções mais específicas via relação de herança.

No tangente aos resultados encontrados com a comparação trilingue, tem-se que das 66 entradas construcionais do inglês consideradas na análise, 61 possuem correspondência em PB e 62, em sueco. Além disso, 56 apresentam correspondência funcional em PB, enquanto 54, em sueco. Segundo Lyngfelt et al. (2018a), os dados também permitem certa comparação indireta entre PB e sueco, evidenciando que 58 construções do inglês possuem correspondentes em ambas as línguas. Dessas, 52

podem ser deduzidas indiretamente como sendo potencialmente equivalentes em relação à função. Esses números são, contudo, potenciais, visto que são derivados de uma comparação feita via inglês e não direta entre PB e sueco. Apesar disso, conforme afirmam Lyngfelt et al. (2018a), os números relativamente altos de casos possíveis nas duas línguas são bastante promissores para trabalhos futuros de conexão entre os dois recursos.

Em relação aos aspectos formais das construções, em ambas as línguas há mais construções que apresentam diferenças formais – 23 em PB e 31 em sueco. Por outro lado, 18 construções são formalmente semelhantes tanto em PB quanto em sueco, e 20 apresentam diferenças externas em PB, enquanto 13, em sueco. Segundo Lyngfelt et al. (2018), embora os resultados mostrem, surpreendentemente, que PB tenha menos diferenças formais que o sueco, isso não significa que este seja menos semelhante ao inglês, uma vez que tais resultados podem ser influenciados por uma série de fatores – efeitos aleatórios da amostra, o fator humano envolvido nas análises, e talvez as diferentes políticas editoriais adotadas pelos constructicons do PB e do sueco.

Lyngfelt et al. (2018a) apresentam os resultados e discussões³⁵ com base no nível de correspondência encontrado entre construções, os quais são propostos como: alta (com pontuação final de 3 e 2.5); média (com pontuação de 2 e 1.5); e baixa (scores de 1 e 0). Construções com alta correspondência são pares que são equivalentes tanto funcional quanto formalmente (score de 3) ou são aqueles em que as diferenças formais são devido a construções mais genéricas. A Figura 30, retirada de Lyngfelt et al. (2018) apresenta o quantitativo de construções comparadas com score elevado.

³⁵ As discussões e exemplos de construções incluídas em cada caso descrito por Lyngfelt et al. (2018) não serão apresentadas em detalhes nesta tese. O objetivo aqui é mostrar, a partir dessa iniciativa em direção a proposição de um constructicon multilíngue, que os critérios utilizados na comparação unidirecional entre as entradas construcionais do constructicon do inglês e o PB e o sueco podem ser aplicados a outras línguas.

Score	Language	
	Brazilian Portuguese	Swedish
3	17	17
2,5	18	13

Figura 30: Quantidade de construções com alta correspondência por língua (LYNGFELT ET AL., 2018a)

Construções com baixa correspondência, por um lado, são aquelas em que não há uma construção alvo correspondente à construção fonte (pontuação final 0) e, por outro, são aquelas que diferem tanto em forma quanto em função (score 1). Há cinco construções sem correspondência em PB e quatro em sueco, e há seis que diferem tanto em forma e função em sueco, e duas em PB, conforme mostra a Figura 31.

Score	Language	
	Brazilian Portuguese	Swedish
0	5	4
1	2	6

Figura 31: Quantidade de construções com baixa correspondência (LYNGFELT ET AL., 2018a)

Por fim, o grupo de construções com média correspondência inclui construções que se diferenciam ou quanto à função ou quanto à forma. Há nesse grupo, portanto, casos que são formalmente semelhantes, mas funcionalmente diferentes (score 2), casos que são funcionalmente equivalentes, mas possuem diferenças formais específicas da construção (score 2), e casos com diferenças funcionais e formais devido a construções mais genéricas (score 1.5). O quantitativo de construções com correspondência média é apresentado na Figura 32.

Score	Languages	
	Brazilian Portuguese	Swedish
2 (functional difference)	1	2
2 (formal difference)	22	24
1.5	1	0

Figura 32: Quantidade de construções com correspondência média (LYNGFELT ET AL., 2018a)

A partir das análises empreendidas, Lyngfelt et al. (2018a) discutem perspectivas para o alinhamento dos três constructicons e de outros eventualmente. A primeira questão apontada nesse empreendimento é relativa à ligação das construções e dos recursos. Nesse aspecto, os resultados apresentados quanto às análises comparativas encorajam futuras tentativas de alinhar os três recursos, mas também apresentam alguns problemas que deverão ser resolvidos no processo. O principal deles é referente ao fato de as correspondências serem unidirecionais, o que indica que, mesmo havendo ligação do inglês para o PB e do inglês para o sueco, isso não significa que a relação é exatamente a mesma na direção oposta, e muito menos entre o PB e o sueco. Além disso, os casos que não apresentam correspondência ou os casos de construções correspondentes que não possuem correspondência funcional são desafiadores e exigirão um tratamento cuidadoso.

A segunda questão que Lyngfelt et al. (2018a) apontam está relacionada à representação. O principal problema apontado pelos autores é como representar a estrutura das construções de uma forma aplicável entre línguas distintas. Esse questionamento emerge do fato de a comparação apresentada ter sido realizada relacionada a unidades linguísticas e a estrutura linguística, e não à estrutura da base de dados dos três recursos. Assim, a questão central está em como modelar as relações linguísticas distinguidas nas análises. Outro problema é como representar entidades (as entradas construcionais) de uma forma útil, ou seja, questiona-se quais propriedades das construções precisam ser representadas e como elas devem ser formalizadas.

Lyngfelt et al. (2018a) expõem que a decisão sobre quais aspectos da construção devem ser modelados e quais tipos de ligações devem ser propostos depende do objetivo do alinhamento. Se o alinhamento é proposto apenas para compreensão de língua, as construções correspondentes devem ser ligadas de forma menos estrita, permitindo que o sistema gere interpretações semânticas a partir do input. Por outro lado, se a tarefa também envolve geração linguística, como nos casos de tradução por máquina, as correspondências parciais também deveriam envolver regras internas da língua ou depender do modelo estatístico da língua alvo para reordenar as sequências de texto.

Em termos computacionais, um constructicon precisa ser implementado em uma base de dados relacional, em que construções e seus constituintes sejam representados como entidades e as relações entre tais entidades, por sua vez, modeladas. O Constructicon da FN-Br já apresenta essa característica e, por esse motivo, mostra-se um passo à frente de outros constructicons. Lönneker-Rodman (2007 apud LYNGFELT ET AL., 2018a) aponta que há duas escalas em que o alinhamento de recursos deve ser avaliado. A primeira refere-se aos princípios subjacentes que guiam o desenvolvimento do recurso; e a segunda, à possibilidade de conectar não apenas os recursos e entidades em si como um todo, mas também suas partes constituintes. Assim, todas as escolhas relacionadas a como conectar construções entre línguas são dependentes de como os recursos representam construções computacionalmente, em termos de quais aspectos da forma e do sentido das construções serão expressos na representação.

Como se pôde perceber, essa primeira etapa da pesquisa, que envolveu a análise contrastiva trilingue, teve o intuito de verificar se o alinhamento entre constructicons de línguas distintas era um caminho viável de se percorrer e, além disso, pretendeu checar quais possibilidades e desafios ele envolve. Um segundo propósito do trabalho foi desenvolver um formato para comparação que fosse útil para futuros trabalhos nesse campo.

Sendo esse trabalho contrastivo referente à fase Linguística da metodologia proposta nesta tese e, portanto, um ponto de partida para o alinhamento entre constructicons – visto que o método utilizado pode ser aplicável à comparação com outras línguas, bem como com outras construções –, a pesquisa empreendida nesta tese busca dar o primeiro passo na representação do alinhamento computacional de recursos. Sobre essa fase Representacional dedica-se a seção a seguir.

5.2 MAPEAMENTO DE RESTRIÇÕES ÀS CONSTRUÇÕES DO PB

Nesta etapa do desenvolvimento do trabalho, o primeiro passo foi selecionar as construções do Constructicon de Berkeley identificadas como tendo correspondência em português para que fossem cadastradas as restrições sintático-semânticas apresentadas por cada uma delas. O objetivo desta tarefa é, tendo mapeado restrições, criar condições para que o sistema identifique, de forma automática, construções de línguas distintas que sejam pareáveis.

Para tanto, antes de mapear as restrições, foi necessário adequar as definições das construções expandidas do inglês na nossa base de dados. A maioria das construções apresentavam, de acordo com a metodologia do Constructicon de Berkeley, uma definição com foco na semântica da construção. No Constructicon do PB, por outro lado, a definição enfoca a sintaxe da construção e deixa a representação da contraparte semântica sob responsabilidade da rede de frames, através da relação de Evocação e de outras restrições semânticas. Diante dessa divergência entre as metodologias adotadas para a definição de uma construção, o primeiro passo, após a inserção das construções do Constructicon de Berkeley no Constructicon do PB, foi adequar as definições.

Na Figura 33, há a definição da construção *Comparison_inequality* – a versão do inglês da qual foi expandida a construção *Comparação_desigualdade* do PB – juntamente com a definição de seus CEs, conforme consta no Constructicon do inglês. Note-se que os CEs *ITEM*, *STANDARD* e *DIFFERENCE* não deixam evidente qual sua especificação sintagmática na construção, não é possível saber, por exemplo, que os CEs *ITEM* e *STANDARD* devem ser preenchidos por um sintagma nominal, enquanto o CE *DIFFERENCE*, por um sintagma adverbial. A única informação que se consegue a partir da definição desses CEs é acerca das suas funções. Para exemplificar, tomemos a tradução das definições dos CEs *ITEM*, *STANDARD* e *DIFFERENCE*. O primeiro é definido como “O *ITEM* que está sendo comparado”. O segundo é “O *PADRÃO* contra o qual o *ITEM* está sendo comparado”. Por último, o CE *DIFFERENCE* tem por definição “A *DIFERENÇA* em valor entre o *ITEM* e o *PADRÃO*”.

Comparison_inequality NoColor NoTag ColorTag summary

- Inherits from Comparison
 - This construction licenses the creation of a complex comparative adjectival predicator and the realization of the arguments of that predicator. The comparative expression indicates the inequality of two values on a scale.
 - The comparison phrase is made up of a marker (**CD_Marker**, also the CEE) and a noncomparative base expression (**CD_Base_expression**). The "CD" prefix ("construction daughter") indicates that the spans so-labeled are components arranged by the construction into a complex phrase with its own valence. The marker indicates the "direction" of comparison (more than or less than), and may be either *more*, *less*, or *-er*. In case of a suppletive comparative form such as *worse*, both the **CD_Marker** and **CD_Base_expression** labels are applied to the word.
 - The comparison phrase has a valence of the **Item** and **Standard**, the two entities which are compared for the values of particular attributes on a scale. The scale is indicated by the base expression (e.g., *X is taller than Y* compares X and Y on a scale of height). The **Item** is normally the external argument, and is compared against a **Standard**, normally expressed as a complement of the Comparison_phrase.
 - The **Standard** covers all varieties of phrase (*than him*, *than he is*, *than expected*) except where it expresses a particular value, *than six feet*. In that case **Standard_value** is used. The element **Difference** indicates the difference in values between the **Item** and **Standard**. It may be a measurement phrase (*three inches*) or a more vague specification (*much*).
-
- **Approximation**(app): How closely the **Item** comes to the measurement expressed by the comparison phrase. This element is usually expressed with approximations such as *almost* and *nearly*. That tree is [_{app}nearly] taller than mine.
 - **CD_Base_expression**(bas): The non-comparative expression that forms the basis for the comparison phrase, and which indicates the scale on which the Item and Standard (or their values) are being compared. That tree is several feet [_{bas}taller] than this one.
 - **CD_Marker**(mar): The word that marks the type of comparison being done. In some cases individual words may incorporate comparison meaning, and should be so marked (e.g., *better*). That tree is several feet [_{mar}taller] than this one.
 - **Difference**(dif): The difference in value between the Item and Standard. Expressible as a measure phrase (*two inches*) or with an adverb (*much*). That tree is [_{dif}several feet] taller than this one.
 - **Item**(ite): The **Item** is what is being compared. [_{ite}That tree] is several feet taller than this one.
 - **Multiplicative**(mul): A phrase such as *seven times* or *twice* may precede the **CD_Marker**, indicating that the extent to which **Item** differs from the Standard on the indicated scale. The use of a Multiplicative requires the scale to be (or be construed as) quantifiable. It was [_{mul}forty times] brighter than the sun.
 - **Standard**(sta): The Standard is what the Item being compared against. That tree is several feet taller [_{sta}than this one].
 - **Standard_value**(sta_val): The value of the standard, expressed as a complement of the comparison phrase. That building must be taller than [_{sta_val}two thousand feet]!

Figura 33: Definição da construção *Comparison_inequality* conforme consta no Constructicon do inglês

A Figura 34, por sua vez, traz a construção *Comparação_desigualdade* do PB, tendo suas definições focadas na forma, e não na função. Esta é expressa pela relação com os frames da FN-Br, através da relação de Evocação. A construção *Comparação_desigualdade*, portanto, evoca o frame de *Comparação_avaliativa*, descrito na Figura 35. Tal decisão está ligada ao fato de a FN-Br e o Constructicon do PB terem uma base de dados integrada, dessa maneira, a representação semântica fica a cargo dos frames, sem necessidade de isso seja feito novamente pelo constructicon, gerando, portanto, uma economia para o sistema e uma melhoria na capacidade de representação do recurso, visto que, de um lado, as definições sintáticas são fundamentais para que o *parser* gere resultados coerentes e objetivos, e, de outro, as definições semânticas encontradas na base de frames são refinadas e atestadas em *corpora*.

Comparação_desigualdade []

Definição

Um núcleo adjetival ou adverbial, a **Expressão_base**, codifica o aspecto sob o qual um **Sintagma_Núcleo** é comparado em relação a um **Sintagma_Padrão**, o qual pode estar implícito. A comparação é indicada por um **Marcador** morfológico ou sintático (o CEE).

Exemplo(s)

Elementos da Construção

Conjunção_comparativa [conj_comparative]	Conjunção que introduz o Sintagma_padrao .
Expressão_base [bas]	O núcleo adjetival ou adverbial que determina o aspecto sob o qual o Sintagma_Núcleo e o Sintagma_Padrão estão sendo comparados.
Marcador [mar]	O advérbio ou morfema que marca o tipo de comparação que está sendo feita.
SAdv_grau [app]	Um sintagma adverbial modifica o Marcador de modo a indicar o grau em que o Sintagma_Núcleo pode ser comparado ao Sintagma_padrao .
Sintagma_Núcleo [ite]	O sintagma nominal ou verbal nuclear que está sendo comparado.
Sintagma_Padrão [sta]	O Sintagma_Padrão é aquilo a que o Sintagma_Núcleo está sendo comparado.

Relações

Evoca Comparação_avaliativa
Herda de Comparação

Figura 34: Definição da construção *Comparação_desigualdade*

Comparação_avaliativa

Definição

Este frame é sobre a comparação estática de um **Item_perfilado** com um **Item_padrao**, geralmente com relação a algum **Atributo**. Nesse frame, há uma assimetria inerente entre o **Item_perfilado** e o **Item_padrao** em que os dois não podem ser expressos em conjunto como um assunto. Também são encontradas expressões de **Gräu** que indicam o quanto os itens estão próximos uns dos outros na escala evocada pelo **Atributo**. Além disso, o valor específico do padrão ou item em uma determinada escala pode ser dado pelo **Atributo_padrao** do **Atributo_perfilado**, respectivamente.

Exemplo(s)

Elementos de Frame Nucleares

FE Core:

Atributo [Attribute]	Marca os constituintes que indicam em qual relação Item-1 e Item-2 são iguais entre si.
Atributo_padrao [Standard_attribute] excludes: Atributo	É usado quando o padrão é um valor específico em uma escala.
Atributo_perfilado [Profiled_attribute] excludes: Atributo	O Atributo_perfilado é um valor específico em uma escala que está sendo comparada a um Item_padrao ou um Atributo_perfilado .
Item_padrao [Standard_item]	Em sentenças ativas, o Item_padrao é o elemento gramaticalmente menos proeminente.
Item_perfilado [Profiled_item]	Em sentenças ativas, o Item_perfilado é o gramaticalmente mais proeminente dos itens que são comparados e percebidos como o sujeito.

Figura 35: A definição do frame de *Comparação_avaliativa*

Na Figura 36, há um gráfico retirado da base de dados do Constructicon do PB em que é possível visualizar os CEs da construção *Comparação_desigualdade* relacionados aos FEs do frame de *Comparação_avaliativa*.

Relation type: Evoca

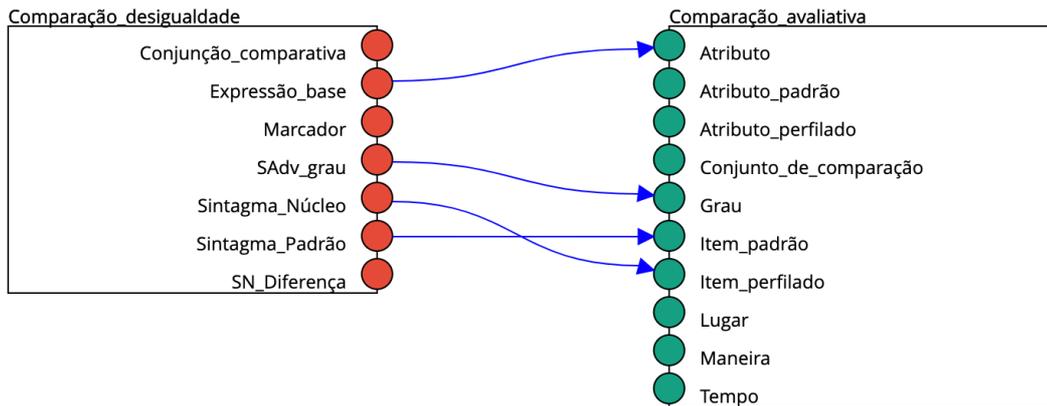


Figura 36: A relação de Evocação aplicada aos ECs e FEs

As adequações acima descritas foram feitas para cada uma das 53 construções do inglês expandidas do Constructicon do inglês para o Constructicon do PB. O passo seguinte foi, então, o mapeamento de restrições.

De modo a nortear o trabalho e basear as análises para o mapeamento de restrições, foram adicionadas para cada uma das 53 construções expandidas do inglês para o PB cinco sentenças-exemplo formuladas exclusivamente para este trabalho. As sentenças-exemplo permitem visualizar as características e particularidades da construção e, por esse motivo, direcionam o mapeamento de restrições, permitindo visualizar se um dado elemento construcional apresenta ou não uma determinada restrição. Como exemplo, vejamos a Figura 36 que contém uma sentença exemplar da construção *Comparação_desigualdade*.

[119465]	AST_MS_APP	NI	Minha	escola	é	sete	vezes	maior	que	a	sua	.
Comparação_desigualdade												
CE			Sintagma_Núcl		SN_Diferen		Expre	Con		Sinta		
Comparação_avaliativa.FE			Item_perfila				Atrib			Item_		
CEE												
CstrPT			NP				AdvP			AdjP		

Figura 37: Sentença-exemplo da construção *Comparação_desigualdade* anotada no Constructicon do PB

No processo de anotação, portanto, conforme visualizado na Figura 37, são anotados, na primeira camada, os CEs encontrados na sentença-exemplo. Na segunda camada, são anotados, de forma automática, os FEs do frame evocado pela construção. É importante lembrar que, no Constructicon do PB, o frame é a representação da contraparte semântica da construção, uma vez que as duas bases de dados estão interligadas. Nesse caso, foi adicionada uma restrição semântica que informa ao sistema que a construção *Comparação_desigualdade* evoca o frame *Comparação_avaliativa*. Em seguida, na anotação, caso haja material lexicalmente especificado que evoque essa construção, será anotado o CEE, o que não ocorre na construção em questão. Na próxima camada, são anotados os Tipos Sintagmáticos do construto.

Tendo a anotação das sentenças como ponto de partida, foram adicionadas manualmente no Constraint Editor – Figura 38 – as restrições de cada elemento de construção. Para tanto, em alguns casos, fez-se necessário inserir novas construções na base de dados, como construções mais genéricas requeridas para mostrar a restrição de Construção (CE>cxn), ou seja, quando o CE em questão é licenciado por outra construção também descrita no Constructicon do PB. Isso aumentou, conseqüentemente, o número de construções consideradas na análise desta pesquisa – das 53 expandidas do inglês, passamos a 91 construções.

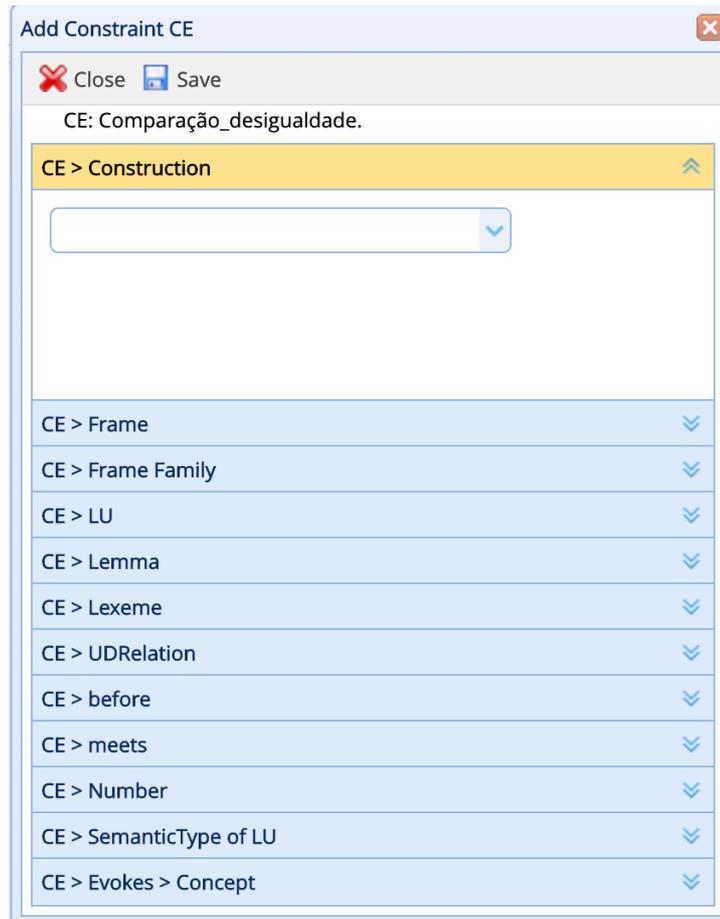


Figura 38: O Editor de Restrições do Constructicon do PB e as restrições constantes na base de dados

No processo de mapeamento de restrições, seguiu-se a ordem em que as restrições são mostradas no Editor, quando aplicáveis. Assim, em geral, a primeira restrição a ser adicionada é a restrição de Construção. Essa restrição só não se aplica quando o CE é muito genérico e pode ser preenchido por qualquer elemento, como é exemplificado na Figura 39, que traz a construção *Coordenação_Em_vez_de* e seus CEs com as restrições aplicadas a eles.

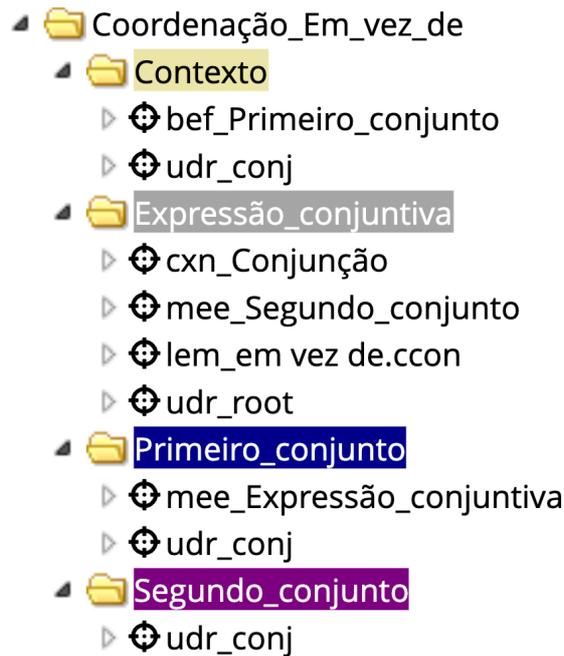


Figura 39: A construção *Coordenação_Em_vez_de* com seus elementos e as respectivas restrições destes no Constructicon do PB

Como se pode ver na Figura 39, o único elemento da construção que apresenta a restrição $CE > cxn$ é o CE *EXPRESSÃO_CONJUNTIVA*. Isso ocorre porque todos os demais não incluem uma categoria específica, podendo ser preenchidos por vários elementos. As sentenças em (7a-c) são licenciadas pela construção *Coordenação_Em_vez_de* e foram anotadas na base de dados do Constructicon do PB. Em (7a), o *PRIMEIRO_CONJUNTO* é um verbo no infinitivo, enquanto o *SEGUNDO_CONJUNTO*, é uma sentença infinitiva. Em (7b), por outro lado, o *PRIMEIRO_CONJUNTO* é uma sentença finita, e o *SEGUNDO_CONJUNTO*, um verbo no infinitivo. Já em (7c), ambos os CEs são ocupados por um nome. Dessa maneira, é difícil afirmar que tais CEs sejam licenciados por uma única construção e então aplicar uma restrição $CE > Cxn$ a eles, já que em cada sentença que exemplifica essa construção eles podem ser licenciados por uma construção diferente.

- (7) a. É melhor **mentir** em vez de **ferir os sentimentos dos outros**.
 b. **Ele dedica tempo demais no trabalho** em vez de **estudar**.
 c. Cada nó contém uma **letra** em vez de **uma palavra**.

Por outro lado, a Figura 40 contém a construção *Comparação_desigualdade* com as restrições devidamente aplicadas a seus CEs. Nesse caso, é possível ver a restrição CE>Cxn aplicada a cada um dos CEs da construção. Como se pode ver, por exemplo, o CE *CONJUNÇÃO_COMPARATIVA* é licenciado pela construção *Conjunção*, mais genérica. Além disso, ele apresenta a restrição *lex_que*, que nos informa que tal *CONJUNÇÃO_COMPARATIVA* é, necessariamente, o lexema “que”. Na Figura 40, é possível conferir as demais restrições aplicadas a cada um dos CEs da construção.

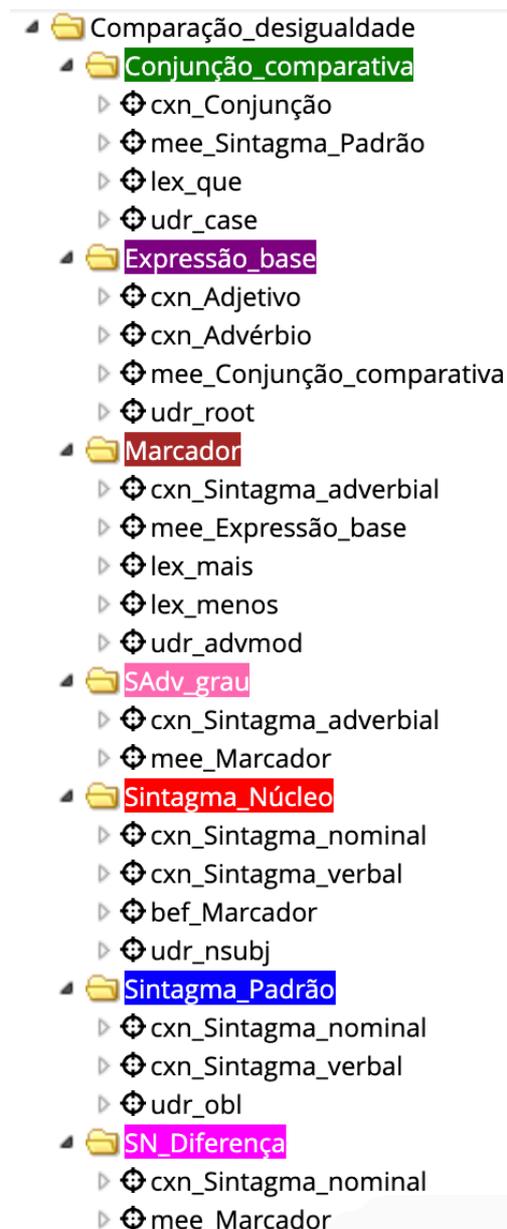


Figura 40: A construção *Comparação_desigualdade* com seus CEs e as respectivas restrições destes.

Como mencionado anteriormente, o mapeamento de restrições foi feito para 91 construções das que estão contidas na base de dados do Constructicon do PB. A Tabela 2 apresenta os quantitativos da modelagem. Nele, é possível ver, na primeira coluna o tipo da restrição, conforme se encontra no Editor de Restrições, e, na segunda coluna, a quantidade de elementos de construção a que cada tipo foi aplicado. Para uma visualização completa de todas as construções e todas as restrições aplicadas a cada um de seus elementos, é possível consultar o Portal da FrameNet Brasil³⁶.

Restrição	QTD.
CE_before	28
CE_construction	178
CE_frame	6
CE_lemma	7
CE_lexeme	39
CE_meets	34
Total Geral	292

Tabela 2: Quantitativos da modelagem de restrições às construções do Constructicon do PB

Com o mapeamento de tais relações na base de dados do constructicon, objetiva-se, portanto, automatizar o reconhecimento de correspondências entre construções de línguas distintas e, conseqüentemente, facilita-se a proposição de um alinhamento entre recursos construcionais. Para tanto, contudo, notamos que apenas as restrições já modeladas antes desta tese poderiam não ser suficientes, visto que podem conter características que são específicas de cada língua. Nesse sentido, propusemos inserir na base de dados as relações das Universal Dependencies. A seção a seguir detalha o trabalho realizado.

5.3 MODELAGEM DE UDS NO CONSTRUCTICON DO PB

No processo de alinhamento construcional, a similaridade entre construções é definida, entre outros aspectos, pelas características dos elementos da construção. Por essa razão, as restrições aplicadas aos CEs são de grande relevância nesse

³⁶ <http://webtool.framenetbr.ufjf.br/index.php/webtool/main>

processo. Especialmente no que tange a esta pesquisa, em que o que se espera é que se identifiquem construções similares entre línguas muito distintas entre si, é fundamental que os CEs sejam definidos com base em características consideradas universais, isto é, que sejam aplicadas igualmente para línguas diferentes. Com isso em mente, buscamos inserir na base de dados as etiquetas das Universal Dependencies. Estas possuem etiquetas para POS (Part-of-speech) – classes de palavras – e relações sintáticas. Embora úteis, as POS são insuficientes para capturar a constituição de construções em línguas distintas. Nesse sentido, foram adicionadas ao Editor de Restrições as etiquetas das UD que identificam relações sintáticas como um fator adicional para caracterizar a estrutura da construção.

A inclusão de UD no Constructicon do PB seguiu o processo que se descreve nas próximas linhas. Inicialmente, as etiquetas de POS e as etiquetas sintáticas foram adicionadas à base de dados do Constructicon do PB. Em seguida, ao anotar cada uma das cinco sentenças-exemplo adicionadas para as construções expandidas do inglês utilizadas nesta pesquisa, foram incluídas duas camadas de anotação – uma referente às etiquetas de POS e outra às relações sintáticas das UD. Isso foi feito a fim de testar se tais etiquetas seriam de fato aplicáveis às construções em análise nesta pesquisa e a fim de direcionar o mapeamento das restrições de relações sintáticas das UD no Editor de Restrições do Constructicon do PB, tornando possível verificar quais relações sintáticas se aplicariam a cada CE. A Figura 41 demonstra a anotação completa de uma sentença-exemplo da construção *Comparação_desigualdade*, contendo as duas camadas de anotação destinadas às etiquetas das UD.

[119465] ■ AST_MS_APP	NI	M	i	n	h	a		e	s	c	o	l	a		é		s	e	t	e		v	e	z	e	s		m	a	i	o	r		q	u	e		a		s	u	a
Comparação_desigualdade		Sintagma_Núc											SN_Diferen				Expres		Con	Sinta																						
Comparação_avaliativa.FE		Item_perfila											Atrib				Item_																									
CEE																																										
CstrPT		NP											AdvP				AdjP																									
UDPos		PRON		NOUN									V	NUM		ADV		ADJ		SCD	PRO																					
UDRelation		nsubj											cadvmod				root		casobl																							

Figura 41: Sentença-exemplo da construção *Comparação_desigualdade* anotada no Constructicon do PB

Como se pode ver na Figura 41, o CE *EXPRESSÃO_BASE* recebe a etiqueta *root*, indicando que ele é o núcleo dessa construção, uma vez que é o elemento que torna possível a comparação. Sendo assim, todos os demais CEs receberão etiquetas de relação sintática em relação ao CE *EXPRESSÃO_BASE*. O CE *SINTAGMA_NÚCLEO*, por exemplo, tem etiqueta *nsubj*, o que indica que é um sujeito nominal; o CE *SN_DIFERENÇA* é um modificador adverbial, por isso, recebe a etiqueta *advmod*; o CE *CONJUNÇÃO_COMPARATIVA* recebe a etiqueta *case*, porque apresenta uma relação de dependência em relação ao elemento que introduz; finalmente, o CE *SINTAGMA_PADRÃO* recebe a etiqueta *obl*, cuja relação é usada para expressões nominais que funcionam como um argumento ou adjunto não-nuclear. Os demais termos incluídos na sentença, como o verbo de ligação “é” recebem etiquetas UD, visto que as UD são aplicadas a todos os termos da sentença, porém ele não é relevante para a descrição da construção.

Após a anotação das sentenças-exemplo, foi feito o mapeamento das restrições de CE>UDRelation a cada uma das construções em análise. A anotação se tornou muito importante nesse processo, pois foi possível verificar CEs que podem exercer diferentes relações sintáticas a depender do contexto da frase em que se encontram. Isso fez com que fosse necessário incluir mais de uma restrição CE>UDRelation para tal elemento, como se pode ver na Figura 42, que apresenta a construção *Predicativa_nominal_recíproca*, seus CEs e as restrições aplicadas a eles.

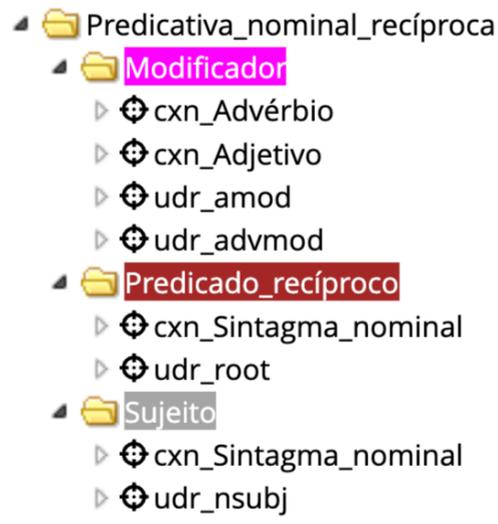


Figura 42: A construção *Predicativa_nominal_recíproca* com seus CEs e as respectivas restrições destes na base de dados do Constructicon do PB.

Como se pode ver na Figura 42, o CE *MODIFICADOR* pode ter uma relação sintática de modificação adjetival com o núcleo nominal, marcado com a restrição *udr_root*, como também pode ter uma relação de modificação adverbial com o núcleo. Na Figura 43, temos uma sentença-exemplo anotada da construção *Predicativa_nominal_recíproca* em que o *MODIFICADOR* é um modificador adjetival. Enquanto isso, a Figura 44 mostra uma sentença anotada tendo o CE *MODIFICADOR* como um modificador adverbial.

[119448]	AST_MS_APP	NI	Júlia e Ana são amigas íntimas?
Predicativa_nominal_recíproca			
CE		Sujeito	Predic Modific
CEE		CEE	
CstrPT		NP	AdjP
UDPos		PRON C PRO AUX NOUN ADJ	
UDRelation		nsubj aux root amod	

Figura 43: Sentença-exemplo da construção *Predicativa_nominal_recíproca* com modificador adjetival anotada no Constructicon do PB

[119445]	AST_MS_APP	NI	Nossas filhas são amigas desde a escola.
Predicativa_nominal_recíproca			
CE		Sujeito	Predic Modificador
CEE		CEE	
CstrPT		NP	AdjP
UDPos		PRON NOUN AUX ADJ SCONJ D NOUN	
UDRelation		nsubj aux root advmod	

Figura 44: Sentença-exemplo da construção *Predicativa_nominal_recíproca* com modificador adverbial anotada no Constructicon do PB

Com o mapeamento de restrições e de modelagem de UDs às construções do Constructicon do PB, é possível obter análises mais coerentes e objetivas, facilitando o processo de alinhamento de construções de línguas distintas. A Tabela 3 apresenta todas as restrições adicionadas à base de dados da FrameNet Brasil no âmbito desta tese, incluindo, agora, as restrições das relações das UDs, e o quantitativo de a quantos CEs tais restrições foram mapeadas.

Restrições	QTD.
CE_before	28
CE_construction	178
CE_frame	6
CE_lemma	7
CE_lexeme	39
CE_meets	34
CE_UD-relation	173
Total Geral	465

Tabela 3: Quantitativos da modelagem de restrições às construções do Constructicon do PB, incluindo as relações de UD

A seção a seguir apresenta uma comparação trilingue, tendo como exemplo e ponto de partida uma proposta de alinhamento das construções de *DCI* do PB, do inglês e do sueco.

5.4 COMPARAÇÃO PB, INGLÊS E SUECO: ALINHAMENTO DAS CONSTRUÇÕES DE DCI

A fim de demonstrar a proposta de alinhamento de constructicons a partir da metodologia aplicada nesta pesquisa, tomamos as construções de *Dativo_com_infinitivo (DCI)* – amplamente estudadas no âmbito do Constructicon do PB – e propusemos uma análise comparativa entre elas e suas equivalentes no sueco e no inglês.

Na versão do PB, as construções de DCI são diferenciadas pela contraparte semântica e, por essa razão, no Constructicon do PB, são representadas por duas construções distintas: a *Dativo_com_infinitivo_Obrigaçao (DCI_Obrigaçao)* e a *Dativo_com_infinitivo_Suficiencia (DCI_Suficiencia)*, filhas da mesma construção, a *Cláusula_relativa_infinitiva_para*.

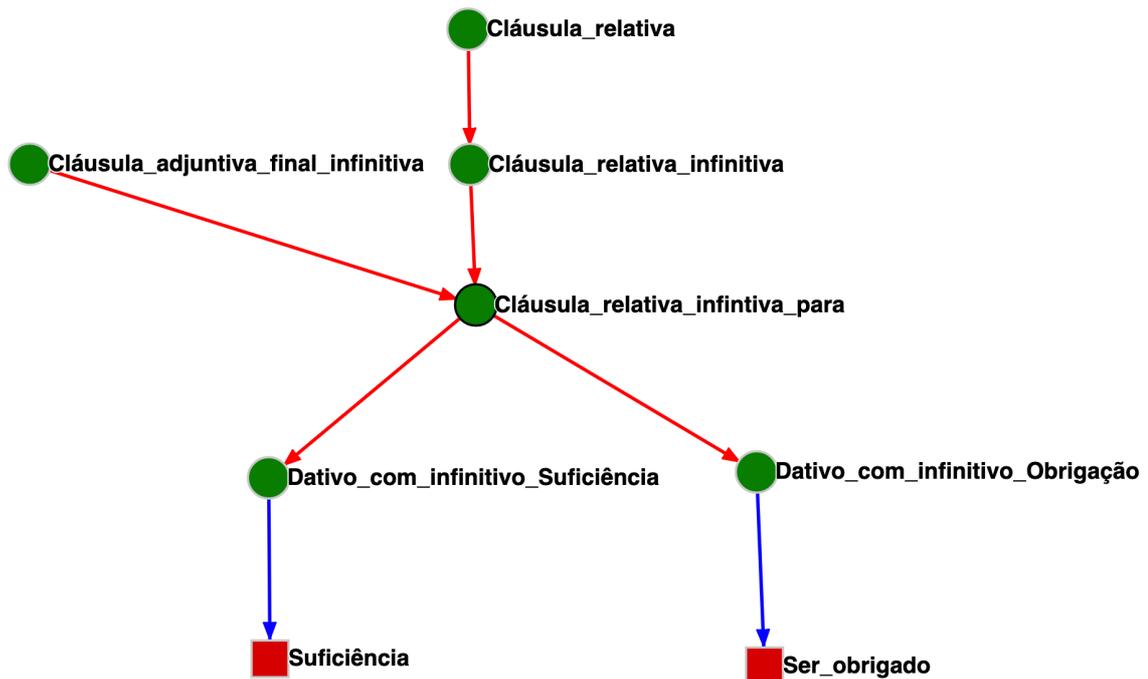


Figura 45: As construções de *DCI_Suficiência* e *DCI_Obrigaçãõ* com as devidas relações de Herança e de Evocação como consta no Constructicon do PB

Na Figura 45, há as relações de Herança_construcional, indicadas pelas setas vermelhas, e de Evocação, indicadas pelas setas azuis, conforme apresentadas no *Grapher* do Constructicon do PB. As relações de herança indicam que tais construções herdam características estruturais de outras construções, sendo versões mais específicas da construção mãe. A construção *Cláusula_relativa*, por exemplo, é uma construção genérica que dá origem a várias outras construções relativas do PB, como a *Cláusula_relativa_infinitiva*, que é uma construção com um aspecto um pouco mais específico que a construção da qual herdou. Por sua vez, ela tem como filha a *Cláusula_relativa_infinitiva_para*, que adiciona uma especificidade às características herdadas da construção mãe: o fato de ser uma relativa introduzida pela preposição *para*, diferentemente de outras cláusulas relativas infinitivas do PB, as quais podem ser introduzidas pela preposição *de*, por exemplo, em expressões como *poder de decidir*. Por fim, da *Cláusula_relativa_infinitiva_para* herdam duas construções – a *DCI_Suficiência* e a *DCI_Obrigaçãõ*. Ambas apresentam dois CEs, o primeiro deles sendo o *NP_NÚCLEO* – um sintagma nominal que ocupa a posição de núcleo – e o segundo, o *PARA_SINF* – uma sentença infinitiva introduzida pela preposição *para*.

Como se pode notar, portanto, no que concerne à natureza sintagmática de seus constituintes, tais construções seriam equivalentes e não haveria a necessidade de separá-las em duas outras construções. O que nos levou à distinção de tais construções foi encontrar a mesma estrutura (pelo menos, à primeira vista) evocando frames distintos. Assim, formalizamos no Constructicon do PB as duas construções, uma evocando o frame de *Suficiência* e a outra evocando o frame de *Ser_obrigado*, como indicam as setas azuis na Figura 45.

Embora sintagmaticamente sejam semelhantes, ao se mapearem restrições, é possível perceber que há também distinção formal entre elas. Sendo assim, em ambas as construções, o *NP_NÚCLEO* apresenta a restrição *cxn_Sintagma_nominal*, o que indica que este elemento é licenciado pela construção *Sintagma_nominal*, e o *PARA_SINF* possui a restrição *cxn_Sentença_infinitiva_preposicionada*, indicando que tal CE é licenciado pela construção *Sentença_infinitiva_preposicionada*. No entanto, diferenciam-se no que tange à restrição *CE>before* que é apresentada pelo CE *NP_NÚCLEO* da *DCI_Obrigaçao* e não pelo *NP_NÚCLEO* da *DCI_Suficiência*. Isso significa que, na *DCI_Obrigaçao*, é necessário que o *NP_NÚCLEO* esteja sempre posicionado anteriormente ao CE *PARA_INF*, em outras palavras, o núcleo nominal deve estar antes da sentença infinitiva preposicionada, para que o sentido evidenciado no frame evocado seja mantido, conforme se nota em (8a-b). Tal obrigatoriedade não é observada na *DCI_Suficiência*, visto que é possível que haja inversão entre os dois elementos na construção sem que esta sofra alteração em seu significado ou em sua função, o que se pode observar em (9a-b). Tal diferença é produto do fato de que, geralmente, na *DCI_Suficiência*, o *NP_NÚCLEO* codifica um elemento que, em relação ao verbo infinitivo, exerce função adjuntiva, enquanto que, na *DCI_Obrigaçao*, tal elemento relaciona-se com o verbo infinitivo em função argumental.

- (8) a. Eu tenho louça para lavar.
 b. *Para lavar, eu tenho louça.

- (9) a. Eu tenho dinheiro para viajar.
 b. Para viajar, eu tenho dinheiro.

No que diz respeito às restrições de UD, o elemento *NP_NÚCLEO* é marcado com a restrição *udr_root*, uma vez que ele é o núcleo da construção de relativização e a ele se relacionam os demais elementos. Por sua vez, o elemento *PARA_SINF* – uma oração relativa infinitiva preposicionada – recebe a restrição *udr_acl*, uma vez que essa relação é utilizada para orações finitas ou não-finitas que modificam um elemento nominal. No caso da *DCI_Suficiência*, o núcleo nominal está sendo modificado pela oração infinitiva preposicionada.

Uma das vantagens de se utilizar as UDs nessa função deriva do fato de estas serem definidas de maneira bastante genérica, uma vez que foram pensadas para dar conta da análise de qualquer língua. Como consequência, o fato de a cláusula introduzida por *para*, no caso das construções de *DCI*, ter um comportamento de oração relativa e um aspecto de oração adverbial final não é problema na escolha da etiqueta de UD a ser aplicada, uma vez que, para ambos os casos, tal etiqueta seria *udr_acl*.

Após a aplicação das restrições, as construções de *DCI_Suficiência* e *DCI_Obrigaçã* podem ser visualizadas na base de dados do Constructicon conforme mostra a Figura 46.

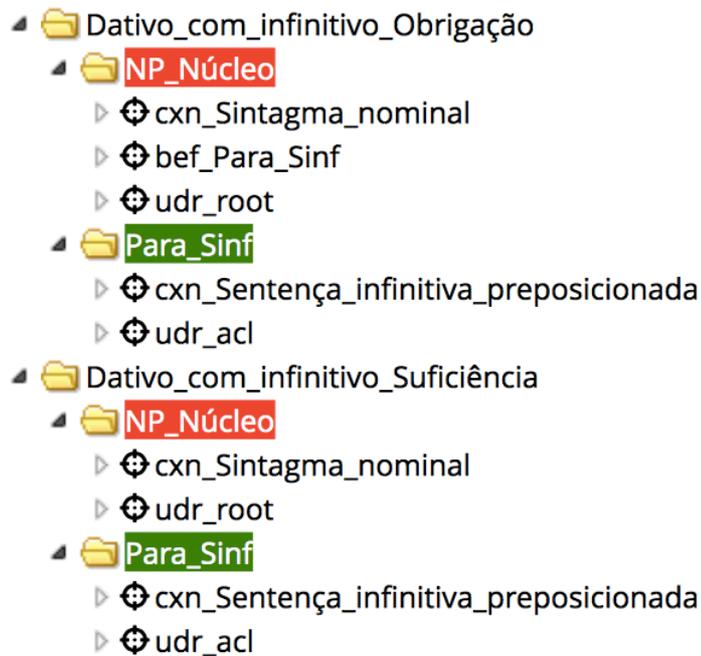


Figura 46: As construções de *Dativo_com_infinitivo_Obrigaçã* e *Dativo_com_infinitivo_Suficiênciã* e suas respectivas restrições na base de dados do Constructicon do PB.

Embora a construção *Cláusula_relativa_infinitiva_para*, no Constructicon do PB, tenha dado origem a duas construções diferenciadas pela contraparte semântica, em inglês um estudo voltado para a identificação de uma construção correspondente à *DCI_Obrigaçao* não foi realizado até o momento. Entretanto, a *DCI_Suficiência* apresenta correspondência à construção *Infinitival_Relative_modal (IRM)* do inglês. De acordo com os critérios abordados na seção anterior, a *IRM* apresenta score 3 em relação ao PB.

Essa construção foi estudada por Fillmore (2008) o qual a denominou construção *Wherewithal*, que, em português, significa “recursos”. O nome escolhido relaciona-se à semântica da construção, uma vez que, assim como na construção de *DCI_Suficiência*, é prevista a presença (ou ausência) de um recurso – realizado pelo elemento nominal – que habilita a realização de uma atividade subsequente. Tal atividade é expressa pela oração infinitiva introduzida, em inglês, por *to*. As sentenças em (10a) e (10b) exemplificam essa explicação. Na sentença (10a), o dinheiro (*the money*) é o elemento nominal nuclear, semanticamente identificado como o recurso que permite que se alcance o objetivo expresso pela oração infinitiva (sair de férias – *to take a vacation*). Em (10b), o verbo faltar (*lack*) indica que o recurso expresso pelo nome funcionários (*staff*), necessário para que o objetivo contido na oração infinitiva (assumir tal trabalho – *to take on such a job*) seja alcançado, está em falta.

(10) a. *I have the money to take a vacation.*

Eu tenho o dinheiro para sair de férias.

b. *We lack the staff to take on such a job.*

Nos faltam os funcionários para assumir tal trabalho.

A construção *Wherewithal* foi então inserida no Constructicon de Berkeley sob o nome de *Infinitival_relative_modal (IRM)*, como se pode ver na Figura 47. A definição contida na entrada construcional da *IRM* indica, inicialmente, que ela herda de uma construção mais genérica, a *Relative_clause* – relação que é

possível visualizar na Figura 48³⁷. A definição apresenta, ainda, os dois CEs que compõem a *IRM* e que são semelhantes aos descritos para a *DCI_Suficiência*. Há, portanto, o CE *HEAD* (que corresponde ao *NP_NÚCLEO* na *DCI_Suficiência*) e o CE *INFINITIVAL_RELATIVE* (correspondente ao *PARA_SINF* na versão do PB).

Infinitival_relative_modal

Inherits *Relative_clause*.

- A noun phrase, the Head, is modified by a *to*-marked non-finitve VP, the *Infinitival_relative*.

ex.: We finally found [_{head}a book] [_{inf}to read to our daughter]. TRANSLATIONS 1 2

- **Head(head):** The Head is a noun phrase modified by the relative clause.

ex.: We finally found [_{head}a book] to read to our daughter. TRANSLATIONS 1 2

- **Infinitival_relative(inf):** The *Infinitival_relative* is a *to*-marked infinitival clause that modifies the Head, and which indicates some variety of modality.

ex.: We finally found a book [_{inf}to read to our daughter]. TRANSLATIONS 1 2

Figura 47: A construção *Infinitival_relative_modal* conforme descrita no Constructicon de Berkeley

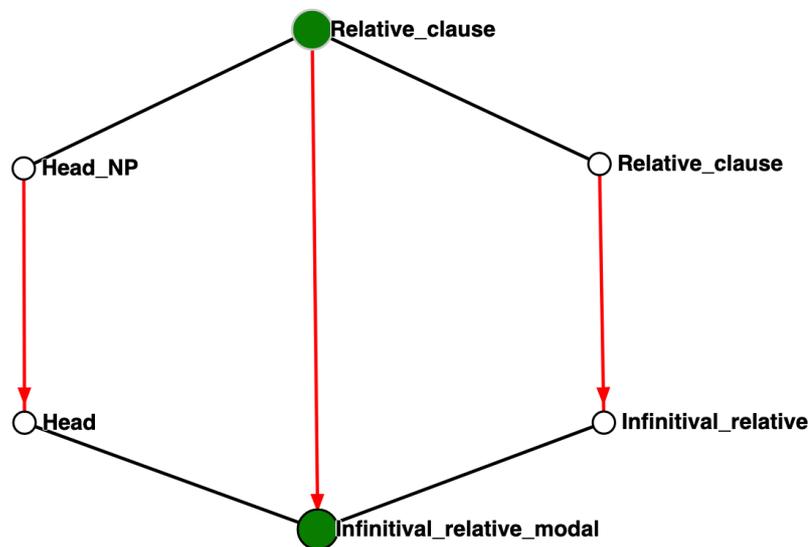


Figura 48: A relação de herança entre a construção *Infinitival_relative_modal* e a construção *Relative_clause*

³⁷ No Constructicon do inglês, essa relação não está modelada. Como se vê na Figura 47, a informação aparece na descrição da construção *Infinitival_relative_modal*, porém, a construção *Relative_clause* não está sequer descrita na base de dados do inglês. Para este trabalho, inserimos a construção *Relative_clause*, com seus respectivos CEs, na base de dados da FrameNet Brasil, e mapeamos a relação entre as duas construções, representada na Figura 48.

Da mesma forma como foi feito para as construções de *DCI* do PB, ao inserir a construção *Infinitival_relative_modal* em nossa base de dados, mapeamos restrições também aos seus CEs. Isso foi possível uma vez que o aparato de software da FrameNet Brasil conta com uma funcionalidade, desenvolvida pelo pesquisador Ely Matos durante estágio pós-doutoral realizado no âmbito da cooperação Brasil-Suécia de que esta tese também faz parte, que permite a inserção de construções de línguas diversas.

Na Figura 49, portanto, estão os dois CEs, *HEAD* e *INFINITIVAL_RELATIVE*, seguidos cada um pelas restrições cabíveis a eles. O CE *HEAD* é licenciado pela construção *Noun_Phrase*, construção equivalente à *Sintagma_nominal* que licencia o CE *NP_NÚCLEO* nas construções de *DCI*. Também de maneira equivalente, deve vir em posição anterior ao CE *INFINITIVAL_RELATIVE*, e recebe a mesma etiqueta das relações de UD's que as construções do PB, *udr_root*, uma vez que as etiquetas UD são feitas para aplicação a qualquer idioma. O segundo elemento, por sua vez, é licenciado pela construção *Infinitival_sentence* e tem por relação sintática das UD's, a etiqueta *udr_acl*, já que exerce a função de cláusula adjetival, modificando o sintagma nominal núcleo. Nota-se que, enquanto o CE *PARA_SINF* das construções de *DCI* do PB é licenciado pela construção *Sentença_infinitiva_preposicionada*, visto que a sentença infinitiva é introduzida pela preposição *para*, em inglês, ela é introduzida por *to* que, nesse caso, é um marcador de infinitivo, e não uma preposição.

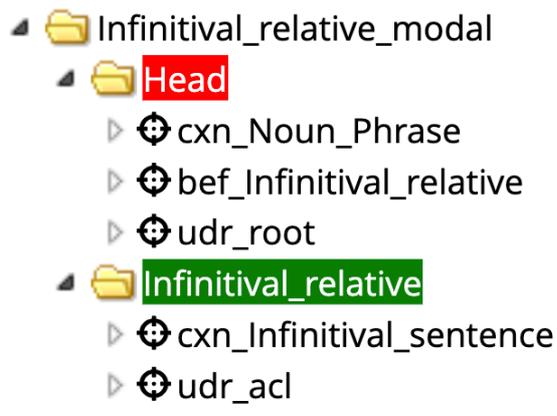


Figura 49: Restrições mapeadas à construção *Infinitival_relative_modal*

A versão sueca da construção *Infinitival_relative_modal* do inglês – correspondente à *Cláusula_relativa_infinitiva_para* do PB – é a construção *attributivinfinitive*. Esta é uma construção genérica, da qual outras quatro são herdadas. O grafo na Figura 50 apresenta as construções do sueco e as relações entre elas. Tal grafo foi gerado a partir de análises produzidas em conjunto com os pesquisadores suecos³⁸ acerca da possibilidade de correspondência entre as construções de *DCI* do PB, a *IRM* do inglês e a contraparte do sueco. Com base nas discussões, chegou-se às relações expostas na Figura 48, ainda não implementadas no Constructicon do sueco até o momento³⁹.

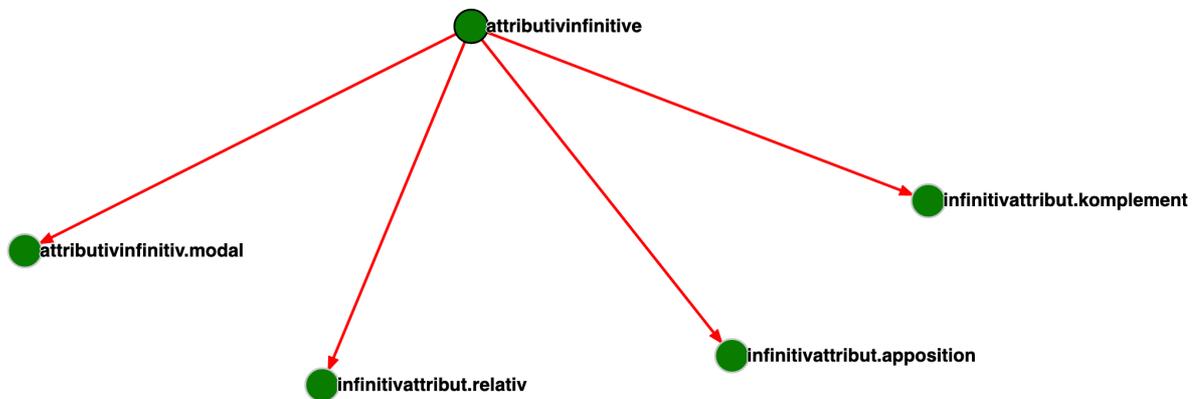


Figura 50: As construções do sueco e a relação de herança entre elas e a construção *attributivinfinitive*

A primeira delas é a *attributivinfinitiv.modal*, a qual apresenta a semântica de suficiência, de forma semelhante à *DCI_Suficiência* do PB. A sentença em (11) exemplifica tal construção. Nela, “*lust eller ork*” (nem a inclinação nem a energia) é o núcleo nominal que habilita a realização da ação em “*att göra dett*” (para fazer isso).

³⁸ Essas discussões foram realizadas durante o período de doutorado sanduíche na Universidade de Gotemburgo sob supervisão do professor Benjamin Lyngfelt, também responsável pelo desenvolvimento do constructicon do sueco.

³⁹ Novamente, como o mapeamento e a representação das relações entre as construções do sueco ainda não haviam sido realizados no Constructicon do sueco, inserimos as construções na base da FrameNet Brasil – que permite a inserção de construções de línguas variadas – e fizemos o mapeamento das relações e restrições de modo a viabilizar o alinhamento de tais construções.

- (11) Jag har varken [[lust eller ork] [att göra det]].
 ‘I have neither the inclination nor the energy to do it’
 ‘Eu não tenho nem a inclinação nem a energia para fazer isso’.

A Figura 51 mostra a construção *attributivinfinitiv.modal* com seus CEs e as restrições aplicadas a eles, conforme foi feito para o PB e para o inglês.



Figura 51: Restrições mapeadas à construção *attributivinfinitiv.modal*

Como se pode ver na Figura 51, as restrições aplicadas ao CE *HEAD* são equivalentes àquelas aplicadas aos CEs na mesma posição nas construções de *DCI_Suficiência* e *Infinitival_relative_modal*, visto que em todos os casos esse CE é licenciado por um sintagma nominal. Entretanto, no que diz respeito ao CE *VP_INF*, nota-se uma diferença quanto à restrição que indica a construção que licencia esse CE. Enquanto, nas construções do PB, esse CE é licenciado pela construção *Sentença_infinitiva_preposicionada*, na construção do inglês e nas construções do sueco, ele é licenciado pela construção *Infinitival_sentence* e *infinitivfras* respectivamente. Isso se deve ao fato de que, nas construções do inglês e do sueco, os elementos que introduzem a sentença infinitiva – “to” e “att” – não são preposições, e, sim, marcadores de infinitivo, diferentemente do que se passa com as construções do PB que apresentam a preposição “para” introduzindo a sentença infinitiva.

Além da construção modal, há a *infinitivattribut relativ* – que abarca os casos que evocam o frame de Obrigação. Segundo a definição proposta

pelo professor Benjamin Lyngfelt durante as análises empreendidas acerca da construção em questão, a *infinitivattribut.relativ* indica a possibilidade, a necessidade ou a obrigação em realizar uma ação, descrita pelo sintagma verbal infinitivo, envolvendo a entidade referenciada pelo núcleo nominal. Este corresponde a um *gap* dentro do sintagma infinitivo, o qual pode ser preenchido pelo núcleo nominal. O exemplo em (12) foi dado pelos pesquisadores suecos para exemplificar a construção *infinitivattribut.relativ*. Segundo eles, “att investera” (para investir) apresenta um *gap* que é preenchido por “pengar” (dinheiro), indicando que dinheiro é a entidade que será investida. Embora a definição da construção aponte para uma obrigação, ou uma necessidade, há uma ambiguidade no sentido dessa sentença em especial, gerada devido à semântica do verbo “investera” (investir). Este pode indicar que dinheiro é necessariamente algo que deve ser investido, cobrindo o sentido de obrigação, mas é possível também interpretar o dinheiro como sendo o recurso que habilita o investimento, o que nos leva ao sentido de suficiência.

- (12) Men i all enkelhet kan jag säga att jag har [[pengar]i [att investera Øi]].
 ‘But I can simply say I have money to invest’
 ‘Mas eu posso simplesmente dizer que eu tenho dinheiro para investir’

Observando os exemplos em (11) e (12), à primeira vista, poderíamos dizer que são instâncias da mesma construção, da mesma forma que ocorre com os exemplos da *DCI_Suficiência* e da *DCI_Obrigação*. A diferença entre elas, segundo a análise dos pesquisadores suecos, é a existência, na construção *infinitivattribut.relativ*, de um *gap* argumental no sintagma verbal infinitivo que é preenchido pelo núcleo nominal e se relaciona com o verbo infinitivo em função argumental. Isso é o que garante, de acordo com os pesquisadores suecos, a noção de obrigação nessa construção. Apesar de no PB ser possível encontrar esse *gap* argumental em algumas das sentenças que exemplificam a *DCI_Obrigação*, essa característica não assegura a semântica de obrigação. No exemplo em (12), isso fica evidente na ambiguidade do verbo “investera” (investir), o qual, mesmo tendo um *gap* sintático – possível de ser preenchido pelo núcleo nominal “pengar” (dinheiro) -, pode ter uma leitura de obrigação ou de suficiência. Vejamos outro exemplo dessa construção do sueco em (13).

- (13) Alla har [sitt [jobbi] [att göra Ø]]
 ‘Everyone has his work to do’.
 ‘Todo mundo tem seu trabalho para fazer’.

Em (13), “ter trabalho para fazer” evoca mais claramente o frame de Obrigação. O que se tem nessa frase é a ideia de que todas as pessoas têm um trabalho que deve ser feito necessariamente. Como se nota, em “att göra” (para fazer), há um gap argumental que se preenche com “jobb” (trabalho), como no exemplo em (12), porém a semântica de obrigação, nessa frase, está muito mais evidente devido ao sentido individual de cada termo que preenche os CEs e não, como propõem os pesquisadores suecos, ao *gap* sintático. Esses exemplos confirmam a importância do mapeamento de restrições e relações para confirmar as impressões do linguista e garantir análises mais objetivas e confiáveis, visto que apenas a análise intuitiva do linguista pode não abarcar todas as especificidades das construções e pode fornecer uma interpretação limitada de aspectos relevantes das construções. A Figura 52, portanto, traz a construção *infinitivattribut.relattiv* com seus CEs e as restrições aplicadas a cada um deles. Note-se que sintagmaticamente essa construção é de fato idêntica à anterior, a qual apresenta semântica de suficiência. A diferença, contudo, é evidenciada pela evocação de frames distintos. Perceba-se que, na atribuição de restrições, não propusemos qualquer definição quanto à natureza argumental ou não do *gap*.

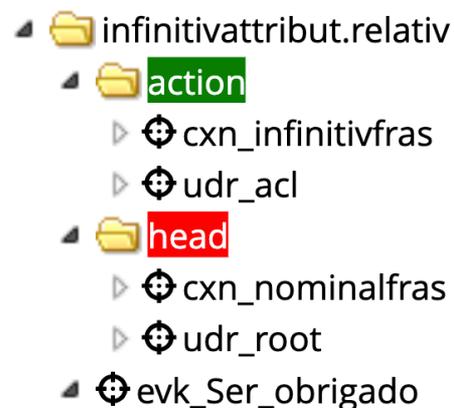


Figura 52: Restrições mapeadas à construção *infinitivattribut.relattiv*

Há ainda, herdando da construção *attributivinfinitiv*, as construções *infinitivattribut.apposition* e *infinitivattribut.komplement*. Como nas demais, em ambas, há um núcleo nominal seguido por um sintagma verbal infinitivo. No entanto, na primeira, conforme foi apresentado pelos pesquisadores suecos, este sintagma infinitivo é uma especificação correferencial do núcleo nominal e, na segunda, um complemento do núcleo, o que implica diferentes etiquetas de relações sintáticas das UDs nos CEs VPinf. A *infinitivattribut.apposition* está exemplificada em (14) e a *infinitivattribut.komplement*, em (15). A Figura 53, por sua vez, demonstra ambas as construções com seus CEs e as restrições mapeadas a cada um deles.

- (14) idén att åka till Brasilien
'the idea to go to Brazil'
'A ideia de ir ao Brasil'
- (15) beslutet att ställa in konserten
'the decision to cancel the concert'
'A decisão de cancelar o concerto'

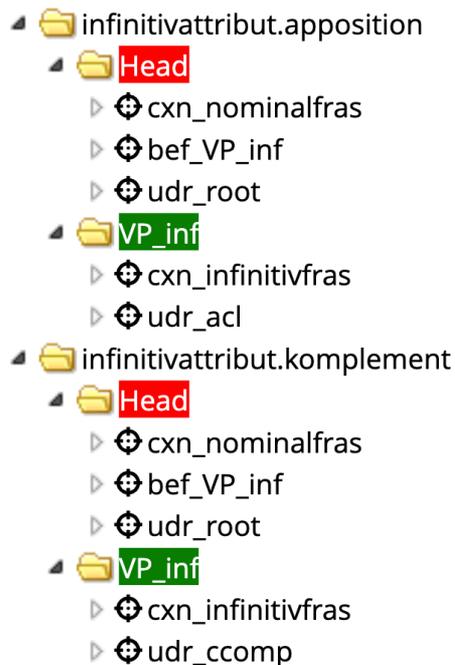


Figura 53: Restrições mapeadas às construções *infinitivattribut.apposition* e *infinitivattribut.komplement*

Como se pode ver nos exemplos em (15), ambos são traduzidos para o português com o uso da preposição “de”, ao invés da preposição “para” – caso das outras duas construções abordadas acima –, o que indica, seguindo o Princípio da Não-Sinonímia (GOLDBERG, 1995), que se trata de construções diferentes no PB, mas ainda assim relacionadas às construções de *DCI* através das construções mães, como a *Cláusula_relativa_infinitiva*. Isso pode ser afirmado com base nas análises de Torrent (2015) com dados do Português do século 13, que demonstram que a construção *Possessiva_habilitativa* (*DCI_Suficiência*) já ocorreu com o esquema [de Vinf]. Em Laviola (2015), foi desenvolvido um teste de similaridade semântica contendo sentenças-exemplo das construções de *DCI* destinadas ao julgamento de falantes nativos do PB. Estes deveriam julgar em uma escala de 1 a 7 cada uma das sentenças-exemplo, em que 1 significaria pouco aceitável e 7 muito aceitável. Cada sentença, portanto, teria um *score* máximo de 7 pontos. Um dos testes foi a substituição da preposição “para” pela preposição “de” nas sentenças-exemplo das *DCI*, conforme se pode ver em (16) e em (17) respectivamente.

- (16) Eu não tenho tempo **para** assistir televisão.
 (17) Eu não tenho tempo **de** assistir televisão.

Esse teste retornou um *score* médio elevado nas nossas análises – o segundo maior, depois do padrão com a preposição “para”. Isso indica que, em muitas sentenças, há a possibilidade de paráfrase da preposição sem que isso cause prejuízos para o julgamento dos falantes nativos. Além de isso ser devido à possibilidade de alguns nomes admitirem mais de uma preposição, como fica evidente nos exemplos (16) e (17) com o nome “tempo”, há, ainda, a trajetória histórica da construção de *DCI*, conforme demonstrado por Torrent (2015).

Conforme demonstrado nessa análise, pode-se assumir que as construções de *Dativo_com_infinitivo* do PB, são comparáveis às construções do inglês e do sueco que abarcam o mesmo padrão, e, por consequência, é possível alinhá-las computacionalmente, ainda que de modo parcial, independente de qualquer diferença metodológica que possa haver entre os recursos ou quaisquer outras dificuldades.

Discutir esses impasses, entretanto, é relevante, visto que a busca por estratégias que possam solucioná-los aumenta a objetividade e a confiabilidade do

alinhamento. Neste trabalho, o objetivo é demonstrar que é possível realizar o alinhamento entre construções de línguas distintas, ainda que haja diferenças metodológicas entre os constructicons. Apesar disso, cabe apontar aqui quais impasses foram identificados nas análises empreendidas. O primeiro deles é fato de que, para a *Infinitival_relative_modal* do inglês, ainda não foi feita uma análise completa, e ainda não se tem nenhum estudo que indique que esse padrão construcional tenha uma construção irmã que evoque o frame de Obrigação, como ocorre com a *DCI_Obrigação* do PB e a *infinitivattribut.relativ* do sueco, em alguns casos. Além disso, embora o Constructicon do inglês mantenha uma relação estreita com a FrameNet, de forma semelhante ao Constructicon do PB, muitas alterações no que tange à perspectiva analítica foram feitas no Constructicon do PB que não ocorreram da mesma forma no recurso do inglês.

Nota-se, portanto, que um dos maiores desafios ao se propor um alinhamento entre constructicons está relacionado às diferentes perspectivas analíticas adotadas pelos recursos de cada língua. Como mencionado, o Constructicon do PB e o do inglês são recursos dependentes das framenets de suas línguas, uma vez que a contraparte semântica das construções é, em sua maioria, fornecida pela relação de Evocação de frames. O Constructicon do sueco, por sua vez, é um recurso independente, ou seja, embora exista uma framenet para o sueco, os recursos não compartilham a mesma base de dados – como o faz a FN-Br e o Constructicon do PB. Isso, entretanto, não tem impedido que o Swe Ccn insira, na definição de construções, links para os frames evocados.

Devido às diferentes perspectivas analíticas dos constructicons de cada uma das línguas, bem como às diferentes estratégias adotadas na diferenciação e inserção de construções nos respectivos recursos, faz-se necessário adotar representações intermediárias – como UDs e as framenets –, como foi apresentado neste trabalho. Assim, é possível comparar e alinhar construções de línguas distintas, sem que tais perspectivas e estratégias diferentes limitem o processo.

A Figura 54 apresenta um grafo gerado na base de dados da FrameNet Brasil o qual mostra o alinhamento das construções de *DCI* do PB e suas contrapartes no inglês e no sueco. No grafo, as setas vermelhas indicam a relação de Herança_construcional. Note-se, portanto, que as construções do PB, *Dativo_com_infinitivo_Suficiência* e

Dativo_com_infinitivo_Obrigaçã herdam da construção *Cláusula_relativa_infinitiva_para*, a qual, por sua vez, herda das construções *Cláusula_relativa_infinitiva* e *Cláusula_adjuntiva_final_infinitiva*. Da mesma forma, as construções do sueco herdam todas da mesma construção, a *attributivinfinitive*. Já a construção do inglês, a *Infinitival_relative_modal* herda da construção *Relative_clause*. As setas pretas indicam a relação Elemento_de, ou seja, conectam um determinado elemento a uma construção ou a um frame. Estando a seta preta conectada a um círculo verde – representativo das construções –, isso indica que o elemento em questão é um elemento de construção. Se, por outro lado, estiver conectada a um quadrado vermelho – que representa os frames –, significa que é um elemento de frame.

As setas azuis mostram a relação de Evocação e se conectam a um quadrado vermelho, que é o frame evocado pela construção. Dessa maneira, as construções de *Dativo_com_infinitivo_Obrigaçã*, do PB, e *infinitivattribut.relativ*, do sueco, as quais apresentam correspondência entre si, evocam todas o mesmo frame, *Ser_obrigado*. As construções *Dativo_com_infinitivo_Suficiênci*, do PB, *Infinitival_relative_modal*, do inglês, e *attributivinfinitiv.modal*, do sueco, têm setas azuis direcionadas ao frame de *Suficiênci*. As demais construções do sueco, *infinitivattribut.apposition* e *infinitivattribut.komplement*, não se conectam a nenhum frame no grafo, pois não evocam frame, já que são construções cujas contrapartes funcionais ainda não foram tratadas em termos de frames pela FrameNet.

As setas amarelas ligam os Elementos de Construção às etiquetas das Universal Dependencies que marcam a relação sintática de cada um deles nas construções em questão. Note-se, portanto, que todos os núcleos nominais de todas as construções nas três línguas, tanto os CEs *NP_NÚCLEO* das construções de *DCI* do PB, quanto os CEs *HEAD* das construções do inglês e do sueco apresentam a etiqueta *udr_root*, já que são os núcleos das construções. Já o elemento da construção preenchido pela sentença infinitiva nas duas construções do PB, na construção do inglês e em três das construções do sueco é marcado com a etiqueta das UD's *udr_acl*, visto que é uma oração que modifica um nome. O CE *VP_INF* da

construção *infinitivattribut.komplement*, por sua vez, tem uma seta amarela em direção à etiqueta *udr_ccomp*, mostrando que é um complemento oracional.

Como se pode perceber com o grafo, as construções de *DCI_Suficiência* e *DCI_Obrigaçã* do PB e as equivalentes a ela no inglês (*Infinitival_relative_modal*) e no sueco (*attributivinfinitiv.modal* e *infinitivattribut.relativ*) convergem nas estruturas de representação intermediárias eleitas para esta proposta de alinhamento. Nesse sentido, no que tange à semântica, a *DCI_suficiência*, a *IRM* e a *attributivinfinitiv.modal* evocam o mesmo frame, o que também ocorre com a *DCI_Obrigaçã* e a *infinitivattribut.relativ*. Da mesma maneira, no que diz respeito à forma, todas essas construções apresentam as mesmas etiquetas das UD's, apresentando, portanto, as mesmas relações sintáticas. Essa convergência confirma, portanto, a validade das categorias de frame e UD para o alinhamento de construções proposto nesta tese.

Dessa maneira, conclui-se que, com as análises empreendidas e as discussões que foram realizadas acerca das construções de *DCI* do PB e suas contrapartes no sueco e no inglês, é viável realizar o alinhamento de tais construções, como se pode visualizar na Figura 54. Entretanto, cabe também neste trabalho apontar para as limitações percebidas no estudo dessa pequena amostra de construções.

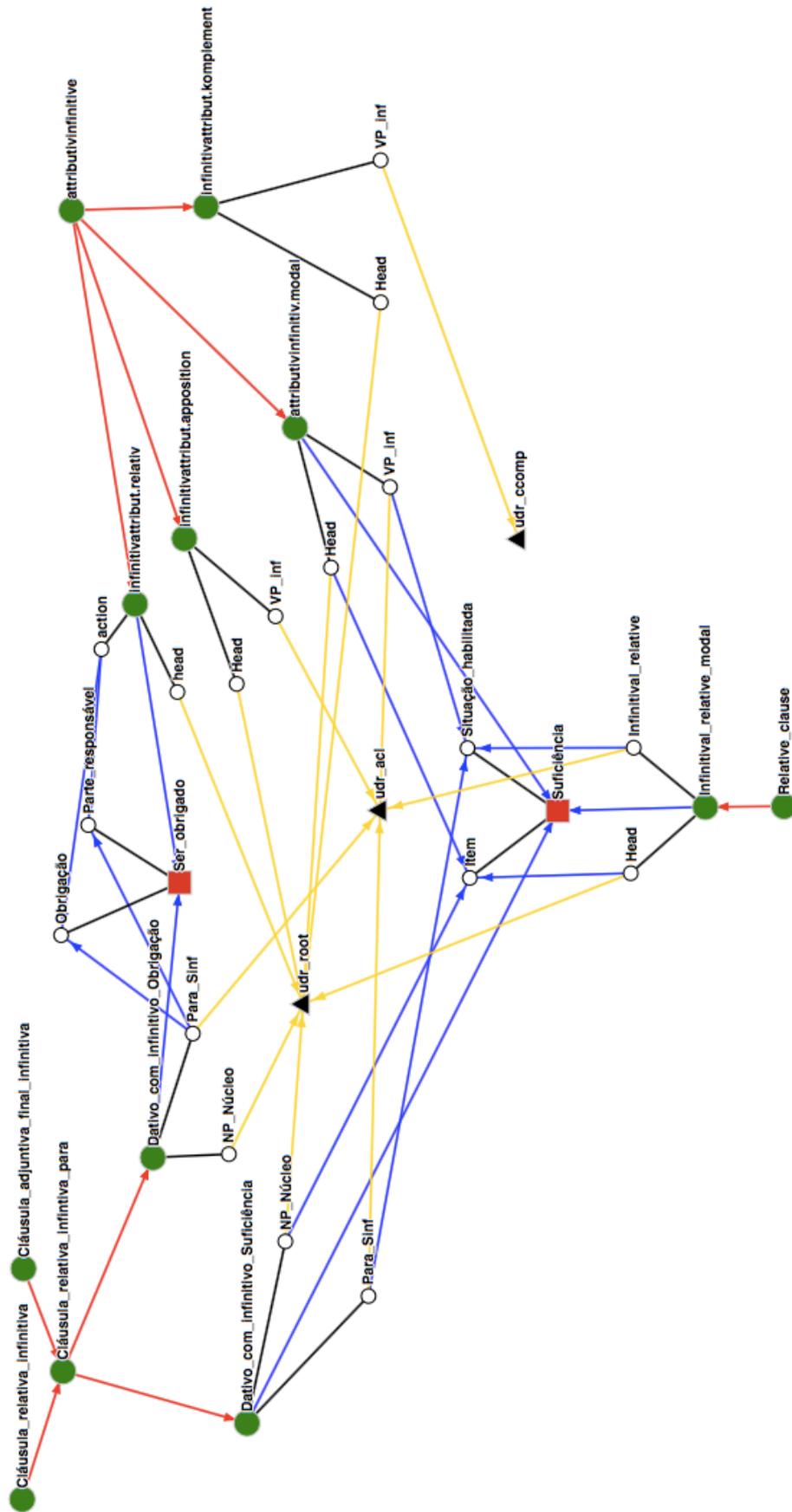


Figura 54: Alinhamento das construções de DCI e suas contrapartes em inglês e sueco.

Primeiramente, percebe-se que a decisão inicial de não usar as UD POS como restrições modeladas e, potencialmente, como pontos de alinhamento, precisa ser investigada em maior profundidade. Quando tal decisão foi tomada, ela se baseou no fato de que a informação sobre classes de palavras poderia se apresentar como uma restrição tão inconsistente – já que alguns CEs podem ser preenchidos por uma grande diversidade de estruturas sintagmáticas –, que seria inútil. Porém, ao observarmos a Figura 54, é impossível não perceber que a aplicação de uma restrição UD-POS “NOUN” aos CEs núcleo das construções das três línguas traria informações importantes para o alinhamento. Por outro lado, em termos de implementação, essa restrição precisaria ser feita no CE *NOME* da construção de *Sintagma_nominal*, a qual, por sua vez, é a restrição CE>Cxn do CE *NP_NÚCLEO* da construção de *DCI*. Isto significa dizer que aplicaríamos uma restrição a um dos elementos de uma construção que, por sua vez, é também uma restrição imposta a um elemento da construção para a qual aquela restrição importa em termos de alinhamento. Uma possível solução, a ser investigada em trabalhos futuros, seria a aplicação das UD-POS às construções de classes de palavras, tais como *Nome*, *Verbo* e *Adjetivo*.

Em segundo lugar, para que a proposta de alinhamento aqui apresentada seja eficaz, há a necessidade futura de se empreender um esforço conjunto dos projetos de cada língua no que tange ao mapeamento de restrições sintáticas, com base nas relações das UDs, além de restrições semânticas, com base nos frames evocados pelas construções e pelos CEs, para assegurar o alinhamento das demais construções contidas nos três recursos.

6 CONCLUSÕES

O objetivo desta pesquisa centrou-se em buscar uma metodologia que possa ser aplicada aos esforços de alinhamento entre recursos construcionais, levando em consideração as especificidades e envezamentos analíticos próprios de cada constructicon. Para tanto, foi assumida a hipótese de que representações intermediárias e interlinguais, como as *framenets* e as *Universal Dependencies*, poderiam ser úteis nesse empreendimento.

Nesse contexto, considerando as etapas propostas no capítulo de metodologia deste trabalho – as quais foram baseadas nas fases que Dias-da-Silva et al (2007) propõem para o desenvolvimento de pesquisas linguísticas com foco computacional –, na **fase linguística** deste trabalho, foram realizadas análises comparativas entre construções registradas na base de dados do constructicon do inglês e possibilidades de correspondência em PB e em sueco, a partir da definição de critérios de análise. Tais critérios consideraram aspectos como: (1) correspondência geral na língua alvo; (2) correspondência funcional; (3) correspondência formal; (4) diferenças formais externas. A partir da análise contrastiva trilingue, chegou-se à conclusão preliminar de que os recursos podem ser comparáveis e que, portanto, seria possível alinhá-los computacionalmente.

Contudo, uma vez que esse trabalho foi realizado de forma unidirecional, ou seja, partindo do inglês para o sueco e o PB, e não tendo análises na direção oposta ou entre o PB e o sueco diretamente, para se propor um alinhamento computacional de fato, outras ações se tornam imprescindíveis. Assim, na **fase representacional**, este trabalho focou-se em mapear, na base de dados do Constructicon do PB, restrições às construções do PB que foram identificadas como equivalentes no trabalho comparativo da fase linguística. Essas restrições especificam características de cada elemento construcional no que tange a aspectos lexicais, semânticos e sintáticos, tornando a leitura dessas especificações mais claras para a máquina e, ao mesmo tempo, mais amigáveis ao usuário – em comparação à leitura de uma AVM. Além disso, as restrições estabelecem relações entre construções, visto que muitos elementos de construção são licenciados por outras construções.

No mapeamento de restrições no Constructicon do PB, é também possível identificar o CE com alguma etiqueta de relação sintática das *Universal Dependencies*, as quais, por terem um aspecto interlingual, ou seja, foram idealizadas para serem

aplicadas a qualquer língua, garantem certa generalização, favorecendo o alinhamento multilíngue. Dessa maneira, neste trabalho, foram analisadas 91 construções para cujos CEs foram mapeadas 465 restrições. Destas, as que foram mapeadas em maior quantidade foram as restrições do tipo CE>construction e CE>UD_relation - a primeira mapeada a 178 CEs e a segunda a 173 CEs.

Após o mapeamento dessas restrições e relações para as construções do PB, inserido na **fase implementacional**, foi proposto um estudo das construções de DCI do PB em comparação com suas contrapartes em inglês e em sueco, do qual se pode resumir que:

- A) O sueco apresenta quatro construções que herdam de uma construção mais genérica, a *attributivinfinitive*, assim como as construções de DCI do PB que são filhas da construção *Cláusula_relativa_infinitiva_para,que*, por sua vez, herda da *Cláusula_relativa_infinitiva*.
- B) A construção do sueco que os pesquisadores do Swedish Constructicon apontam como tendo semântica de obrigação, a *infinitivalrelative*, apresenta um *gap* argumental no sintagma verbal infinitivo que é preenchido pelo núcleo nominal, o que é apontado por eles como o critério que a diferencia das demais construções herdeiras da *attributivinfinitive*. Assim, segundo a análise do sueco, é o *gap* argumental que evidencia a semântica de obrigação, o que, como foi demonstrado, não é garantido para todos os outros casos, visto que, a depender da semântica individual dos termos que compõem os CEs da construção, ainda que haja um *gap* sintático, pode-se ter uma leitura ambígua entre obrigação e suficiência.
- C) As demais construções do sueco apontam para construções relativas do PB que sejam introduzidas pela preposição “de”.
- D) O inglês apresenta apenas uma construção descrita em sua base de dados e essa dá conta apenas do sentido de suficiência, a *Infinitival_relative_modal*, não havendo descrito no recurso do inglês um padrão que evoque o frame de *Ser_obrigado*, como ocorre no PB e no sueco. Isso aponta para a necessidade de um estudo mais aprofundado desse padrão construcional no inglês.

E) Embora cada um dos recursos adote diferentes estratégias e perspectivas analíticas, é possível alinhar construções de línguas distintas, ainda que estas adotem metodologias de trabalho díspares, conforme foi demonstrado na seção 5.4. Entretanto, esse alinhamento só será possível para todas as construções nos três recursos, se os projetos se dispuserem a mapear restrições às suas construções em suas bases de dados de modo a aplicar representações intermediárias que possibilitem relacionar suas construções sem impedir que mantenham suas estratégias e perspectivas analíticas distintas. Certamente, não se trata de tarefa simples, haja vista as mais de 400 restrições modeladas para as 91 construções do PB que foram implementadas no Constructicon da FrameNet Brasil, no âmbito desta tese. Entretanto, acredita-se que, uma vez incorporadas ao fluxo de criação de construções, a adição de restrições pode se tornar um padrão.

A partir das análises e propostas feitas nesta tese de doutorado, pode-se depreender dela as seguintes contribuições:

- (i) As análises contrastivas trilingües desenvolvidas neste trabalho deram origem a quatro critérios de comparação que podem ser utilizados por pesquisas futuras de contraste entre construções de línguas distintas.
- (ii) As discussões teórico-metodológicas realizadas neste trabalho propõem a utilização de representações intermediárias, como UD's e framenets, para a realização do alinhamento de construções e constructicons, permitindo que vários projetos de línguas distintas realizem alinhamentos construcionais em direção a um constructicon multilíngue, ainda que adotem metodologias e estratégias analíticas muito diferentes.
- (iii) Esta pesquisa mostrou ainda como as três fases propostas por Dias-da-Silva et al (2007) para o desenvolvimento de pesquisas linguístico-computacionais podem ser aplicadas em um trabalho construcionista. A primeira fase incluiu um trabalho linguístico de análise contrastiva entre construções do sueco, do inglês e do PB a partir de critérios de comparação. A segunda fase envolveu a proposição de representações intermediárias para a realização do alinhamento construcional. Por último, a terceira fase que propõe a implementação de tais representações na base de dados, fornecendo subsídios para um alinhamento que não dependa exclusivamente da intuição do linguista. Pode-se afirmar, portanto, que esta

pesquisa de doutorado fornece os métodos e abre caminho para que outras pesquisas voltadas para a comparação de construções e o alinhamento de constructicons sejam empreendidas.

Esta pesquisa de tese gerou, ainda, os seguintes produtos acadêmicos:

- (i) Apresentação no 1º Workshop CAPES-STINT na Universidade Federal de Juiz de Fora, no Brasil, em outubro de 2015.
- (ii) Apresentação do trabalho *“The Constructicon and its development in the context of FrameNet Brasil”* na 6th Cognitive Linguistics Conference na Universidade de Bangor, no Reino Unido, em julho de 2016.
- (iii) Apresentação do trabalho intitulado *“Comparing constructicons across languages: Brazilian Portuguese and Swedish counterparts to the construction entries in the FrameNet English Constructicon”* no 9th International Conference on Construction Grammar na Universidade Federal de Juiz de Fora, no Brasil, em outubro de 2016.
- (iv) Apresentação no Constructicon Symposium em Moscou, na Rússia, do trabalho intitulado *“The FrameNet Brasil Constructicon: recent changes and next steps”*, em março de 2017.
- (v) Apresentação do trabalho *“The advances in the Brazilian Portuguese Constructicon towards a multilingual constructicography”* no Linguistic Research Seminar da Universidade de Gotemburgo, na Suécia, em março de 2017.
- (vi) Apresentação do trabalho de título *“Comparing the use of Universal Dependencies and FrameNet Grammatical Functions for multilingual alignment of Constructions”* na International Cognitive Linguistics Conference, na Universidade de Tartu, na Estônia, em julho de 2017.
- (vii) Apresentação no Research Colloquium do Institute of German Language and Literature da Universidade de Düsseldorf, na Alemanha, em novembro de 2018.
- (viii) Apresentação no 10th International Conference on Construction Grammar do trabalho intitulado *“The Brazilian Portuguese Constructicon and the use of Universal Dependencies for multilingual alignment of Constructicons”* na Universidade de Sorbonne, na França, em julho de 2018.
- (ix) Publicação de artigo na Revista DELTA (TORRENT ET AL., 2017).

- (x) Publicação de dois capítulos de livro no volume *Constructicography*, pela editora John Benjamins (LYNGFELT ET AL., 2018a; TORRENT ET AL., 2018).
- (xi) Publicação nos Proceedings of The AAAI 2017 Spring Symposium on Computational Construction Grammar and Natural Language Understanding (LAVIOLA ET AL., 2017; MATOS ET AL., 2017).

Por fim, como trabalhos futuros, vê-se a possibilidade de explorar a contribuição das outras restrições modeladas – CE>Construção, CE>Frame, CE>Frame_family, etc. – para o alinhamento de construções, ou mesmo a adição de outras representações interlinguais, como os Comparative Concepts (CROFT, no prelo), ao quadro de restrições do Constructicon da FrameNet Brasil.

REFERÊNCIAS BIBLIOGRÁFICAS

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

BÄCKSTRÖM, L., LYNGFELT, B. & SKÖLDBERG, E. Towards Interlingual Constructicography: on correspondence between constructicon resources for English and Swedish. In: 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.

BERGEN, B. & CHANG, N. Embodied Construction Grammar. In: HOFFMAN, T. & TROUSDALE, G. **The Oxford Handbook of Construction Grammar** (pp.169- 190). New York: Oxford University Press, 2013.

BOAS, H. C. Bilingual FrameNet Dictionaries for Machine Translation. In: RODRIGUEZ, M. G. & ARAUJO, C. P. S. **Proceedings of The Third International Conference on Language Resources and Evaluation**. Las Palmas, Espanha, 2002.

BOAS, Hans C. Cognitive Construction Grammar. In: HOFFMANN, T. and TROUSDALE, G. (eds.), **The Oxford Handbook of Construction Grammar**, 233–254. Oxford: Oxford University Press.

BOAS, H. C. & SAG, I.A. (eds.) **Sign-Based Construction Grammar**. Stanford: CSLI Publications. 2012.

BOAS, H. C. & ZIEM, A. Constructing a constructicon for German: Empirical, theoretical, and methodological issues. In: BORIN, L.; LYNGFELT, B.; OHARA, K. & TORRENT, T. T. **Constructicography: Constructicon Development across Languages**. Amsterdam: John Benjamins, 2018.

BORIN, L., M. FORSBERG & L. LÖNNGREN. Saldo: a touch of yin to WordNet's yang.

In: **Language Resources and Evaluation**. 2013.

BRODBECK, R.C.M.S. **Um monte de problemas gera uma chuva de respostas: estudo de um caso de desencontro na quantificação nominal em Português**. Tese de doutorado em Linguística apresentada à Coordenação do Curso de Pós-Graduação em Linguística: área de concentração em Linguagem e Cognição da Faculdade de Letras da Universidade Federal de Juiz de Fora, 2010.

CHOMSKY, N. Three Models for the Description of Language. **IRE Transactions on Information Theory** IT-2, no. 3 (September 1956): 113-24, 1956.

_____. **Syntactic Structures**. The Hague: Mouton. 1957.

_____. **The minimalist program**. MIT Press, Cambridge, Mass. 1995.

CROFT, W. **Radical Construction Grammar**. Syntactic theory in typological perspective. Oxford: Oxford University Press. 2001.

DIAS-DA-SILVA, B. C., MONTILHA, G., RINO, L. H. M., SPECIA, L., NUNES, M. D. G. V., DE OLIVEIRA JR, O. N. & PARDO, T. A. S. Introdução ao Processamento das Línguas Naturais e Algumas Aplicações. **Série de Relatórios do Núcleo Interinstitucional de Lingüística Computacional**, 3, 2007.

FILLMORE, C. J. The Case for Case. **Universal Linguistic Theory**, p. 1-89, 1968.

_____. The case for case reopened. **Syntax and semantics**, 8, p. 59-82, 1977.

_____. Frame Semantics. In: **The Linguistic Society of Korea (org.)**. Linguistics in The Morning Calm. Seoul: Hanshin, 1982.

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

_____, KAY, P. & O'CONNOR, M. C. Regularity and idiomaticity in grammatical constructions: The case of let alone. **Language**, p. 501-538, 1988.

_____, LEE-GOLDMAN, R. & RHOMIEUX, R. The FrameNet Constructicon. In BOAS, H. C. & SAG, I. A. (Eds.). **Sign-Based Construction Grammar**. Stanford: CSLI Publications, 2012.

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.

_____, **Constructionist approaches**. In Hoffmann & Trousdale (eds.), (pp. 15–31). 2013

GRUZITIS, N.; NESPORE-BERZKALNE, G.; SAULITE, B. Creation of Latvian FrameNet based on Universal Dependencies. In: **Proceedings of the International FrameNet Workshop**. Myiazaki, Japan, 2018. p. 23-27.

JANDA, L. A., LYASHEVSKAYA, O., NESSET, T., RAKHILINA, E., TYERS, F. M. A Constructicon for Russian: Filling in the Gaps. In: BORIN, L.; LYNGFELT, B.; OHARA, K. & TORRENT, T. T. **Constructicography: Constructicon Development across Languages**. Amsterdam: John Benjamins, 2018.

JURAFSKY, D. & MARTIN, J. H. **Speech and Language Processing** (2nd edition). Prentice-Hall, Inc. Upper Saddle River, NJ, USA, 2009.

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

LAGE, L. M. **Frames e construções**: A implementação do constructicon na FrameNet Brasil. Dissertação (Mestrado em Linguística). Universidade Federal de Juiz de Fora, Juiz de Fora, 2013.

_____. **Modelagem Linguístico-Computacional das relações entre construções e frames no constructicon da FrameNet Brasil**. Tese de Doutorado em Linguística. Universidade Federal de Juiz de Fora. Juiz de Fora, 2018.

LAVIOLA, A. B. **Frames e Construções em Contraste**: uma análise comparativa Português - Inglês no tangente à implementação de Constructicons. Dissertação de Mestrado em Linguística. Universidade Federal de Juiz de Fora. Juiz de Fora, 2015.

_____. & TORRENT, T. T. Construções em Contraste: uma análise da Construção Habilitativa em Para Infinitivo do Português do Brasil e da Construção Wherewithal do Inglês. In: **Anais do XI Encontro de Linguística de Corpus (ELC 2012)**. São Carlos: USP, 2012.

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

LANGACKER, R.W. Interview with Ronald W. Langacker. Entrevista concedida a Diogo Pinheiro. **Revista Linguística**. Vol. 14, n. 1, p. 35-47. Rio de Janeiro, 2018.

LEE-GOLDMAN, R. & PETRUCK, M. R. L. The FrameNet Constructicon in Action. In: BORIN, L.; LYNDFELT, B.; OHARA, K. & TORRENT, T. T. **Constructicography**: Constructicon Development across Languages. Amsterdam: John Benjamins, 2018.

LÖNNEKER-RODMAN, B. **Multilinguality and FrameNet**. Berkeley, CA: ICSI, 2007.

LYASHEVSKAYA, O., & KASHKIN, E. FrameBank: a database of Russian lexical constructions. In: KHACHAY, M. Y., KONSTANTINOVA, N., PANCHENKO, A., IGNATOV, D. I. & LABUNETS, G. V. (Eds.), **Analysis of images, social networks and texts**. Fourth international conference, AIST 2015, Yekaterinburg, Russia, April 9-11, 2015, Revised selected papers (= Communications in computer and information science, Vol. 542) (pp. 337-348). Springer. 2015.

LYNGFELT, B.; BORIN, L.; FORSBERG, M.; PRENTICE, J.; RYDSTEDT, R.; SKÖLDBERG, E. & TINGSELL, S. Adding a construction to the Swedish resource network of Språkbanken. In: JANCSARY, J. Ed. **Empirical Methods in Natural Language Processing: Proceedings of the Conference on Natural Language Processing**. Vienna: ÖGAI, 2012.

LYNGFELT, B. Introduction: constructions and constructicography. In: BORIN, L.; LYNGFELT, B.; OHARA, K. & TORRENT, T. T. **Constructicography: Constructicon Development across Languages**. Amsterdam: John Benjamins, 2018.

LYNGFELT, B.; TORRENT, T. T.; LAVIOLA, A. B.; BÄCKSTRÖM, L., HANNESDÓTTIR, A. H. & MATOS, E. E. Aligning constructions across languages: A trilingual comparison between English, Swedish, and Brazilian Portuguese. In: BORIN, L.; LYNGFELT, B.; OHARA, K. & TORRENT, T. T. **Constructicography: Constructicon Development across Languages**. Amsterdam: John Benjamins, 2018a.

LYNGFELT, B.; BÄCKSTRÖM, L.; BORIN, L.; EHRLEMARK, A. & RYDSTEDT, R. Constructicography at work: Theory meets practice in the Swedish Constructicon. In: BORIN, L.; LYNGFELT, B.; OHARA, K. & TORRENT, T. T. **Constructicography: Constructicon Development across Languages**. Amsterdam: John Benjamins, 2018b.

MCSHANE, M. Natural Language Understanding (NLU, not NLP) in Cognitive Systems. In: **Association for the Advancement of Artificial Intelligence**. p. 43-56, 2017.

MATOS, E. E.; TORRENT, T. T.; ALMEIDA, V. G.; SILVA, A. B. L.; LAGE, L. M.; MARCAO, N. D. & TAVARES, T. S. Constructional Analysis Using Constrained

Spreading Activation in a FrameNet-Based Structured Connectionist Model. In: **The AAAI 2017 Spring Symposium on Computational Construction Grammar and Natural Language Understanding Technical Report SS-17-02**. Palo Alto, CA: AAAI Publications, v.17, p.222-229, 2017.

NIVRE, J. Towards a universal grammar for natural language processing. In GELBUKH, A., editor, **Computational Linguistics and Intelligent Text Processing**, pages 3-16. Springer, 2015.

_____. **Dependency grammar and dependency parsing**. Technical Report MSI 05133, Växjö University, School of Mathematics and Systems Engineering, 2005.

OHARA, K. H. Toward Constructicon Building for Japanese in Japanese FrameNet. **Veredas (UFJF. Online)**, v. 17, p. 11-27, 2013.

OHARA, K. H. Relations between Frames and Constructions: A Proposal from Japanese FrameNet Constructicon. In: BORIN, L.; LYNGFELT, B.; OHARA, K. & TORRENT, T. T. **Constructicography: Constructicon Development across Languages**. Amsterdam: John Benjamins, 2018.

OTHERO, G. A. **Teoria X-barra: descrição do português e aplicação computacional**. São Paulo: Contexto, 2006. 160p.

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

PETRUCK, M. R. L. Frame Semantics. In: VERSCHUEREN, J. OSTMAN, J. & BLOMMAERT, J. (Eds.) **Handbook of Pragmatics**. Amsterdam & Philadelphia: John Benjamins, 1996.

RUPPENHOFER, J.; ELLSWORTH, M.; PETRUCK, M. R. L.; JOHNSON, C. R. & SCHEFFCZYK, J. **FrameNet II: Extended theory and practice**. Berkeley: International Computer Science Institute, 2016.

SALOMÃO, M. M. M. The Metaphoric Grounding of Grammar: the Modal Construction with 'give' in Brazilian Portuguese. In: **The 33rd Annual Meeting of The Berkeley**

Linguistics Society. University of California, Berkeley. 2007.

_____. Tudo certo como dois e dois são cinco: Todas as construções de uma língua. In: MIRANDA, N. S. & SALOMÃO, M. M. M. (Org.). **Construções do Português do Brasil: da gramática ao discurso**. Belo Horizonte: UFMG, 2009

_____, TORRENT, T. T. & SAMPAIO, T. F. A Linguística Cognitiva Encontra a Linguística Computacional: notícias do projeto FrameNet Brasil. **Cadernos de Estudos Linguísticos**, 53(2), 2013.

SPÄRCK-JONES, K. Computational linguistics: what about the linguistics? **Computational linguistics**, v. 33, n. 3, p. 437-441, 2007.

TAVARES, T. S. **Construção Binominal de Quantificação Indefinida no PB: uma abordagem construcionista**. Dissertação de Mestrado em Linguística. Universidade Federal de Juiz de Fora. Juiz de Fora, 2013.

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.

_____; SALOMÃO, M. M. M.; CAMPOS, F.C.A.; BRAGA, R. M. M.; MATOS, E. E. S.; GAMONAL, M. A.; GONÇALVES, J.; SOUZA, B. P.; GOMES, D. S.; PERON-CÔRREA, S. R. Copa 2014 FrameNet Brasil: a frame-based trilingual electronic dictionary for the Football World Cup. In: **Proceedings of COLING 2014, The 25th International Conference on Computational Linguistics: System Demonstrations** (pp. 10–14). Dublin, Ireland: ACL. 2014.

TORRENT, T. T.; LAGE, L. M.; TAVARES, T. S.; SILVA, A. B. L. Relações de Herança entre Construções e entre Frames: desafios da extensão do modelo construcionista para o domínio computacional no âmbito da FrameNet Brasil. **DELTA**, v. 33(1), p. 45-77, 2017.

_____.; MATOS, E. E.; LAGE, L. M.; LAVIOLA, A. B.; TAVARES, T. S. & ALMEIDA, V. G. Towards the Continuity Between the Lexicon and the Constructicon in FrameNet Brasil. In: BORIN, L.; LYNGFELT, B.; OHARA, K. & TORRENT, T. T. **Constructicography: Constructicon Development across Languages**. Amsterdam: John Benjamins, 2018.

UNIVERSAL DEPENDENCIES PROJECT. **Universal Dependencies**. Website do projeto Universal Dependencies. Disponível em: <http://www.universaldependencies.org>. Acesso em: 14 jan. 2019.

VOSSSEN, R. W. Relative Strengths and Weaknesses of Small firms in innovation. In: **International Small Business Journal**, 16(3), 7. 1998.

WINOGRAD, T. Understanding natural language. In: **Cognitive Psychology**, 3(1), 191. 1972.

ZIEM, A. & BOAS, H. C. Towards a Constructicon for German. In **The AAAI 2017 Spring Symposium on Computational Construction Grammar and Natural Language Understanding Technical Report SS-17-02**. Palo Alto, CA: AAAI Publications, v.17, p. 274-277, 2017.