

RELATÓRIO DE PESQUISA

Construções comparativas em português e sua anotação usando a sintaxe de dependências



OPEN ACCESS

EDITADO POR

- Marta Deysiane Alves Faria Sousa (UFS)
- Adriana Pagano (UFMG)
- Jorge Baptista (INESC-Lisboa)

AVALIADO POR

- Josilene de Jesus Mendonça (UFS)
- Rosana Helena Nunes (USP)

SOBRE OS AUTORES

- Magali Sanches Duran
Análise formal. Investigação. Metodologia. Escrita do manuscrito original. Escrita – revisão e edição.
- Lucelene Lopes
Software. Visualização. Curadoria de dados.
- Thiago Alexandre S. Pardo
Administração do projeto. Supervisão.
- Maria das Graças Volpe Nunes
Escrita – revisão e edição. Curadoria de dados. Validação.

DATAS

- Recebido: 20/11/2022
- Aceito: 20/07/2023
- Publicado: 31/12/2023

COMO CITAR

Duran, M. S.; Lopes, L.; Pardo, T. A. S.; Nunes, M. G. V. (2023). Construções comparativas em português e sua anotação usando a sintaxe de dependências. *Revista da Abralín*, v. 22, n. 2, p. 272-307, 2023.

Magali Sanches DURAN

Núcleo Interinstitucional de Linguística Computacional (NILC-ICMC-USP)

Lucelene LOPES

Núcleo Interinstitucional de Linguística Computacional (NILC-ICMC-USP)

Thiago Alexandre Salgueiro PARDO

Núcleo Interinstitucional de Linguística Computacional (NILC-ICMC-USP)

Maria das Graças Volpe NUNES

Núcleo Interinstitucional de Linguística Computacional (NILC-ICMC-USP)

RESUMO

A capacidade mental de fazer comparações é comum a todos os seres humanos, porém as formas de expressar comparações nas diversas línguas são muito variadas e a gramaticalização de algumas formas deu origem às chamadas “construções comparativas”, que são estruturas complexas, altamente frequentes e versáteis, usadas para expressar avaliações sobre diversos tópicos. Atualmente, devido ao grande volume de avaliações feitas por usuários na Web, cresceu o interesse pelo processamento automático das construções comparativas, com o objetivo de inferir, por exemplo, qual o tópico da comparação, com o que esse tópico está sendo comparado, qual o parâmetro de comparação e se o que está sendo expresso pela comparação é positivo ou negativo. O primeiro requisito para facilitar o processamento automático das construções comparativas, de modo a permitir a extração automática de informações, é que elas sejam anotadas de uma forma lógica e consistente. Com foco nesse objetivo, apresentam-se, neste artigo, diretrizes de anotação de estruturas comparativas em língua portuguesa usando

o conjunto de etiquetas proposto pelas diretrizes de anotação das Universal Dependencies. Os estudos que embasaram essa proposta envolveram a revisão de trabalhos de tipologia das línguas e trabalhos específicos da língua portuguesa. As diretrizes foram testadas no cópulus Porttinari-base (Pardo et al., 2021; Duran et al. 2023), e aperfeiçoadas até alcançarem o estágio aqui apresentado. O conjunto de 122 sentenças anotadas está disponibilizado na Web, dentro do ambiente de anotação Arborator-Grew-NILC (Miranda & Pardo, 2022).

ABSTRACT

The mental ability to make comparisons is common to all humans. However, the ways of expressing comparisons in different languages may vary considerably and the grammaticalization of some forms gave rise to the so-called “comparative constructions”, which are complex, highly frequent, and versatile structures used to express evaluations about various topics. Currently, due to the large volume of evaluations produced by users on the Web, interest has grown in the automatic processing of comparative constructions, with the goal of inferring, for example, which is the comparison topic, what this topic is being compared to, what is the comparison parameter, and whether the comparison is positive or negative. The first requirement to facilitate automatic processing of comparative constructions is to properly tag them in a logical and consistent way. Focusing on this goal, this paper presents guidelines for the syntactic annotation of comparative structures in Portuguese using the set of labels proposed by the Universal Dependencies annotation guidelines. The studies that formed the basis for this proposal involved the review of works on typology of languages and works specific to the Portuguese language. The guidelines were tested on the Porttinari-base corpus (Pardo et al., 2021; Duran et al. 2023) and refined until they reached the stage presented here. The set of 122 annotated sentences is available at Arborator-Grew-NILC framework for corpus annotation (Miranda & Pardo, 2022).

PALAVRAS-CHAVE

Construções comparativas. Relações de dependência. Língua portuguesa. Universal dependencies.

KEYWORDS

Comparative constructions. Dependency relations. Portuguese language. Universal dependencies.

RESUMO PARA NÃO ESPECIALISTAS

Analisar e anotar construções comparativas de uma forma lógica é muito relevante para o Processamento de Linguagem Natural, pois o cóp-us anotado fornece uma base para várias aplicações computacionais, como a extração de informações, a tradução automática e sistemas de perguntas e respostas. Uma construção comparativa é constituída de duas orações: uma principal, contendo o termo que é comparado, e uma subordinada, contendo o termo que é padrão da comparação. Se uma construção comparativa explicitasse todas as palavras, teríamos algo como: “João gosta mais de comer maçãs do que eu gosto de comer maçãs”. No entanto, as palavras repetidas (“gostar de comer maçãs”), que constituem o parâmetro de comparação, são redundantes, e por isso mesmo são suprimidas na oração comparativa, resultando em “João gosta mais de comer maçãs do que eu”. Como muitas vezes (mas não sempre) o parâmetro de comparação abarca o verbo da oração, fica difícil identificar a oração subordinada após a elipse do verbo. Dois marcadores estão presentes na maioria das construções comparativas: o marcador de grau do parâmetro de comparação (no exemplo, o advérbio “mais”) e o marcador de padrão, que introduz a oração subordinada comparativa (no exemplo, a locução “do que”). Identificar as partes de uma construção comparativa é essencial para que se identifique o que está elíptico e possa-se analisar adequadamente o que restou. O presente artigo apresenta diversas variações de construções comparativas na língua portuguesa e propõe uma forma de analisá-las e anotá-las usando a sintaxe de dependências do esquema de anotação de cóp-us Universal Dependencies..

Introdução

Comparar é uma das formas de expressar uma avaliação acerca de um tó-pico e, portanto, a codificação e a decodificação de construções comparativas desempenham importante papel na pragmática das líng-uas. A realização sintática da comparação, contudo, varia muito entre as líng-uas, o que levou muitos linguistas a tentarem estabelecer paralelismo entre as estratégias de comparação (Ultan, 1972; Stassen, 1985, 2012; Dixon, 2008; Beck et al., 2009; Haspelmath, 2017).

Mesmo no âmbito de uma mesma língua, pode haver muita variação nas formas de expressar comparações e, embora seja relativamente simples para um falante identificar uma construção comparativa, sua análise sintática não é trivial, como poderá ser constatado ao longo deste artigo.

Atualmente, há grande interesse corporativo em melhorar o processamento automático de construções comparativas, especialmente para fins de extração e sumarização de opiniões (Ganapathibhotla & Liu, 2008; Varathan et al., 2017). Esse interesse aumenta à medida que proliferam na Web os conteúdos gerados por usuários, grande parte contendo opiniões, pois tornou-se humanamente árduo, quando não impossível, ler todas as opiniões sobre um mesmo tópico.

No entanto, apesar dos avanços no Processamento de Linguagem Natural (PLN), ainda há uma lacuna no que se refere à melhor maneira de anotar sintaticamente as construções comparativas objetivando seu processamento automático. Essa constatação foi feita durante a anotação¹ do *corp*pus Porttinari-base (Pardo et al., 2021; Duran et al. 2023), pois a variabilidade de construções encontradas foi muito maior do que aquela descrita nas gramáticas da língua.

Com vistas a preencher parte dessa lacuna, este artigo reúne subsídios encontrados na literatura e os aplica à análise das construções comparativas em língua portuguesa. A partir dessa análise, propõem-se diretrizes para anotação de construções comparativas na língua portuguesa utilizando a abordagem Universal Dependencies (doravante, UD) (Nivre et al., 2020; De Marneffe et al., 2021).

A UD, baseada na sintaxe de dependências descrita por Tesnière² (2015), vem se mostrando uma forma de anotação de *corp*pus bem sucedida para propósitos de PLN, principalmente quando o intuito é promover o paralelismo entre as línguas.

A metodologia deste estudo é *corp*pus-driven³ (Tognini-Bonelli, 2002; Biber, 2012) pois foram dificuldades de anotação de *corp*pus (ao encontrar casos não descritos pelas gramáticas) que suscitaram a necessidade de estudar a fundo as construções comparativas e propor maneiras lógicas e consistentes de anotá-las.

Uma vez concluído o estudo, elaborou-se uma proposta de anotação sintática, usando a sintaxe de dependências, e voltou-se ao *corp*pus para reanotar as construções comparativas de acordo com as diretrizes elaboradas. Embora não descrito neste artigo, o processo de aperfeiçoamento das diretrizes foi exaustivo, pois a cada vez que as diretrizes eram aplicadas à anotação do *corp*pus, tornava-se evidente a necessidade de novos ajustes. A proposta aqui apresentada é, portanto, fruto de algumas etapas de retroalimentação entre a elaboração das diretrizes e sua aplicação na anotação do *corp*pus. Nas 8.418 sentenças do *corp*pus Porttinari-base, foram identificadas 122 construções comparativas, sendo 60 de

¹ Anotação de *corp*pus é uma atividade essencial para o PLN, na qual etiquetas são atribuídas a partes do texto, de forma a refletir uma análise humana, como, por exemplo, etiquetas morfossintáticas, etiquetas sintáticas e etiquetas de papéis semânticos. O resultado, um *corp*pus anotado num formato legível por máquina, tem por finalidade servir para treinamento de algoritmos de aprendizagem de máquina, que resultam em classificadores que reproduzem automaticamente uma análise semelhante à análise humana.

² Embora Tesnière seja citado como o precursor da sintaxe de dependências pelos idealizadores das Universal Dependencies, sabe-se que essa teoria teve início muito antes, conforme discutem Imrènyi & Mazziotta (2020).

³ A abordagem *corp*pus-driven é aquela em que hipóteses são levantadas a partir de fenômenos observados em *corp*pus. Essa abordagem contrasta com a abordagem *corp*pus-based, na qual hipóteses levantadas por estudiosos buscam comprovação em *corp*pus. Essa abordagem contrasta com a abordagem *corp*pus-based, na qual hipóteses levantadas por estudiosos buscam comprovação em *corp*pus. Pode-se dizer que o conhecimento produzido com a abordagem *corp*pus-driven é incidental, enquanto o conhecimento produzido com a abordagem *corp*pus-based é intencional.

igualdade e 62 de desigualdade. Esse conjunto de sentenças anotadas com sintaxe de dependências está disponível para consulta no ambiente de anotação Arborator-Grew-NILC⁴ (Miranda & Pardo, 2022).

O esquema mostrado na Figura 1 mostra o percurso da pesquisa.

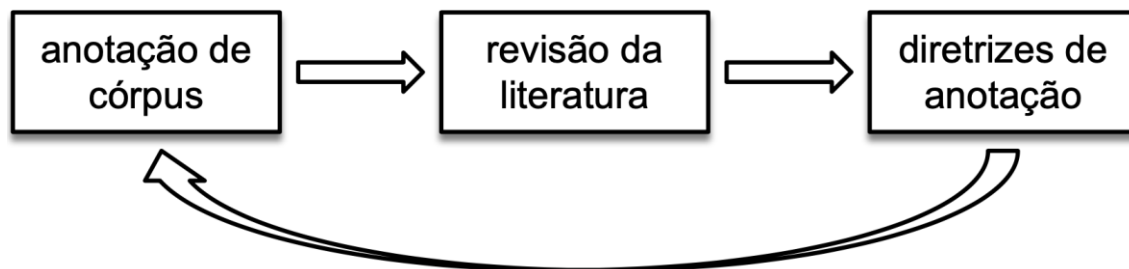


Figura 1 - percurso de pesquisa.

Fonte: elaboração própria

O artigo foi dividido em cinco seções. A Seção 1 tem por objetivo descrever a metodologia de trabalho e fornecer uma breve descrição sobre a gramática de dependências e o esquema de anotação da UD, pano de fundo para compreender as discussões e propostas apresentadas no artigo. Na Seção 2, apresenta-se a revisão bibliográfica dos trabalhos sobre construções comparativas que forneceram informações inspiradoras para este estudo, tanto aqueles que apresentam abordagens de tipologia linguística quanto aqueles que tratam especialmente da língua portuguesa. Na Seção 3, são analisadas as construções comparativas de língua portuguesa e, na Seção 4, é apresentada uma proposta de anotação dessas construções usando a sintaxe de dependências da abordagem UD. Na Seção 5 são feitas reflexões finais e sugestões de trabalhos futuros.

1. Metodologia e fundamentação teórica

O estudo relatado neste artigo foi motivado por dificuldades na anotação de construções comparativas no cópulus Porttinari-base, usando o esquema de anotação da abordagem Universal Dependencies (UD). Isso significa que nem o cópulus nem a abordagem de anotação sintática foram escolhidos para estudar construções comparativas, muito pelo contrário: a construção e a anotação do cópulus precederam e motivaram o estudo das construções comparativas.

O percurso da pesquisa iniciou-se com a recolha de exemplos de construções comparativas durante a tarefa de anotação do cópulus Porttinari-base. Esse cópulus é o primeiro do projeto Porttinari (Pardo et al., 2021; Duran et al. 2023) cuja anotação foi completada. O Porttinari-base é constituído de

⁴ https://arborator.icmc.usp.br/#/projects/Construcoes_Comparativas_Corpus_Porttinari

8418 sentenças avulsas (e não de textos completos) extraídas do *cópus Folha-Kagle*⁵. Seu gênero é jornalístico e pertence a variedade brasileira da língua portuguesa.

Em seguida, foi feita uma revisão crítica da literatura sobre construções comparativas, em busca de subsídios que auxiliassem na análise sintática dos exemplos coletados (Seção 2). Esses subsídios foram utilizados para analisar exemplos construídos para representar o protótipo de cada um dos tipos de construção comparativa do português encontrados no *cópus Portinari-base* (Seção 3). Embora os exemplos construídos sejam reduzidos, com o propósito de destacar as construções comparativas, não houve, e nem haveria sentido em haver, simplificação das estruturas comparativas.

Uma vez compreendida a lógica das construções comparativas no português, o conhecimento foi transposto para a análise de dependências sintáticas das sentenças recolhidas preliminarmente durante a tarefa de anotação de *cópus* (Seção 4). Nessa etapa, foram estabelecidas diretrizes para garantir uniformidade e consistência na anotação de construções comparativas ao longo do *cópus*. Exemplos das sentenças de *cópus* analisadas são comentados e ilustrados com as árvores de dependência produzidas.

Nesta seção apresentam-se breves descrições da gramática de dependências (1.1) e da abordagem UD de anotação de dependências (1.2), pois constituem fundamentos teóricos para todo o trabalho relatado no artigo.

1.1 A gramática de dependências

A gramática de dependências é uma proposta para representação gráfica da análise sintática de uma sentença. A proposta mais completa de gramática de dependências de que se tem notícia foi desenvolvida por Tesnière (Tesnière, 1959 e Tesnière, 2015), na França, no período entre a I e a II guerras mundiais. O linguista faleceu em meados do século XX, sem que suas ideias tivessem tido grande impacto fora da Europa.

A gramática de dependências, ao contrário da gramática de constituintes, não trabalha com o conceito de sintagmas, mas sim de relações entre palavras. As relações são direcionais, partindo do *head* ou governante da relação em direção ao dependente. A representação gráfica gera uma árvore sintática a partir de uma raiz, o *root* da sentença, que coincide com o predicado da oração principal da sentença. A partir do *root* saem arcos (ou “galhos”) apontando para as palavras que constituem termos essenciais da oração, como sujeito e objeto, assim como para modificadores, como advérbios e orações adverbiais. Essas palavras, dependentes do *root*, podem ser, por sua vez, *head* de outras relações de dependência com seus próprios modificadores, como a relação entre um substantivo e seus determinantes ou entre um substantivo e seus adjuntos adnominais. O ressurgimento e a revalorização da gramática de dependências começaram a ocorrer a partir do final do século XX, pois percebeu-se que sua proposta de

⁵ <https://www.kaggle.com/datasets/marlesson/news-of-the-site-folhaol>

representação é mais adequada para o processamento automático de linguagem natural do que outros tipos de representação, como a gramática de constituintes, por exemplo.

1.2 A abordagem da anotação universal dependencies

Em função da aplicabilidade computacional da análise de dependências, surgiu em 2013 o primeiro esforço internacional no sentido de criar um conjunto único de etiquetas para analisar dependências sintáticas em várias línguas, ou seja, uma abordagem multilíngue de anotação sintática (McDonald et al., 2013), contendo etiquetas para nomear as categorias morfossintáticas ou PoS (de *Part-of-Speech*) e as relações de dependência sintática ou *deprel* (de *dependency relation*). Essa iniciativa, realizada por um consórcio de cientistas da linguística computacional de diversos países, evoluiu para um esquema de anotação que passou a ser conhecido como Universal Dependencies (UD) (Nivre et al., 2020; De Marneffe et al., 2021). As diretrizes de anotação da UD são constantemente atualizadas em seu site⁶. Em 2023, o site da UD contém em torno de 200 corpúsculos anotados de 141 diferentes línguas, incluindo português e onze línguas indígenas brasileiras (apuriná, bororo, guajajara, caapor, karo, madi, makurap, mundurucu, nheengatu, tupinambá e xavante). Em sua versão 2, o esquema de anotação UD possui dezessete etiquetas morfossintáticas (PoS tags), sendo seis para palavras de conteúdo (ADJ, ADV, INTJ, NOUN, PROPN, VERB), oito para palavras funcionais (ADP, AUX, CCONJ, DET, NUM, PART, PRON, SCONJ) e três artificiais (PUNCT, SYM, X). Também possui 37 etiquetas de relações de dependência, *deprel*, sendo as principais: **root** (que aponta para a raiz da árvore sintática), **nsubj** (sujeito), **obj** (objeto direto), **obl** (oblíquo, que serve tanto para objetos indiretos quanto para adjuntos adverbiais preposicionados), **amod** (adjetivo adjunto adnominal), **advmod** (adjunto adverbial constituído de advérbio), **acl** (oração adjetiva), **acl:relcl** (oração relativa), **advcl** (oração adverbial), **ccomp** (oração objetiva direta desenvolvida), **xcomp** (oração objetiva direta cujo sujeito é o sujeito ou o objeto da oração principal), **csubj** (sujeito oracional). A lista completa das etiquetas do esquema de anotação UD e as diretrizes de como e quando aplicá-las na língua portuguesa podem ser encontradas em Duran (2021), para *PoS tags* e Duran (2022), para *deprel*. Publicações, ferramentas de anotação e material de apoio à anotação no esquema UD em português podem ser encontrados no site do projeto POeTiSA⁷.

2. Comparativas: revisão crítica da literatura

As construções comparativas já foram alvo de análise em diversas línguas e em diferentes Tradições linguísticas. No português, essas construções são abordadas por gramáticas normativas (Rocha Lima, 2010; Cegalla, 2020; Cunha & Cintra, 2021) dentro do tópico das orações subordinadas adverbiais e, com maior detalhamento e exemplificação, por gramáticas descritivas (Neves, 2000; Castilho, 2010), que classificam essas construções como correlativas.

⁶ <https://universaldependencies.org/>

⁷ <https://sites.google.com/icmc.usp.br/poetisa>

Segundo as gramáticas consultadas, os tipos de comparação possíveis são:

- | | |
|-------------------------------|---|
| (a) de superioridade absoluta | (ex: João é o mais alto de todos) |
| (b) de inferioridade absoluta | (ex: João é o menos inteligente de todos) |
| (c) de superioridade relativa | (ex: João é mais alto do que eu) |
| (d) de inferioridade relativa | (ex: João é menos inteligente do que eu) |
| (e) de igualdade | (ex: João é tão inteligente quanto eu) |

As construções comparativas, desde que entendidas como formas sintéticas e versáteis de expressar comparação entre dois elementos avaliados⁸ referem-se principalmente às comparações de tipo (c) e (d), também chamadas de comparativas de desigualdade, e às comparações de tipo (e), também chamadas de comparativas de igualdade. Portanto, um primeiro foco de distinção que será retomado ao longo deste artigo é esse:

- Comparativas de Igualdade
- Comparativas de Desigualdade

Essa distinção, no entanto, não significa que as comparativas de igualdade sejam totalmente diferentes das comparativas de desigualdade, inclusive porque a negação nas comparativas inverte seu significado, ou seja, a construção de desigualdade torna-se uma igualdade e a construção de igualdade torna-se uma desigualdade:

1. João não é tão alto quanto Luiz. (= João é menos alto que Luiz)
2. João não é menos inteligente que Luiz. (= João é tão inteligente quanto Luiz)

Na literatura do português, contudo, não foi encontrada uma análise sobre as partes constituintes de uma construção comparativa. Esse tipo de análise encontra-se em trabalhos na área da linguística conhecida como Tipologia (Croft, 2004).

Conforme constata Treis (2018) em sua revisão dos trabalhos tipológicos relativos a comparações, os teóricos apregoam a existência de cinco elementos importantes para as construções comparativas (embora possam utilizar diferentes termos para se referirem a esses elementos):

- o termo comparado (doravante “comparado”)
- o parâmetro de comparação (doravante “parâmetro”)
- o termo ou padrão de comparação (doravante “padrão”)
- o marcador que atribui grau ao parâmetro (doravante “marcador de grau”)

⁸ a comparação de superioridade também pode ser considerada uma construção, porém não apresenta variações e não constitui um desafio para a análise sintática.

- o marcador que introduz o termo ou padrão de comparação (doravante “marcador de padrão”)

Na Tabela 1, cada uma dessas cinco categorias de elementos está ilustrada, tomando como exemplos as construções comparativas:

3. João é mais forte do que Luiz.
4. João é tão forte quanto Luiz.

comparado	marcador de grau	parâmetro	marcador de padrão	padrão
João	Mais	[é] forte	do que	Luiz
João	Tão	[é] forte	quanto	Luiz

Tabela 1 - Os cinco elementos das construções comparativas para as sentenças “João é mais forte do que Luiz” e “João é tão forte quanto Luiz”.

Fonte: elaboração própria

Autores como Dixon (2008), Beck et al. (2009) e Haspelmath (2017) utilizam esses cinco elementos para comparar a forma como dezenas e até centenas de línguas realizam a comparação. Eles encontraram, por exemplo, línguas que simplesmente não possuem uma estrutura gramaticalizada para expressar comparações, línguas que não possuem marcadores de grau, línguas que apresentam uma ordem invertida dos cinco elementos, línguas que possuem um único marcador de padrão e línguas que possuem vários deles. Segundo Haspelmath (2017), a única variação não observada em nenhuma língua que apresenta construções comparativas gramaticalizadas é a ausência de marcador de padrão.

A análise tipológica é inspiradora no sentido de apresentar perguntas que podem ser feitas ao se analisar as comparações em português. Uma dessas perguntas é se é possível fazer comparações em português sem usar as construções comparativas gramaticalizadas. A resposta é sim, é possível, e as estratégias são muito semelhantes às encontradas em outras línguas, fortemente dependentes do léxico, como nos exemplos a seguir em que se compara a altura do falante com a do interlocutor:

5. Sua altura **equipara-se** à minha.
6. Você **equipara-se (igual-se)** a mim em altura.
7. Em altura, você **equipara-se** a mim.
8. Sua altura é **igual** à minha.
9. Sua altura e a minha são **iguais**.
10. Você me **supera** em altura.
11. Sua altura **supera (excede)** a minha.
12. Eu sou alto **comparado** a você.

No entanto, os exemplos anteriores não apresentam dificuldades de anotação sintática. A dificuldade está precisamente nas construções comparativas, nas quais grandes porções de informações estão elípticas.

Entender as elipses nas construções comparativas é crucial para identificar os cinco elementos que as constituem. A questão das elipses nas comparativas já foi tema de muitos estudos (por exemplo, Pulman, 1991; Lechner, 2008; Kántor & Bacstai-Atkari, 2012). As elipses nas construções comparativas justificam-se pelos conceitos de “dado” e “novo” explorados pela Pragmática: as informações já dadas na primeira parte da comparação, e que estariam também presentes na segunda parte da comparação, sofrem elipse. Essa elipse tanto atende a princípios de economia (elimina as redundâncias) quanto confere maior saliência às informações novas. Esses estudos, contudo, não identificaram a função das informações elípticas dentro da estrutura comparativa (como será visto mais adiante, as informações elípticas correspondem ao parâmetro de comparação, ou seja, a parcela comum entre as duas partes da comparação).

A discussão que permeia os estudos sobre elipses é se as palavras remanescentes na segunda parte de uma construção comparativa constituem uma oração ou se são um simples sintagma. Essa questão é explorada detalhadamente por Lechner (2008). O autor observa que, se apenas a superfície for considerada, o segundo termo das construções comparativas ora será uma oração, ora será um sintagma, o que, segundo ele, não é satisfatório. Pulman (1991), em um estudo voltado para necessidades da Linguística Computacional, defende que as construções comparativas sejam tratadas como um único sintagma nominal (e não duas orações), embora altamente modificado por elipses.

Em português, Pereira, Pinto e Pratas (2014) utilizam o pronome “eu” para resolver essa questão: uma vez que “eu” é um pronome exclusivo do caso reto, “eu” só pode ser sujeito e, se há um sujeito, há uma oração, mesmo que a maior parte da oração esteja elíptica⁹. Os autores, portanto, argumentam que o segundo termo de uma construção comparativa é sempre uma oração¹⁰:

13. Você é mais forte do que **eu sou forte**.

Outras perguntas inspiradas nos achados dos estudos tipológicos são:

- (a) Quais as funções sintáticas desempenhadas pelos cinco elementos das construções comparativas no português?
- (b) A ordem dos elementos das construções comparativas no português é fixa ou pode mudar?
- (c) Todas as construções comparativas do português apresentam os cinco elementos (comparado, parâmetro, padrão, marcador de grau e marcador de parâmetro)?

⁹ As partes elípticas de uma oração serão representadas por texto tachado em todos os exemplos.

¹⁰ Ao contrário do inglês, por exemplo, em que o pronome oblíquo é utilizado na comparação: *You are stronger than me*.

Com base na revisão teórica aqui apresentada, serão analisadas, na Seção 3, construções comparativas no português identificadas durante a anotação do corpus Porttinari-base. Para fins de didatismo, as sentenças do corpus contendo essas construções foram reduzidas, de forma a ressaltar as estruturas de comparação, foco da discussão. Essa análise buscará respostas às perguntas levantadas, que, no melhor do nosso conhecimento, não foram ainda apropriadamente respondidas na literatura da área.

3. Análise das construções comparativas do português

Observando-se as construções comparativas, percebe-se que, na verdade, elas são constituídas de duas partes simétricas. A fim de ilustrar isso, apresentam-se a seguir exemplos em que as partes simétricas de cada oração estão escritas na mesma cor: comparado e padrão em vermelho e parâmetro em azul. O parâmetro, por ser a base da comparação, é idêntico nas duas partes e, por isso, encontra-se elíptico na segunda parte. Os marcadores de grau e de padrão, que formam a estrutura das construções, estão em verde.

- | | |
|----------------------------------|--------------------------------|
| 14. João é mais simpático | do que Luiz é simpático. |
| 15. João gosta mais de frutas | do que Luiz gosta de frutas. |
| 16. João tem mais dinheiro | do que Luiz tem dinheiro. |
| 17. João trabalha mais | do que Luiz trabalha. |
| 18. João tem mais apartamentos | do que Luiz tem casas. |
| 19. João é mais simpático | do que João é confiável. |
| 20. João gosta mais de frutas | do que João gosta de legumes. |
| 21. João tem mais dinheiro | do que João tem talento. |
| 22. João dorme mais | do que João trabalha. |
| 23. João tem mais medo de aranha | do que João tem medo de cobra. |
| 24. João é mais irmão | do que João é amigo. |

Como pode ser observado, a primeira parte da comparação contém o comparado, o parâmetro e o marcador de grau, ao passo que a segunda parte contém o padrão, o parâmetro (elíptico) e o marcador de padrão.

Apresenta-se a seguir a análise desses cinco elementos que compõem as construções comparativas em português, observando sua correlação com as funções sintáticas (3.1), a alternância em sua ordem de ocorrência (3.2) e casos de supressão de um dos elementos (3.3).

3.1 A função sintática dos cinco elementos das comparativas

O fato de as construções comparativas serem compostas de duas partes simétricas, mesmo que parte delas esteja elíptica, facilita sua análise sintática. Nos próximos subtópicos, cada um dos cinco elementos das construções comparativas é analisado sob o ponto de vista de sua realização sintática.

3.1.1 Comparado

Pelo que se pôde observar durante a tarefa de anotação sintática do *cópus Porttinari-base* (Pardo et al., 2021; Duran et al. 2023), o comparado frequentemente encontra-se na posição de sujeito da oração matriz (como nos exemplos 14, 15, 16 e 17), porém pode ocorrer em qualquer outra função sintática, como no predicado verbal (22) ou nominal (19 e 24) e nos complementos verbais (20 e 21) ou nominais (23). É possível, inclusive, que o comparado seja constituído por palavras de diferentes funções sintáticas (como no exemplo 18, em que o comparado é o sujeito e o complemento do predicado juntos).

Seja qual for a função sintática do comparado, ela será a mesma função exercida pelo padrão na segunda parte da construção comparativa. Assim, se o comparado é sujeito, o padrão é sujeito, como “João” e “Luiz” nas sentenças 14, 15, 16 e 17; se o comparado é um predicado nominal, como nos exemplos 19 e 24, o padrão também é um predicado nominal, como “simpático” e “confiável” na sentença 19 e “irmão” e “amigo” na sentença 24.

3.1.2 Padrão

O padrão tem, na oração subordinada, a mesma função sintática que o comparado tem na oração matriz, porém muitas vezes sua função não fica explícita, pois na oração subordinada quase sempre só o padrão resiste à elipse, já que ele é a única informação nova¹¹. Quando um predicado faz parte do padrão (como “trabalha”, no exemplo 22), o que resiste à elipse na segunda parte da construção comparativa ainda é facilmente identificável como uma oração. Porém, quando o padrão é um sintagma nominal (“Luiz”, “talento” e “amigo” nos exemplos 14, 15, 16, 17, 21 e 24), um sintagma preposicionado (como “de legumes” e “de cobra”, nos exemplos 20 e 23) ou um adjetivo predicativo (“confiável”, no exemplo 19), o que resiste à elipse é mais dificilmente identificado como uma oração, pois justamente o verbo está elíptico. No caso em que o que resiste à elipse são claramente dois sintagmas nominais com funções diferentes (“Luiz” e “casas”, no exemplo 18, respectivamente sujeito e objeto direto), a elipse do verbo fica mais evidente.

3.1.3 Parâmetro

¹¹ Neves (2000) diz que a oração subordinada inteira pode se encontrar elíptica, quando o contexto é suficiente para recuperar o padrão de comparação. Ex: “Isso é mais interessante.” (mais interessante do que algo que foi mencionado antes).

O parâmetro é aquilo que os dois termos comparados (comparado e padrão) têm em comum e serve de base para a comparação, podendo ocorrer em qualquer função sintática. Por ser a base da comparação, o parâmetro é idêntico nas duas partes da construção comparativa (palavras em azul nos exemplos de 14 a 24). Porém, exatamente por ser idêntico, é uma informação redundante (informação dada) e sofre elipse na segunda parte da construção comparativa (palavras tachadas nos exemplos de 14 a 24). As informações elípticas na segunda oração, portanto, correspondem ao parâmetro da comparação. Essa simetria é importante pois, conhecendo o parâmetro, pode-se inferir o que está elíptico (isso pode ser importante para o processamento automático de línguas naturais) ou, inferindo-se o que está elíptico na segunda parte, identifica-se com mais facilidade o parâmetro na primeira parte.

Assim, por exemplo, na sentença 14, o parâmetro é o predicado nominal “é simpático” (compara-se o quanto João e Luiz são simpáticos). Na sentença 15, o parâmetro é o predicado verbal, incluindo verbo e objeto: “gosta de frutas” (compara-se o quanto João e Luiz gostam de frutas). Na sentença 18, o parâmetro é apenas o verbo, “tem” (compara-se simultaneamente a quantidade e o tipo de propriedades que João e Luiz têm). Na sentença 22, o parâmetro é o sujeito “João” (compara-se a quantidade de tempo que João dedica às ações de “dormir” e “trabalhar”). Na sentença 23, o parâmetro é o sujeito juntamente com o predicado verbal (uma construção de verbo suporte): “João tem medo” (comparam-se as coisas das quais João tem medo).

3.1.4 Marcador de grau

O marcador de grau sempre é um advérbio quando estiver modificando um parâmetro expresso por verbo, adjetivo ou advérbio. Porém, quando o parâmetro é um substantivo, o marcador de grau pode ser também um pronome determinante quantificador, como “mais dinheiro”, no exemplo 16, e “mais apartamentos”, no exemplo 18, que significa “uma quantidade maior de dinheiro” e “uma quantidade maior de apartamentos”.

No exemplo 24, contudo, apesar de estar modificando um substantivo (“mais amigo”), “mais” tem sentido de “em maior grau” e não “uma quantidade maior”, tanto que o substantivo não está no plural. Nesse caso, como o substantivo está em posição de predicativo, é justificável sua modificação por um advérbio¹².

O marcador de grau também pode estar “embutido” em um adjetivo ou advérbio que marca o grau morfológicamente, como é o caso, no português, dos adjetivos “maior”, “menor”, “melhor” e “pior” e dos advérbios “melhor” e “pior” (que modificam exclusivamente verbos). Nesses casos, o marcador de grau passa a fazer parte do parâmetro e não pode ser identificado isoladamente.

¹² Ou, caso o verbo de cópula fosse considerado root do predicado nominal, poderia-se dizer que o “mais” estaria modificando o verbo de cópula. Esse não é o caso da anotação UD, adotada nos trabalhos descritos neste artigo.

Os marcadores de grau das comparativas de desigualdade são: “mais” e “menos”. Nas comparativas de desigualdade, os marcadores de grau são: “tão”, e “assim”¹³. A seleção lexical do marcador de grau determina a seleção do marcador de padrão, pois ambos constituem uma mesma estrutura, por exemplo: “mais. . . do que”, “menos. . . do que”, “assim. . . como”, “tão. . . quanto”.

Um destaque importante é para a palavra “tanto”, que só é marcador de grau em sentido estrito quando é advérbio posposto a um verbo (25) ou quando é pronome quantitativo anteposto a um substantivo (26):

25. Você cresceu **tanto** quanto eu ~~eresci~~.
 26. Você tem **tanto** dinheiro quanto eu ~~tenho dinheiro~~.

“Tanto” também participa das construções comparativas como parte da locução conjuntiva subordinativa “tanto quanto”, que é marcador de padrão em comparativas de igualdade (27).

27. João é simpático **tanto quanto** Luiz. (João é simpático no mesmo grau que Luiz)

Nas ocorrências em que “tanto” é uma palavra invariável que se antepõe a um verbo, substantivo ou adjetivo, ela passa a funcionar como conjunção coordenativa aditiva que marca o primeiro termo de uma coordenação, sendo o segundo termo marcado pelas também conjunções coordenativas “quanto” (28, 29 e 30) ou “como” (31, 32 e 33)¹⁴.

- | | |
|--|--------------------------------|
| 28. João tanto trabalha quanto estuda. | (João trabalha e estuda) |
| 29. Tanto João quanto Luiz são simpáticos. | (João e Luiz são simpáticos) |
| 30. João é tanto esperto quanto inteligente. | (João é esperto e inteligente) |
| 31. João tanto trabalha como estuda. | (João trabalha e estuda) |
| 32. Tanto João como Luiz são simpáticos. | (João e Luiz são simpáticos) |
| 33. João é tanto esperto como inteligente. | (João é esperto e inteligente) |

3.1.5 Marcador de padrão

O marcador de padrão é a principal marca de uma construção comparativa. Admitindo-se que a segunda parte de uma comparação é sempre uma oração, o marcador de padrão tem sempre a função de conjunção subordinativa, seja ele constituído de uma única conjunção (como “que” ou “quanto”) ou de uma locução conjuntiva (como “do que”, “assim como” e “tanto quanto”).

¹³ É interessante observar que atualmente o advérbio “assim” não é mais usado como marcador de grau de igualdade, mas sobrevive na oração do “Pai Nosso”: “seja feita a Vossa vontade assim na terra como no céu”.

¹⁴ Neves (2000) classifica essas estruturas aditivas como “comparativas não correlativas”.

O conjunto de marcadores de padrão no português é diferente para comparativas de igualdade e comparativas de desigualdade. A escolha do marcador de padrão normalmente está atrelada à escolha do marcador de grau. Nas comparativas de desigualdade, os marcadores de padrão são “que” e “do que”, além da locução “mais do que¹⁵”, que dispensa o marcador de grau. Nas comparativas de igualdade, há uma variedade maior: “quanto” (sempre em combinação com o marcador de grau “tão”), “como” (com o marcador de grau “assim”) e as locuções que dispensam a presença de marcador de grau, como “assim como”, “bem como”, “tanto quanto” e “que nem”. Alguns marcadores de padrão foram abordados nos estudos de Ranchhod (1991) e Ranchhod & Gioia (1996), que analisaram o uso do “como” nas comparativas de igualdade, e Pereira, Pinto & Pratas (2014), que investigaram as restrições de seleção do “que” e “do que” das comparativas de desigualdade.

Uma das conclusões a que se chega é a de que o papel sintático dos três elementos variáveis das construções comparativas (comparado, padrão e parâmetro) varia muito e, portanto, não existe correlação entre as funções sintáticas e as funções dos cinco elementos das construções comparativas. Isso pode ser observado na Tabela 2, que fixa as partes estruturais (marcador de grau e de parâmetro, destacados em verde) dos exemplos de comparativas já citados, de 14 a 24, e aponta as funções sintáticas do comparado (em vermelho), do padrão (em vermelho) e do parâmetro¹⁶, (em azul).

nro da sentença	sujeito	marcador de grau	predicado	complemento	marcador de padrão	sujeito	predicado	complemento
14	João	mais	simpático		do que	Luiz	simpático	
15	João	mais	gosta	de frutas	do que	Luiz	gosta	de frutas
16	João	mais	tem	dinheiro	do que	Luiz	tem	dinheiro
17	João	mais	trabalha		do que	Luiz	trabalha	
18	João	mais	tem	apartamentos	do que	Luiz	tem	casas
19	João	mais	simpático		do que	João	confiável	
20	João	mais	gosta	de frutas	do que	João	gosta	de legumes
21	João	mais	tem	dinheiro	do que	João	tem	talento
22	João	mais	dorme		do que	João	trabalha	
23	João	mais	tem medo	de aranha	do que	João	tem medo	de cobra
24	João	mais	irmão		do que	João	amigo	

Tabela 2 - Análise das funções sintáticas do comparado, do parâmetro e do padrão.

Fonte: elaboração própria

É importante ressaltar que, na Tabela 2, verbos de cópula foram suprimidos, verbos suporte aparecem na mesma célula do predicador (“ter medo”) e o marcador de grau (“mais”) foi movido para a mesma posição em todas as sentenças, embora na língua possa ocorrer em diferentes posições, dependendo da natureza da comparação.

¹⁵ Apenas nas comparativas antepostas.

¹⁶ Para fins de comparação, omitiram-se os verbos de cópula.

A lista de exemplos apresentada (de 14 a 24) é uma amostra que tem por objetivo apenas ilustrar a alternância das funções sintáticas do comparado (consequentemente, também do padrão) e do parâmetro. Foram utilizados exemplos simplificados, pois os exemplos de *córpus* são mais longos. Na verdade, existe uma infinidade de formas de usar a estrutura das construções comparativas. Neves (2000, p. 893 a 913), apresenta uma extensa lista de exemplos de construções comparativas, inclusive um em que o comparado e o padrão são sujeitos oracionais¹⁷.

34. *E mencionar um cadáver ensanguentado causa menos impacto do que mostrá-lo.* (Neves, 2000, p. 912).

Nesse exemplo, os cinco elementos são:

comparado: mencionar um cadáver ensanguentado

marcador de grau: menos

padrão: mostrá-lo (mostrar um cadáver ensanguentado)

marcador de padrão: do que

parâmetro: causar impacto

Um tipo de comparação muito comum no *córpus* Porttinari-base, mas não mencionado nas gramáticas consultadas, é a comparação entre o real (fato) e o irreal (desejo ou suposição).

35. Acho que *isso vai ser mais difícil do que eu imaginava que isso seria difícil.*

36. *As coisas pioraram mais do que nós pensávamos que as coisas piorariam.*

37. *Você comprou mais coisas do que você precisava comprar coisas.*

38. *O sucesso foi tão grande quanto ele calculava que o sucesso seria grande.*

Essas construções têm duas partes quase simétricas. O real é o comparado e o irreal é o padrão. Contudo, o comparado confunde-se com o parâmetro: enquanto o comparado é uma afirmação contida na oração matriz, o parâmetro é essa mesma afirmação, só que atemporal. Por exemplo, no exemplo 35, o comparado é “isso vai ser difícil”, o padrão é “imaginávamos que isso seria difícil” e o parâmetro é “isso ser difícil”. A principal diferença entre as duas partes desse tipo de comparação é a presença do verbo que marca o irreal na segunda parte (“imaginar”, “pensar”, “precisar”, “calcular”). A estrutura dessas construções mantém-se regular, ou seja, apresenta o marcador de grau e o marcador de padrão.

3.2 A ordem dos elementos das comparativas

¹⁷ A autora apresenta esse exemplo no item 3.2.5 Comparação entre predicções. Ela não emprega os termos “comparado” e “padrão” aqui adotados.

A ordem canônica das construções comparativas é:

comparado → marcador de grau → parâmetro → marcador de padrão → padrão

Porém há casos de alternância em que a oração que contém o padrão aparece anteposta (39) ou intercalada (40) à oração que contém o comparado.

39. **Você, mais do que eu, sabe de quem é a culpa.**

comparado (você), marcador de padrão (mais do que), padrão (eu), parâmetro (saber de quem é a culpa)

40. **Assim como eu, você desistiu.**

marcador de padrão (assim como), padrão (eu), comparado (você), parâmetro (desistir)

Nesses casos, observa-se que o marcador de grau move-se juntamente com o marcador de padrão, desempenhando uma única função (a de marcar o padrão).

Uma consequência da anteposição do padrão é a impossibilidade de transformar a comparativa de igualdade em uma desigualdade pelo uso da negação, pois a negação afeta ambos, comparado e padrão:

41. **Assim como eu, você não desistiu.** (eu e você não desistimos)

A negação na ordem canônica ficaria ambígua:

42. **Você não desistiu assim como eu.** (você não desistiu, mas eu desisti/não desisti?)

Outra alternância de ordem observada é a do marcador de grau. Normalmente, o marcador de grau antecede o parâmetro, porém, quando o parâmetro inclui um predicado verbal, a posição do marcador pode ser anteposta (43) ou posposta (44) ao verbo modificado.

43. João **mais dorme** do que trabalha.

44. João **dorme mais** do que trabalha.

Quando comparado e padrão são coordenados na função de sujeito, a construção comparativa deixa de existir, pois as duas orações que a caracterizam se transformam em uma única oração:

45. Tanto João quanto Luiz são inteligentes. (João e Luiz são inteligentes)

Nesses casos, os marcadores de grau e de padrão (“tanto” e “quanto”) mudam de função e passam a ser conjunções coordenativas. Uma consequência disso é que a negação, por aplicar-se a uma única oração, afeta igualmente ambos os termos coordenados:

46. Tanto João quanto Luiz não são bonitos. (João e Luiz não são bonitos)

3.3 A ausência de um dos elementos das comparativas

Em língua portuguesa, é comum a ocorrência do marcador de grau sem que seja apresentado um padrão de comparação. Na ausência da segunda oração da construção comparativa, com o marcador de padrão e o padrão, pode-se dizer que a construção comparativa não se realizou:

47. Isso é menos importante agora. (menos importante do que o quê?)

Contudo, há ocorrências de construções comparativas em que um dos cinco elementos está elíptico. O caso mais comum é a ausência de marcador de grau nas comparativas de igualdade. Como o grau dessas comparativas é igual, ou seja, nem “mais” nem “menos”, o próprio marcador de padrão é suficiente para caracterizar a comparação (48, 49):

48. Ele descasca laranja que nem eu descasco laranja.

49. Seja obediente como seu irmão é obediente¹⁸.

O marcador de grau também desaparece de sua posição e combina-se com o marcador de padrão em construções de desigualdade que apresentam o padrão antes do comparado (50).

50. Mais do que ninguém devia dar o exemplo de bom comportamento, eles deviam dar o exemplo de bom comportamento.

Esse mesmo fenômeno de “junção” de marcadores de grau e de padrão é observado em algumas construções comparativas de igualdade independentemente de o padrão ocorrer antes (49, 50) ou depois do comparado (53, 54).

51. Tanto quanto ele é seu irmão, eu sou seu irmão.

52. Assim como os jogadores ficaram decepcionados, o técnico ficou decepcionado.

53. Eu sou seu irmão tanto quanto ele é seu irmão.

54. O técnico ficou decepcionado, assim como os jogadores ficaram decepcionados.

¹⁸ Nesta sentença, a elipse do sujeito ocasiona a elipse do comparado.

Outro caso encontrado no *cópus* Porttinari-base foi de elipse do padrão, situação em que a oração subordinada fica reduzida ao marcador de padrão:

55. Sou teimoso, mas **ele** é **tão teimoso quanto eu sou teimoso**.

(Sou teimoso, mas ele é tão teimoso quanto.)

Esse caso de elipse justifica-se porque há uma oração coordenada anterior que antecipa o padrão (eu) e o parâmetro (sou teimoso) que pertenceriam à segunda oração da construção comparativa.

Por fim, foram encontrados casos de elipse na oração matriz, em que uma oração relativa foi reduzida ao predicativo apenas, passando a funcionar superficialmente apenas como adjetivo.

56. Ouvimos o batuque do Oludum, **que é tão irreverente quanto as imagens criadas por o americano Andy Warhol são irreverentes**.

(Ouvimos o batuque do Oludum, tão irreverente quanto as imagens criadas pelo americano Andy Warhol.)

57. Comprou uma bicicleta **que é mais moderna do que a minha bicicleta é moderna**.

(Comprou uma bicicleta mais moderna do que a minha.)

Nesses casos, o ideal seria haver coerência: se a elipse no segundo termo da comparação é levada em conta para considerá-lo uma oração, a elipse no primeiro termo da oração também deveria sê-lo. Desta forma, ambos os termos da comparação seriam interpretados como orações. Contudo, no *cópus* Porttinari-base, ainda não se adotou esse procedimento. Os adjetivos modificados pelos marcadores de grau (“mais” e “menos”) e os adjetivos comparativos (“melhor”, “pior”, “maior”, “menor”) ainda são anotados como modificador adjetival e não como oração adjetiva, sempre que não estiverem em posição de predicativo, ou seja, sempre que não forem acompanhados de verbo de cópula. Na verdade, isso implicaria reconhecer a existência de orações adjetivas reduzidas de predicativo, ou seja, orações em que o pronome relativo e o verbo de cópula estariam elípticos. Restaria definir como distinguir adjetivos predicativos e adjetivos não predicativos se superficialmente houver apenas um substantivo e um adjetivo. A vírgula pode ser marca da elipse, como no exemplo 56, porém há casos, como no exemplo 57, que não apresentam vírgula entre o substantivo e o adjetivo. A marca que indicaria uma oração elíptica, nesse caso, teria que ser o simples fato de o adjetivo ser modificado por um advérbio.

4. Proposta

A proposta de anotação de construções comparativas usando a sintaxe de dependências aqui apresentada emprega o conjunto de etiquetas do modelo Universal Dependencies. São 37 etiquetas de relações de dependências e 17 etiquetas de categorias morfossintáticas. Essas etiquetas e as instruções de seu uso encontram-se descritas de forma concisa em Duran et al. (2022a) e de forma detalhada em Duran (2022b) e Duran (2021).

No grupo de estudos da UD dedicado às construções comparativas¹⁹ há muita discussão sobre onde ancorar o *head* da relação que une os dois segmentos dessas construções (utiliza-se aqui o termo “segmentos” pois não há consenso na UD de que se trata de duas orações, pois as línguas diferem na forma de realização dos elementos da comparação). Grupos de algumas línguas afirmam que se trata de uma relação entre duas orações; outros grupos defendem que se trata de uma relação entre uma oração e um sintagma; e outros ainda que se trata de uma relação entre dois sintagmas. Além do motivo natural dessas divergências, que é a diferença entre as línguas, existe o agravante de alguns decidirem considerar as elipses para classificar um segmento como oração e outros decidirem desconsiderá-las, julgando apenas o que está na superfície para classificar os segmentos.

Em português, pelo que foi visto até agora, as construções comparativas são constituídas de duas orações, com ressalva à situação em que um adjetivo claramente predicativo é anotado como simples modificador adjetival por faltarem marcas superficiais que induzam a sua interpretação como oração²⁰.

Sendo assim, a única relação de dependência da UD adequada para unir as duas partes da construção comparativa em português é a relação **advcl** (adverbial clause), usada para anotar orações subordinadas adverbiais.

Contudo, a dúvida comum dos anotadores é sobre qual token da construção deve ser o *head* e qual token deve ser o dependente dessa relação **advcl**. Não se sabe se essa dúvida é gerada pelo “descasamento” entre a função sintática dos tokens da sentença e sua função na construção comparativa ou se é motivada pelo grande número de elipses.

Acredita-se que a análise dos cinco elementos das construções comparativas possa auxiliar o anotador na tarefa de identificar o *head* e o dependente da **advcl**. A fim de facilitar a anotação, estipulou-se uma regra clara: o *head* da relação **advcl** é o predicado da oração que contém o comparado e o dependente da relação **advcl** é o predicado da oração que contém o padrão. No caso de o predicado de uma oração estar elíptico, a relação de dependência é atribuída ao token mais importante remanescente na oração.

Aqui cabe um esclarecimento: na UD, existe uma regra de “herança” do *head* das orações que sofrem elipse do predicado. O “herdeiro” natural é o verbo auxiliar ou o verbo de cópula e, na ausência destes, a ordem de prioridade é a seguinte:

¹⁹ <https://universaldependencies.org/workgroups/comparatives.html>

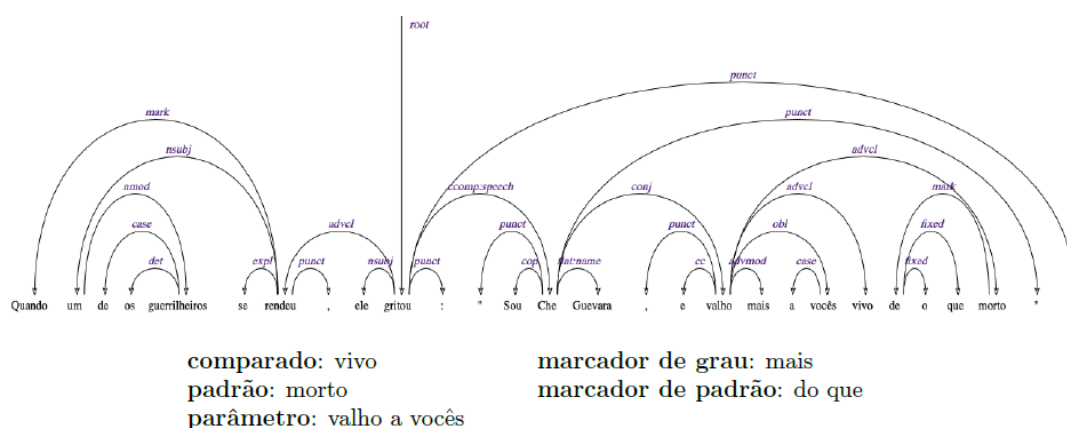
²⁰ As principais marcas usadas para identificar um adjetivo predicativo são a presença do verbo de cópula e/ou a introdução feita por conjunção subordinativa. Exemplo desse último caso são as orações adverbiais concessivas como: “Ainda que pobre, anda bem vestido” (Ainda que seja pobre, anda bem vestido).

nsubj > obj > iobj > obl > advmod > csubj > xcomp > ccomp > advcl

(respectivamente: sujeito, objeto direto, objeto indireto, complemento oblíquo, adjunto adverbial, sujeito oracional, complemento verbal oracional aberto, complemento verbal oracional fechado, oração adverbial).

Para ilustrar a aplicação dessas regras, diferentes árvores de dependências do corpus Porttinaribase são apresentadas a seguir, acompanhadas da análise dos cinco elementos das construções comparativas, com comentários na sequência.

Na sentença da Figura 2 (*Quando um dos guerrilheiros se rendeu, ele gritou: “Sou Che Guevara, e valho mais a vocês vivo do que morto.”*), observa-se que o predicado da oração matriz, que contém o comparado (“vivo”), é “valho”, e o predicado da oração subordinada, que contém o padrão (“morto”), seria “valho”, porém, devido à elipse desse verbo, a relação de dependência é ancorada em “morto”. Destaca-se também a anotação do marcador de grau, “mais”, como **advmod**, que é a relação que liga verbos a modificadores adverbiais. O marcador de padrão, a locução conjuntiva “do que”, tem seus tokens reunidos pela relação **fixed**, que é a relação usada na UD para constituir multipalavras funcionais.



Quando um de os guerrilheiros se rendeu , ele gritou : “~~eu~~ Sou Che Guevara , e ~~eu~~ valho mais a vocês vivo de o que ~~eu valho a vocês~~ morto ” .

Figura 2 - Anotação da sentença FOLHA DOC000018 SENT051.

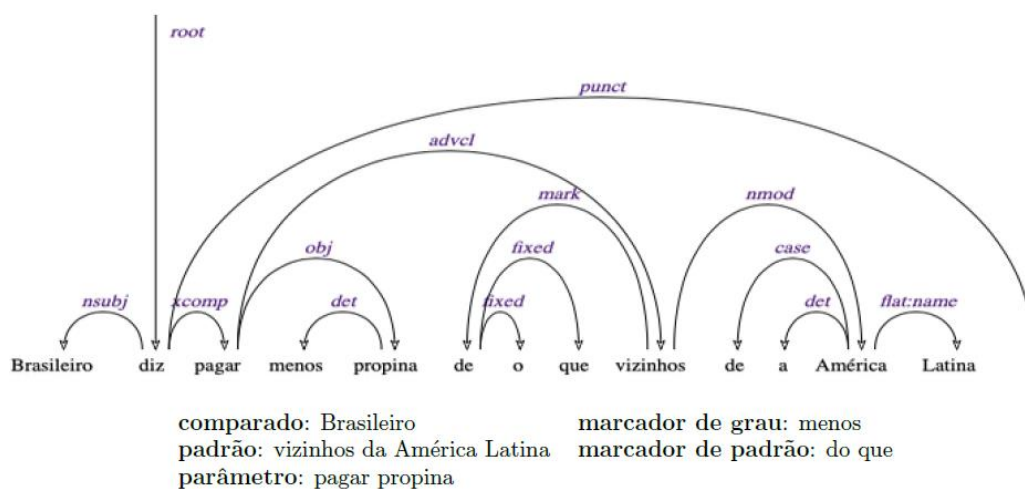
Fonte: elaboração própria

Na sentença da Figura 3 (*Brasileiro diz pagar menos propina de o que vizinhos de a América Latina.*), observa-se que o comparado é sujeito do verbo “diz” e, por consequência, também sujeito do verbo “pagar”²¹. O verbo “pagar” faz parte do parâmetro e, por isso, está elíptico na oração que contém o padrão (“vizinhos da América Latina”). O sujeito dessa oração é “vizinhos”, que assume o *head* da oração

²¹ Os verbos “dizer” e “pagar” estão unidos pela relação **xcomp**, usada para unir orações que têm o sujeito do segundo controlado pelo sujeito do primeiro.

subordinada pelo fato de o predicado, “pagar”, estar elíptico. Observa-se também na Figura 3 que o marcador de grau “mais” não está marcando grau, mas sim quantidade, e por isso é anotado com a relação **det**, que une os nominais a seus determinantes.

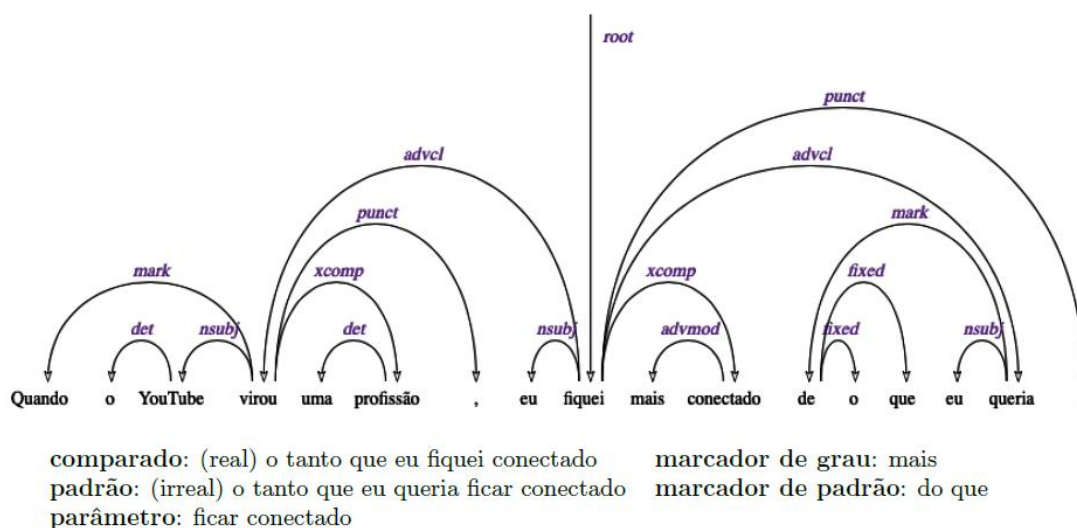
A Figura 4 (*Quando o YouTube virou uma profissão, eu fiquei mais conectado do que eu queria.*) ilustra uma comparação entre o real (o tanto que eu fiquei conectado) e o irreal (o tanto que eu queria ficar conectado). O verbo que marca o irreal é o volitivo “querer” e, por ser *head* da oração subordinada, é o dependente da relação **advcl** que caracteriza a construção comparativa.



Brasileiro diz pagar menos propina de o que vizinhos de a América Latina ~~pagam propina.~~

Figura 3 - Anotação da sentença FOLHA DOC000059 SENT035.

Fonte: elaboração própria

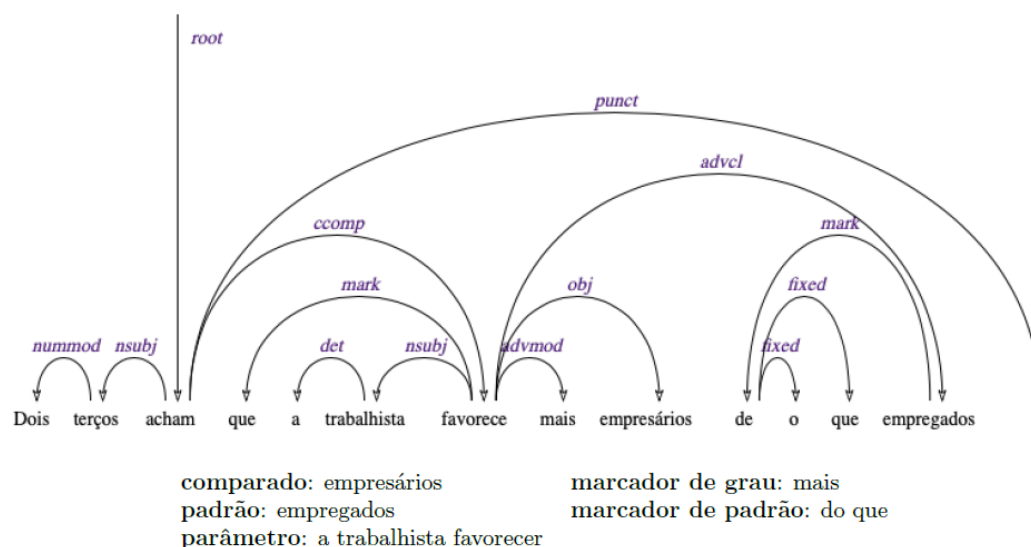


Quando o YouTube virou uma profissão, eu fiquei mais conectado de o que eu queria ficar conectado.

Figura 4 - Anotação da sentença FOLHA DOC000097 SENT028.

Fonte: elaboração própria

Na Figura 5 (Dois terços acham que a trabalhista favorece mais empresários do que empregados.), observa-se que o marcador de grau “mais” está modificando o verbo “favorece” e não o substantivo “empresários”, e por isso é anotado como **advmod**. O comparado e o padrão são objetos diretos de suas respectivas orações, mas só o padrão é *head* de oração, devido à elipse do sujeito e do verbo na oração subordinada. Sendo assim, a relação **advcl** que configura a construção comparativa se dá entre o verbo “favorece”, predicado da oração matriz, e o objeto da oração subordinada, “empregados”.



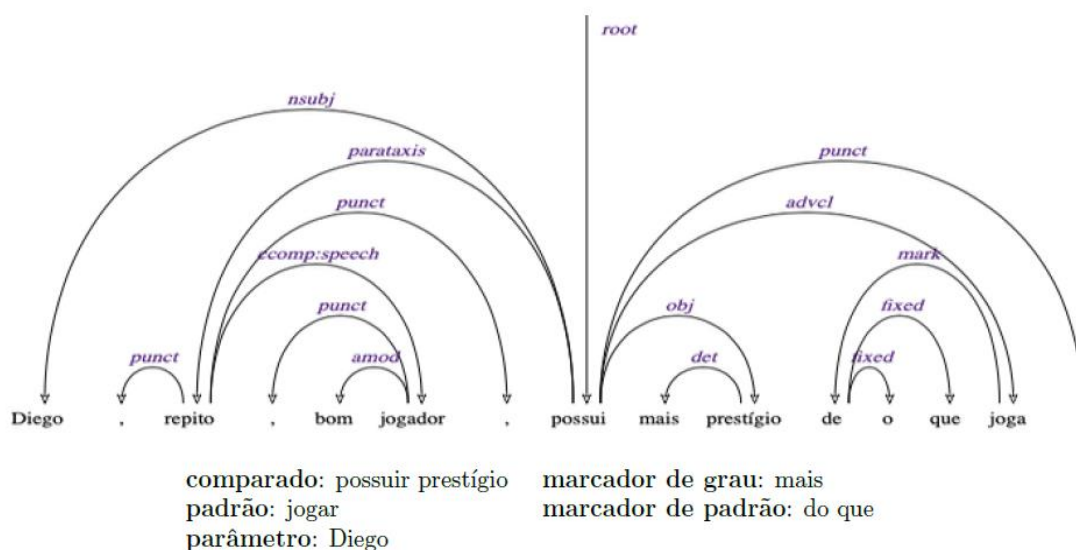
Dois terços acham que a trabalhista favorece mais empresários de o que a trabalhista favorece empregados.

Figura 5: Anotação da sentença FOLHA DOC001771 SENT030.

Fonte: elaboração própria

A Figura 6 (*Diego, repito, bom jogador, possui mais prestígio do que joga.*) ilustra um caso raro no *cópus*, em que o sujeito é o parâmetro e o comparado e o padrão têm naturezas diferentes: “possuir prestígio” e “jogar”. A construção comparativa apresenta verbos nas duas orações e, por isso, não constitui dificuldade para a anotação. Destaca-se aqui a ocorrência do marcador de grau “mais” entre o verbo e o objeto direto, interpretado como quantificador do substantivo “prestígio” (e não como advérbio modificador de “possui”) e, por isso, anotado com a relação **det**.

Figura 7 (*Nunca, quem teve doze pontos a menos do que o líder conseguiu ser campeão.*) ilustra uma construção comparativa dentro de uma oração relativa que modifica o sujeito da oração principal da sentença. As duas orações que participam da construção comparativa têm o mesmo verbo, contudo, na oração subordinada, o verbo está elíptico e o sujeito, “líder”, passa a ser o *head* da oração. Sendo assim, a relação subordinada **advcl** se dá entre o verbo “teve”, da oração matriz, e o sujeito “líder”, da oração subordinada.

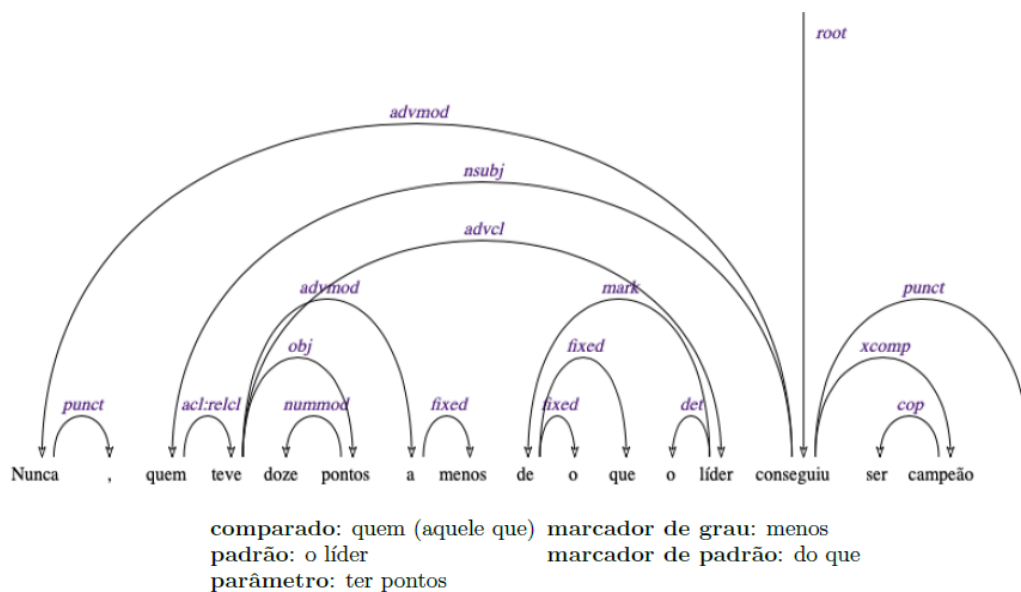


Diego, repito, bom jogador, possui mais prestígio de o que Diego joga.

Figura 6 - Anotação da sentença FOLHA DOC001186 SENT013.

Fonte: elaboração própria

Merece destaque a anotação do marcador de grau, “menos”, que ocorre precedido de uma preposição e, por isso, constitui uma expressão multpalavra funcional (“a menos”), cujos tokens são unidos pela relação **fixed**.

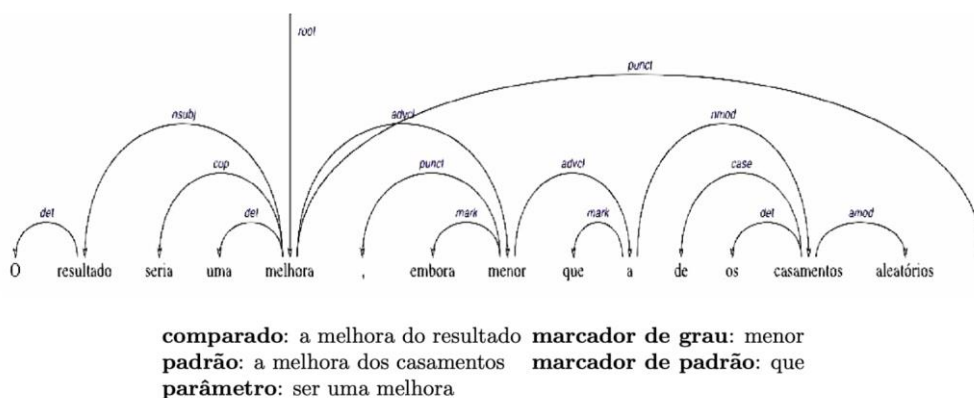


Nunca, quem teve doze pontos a menos de o que o líder ~~teve pontos~~ conseguiu ser campeão.

Figura 7 - Anotação da sentença FOLHA DOC003023 SENT049.

Fonte: elaboração própria

A Figura 8 (O resultado seria uma melhora, embora menor que a melhora dos casamentos aleatórios.) ilustra a elipse do verbo de cópula tanto na oração matriz quanto na oração subordinada. A elipse na oração matriz se justifica pelo fato de ela constituir uma oração adverbial concessiva, introduzida pela conjunção subordinativa “embora”, marca suficiente para configurar uma oração, mesmo na ausência de um verbo. O adjetivo comparativo “menor”, além de atuar como marcador de grau, é também o head da relação **advcl** comparativa, por ser o predicado nominal da oração matriz da construção comparativa.



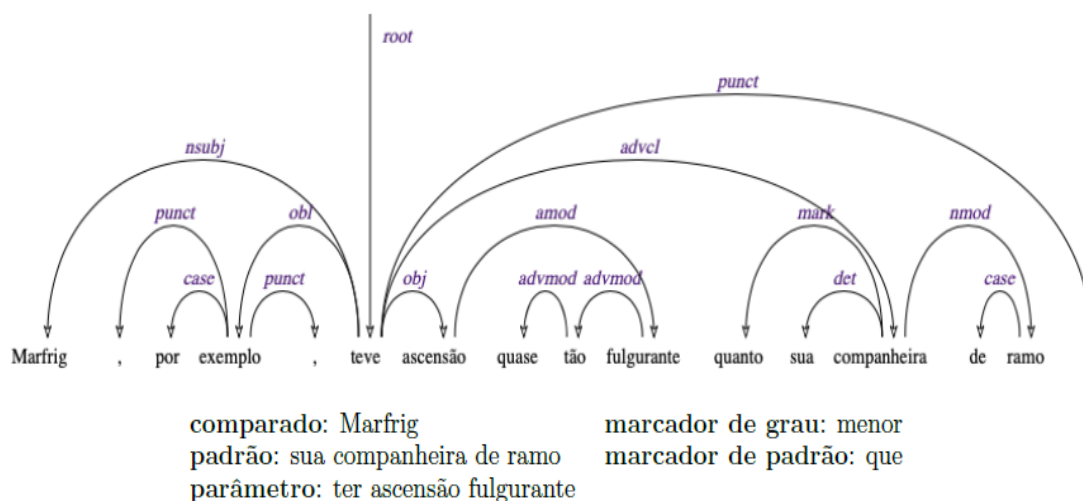
O resultado seria uma melhora, embora ~~o resultado fosse uma melhora~~ menor que a melhora de os casamentos aleatórios ~~seria~~.

Figura 8 - Anotação da sentença FOLHA DOC003118 SENT027

Fonte: elaboração própria

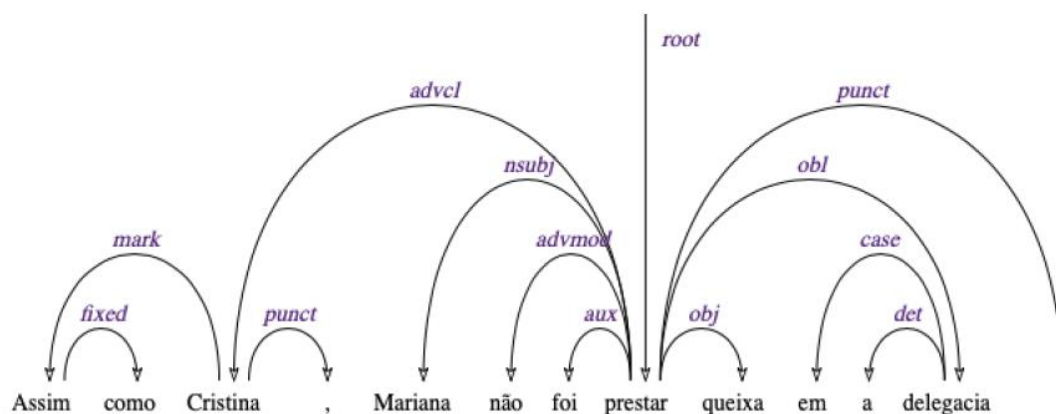
A Figura 9 (*Marfrig, por exemplo, teve ascensão quase tão fulgurante quanto sua companheira de ramo.*) ilustra uma construção comparativa prototípica, com ordem tradicional dos elementos da comparação, ou seja, com o comparado antes do padrão. A construção apresenta marcador de grau e marcador de padrão separados. O *head* da relação *advcl* é o verbo “teve” e o dependente é substantivo “companheira”, token que é sujeito da oração subordinada e herdeiro da função de *head* da oração por causa da elipse do predicado “teve”.

A Figura 10 (*Assim como Cristina, Mariana não foi prestar queixa na delegacia.*) exhibe a anotação de uma construção comparativa de igualdade na qual a oração subordinada encontra-se anteposta à oração principal. Nas comparativas de igualdade, como já discutido, é comum haver supressão do marcador de grau. O que outrora devia ser marcador de grau, “assim”, passou a fazer parte do marcador de padrão “assim como”, anotado como uma multipalavra funcional, cujos tokens são unidos pela relação *fixed*, seguindo as diretrizes da UD.



Marfrig, por exemplo, teve ascensão quase tão fulgurante quanto sua companheira de ramo ~~teve ascensão fulgurante~~.

Figura 9 - Anotação da sentença FOLHA DOC003203 SENT017
 Fonte: elaboração própria



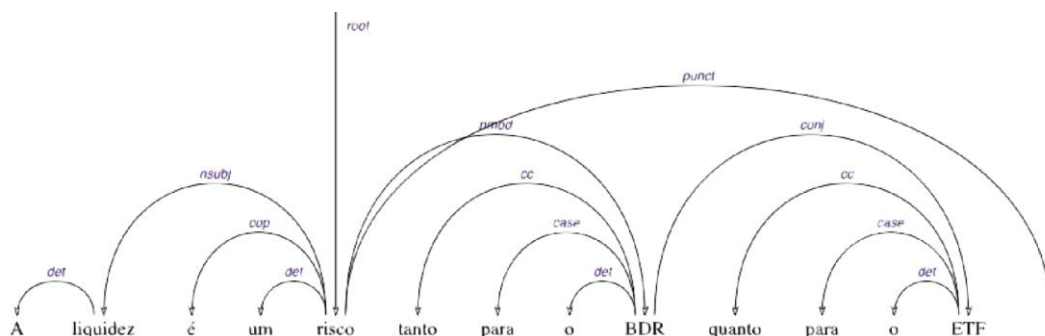
comparado: Mariana
 padrão: Cristina
 parâmetro: ir prestar queixa na delegacia

marcador de grau: inexistente
 marcador de padrão: assim como

Assim como Cristina não foi prestar queixa na delegacia, Mariana não foi prestar queixa em a delegacia .

Figura 10 - Anotação da sentença FOLHA DOC001176 SENT044
 Fonte: elaboração própria

A Figura 11 (A liquidez é um risco tanto para o BDR quanto para o ETF.) ilustra uma construção que fica entre uma comparativa de igualdade e uma coordenação aditiva. Optou-se por anotar casos assim como coordenações aditivas, considerando os marcadores “tanto” e “quanto” como conjunções aditivas, dependentes da relação **cc**. A construção pode ser parafraçada por uma coordenação aditiva prototípica: “A liquidez é um risco para o BDR e para o ETF”.



comparado: BDR
 padrão: ETF
 parâmetro: A liquidez é um risco

marcador de grau: inexistente
 marcador de padrão: tanto quanto

A liquidez é um risco tanto para o BDR quanto a liquidez é um risco para o ETF.

Figura 11 - Anotação da sentença FOLHA DOC000099 SENT036
 Fonte: elaboração própria

5. Conclusão

Pelo que foi exposto, a revisão da literatura, especialmente dos estudos tipológicos, auxiliou a análise das partes integrantes das construções comparativas. Os cinco elementos constituintes das construções comparativas (comparado, marcador de grau, parâmetro, marcador de padrão e padrão) foram utilizados para analisar diversas ocorrências de orações comparativas em português. Essa análise levou à conclusão de que a estrutura comparativa gramaticalizada é extremamente versátil e possibilita a realização de diversos tipos de comparações, o que dificulta a identificação de padrões. Em tal estrutura, o marcador de grau e, principalmente, o marcador de padrão são elementos lexical e sintaticamente constantes, ao passo que o comparado, o parâmetro e o padrão são os elementos lexical e sintaticamente variáveis.

O comparado e o marcador de grau encontram-se na oração matriz, ao passo que o padrão e o marcador de padrão encontram-se na oração subordinada. O parâmetro, por sua vez, por ser a base da comparação, encontra-se repetido nas duas orações, porém é sempre explícito na oração matriz e (quase sempre) elíptico na oração subordinada. Além disso, comparado e padrão são elementos simétricos e desempenham a mesma função sintática em suas respectivas orações.

A conclusão mais importante para fins de processamento de línguas naturais foi a de que os cinco elementos das construções comparativas, úteis para extração de informações, parecem não ter correlação com as funções sintáticas. Essa não correlação entre esses dois tipos de análise, aliás, sugere que os elementos das construções comparativas devam ter uma “camada” própria para anotação²².

Com base nessas conclusões, podem ser tomadas decisões de anotação sintática (no caso do presente artigo, na anotação do *corp*us Porttinari-base), sendo a principal delas a de anotar todas as sequências introduzidas por locuções conjuntivas comparativas de desigualdade como orações adverbiais, usando a relação de dependência **advcl**. Quanto às construções de igualdade, decidiu-se tratar como coordenações aditivas aquelas em que os marcadores “tanto/como” e “tanto/quanto” podem ser, sem prejuízo semântico, substituídos pela conjunção “e”. Outra decisão importante foi a de anotar como *head* da **advcl** sempre o predicado da oração matriz da construção comparativa. Nos casos em que um adjetivo é possivelmente uma relativa reduzida, com pronome relativo e verbo de cópula elípticos, decidiu-se reconhecer o adjetivo como *head* da relação **advcl**.

Um importante esclarecimento a ser feito é o fato de que todas as decisões de anotação respeitaram princípios básicos da abordagem UD. Essa abordagem é criticada, contudo, por não admitir a atribuição de *head* de relações de dependência a palavras funcionais (Kim & Gerdes, 2019). A eliminação dessa restrição poderia alterar significativamente a forma de anotação das construções comparativas, porém essa discussão foge ao escopo do presente artigo.

²² Na UD, diversas informações semânticas podem ser combinadas às funções sintáticas, com o uso de uma extensão no nome da relação de dependência, como é o caso de **oblagent** para marcar o agente das passivas, porém esse recurso não atende às necessidades das construções comparativas.

Além das árvores aqui apresentadas, este artigo disponibiliza o conjunto das construções comparativas anotadas no corpus Porttinari-base (link na declaração de disponibilidade de dados). Um trabalho futuro será expandir esse conjunto a fim de utilizá-lo como corpus de treinamento para observar se a anotação proposta é lógica o suficiente para que a automatização da tarefa seja bem sucedida. Além disso, espera-se que as análises dos cinco elementos que constituem a lógica das construções comparativas possam inspirar trabalhos de anotação dedicados a esse tipo de construção.

Informações complementares

Avaliação e resposta dos autores

Avaliação: <https://doi.org/10.25189/rabralin.v22i2.2123.R>

Resposta dos autores: <https://doi.org/10.25189/rabralin.v22i2.2123.A>

Editores

Marta Deysiane Alves Faria Sousa

Afiliação: Universidade Federal de Sergipe

ORCID: <https://orcid.org/0000-0003-1297-2037>

Adriana Pagano

Afiliação: Universidade Federal de Minas Gerais

ORCID: <https://orcid.org/0000-0002-3150-3503>

Jorge Baptista

Afiliação: Universidade do Algarve - INESC-ID Lisboa

ORCID: <https://orcid.org/0000-0003-4603-4364>

RODADAS DE AVALIAÇÃO

Avaliadora 1: Josilene de Jesus Mendonça

Afiliação: Universidade Federal de Sergipe

ORCID: <https://orcid.org/0000-0001-8291-6386>

Avaliador 2: Hadinei Ribeiro

Afiliação: Universidade de São Paulo

ORCID: <https://orcid.org/0000-0002-3157-6366>

AVALIADORA 1

O artigo Construções comparativas em português e sua anotação usando a sintaxe de dependências se propõe a apresentar diretrizes de anotação de estruturas comparativas usando o conjunto de etiquetas do modelo Universal Dependencies, contribuindo para o processamento automático das comparações em português. No entanto, para que o trabalho atinja esse objetivo, algumas lacunas informacionais precisam ser preenchidas no texto. Sugiro alguns pontos que precisam ser melhor explorados:

- A sintaxe de dependências é apontada já no título do trabalho como o aporte teórico que será usado para o desenvolvimento das diretrizes de anotação. Porém, essa teoria não é apresentada, nem mesmo superficialmente. Uma apresentação simples desse suporte teórico pode contribuir para a compreensão da pesquisa por pessoas não especialistas da área.

- No mesmo sentido, uma explicação a respeito da abordagem Universal Dependencies contribuirá para a compreensão da proposta. Na seção Proposta, os autores mencionam que o modelo UD é composto por um conjunto de etiquetas, porém essas etiquetas não são apresentadas. Um detalhamento da abordagem e das etiquetas contribuirá para a possibilidade de replicação do procedimento por outros pesquisadores.

Ainda em relação ao modelo de UD, um questionamento que fica da leitura do texto é se outras abordagens foram testadas. Esse questionamento surge, principalmente, a partir da leitura da introdução, em que os autores se propõem a apresentar maneiras lógicas e consistentes de anotação das estruturas comparativas.

- O que os autores entendem por anotação no âmbito do trabalho desenvolvido? Como a anotação ajuda no processamento automático de informações? Uma explicação dessa perspectiva de trabalho, mesmo que simples, pode fortalecer a importância do estudo.

- Corpus Porttinari-base: seria interessante que o corpus fosse apresentado e sua escolha justificada. Algumas questões sobre o corpus: são dados de escrita? Quantos textos compõem o corpus? Qual variedade do português o corpus representa?

Essas informações podem contribuir para a compreensão dos procedimentos metodológicos adotados para o desenvolvimento das diretrizes de anotação das estruturas comparativas, o que contribuirá também para a possibilidades de uso das diretrizes por outros pesquisadores.

RESUMO

Não ficou clara a metodologia usada para o desenvolvimento das diretrizes de anotação das estruturas comparativas. Os resultados apresentados não dialogam com o objetivo do estudo.

Ainda no resumo, os autores informam que o conjunto de sentenças anotadas está disponibilizado na web. Aqui seria interessante apresentar o nome do repositório em que as sentenças estão disponibilizadas, pois web é muito vago.

TEXTO

Na introdução, os autores mencionam o uso da metodologia corpus-driven. Em que consiste esse procedimento? Como o uso dessa metodologia contribuiu para a elaboração das diretrizes de anotação das estruturas comparativas?

Em termos de apresentação do texto, sugiro rever a legenda das ilustrações (tabela 1 e tabela 2). Nesse caso, não se trata de tabelas, mas de quadros, pois não apresentam dados quantitativos.

Ainda em relação à apresentação do texto, rever a numeração dos exemplos. Na pág. 13, a numeração passa de (24) para (34).

No tocante aos exemplos, por que até o exemplo (24) não foram usadas sentenças do corpus sob análise? Os autores explicam, na página 13, que foram usados exemplos simplificados, o que nos possibilita um questionamento: Se a proposta do trabalho é apresentar diretrizes lógicas e consistentes de anotação, essas diretrizes não deveriam se aplicar a qualquer construção comparativa do português?

Em relação às conclusões e decisões apresentadas na seção Conclusão, sugiro que sejam sistematizadas em forma de diretrizes, em um quadro, por exemplo, de modo a deixar claro quais são as diretrizes propostas, possibilitando a replicação dessas diretrizes em outros corpora.

Por fim, sugiro rever com atenção as normas de organização e formatação das diretrizes para autores da revista da Abralín. Nesse tocante, é importante observar a ausência de algumas seções obrigatórias: i) resumo para não-especialistas, ii) conflito de interesse, iii) protocolo e pré-registro de pesquisa e iv) declaração de disponibilidade de dados.

Na nota 21, os autores apresentam um link para um repositório onde o corpus anotado está disponibilizado. Apresentar essa informação na seção declaração de disponibilidade de dados, conforme regras de submissão da revista.

AVALIADOR 2

O artigo objetiva apresentar diretrizes de anotação de estruturas comparativas em língua portuguesa usando o conjunto de etiquetas proposto pelas diretrizes de anotação das Universal Dependencies. O texto, escrito de forma bastante clara, aponta os objetivos do estudo, revisão da literatura, proposta de análise e conclusões. Com base em literatura precedente, os autores lançam mão de cinco elementos de identificação de estruturas comparativas de forma a elencar os padrões das construções comparativas do português. Os resultados mostram-se bastante consistentes e contribuem sobremaneira para tarefas de etiquetagem/anotação de corpus de forma automatizada. O estudo, inclusive, aponta para novas investigações, em que se busca aplicar os achados em corpus de treinamento. Há poucas questões gramaticais para serem revisadas no artigo. As referências são atuais e pertinentes à argumentação desenvolvida.

RESPOSTA DOS AUTORES

Agradecemos as sugestões da Dra. Josilene de Jesus Mendonça e gostaríamos de comentar quais foram acatadas na versão revisada que ora submetemos e quais não foram (e por que não foram):

SUGESTÕES ACATADAS:

Incluimos uma seção de "Metodologia e Fundamentos Teóricos", na qual explicamos maiores detalhes que não haviam ficado claros na primeira versão. Nessa seção fizemos uma breve descrição da Gramática de Dependências e da abordagem de anotação Universal Dependencies, bem como fornecemos informações sobre o cópulus Porttinari-base. Com isso, acreditamos ter suprido as "lacunas informacionais" apontadas pela parecerista.

Substituímos o termo "sentenças simplificadas" por "sentenças reduzidas", pois não simplificamos as construções comparativas: nós construímos exemplos reduzidos para dar destaque às construções comparativas e para podermos explicar didaticamente casos muito complexos.

Incluimos as seções:

Resumo para não Especialistas,

Conflito de Interesse.

Protocolo e Pré-Registro de Pesquisa,

Declaração de Disponibilidade de Dados.

Como utilizamos o template do Latex e ele não contém essas seções, acabamos não as colocando na primeira versão, por pura falta de atenção. Sugerimos que o template do Latex seja atualizado pela Abralín de forma a incluir todas as seções obrigatórias, pois evitará que outros autores possam cometer o mesmo erro.

Incluimos notas de rodapé explicando o que é "anotação de cópulus" e o que é a abordagem corpus-driven.

Incluimos, no resumo, o nome do ambiente onde está hospedado o conjunto de dados disponibilizado, bem como fornecemos o link para acesso na Declaração de Disponibilidade de Dados.

SUGESTÕES NÃO ACATADAS:

Não trocamos a palavra "tabela" por "quadro" pois o template do Latex da Abralín só admite 3 tipos de inserções: Tabela, Gráfico e Figura. Não há como criar "Quadro" pois se tornaria um objeto não referenciável (sem possibilidade de incluir legenda e numeração), por isso optamos por "Tabela". Seria interessante que a Abralín alterasse o modelo Latex para admitir "Quadro" também.

A parecerista teve a impressão de que a numeração dos exemplos pulou do 24 para o 34, porém os exemplos de números 25 a 33 estão na ordem, na página 15. Provavelmente a confusão ocorreu porque a Tabela 2, que traz uma cópia das sentenças de 14 a 24, está alocada antes da sentença 34 (a alocação de objetos grandes referenciáveis quase sempre se dá fora da sequência do texto que os referencia, infelizmente).

Não fizemos, na Conclusão, um quadro com as diretrizes de anotação propostas, pois as figuras das árvores sintáticas são essenciais para explicá-las e acreditamos que a forma como as propostas foram apresentadas é a ideal para que outros anotadores de córpus em UD (Universal Dependencies) possam replicá-las.

Como o estudo apresentado se deu em função da anotação de córpus com o esquema de anotação UD (tarefa principal do projeto em que trabalhamos) não vimos razão para tentarmos outros esquemas de anotação além da sintaxe de dependências. Para deixar claro que as escolhas do córpus e da abordagem de anotação precederam o estudo, inserimos explicitamente comentário a respeito na seção de Metodologia e Fundamentos Teóricos. Esclarecemos também, no texto, que a sintaxe de dependências vem se consagrando há alguns anos como a mais apropriada para fins de Processamento de Linguagem Natural.

Conflito de Interesse

As autoras e o autor não têm conflitos de interesse a declarar.

Protocolo e pré-registro de pesquisa

Não se aplica.

Declaração de Disponibilidade de Dados

Este artigo disponibiliza livre acesso ao conjunto das construções comparativas anotadas do córpus Portinari-base no seguinte endereço eletrônico: https://arborator.icmc.usp.br/#/projects/Construcoes_Comparativas_Corpus_Portinari.

Fontes de financiamento

Este trabalho foi realizado no âmbito do Centro de Inteligência Artificial da USP (C4AI <http://c4ai.inova.usp.br/>), com o apoio da IBM e da FAPESP (processo 2019/07665-4). Este projeto também foi apoiado pelo Ministério da Ciência, Tecnologia e Inovações, com recursos da Lei nº 8.248, de 23 de outubro de 1991, no âmbito do PPI-Softex, coordenado pela Softex e publicado como Residência em TIC 13, DOU 01245.010222/2022-44.

REFERÊNCIAS

BECK, Sigrid; KRASIKOVA, Sveta; FLEISCHER, Daniel; GERGEL, Remus; HOFSTETTER, Stefan; SAVELSBERG, Christiane; VANDERELST, John; VILLALTA, Elisabeth. Crosslinguistic variation in comparison constructions. **Linguistic Variation Yearbook**, v. 9(1), p. 1-66, jan. 2009. DOI <https://doi.org/10.1075/livy.9.01bec>. Acesso em: 13 outubro 2022.

BIBER, Douglas. Corpus-Based and Corpus-driven Analyses of Language Variation and Use. In: HEINE, Bernd; NARROG, Heiko. **The Oxford Handbook of Linguistic Analysis**. Oxford: Oxford Academic, 2012. DOI <https://doi.org/10.1093/oxfordhb/9780199544004.013.0008>. Acesso em: 13 outubro 2022.

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

CEGALLA, Domingos Paschoal. **Novíssima Gramática da Língua Portuguesa**. São Paulo: Companhia Editora Nacional, 2020.

CROFT, W. **Typology and Universals** (2nd ed., Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press, 2002. DOI <https://doi.org/10.1017/CBO9780511840579>. Acesso em: 13 outubro 2022.

CUNHA, Celso Ferreira; LINDLEY CINTRA, Luis Filipe. **Nova gramática do Português contemporâneo**. Rio de Janeiro: Lexikon Editora Digital, 7a edição, 2017. UR: <https://ia800706.us.archive.org/12/items/NovaGramaticaDoPortugusContemporneo>. Acesso em: 13 outubro 2022.

DE MARNEFFE, Marie-Catherine; MANNING, Christopher D.; NIVRE, Joakim; ZEMAN, Daniel. Universal Dependencies. **Computational Linguistics**, v. 47(2), p. 255-308, 2021. DOI https://doi.org/10.1162/coli_a_00402. Acesso em: 13 outubro 2022.

DIXON, R.M.W. Comparative constructions: A cross-linguistic typology. **Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"**, v. 32(4), p. 787-817, 2008. DO: <https://doi.org/10.1075/sl.32.4.02dix>. Acesso em: 13 outubro 2022.

DURAN, M.S. **Manual de Anotação de PoS tags: Orientações para anotação de etiquetas morfossintáticas em Língua Portuguesa, seguindo as diretrizes da abordagem Universal Dependencies (UD)**. Relatório Técnico do ICMC 434. Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo. São Carlos-SP, Setembro, 2021. URL https://drive.google.com/file/d/1BddPswn-_loo-A5GslDA1cO1kqbcCahb. Acesso em: 13 outubro 2022.

DURAN, M.S.; NUNES, M.G.V.; LOPES, L.; PARDO, T.A.S. Manual de anotação como recurso de Processamento de Linguagem Natural: o modelo Universal Dependencies em língua portuguesa. **Domínios de Linguagem**, v. 16(4), p. 1608-1643, 2022. DOI <https://doi.org/10.14393/DL52-v16n4a2022-13>. Acesso em: 13 outubro 2022.

DURAN, M.S. **Manual de Anotação de Relações de Dependência - Versão Revisada e Estendida: Orientações para anotação de relações de dependência sintática em Língua Portuguesa, seguindo as diretrizes da abordagem Universal Dependencies (UD)**. Relatório Técnico do ICMC 440. Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo. São Carlos-SP, Outubro, 2022. URL <https://drive.google.com/file/d/1ile8Wfxu1qdrZOMLGqkvVuQ4fXvHgVMo>. Acesso em: 13 outubro 2022.

DURAN, Magali Sanches; LOPES, Lucelene; NUNES, Maria das Graças Volpe; PARDO, Thiago Alexandre Salgueiro. The Dawn of the Portinari Multigenre Treebank: Introducing its Journalistic Portion. In: **Simpósio Brasileiro de Tecnologia Da Informação E Da Linguagem Humana (STIL)**, 14., 2023, Belo Horizonte/MG. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2023 . p. 115-124. DOI: <https://doi.org/10.5753/stil.2023.233975>.

GANAPATHIBHOTLA, Murthy; LIU, Bing. Mining Opinions in Comparative Sentences. In: **Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)**. Manchester: Coling 2008 Organizing Committee, p. 241-248, 2008. URL <https://aclanthology.org/C08-1031>. Acesso em: 13 outubro 2022.

HASPELMATH, Martin. Equative constructions in world-wide perspective. In: TREIS, Yvonne; VANHOVE, Martine. **Similative and equative constructions: A crosslinguistic perspective**. Amsterdam: John Benjamins, 2017. DOI <https://doi.org/10.1075/tsl.117.02has>. Acesso em: 13 outubro 2022.

IMRÉNYI, András; MAZZIOTTA, Nicolas. **Chapters of Dependency Grammar: A historical survey from Antiquity to Tesnière**. Amsterdam: John Benjamins, 2020. DOI <https://doi.org/10.1075/slcs.212>. Acesso em: 13 outubro 2022.

KÁNTOR, Gergely; BACSKAI-ATKARI, Julia. **Elliptical comparatives revisited**. Budapest: Research Institute for Linguistic, Hungarian Academy of Sciences, 2012. URL https://www.bacskaiatkari.de/pdf/comparatives_bacskai_atkari_kantor.pdf. Acesso em: 13 outubro 2022.

LECHNER, Winfried. **Ellipsis in comparatives**. Berlin, New York: De Gruyter Mouton, 2008. DOI <https://doi.org/10.1515/9783110197402>. Acesso em: 13 outubro 2022.

MCDONALD, Ryan; NIVRE, Joakim; QUIRMBACH-BRUNDAGE, Yvonne; GOLDBERG, Yoav; DAS, Dipanjan; GANCHEV, Kuzman; HALL, Keith; PETROV, Slav; ZHANG, Hao; TACKSTROM, Oscar; BEDINI, Claudia; BERTOMEU CASTELLÓ, Núria; LEE, Jungmee. Universal Dependency Annotation for Multilingual Parsing. In: **Proceedings of ACL**, 2013.

MIRANDA, L.G.M.; PARDO, T.A.S. An Improved and Extended Annotation Tool for Universal Dependencies-based Treebank Construction. In: **Proceedings of the PROPOR Demonstrations Workshop**, 2022, p. 1-3. URL: <https://drive.google.com/file/d/1Gz9k3-SU72zXx6v2a0ITrutHVAtpUMUX>. Acesso em: 18 maio 2023.

NEVES, Maria Helena de Moura. **Gramática de Usos do Português**. São Paulo: Editora Unesp, 2000.

NIVRE, Joakim; DE MARNEFFE, Marie-Catherine; GINTER, Filip; HAJI_C, Jan; MANNING, Christopher D.; PYYSALO, Sampo; SCHUSTER, Sebastian; TYERS, Francis; ZEMAN, Daniel. Universal Dependencies v2: An Evergrowing Multilingual Treebank Collection. In: **Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC 2020)**. Marseille: European Language Resources Association, p. 4034-4043, 2020. URL <https://aclanthology.org/2020.lrec-1.497>. Acesso em: 13 outubro 2022.

OSBORNE, Timothy; GERDES, Kim. The status of function words in dependency grammar: A critique of Universal Dependencies (UD). **Glossa: a journal of general linguistics**, v. 4(1), 17, 2019. DOI <https://doi.org/10.5334/gjgl.537>. Acesso em: 13 outubro 2022.

PARDO, T.A.S.; DURAN, M.S.; LOPES, L.; DI FELIPPO, A.; ROMAN, N.T.; NUNES, M.G.V. Portinari - a large multi-genre-treebank for Brazilian Portuguese. In Proceedings of the XIV Symposium in Information and Human Language (STIL), São Paulo: SBC, pp. 1-10. 2020. DOI <https://doi.org/10.5753/stil.2021.17778>. Acesso em: 13 outubro 2022.

PEREIRA, Sandra; PINTO, Clara; PRATAS, Fernanda. Construções comparativas em português: porque algumas são mais iguais que outras. In: **Textos Selecionados do XXIX Encontro da Associação Portuguesa de Linguística. Lisboa: Associação Portuguesa de Linguística**, 2014. URL <http://hdl.handle.net/10451/33739>. Acesso em: 13 outubro 2022.

PULMAN, S. G. Comparatives and Ellipsis. In: **Proceedings of the 5th European Meeting of the Association for Computational Linguistics**, Berlin, 1991. URL <https://aclanthology.org/E91-1002>. Acesso em: 13 outubro 2022.

RANCHHOD, Elisabete. Frozen Adverbs – Comparative Forms Como C in Portuguese. **Linguisticae Investigationes**, v. 15(1), p. 141 - 170, jan. 1991. DOI <https://doi.org/10.1075/li.15.1.07ran>. Acesso em: 13 outubro 2022.

RANCHHOD, Elisabete; DE GIOIA, Michele. Comparative Romance Syntax. Frozen Adverbs in Italian and in Portuguese. **Linguisticae Investigationes**, v. 20(1), p. 33-85, jan. 1996. DOI <https://doi.org/10.1075/li.20.1.04ran>. Acesso em: 13 outubro 2022.

ROCHA LIMA, Carlos Henrique. **Gramática normativa da língua portuguesa**. São Paulo: Editora José Olympio, 2010.

STASSEN, Leon. **Comparison and Universal Grammar**. Oxford: Basil Blackwell, 1985.

STASSEN, Leon. Comparative constructions. In: DRYER, Matthew S.; HASPELMATH, Martin. **The World Atlas of Language Structures Online**. Leipzig: Max Planck Institute for Evolutionary Anthropology, 2013. URL <https://wals.info/chapter/121>. Acesso em: 13 outubro 2022.

TESNIÈRE, Lucien. **Éléments de Syntaxe Structurale**. Paris: Librairie C. Klincksieck, 1959.

TESNIÈRE, Lucien. **Elements of Structural Syntax**. Tradução de OSBORNE, Timothy; KAHANE, Sylvain. Amsterdam: John Benjamins, 2015.

TOGNINI-BONELLI, Elena. **Corpus Linguistics at Work**. Computational Linguistics, v. 28(4), p. 583, December 2002. DOI <https://doi.org/10.1162/coli.2002.28.4.583a>. Acesso em: 13 outubro 2022.

TREIS, Yvonne. Comparative Constructions: An Introduction. **Linguistic Discovery**, v. 16(1), 2018. DOI <https://doi.org/10.1349/PS1.1537-0852.A.492>. Acesso em: 13 outubro 2022.

ULTAN, Russell Some features of basic comparative constructions. **Working Papers on Language Universals (Stanford)**, v. 9, p. 117-162, 1972.

VARATHAN, Kasturi Dewi; GIACHANOU, Anastasia; CRESTANI, Fabio. Comparative opinion mining: A review. **Journal of the Association for Information Science and Technology**, v. 68(4), p. 811{829, April 2017. DOI <https://doi.org/10.1002/asi.23716>. Acesso em: 13 outubro 2022.