

SIMPÓSIO - AT019

MODELOS LINGUÍSTICOS PARA A CRIAÇÃO DE PARSERS: O MODELO CARTOGRÁFICO, OS ADVÉRBIOS ESPECIFICADORES E AS FASES LINGUÍSTICO COMPUTACIONAIS

SILVA, Erico Monteiro¹

Resumo: O presente artigo trata da utilização dos advérbios especificadores Altos e Baixos com o objetivo de demonstrar modelos gramaticais com propriedades mais discursivas na elaboração de *parsers para* análises e descrições de sentenças do Português Brasileiro. Utilizamos a metodologia e o método por Fases elaborado por Dias-da-Silva (1996). Essas fases compreendem a Fase Linguística, que visa a descrição e a análise do fenômeno ou objeto; a Fase Linguístico Computacional que pretende a formalização teórica linguística, a partir das descrições por modelos formais de cunho gerativo, como os modelos da Gramática-Livre de Contexto em conjunto da Matriz de Atributo-Valor, utilizados na composição estrutural da sentença. Os traços de natureza formal e de funcionalidade, fornecem explicitações dos traços que os licenciam na sentença, entretanto, a análise evidencia algumas dificuldades de definir este item lexical tanto na perspectiva linguística no que se refere a natureza, ordem e funções quanto na computacional com a categorização e marcações deterministas, visível no desenvolvimento de analisadores sintáticos automáticos.

Palavras-chave: Advérbio; Parser; Gerativo; Cartografia; Linguística Computacional.

Abstract: This article deals with the use of specifying adverbs High and Low with the objective of demonstrating grammatical models with more discursive properties in the elaboration of parsers for analysis and descriptions of Brazilian Portuguese sentences. We used the methodology and the method by Phases elaborated by Dias-da-Silva (1996). These phases comprise the Linguistic Phase, which aims at describing and analyzing the phenomenon or object; the Computational Linguistic Phase that intends the linguistic theoretical formalization, from the descriptions by formal models of gerativo, like the models of Gramática-Livre of Context in the Matrix of Attribute-Value, used in the structural composition of the sentence. The traits of a formal nature and of functionality, provide explanations of the traits that license them in the sentence, however, the analysis shows some difficulties of defining this lexical item both in the linguistic perspective regarding the nature, order and functions as in the computational with the categorization and deterministic markings, visible in the development of automatic syntactic analyzers.

Keywords: Adverb; Parse; Generative; Cartography; Computational Linguistics.

¹ Universidade Federal de Pernambuco (UFPE), Departamento de Pós-Graduação em Letras - eric9r@gmail.com

INTRODUÇÃO

Dotar as máquinas com capacidade linguística equivalente ou similar aos humanos é um dos objetivos gerais da Linguística Computacional (doravante, LC). Para isso, utiliza-se modelos linguísticos para o tratamento computacional da língua natural, baseando-se nas teorias e pesquisas linguísticas e computacionais, aspirando o desenvolvimento de aplicativos, aplicações e testes teóricos, tratamento e usos de *corpus*, dentre outros, que estão atreladas as subáreas da LC, o Processamento Natural da Linguagem (PNL) e a Linguística de *Corpus*.

Diante dos desafios impostos neste campo interdisciplinar, os modelos gramaticais que postulam o conhecimento de um indivíduo sobre a sua língua, especialmente os modelos gerativos que especificam a faculdade da linguagem como um componente cognitivo entendido por uma base de processos de informações mentais e computacionais, com isto, a gramática de uma língua é um princípio compatível com algoritmos de *parsing*.

Após um período de largas descobertas de Princípios e Parâmetros básicos das línguas pela gramática gerativa, o modelo foi repensado, tanto as regras quanto as restrições foram reduzidas a um menor conjunto “independentemente motivado por princípios de computação ótima ou mais eficiente” (BERWICK;CHOMSKY, 2017, p. 16), ou seja, um sistema mais simples ou mínimo para os sistemas cognitivos e computacionais da sintaxe, inaugurando o Programa Minimalista.

No Programa Minimalista, uma das operações relevantes utilizada para construir a estrutura necessária para a sintaxe é chamada de *Merge*, uma simples operação computacional. *Merge* é uma operação que forma conjuntos de um objeto sintático X (um item lexical, funcional, ou um outro conjunto formado por *Merge*) e um outro objeto sintático Y, formando um conjunto hierarquicamente estruturado {X, Y}. Este objeto sintático recebe um rótulo algorítmico para satisfazer as condições de computação mínima. Berwick e Chomsky (2017) exemplificam a construção sintática com os itens lexicais *Ler* e

livros, de acordo com os autores, *Merge* combina-os formando {ler, livros}, em seguida rotula os itens após uma busca mínima localizando os traços de núcleo, ou seja, os traços do elemento verbal *ler*, indo ao encontro da noção tradicional dos constituintes que identifica um sintagma verbal na sentença “ler livros”. A continuidade das operações fazem com que essas entidades sintáticas continuem em outras computações da linguagem natural.

1 As fases do desenvolvimento

Assumindo a estratégia definida por Dias da Silva (1996), que propõem equacionar os problemas em três domínios complementares, tais domínios são distribuídos por fases, a fase Linguístico, a fase Linguístico-Computacional e a fase Computacional. O autor defende que “as três fases sejam desenvolvidas sucessiva, progressiva e ciclicamente: as representações parciais resultantes das duas primeiras fases podem ser implementadas e, finalmente, testadas, completando, assim, um ciclo” (DIAS-DA-SILVA, 1996, p. 91). Com isto cada domínio é uma condição subsequente a outra fase.

A fase linguístico diz respeito ao conhecimento da linguagem de forma explícita sobre os fenômenos linguísticos visando o desenvolvimento dos sistemas de PNL. O formalismo desenvolvido no Programa Minimalista (PM) que melhor atende a descrição e define o advérbio para nossos propósitos advém da perspectiva cartográfica (RIZZI, 1994; CINQUE, 1999). Na fase linguístico-computacional, há a elaboração dos modelos formais representativos do domínio linguístico. E, por fim, a fase computacional ou implementacional, é a fase de codificação das fases anteriores por meio de linguagens de programação e o ambiente destinado a implementação.

2 O Modelo Cartográfico dos Advérbios

A cartografia consiste de projeções cindidas à periferia esquerda da sentença, como se desenhasse mapas mais detalhados das projeções, além

de entender que as categorias projetadas são ligadas a traços discursivos. Observando as restrições impostas na hierarquia universal (HU) para as projeções com traços funcionais na sentença. A evidência para determinar a HU se encontra no fato dos advérbios serem especificadores das categorias funcionais que checam os traços do núcleo destas categorias. Abandonando a característica até então aceita dos advérbios como apenas adjuntos constituintes secundários e acessórios.

De acordo com Cinque (1999), o advérbio não é um núcleo adjungido ao VP ou alguma projeção dominando, como complemento, o VP. A evidência está no fato de que “advérbios não bloqueiam movimento de várias formas verbais além do fato de que alguns deles podem passar por movimento de Topicalização e Foco, abertos a XPs, mas não a X^o” (CINQUE, 1999, p.4). Cinque (1999) postula uma diferenciação entre os advérbios, os advérbios circunstâncias de lugar, tempo, modo, meio, companhia, razão, objetivo, dentro outros. Estes advérbios são realizados tipicamente na forma preposicional ou na forma de um NP *nu*. O outro tipo são os denominados advérbios propriamente ditos, estes itens são rigidamente ordenados na estrutura funcional. E estes são denominados Advérbios Baixos e Altos de acordo com a sua posição na sentença, como os Altos se situam no escopo da sentença e os baixos mais próximos aos verbos.

Diante destas observações, Cinque percebe que há uma hierarquia que deve ser levada em consideração, pois o resultado de [*solitamente* > *mica* > *gia*] formalizado com outros advérbios, pode-se chegar a um esquema das ordens hierárquicas definidas na GU. Mas à frente, Cinque (1999) demonstra o esquema para outros advérbios, por exemplo: [*solitamente* > *mica* > *gia* > *piu* > *sempre* > *completamente* > *tutto* > *bene*] e [*francamente* > *fortunatamente* > *evidentemente* > *probabilmente* > *ora* > *forse* > *intelligentemente*]. Diante dessas evidências, ele divide os tipos de Advérbios Baixos (Pré-Verbais - *usually, again, often, already, soon*, etc) Advérbios Altos (Sentenciais – *frankly, unfortunately, apparently, probably*, etc.) de acordo com a posição.

A estrutura resultante que representa a proposta pode ser demonstrada com algumas projeções internas hierarquizadas de acordo com Cinque, o MoodActP representando o Modo Ato de Fala acima do AspRepetitivoP que representa o Aspecto Repetitivo e continua nesta projeção em nível hierárquico.

Esses elementos estão relacionados com as categorias funcionais de Tempo, Aspecto, Modo, Voz e Número, e estão ligados às categorias verbais. As evidências utilizadas por Cinque (1999) foram extraídas de diversas línguas europeias, asiáticas e crioulas, assim, a hierarquia dos especificadores postula uma ordem universal.

Representar uma estrutura que dê conta desses aspectos e formalizar para a decodificação pode ser um tanto complexo se levarmos em conta a pluralidade da categoria tanto em ordem quanto em relação ao escopo na sentença.

3 A Fase Representacional e Implementacional dos Advérbios Altos e Baixos

Após a explanação do conhecimento linguístico e sua representação, o próximo passo é a representação formal e os modelos propostos, no caso, o Cartográfico. Tais modelos são reescritos por uma gramática *G* do tipo gramática livre de contexto (*context-free grammar* - CFG). Seguindo a formalização de Jurafsky e Martin (2010) e por traços de atributo-valor explicitando outros tipos de informações gramaticais e lexicais no próprio item. A estrutura arbórea será delineada de forma mais simplificada, sem a necessidade de ser cindida de forma exaustiva, como normalmente observamos nas estruturas cartográficas.

Ao manter os traços, observarmos o traço TP e as projeções para os advérbios <*Mood speech act, Mood evaluative, Mood evidential, Mood epistemic, Tpast, Tfuture,...* Asp continuative...>. Diante da complexidade em modelar completamente essas funções por criar uma estrutura de regras

gramaticais muito detalhadas, simplificamos as estruturas para abarcar as projeções correspondentes ao item.

Exemplificando, utilizamos, com pequenas modificações a estrutura formalizada por Cinque (1999) e adicionamos a MAV com os traços atribuídos ao advérbio “francamente” representada abaixo. Tais estruturas devem ser unificadas:

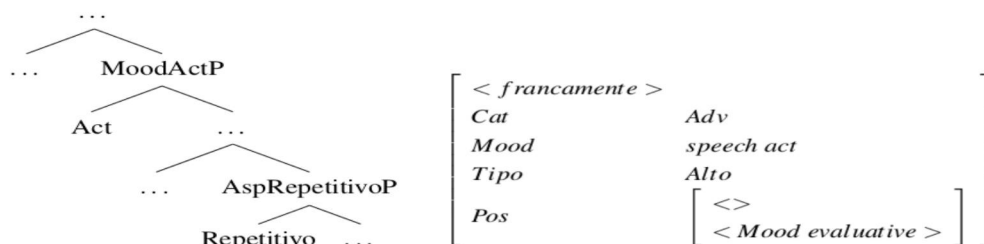


Imagem1: A estrutura Cartográfica e a Matriz de atributo-valor com os traços do advérbio Ato de Fala “francamente”. Fonte: Autor.

Os componentes que adicionamos à MAV demonstram algumas características básicas do item “francamente” e de “provavelmente”. No atributo Cat(egoria) é definido o valor de Adv(érbio) por ser a classe pertencente do item lexical. O atributo, que na verdade é uma projeção, Mood (Modalidade) é atribuído o valor de speech act (Ato de fala) por ser o tipo de projeção pertencente a este advérbio, a sua localização na sentença é definida como um advérbio Alto, o atributo Pos(ição) identifica e localiza a sua ordem hierárquica entre as projeções funcionais, que no caso é “<>” por ser o primeiro valor e não ter nenhum antecessor, não é preenchido, mas o segundo marcador é o <Mood evaluative> (Modalidade avaliativo) que o segue posteriormente, diferente de “provavelmente” que tem os dois itens preenchidos. Os Atos de fala estão em uma camada muito acima das demais, alocados em uma camada funcional específica ou no CP.

Diante do exposto acima, após as representações linguísticas que descrevem e analisam o item lexical e a estrutura, posteriormente são formalizadas os resultados, pode-se implementá-los em modelos teóricos e computacionais, para então, serem decodificados e poderem ser representados em estruturas arbóreas como mostra a imagem abaixo:

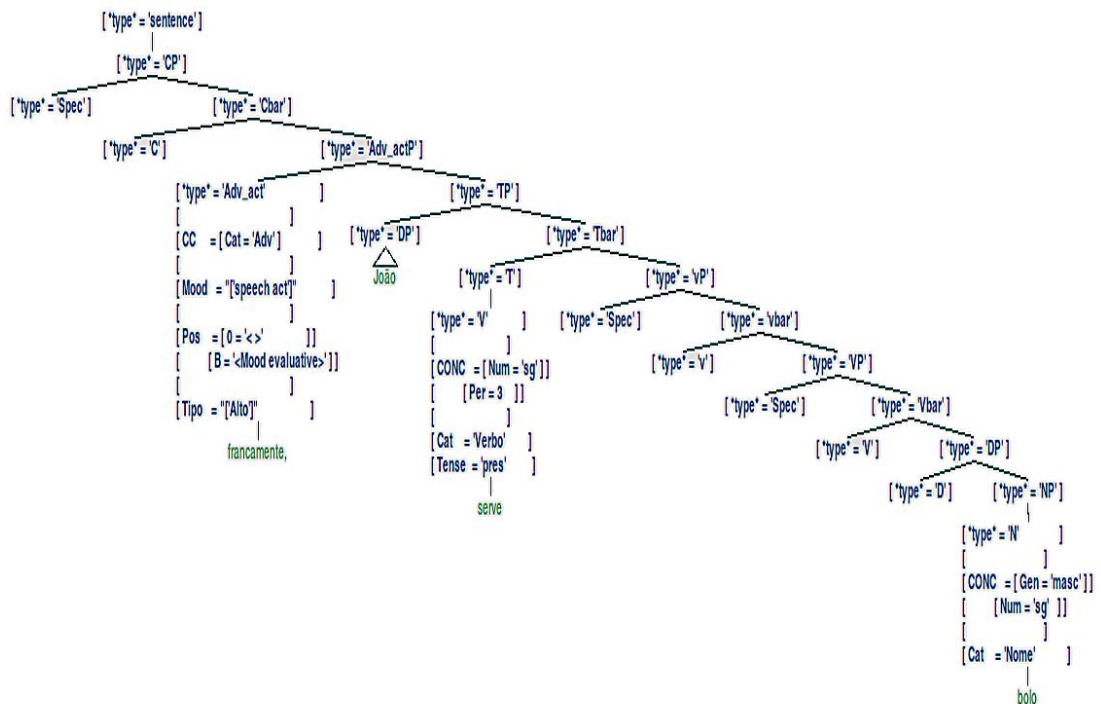


Imagem2: Representação arbórea da sentença com advérbio Ato de Fala “francamente, João serve bolo” possível no Python-NLTK. Fonte: Autor.

Não contrariamos Cinque (1999), mas decidimos, nesta representação, situar o Ato de Fala em um nível mais alto do TP do que dentro do CP. Como também, os traços nos itens lexicais tornam possíveis a identificação dos traços de concordância, tempo e categoria, confirmando a observação sobre os traços MAV enriquecerem a CFG e torná-la mais eficiente na formalização computacional de sentenças.

CONSIDERAÇÕES FINAIS

Os advérbios se apresentam ainda como um fenômeno da língua que necessitam de maiores estudos descritivos e análises detalhadas, tendo em vista este fato, utilizamos os pressupostos da teoria gerativa a partir da perspectiva cartográfica, um empreendimento dentro do Programa Minimalista que visa detalhar (mapear) os traços discursivos das projeções funcionais. A cartografia define os advérbios como especificadores das projeções funcionais, de forma determinada na gramática universal.

Esta característica foi fundamental para utilizamos este modelo para estruturar as regras gerais para a estrutura arbórea e os traços funcionais servirem na matriz de atributo-valor. Os traços de natureza formal e de funcionalidade distribuídos na MAV forneceram explicitações dos traços que os licenciam na sentença, entretanto, a análise evidencia as problemáticas deste item lexical em ambos os campos, além de estar na tumultuada interceção entre as áreas.

REFERÊNCIAS

BRANCO, Antônio; COSTA, Francisco. Hpsg: Arquitetura. In: ALENCAR, Leonel Figueiredo de; OTHERO, Gabriel de Ávila (Ed.). *Abordagens Computacionais da Teoria da Gramática*. Campinas-SP: Mercado de Letras, 2011. p. 219–236.

BERWICK, Roberts C; CHOMSKY, Noam A. *Por que apenas nós? - Linguagem e evolução*. São Paulo:ED. UNESP, 2017.

BRANCO, Antônio; COSTA, Francisco. Hpsg: Arquitetura. In: ALENCAR, Leonel Figueiredo de; OTHERO, Gabriel de Ávila (Ed.). *Abordagens Computacionais da Teoria da Gramática*. Campinas-SP: Mercado de Letras, 2011. p. 219–236.

CINQUE, Guglielmo. *Restructuring and Functional Heads: The cartography of syntactic structures*. New York: Oxford University Press, 2006.

_____. *Adverbs and Functional Heads: A Cross-Linguistic Perspective*. Oxford: Oxford University Press, 1999.

DIAS-DA-SILVA, B. C. *A face tecnológica dos estudos da linguagem: o processamento automático das línguas naturais*. Araraquara, 1996. 272p. Tese (Doutorado em Letras) - Faculdade de Ciências e Letras, Universidade Estadual Paulista.

JURAFSKY, Daniel Saul; MARTIN, James H. *Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition*. 2. ed. New Jersey: Prentice-Hall, 2010.

RIZZI, Luigi. *Relativized minimality*. Cambridge: The MIT Press, 1994.