

ENSAIO TEÓRICO

O modelo de dependências universais: assentando bases teóricas e revisando diretrizes metodológicas

André V. Lopes CONEGLIAN 

Universidade Federal de Minas Gerais (UFMG)

RESUMO

As Dependências Universais (UDs) são um modelo de anotação morfosintática de línguas naturais. Este artigo parte do pressuposto de que na base do modelo ficam previstas a adequação computacional (que se liga ao processamento de língua natural), a adequação tipológica (que se liga à proposta geral do modelo de comparação interlinguística) e a adequação descritiva (que se liga à anotação de línguas individuais). A discussão concentra-se nos dois últimos tipos de adequação. A proposta que se defende aqui é a de que os pressupostos teóricos subjacentes ao modelo das UD necessitam ser claramente explicitados, uma vez que são exatamente tais pressupostos que determinam a construção de quadros categoriais que cancelam as anotações realizadas para línguas individuais. Para tanto, procede-se a uma discussão sobre a concepção do quadro categorial e a formulação da definição das categorias. A avaliação geral a que se chega é que, por um lado, do ponto de vista da proposição de diretrizes gerais, tal quadro categorial deve constituir um conjunto de “conceitos comparativos” (HASPELMATH, 2010), por outro, do ponto de vista da proposição de diretrizes para anotação de línguas individuais, esse quadro deve constituir um conjunto de “categorias descritivas”. Essa dupla visão sobre o quadro categorial das UD pode garantir isonomia e equivalência teórico-metodológica, de modo a tanto otimizar a realização de tarefas de análise e descrição linguísticas assistidas por modelos computacionais quanto a construção de modelos computacionais baseados em anotação linguística consistente do ponto de vista descritivo e tipológico.



OPEN ACCESS

EDITADO POR

- Jorge Baptista (UAIG)
- Adriana Silvina Pagano (UFMG)
- Marta Deysiane Alves Faria Sousa (UFS)

AVALIADO POR

- Adriana Pagano (UFMG)
- Juliano Desiderato Antonio (UEM)

DATAS

- Recebido: 19/11/2022
- Aceito: 20/12/2022
- Publicado: 09/09/2024

COMO CITAR

Coneglian, A. V. L. (2024). O modelo de dependências universais: assentando bases teóricas e revisando diretrizes metodológicas. *Revista da Abralín*, v. 22, n. 2, p. 187-214, 2024.

ABSTRACT

Universal Dependencies (UD) are a framework for morphosyntactic annotation of natural language texts. This article assumes that this framework implies three types of adequacy: computational adequacy (related to natural language processing), typological adequacy (related to the overall proposal of crosslinguistic comparison), descriptive adequacy (related to the annotation of individual languages). The discussion on this paper focuses the last two types of adequacy. The proposal made here is that the underlying theoretical assumptions of UD's framework need to be clearly explicated, because these assumptions determine the establishment of categories based on which annotations for individual languages are made. In this line, I examine the definitions proposed in the framework in order to determine what are the underlying theoretical notions, and I confront general (universal) guidelines for annotation and language-particular guidelines. The overall assessment in this paper is that, on the one hand, from the perspective of proposing general guidelines for annotation, UD's framework should be built as a set of "comparative concepts" (HASPELMATH, 2010); on the other hand, from the perspective of proposing guidelines for the annotation of individual languages, the framework should be built as a set of "descriptive categories". This dual view of the UD's framework can ensure theoretical and methodological equivalence, and its implementation can optimize linguistic analysis and description tasks assisted by computational models.

PALAVRAS-CHAVE

Dependências universais. Tipologia linguística. Conceitos comparativos. Conceitos descritivos.

KEYWORDS

Universal dependencies. Linguistic typology. Comparative concepts. Descriptive concepts.

RESUMO PARA NÃO ESPECIALISTAS

Métodos computacionais que lidam com linguagem natural – mais conhecidos como modelos de processamento de língua natural – têm ganhado cada vez mais espaço na sociedade e estão abrigados, cientificamente, no campo da Linguística Computacional. Um dos modelos que tem ganhado tração no campo é o das Dependências Universais (Universal Dependencies – UD), que leva esse nome porque se pretende como um modelo de análise morfossintática de todas as línguas do mundo. Essa tarefa de

analisar e anotar os textos de todas as línguas do mundo é complicadíssima, porque, como bem se sabe, nenhuma língua é igual a outra, por mais parecidas que elas sejam. É difícil chegar a um modelo de gramática que dê conta da gramática de todas as línguas do mundo. Mas é possível chegar a isso. A proposta deste artigo é discutir qual é a teoria necessária para que essa proposta das UDs continue dando resultados cada vez mais precisos na análise das línguas do mundo. Aí entram em jogo os campos da tipologia e descrição linguísticas, que tem descrito e sistematizado propriedades linguísticas que são comuns a todas as línguas do mundo, apesar de suas diferenças.

Introdução

Dependências Universais (*Universal Dependencies* – UDs, doravante) são um modelo de anotação morfossintática de línguas naturais, projetado para auxiliar o desempenho de tarefas de processamento de língua natural (PLN, doravante). O modelo é relativamente recente, tendo sido proposto em 2014 (NIVRE *et al.*, 2016), mas constrói-se sobre um aparato computacional complexo que vem sendo desenvolvido há décadas (DE MARNEFFE *et al.*, 2014).

No Brasil, mais especificamente, o modelo das UDs tem ganhado espaço, principalmente nos núcleos de pesquisa de linguística computacional e de PLN, sendo usado especialmente para a construção de bancos de dados e para a anotação de construções particulares do português (veja-se Duran *et al.*, neste volume). É virtualmente inexistente a pesquisa em tipologia com recurso ao modelo das UDs, mas um esforço nessa direção tem sido feito na construção de corpora paralelos e na utilização desses bancos de dados para estudos de natureza quantitativa (por exemplo, JING; WIDMER; BICKEL, 2021; LEVSHINA, 2021, 2022).

A proposta que quero defender aqui pode ser resumida na indicação de que as Dependências Universais constituem um método de anotação morfossintática cujos pressupostos teóricos necessitam ser claramente explicitados, uma vez que são exatamente tais pressupostos que determinam a construção de quadros categoriais que cancelam as anotações realizadas. Essa explicitação deve ser feita levando-se em consideração três tipos de adequação que o modelo busca: a adequação computacional (que se liga ao processamento de língua natural), a adequação tipológica (que se liga à proposta geral do modelo de comparação de línguas do mundo) e a adequação descritiva (que se liga à preparação de diretrizes para anotação de línguas individuais). Este artigo concentra-se na adequação tipológica e descritiva do modelo. Nessa direção, discute-se a pertinência de uma visão funcional-tipológica de linguagem e de gramática, visão esta pela qual se configuram as classes e as funções do modelo (Seções 3 e 4).

A discussão deste artigo pressupõe uma certa familiaridade por parte do leitor com o modelo das UD's. No entanto, na medida do possível, certos pontos fundamentais são apresentados para contextualizar a exposição. Este artigo organiza-se em cinco seções: a Seção 1 apresenta um breve histórico da Gramática de Dependências e o seu lugar no campo da linguística teórica, mostrando o modelo das UD's como um 'capítulo' da Gramática de Dependências; a Seção 2 apresenta uma visão geral do modelo das UD's e os princípios que balizam o projeto. As Seções 3 e 4 discutem a adequação tipológica e descritiva do modelo, respectivamente, por meio de um exame do quadro categorial das UD's no que diz respeito à formulação das definições das classes de palavras e das relações de dependência. A Seção 5 traz um balanço geral da avaliação a que se procede neste artigo.

1 O quadro geral: teoria sintática e a Gramática de Dependências

Antes de proceder a uma discussão sobre o modelo das UD's, pode ser interessante circunscrevê-lo no quadro geral da Gramática de Dependências.

Como explica Moravčsik (2019), o principal objetivo de teorias de sintaxe é descrever a seleção e a ordenação linear das palavras que compõem uma sentença de modo que seja possível fazer generalizações a respeito do sistema gramatical das línguas do mundo. Dryer (2006) aponta que, na linguística, costuma-se fazer uma distinção, em certa medida equivocada, entre teorias descritivas, sobre como as línguas são, e teorias explicativas, sobre o porquê as línguas são como são. A Gramática de Dependências, como um modelo genérico de descrição sintática, é naturalmente uma teoria descritiva. Há, no entanto, teorias sintáticas cuja base é o modelo de dependências e que se configuram como teorias explicativas, como é o caso da *Word Grammar* (HUDSON, 2010)¹. As UD's, como um dos "capítulos" mais recentes no desenvolvimento da Gramática de Dependências (nos termos de Mazziotta e Imrényi (2020)), é certamente um modelo mais fortemente descritivo, como se discutirá mais adiante, mas é ledor engano pressupor que não é necessário haver consistência teórica que chancele o modelo (Seções 3 e 4).

Fazer um resgate histórico sobre a Gramática de Dependências está além do propósito deste artigo, mas é necessário fazer uma circunstanciação a respeito de pontos gerais desse modelo de gramática, especialmente no que diz respeito à noção de dependência e à relação entre sintaxe e outros componentes da gramática (particularmente, a semântica).

¹ *Word Grammar* é um modelo de sintaxe que aproveita propostas da ciência cognitiva. Nessa medida, busca responder questões sobre a organização cognitiva do conhecimento linguístico, como esse conhecimento é adquirido, entre outras (Hudson, 2010).

1.1 A Gramática de Dependências no campo da linguística teórica

Não é novidade que a Gramática de Dependências não tem presença marcante no campo da linguística teórica, apesar de já ser verificado um modelo geral de ‘dependência’ em propostas de análise linguística na antiguidade clássica até a Idade Média e a Renascença (COLOMBAT, 2020; LUHTALA, 2020). É no campo da linguística computacional que a Gramática de Dependências tem encontrado terreno fértil desde a década de 1960.

Muito resumidamente, pode-se indicar que, já na modernidade, o trabalho de Lucien Tesnière (2015 [1959]) é considerado a explicitação e o equacionamento de uma gramática com base em um modelo de dependência (veja-se o texto de Pagani neste volume). O modelo de Tesnière ganhou espaço considerável na linguística alemã que produziu importantes dicionários de valência (por exemplo, HELBIG; SCHENKEL, 1991 [1983]). No Brasil, pode-se dizer que o *Dicionário gramatical de verbos do português contemporâneo do Brasil* (Borba, coord., 1990) é, em certa medida², produto dessa tradição de pesquisa em uma gramática de valência (BORBA, 1996).

Do ponto de vista da produção científica contemporânea no campo da sintaxe, apenas em 3 *handbooks* (STEGER; WIEGAND, orgs., 1993; CARNIE *et al.*, orgs., 2014; KERTÉSZ *et al.*, orgs., 2019) constam capítulos sobre a Gramática de Dependências (HERINGER, 1993a, 1993b; HUDSON, 1993; OSBORNE, 2014, 2019a), enquanto há outros *handbooks* que completamente ignoram esse modelo de gramática. Por exemplo, em *The Bloomsbury Companion to Syntax* (LURAGHI; PARODI, orgs., 2013), há apenas uma menção à Gramática de Dependências no que diz respeito aos modelos de sintaxe pelos quais se constroem bancos de dados anotados (FREDDI, 2013).

Vale mencionar a obra *Dependency and valency: an international handbook of contemporary research*, organizada em dois volumes por Ángel *et al.* (2003), os quais somam quase 1.600 páginas e trazem, no total, 122 capítulos sobre os mais diversos temas ligados à Gramática de Dependências. Obviamente, pelo momento em que a obra foi publicada, não há menção às UDs. No primeiro volume, há um conjunto de 9 capítulos dedicados à figura de Tesnière e sua obra.

Como notam Mazziotta e Imrényi (2020), a noção de dependência não está vinculada nem ao formalismo nem ao funcionalismo, talvez os dois grandes paradigmas da linguística contemporânea. Por exemplo, há modelos de base formalista que aproveitam as dependências, como é o caso da *Text-meaning theory* (MEL’ČUK, 1988) e da *Functional Generative Description* (HAJIČOVÁ; SGALL, 2003). Mas também há modelos de base funcionalista e cognitivista que aproveitam as dependências, como é o caso da Gramática Cognitiva (LANGACKER, 1994) e a Gramática de Construções Radical (CROFT, 2022) – esta última especialmente para tarefas de PLN e pedagógicas.

² Apenas em partes, porque acoplada à proposta de valência e estrutura argumental, está uma teoria de base semântica sobre classificação semântica dos verbos (CHAFE, 1979).

1.2 A noção de dependência

Uma teoria de sintaxe se constrói ou sobre a noção de **constituência** ou sobre a de **dependência** (Osborne, 2014). Na linguística moderna e contemporânea, tanto na Europa quanto nos Estados Unidos, teorias sintáticas com base na constituição dominaram o espaço científico da linguística.

Segundo Percival (1990), não há consenso entre os linguistas a respeito da noção de “dependência”. De todo modo, uma definição geral de dependência pode ser encontrada em Mel’čuk (2009, p. 9), para quem a “dependência é, por definição, uma relação assimétrica, semelhante à implicação lógica: um elemento implica outro, mas, de uma maneira geral, não vice-versa”³. Nessa visão, uma estrutura sintática constitui-se de elementos, ou palavras, e das relações de dependência pelos quais os elementos se ligam (Osborne, 2014).

Não é possível determinar se “dependência”, tal como concebida por Mel’čuk (2009), é um fenômeno de natureza sintática ou semântica. Seria possível concluir, já que se fala de um modelo de sintaxe, que “dependência” é um fenômeno sintático. No entanto, como bem mostra Croft (2001, cap. 5), “dependência” tem uma natureza semântica, porque diz respeito às relações significativas que se estabelecem entre as palavras em uma sentença, assim como a “constituência” tem natureza semântica, porque, como explica Croft (2001, p. 186), “contiguidade sintática reflete relações semânticas”.

Uma discussão sobre a natureza sintática ou semântica das dependências é necessária porque permite que seja considerado o modo pelo qual se interfaceiam esses dois componentes da gramática. Note-se que não se trata de postular relações de dependência nos níveis sintático ou semântico, o que é inegável que haja (cf. Mel’čuk, 2009). O ponto que se pretende fazer é outro: é examinar de que modo ficam correlacionadas, se é que ficam, sintaxe e semântica no modelo da Gramática de Dependências. Discernir tal correlação, ou a ausência dela, implica considerar o quão autônoma é a sintaxe em relação à semântica.

A esse respeito, considere-se a interpretação de Van Langendonck (2003) da noção de dependência. Para o autor, “dependência é, na verdade, um caso especial de acompanhamento (do alemão, *Konkomitanz*). Palavras que ocorrem juntas podem revelar uma relação sintático-semântica sancionada como acompanhamento pela gramática”⁴ (p. 171). Note-se que o autor estabelece uma relação causal entre a co-ocorrência de palavras na sentença e o tipo de ligação sintático-semântica entre elas. Uma discussão semelhante se encontra na gramática de Osborne (2019b, p. 36). Segundo esse último autor, as palavras em uma sentença não se arranjam arbitrariamente, mas “de tal modo que significado seja construído”. Fica pressuposta aí uma relação causal direta entre semântica (o significado que uma sentença constrói) e sintaxe (a disposição sentencial das palavras).

³ Tradução deste autor. Texto original: “Dependency is by definition a non-symmetrical relation, of the same type as logical implication: one element implies in some sense the other, but, generally speaking, not vice-versa”.

⁴ Tradução deste autor. Texto original: “Dependency is in fact a special case of companionship (German: *Konkomitanz*). Words that occur together may display a semantic- syntactic relationship sanctioned as companionship by the grammar.”

Essa concepção fica implicada nas maneiras pelas quais se representam as relações de dependência entre as palavras em uma sentença. Explica Osborne (2019a) que uma das vantagens de representar as relações entre palavras por meio de relações de dependência é a fácil identificação de termos que são núcleos e aqueles que são dependentes.

Há distintas maneiras de representação das relações de dependência. Uma delas é por meio do uso de colchetes – recurso que geralmente está associado à representação de relações de constituição. Pode-se ver a diferença entre relações de dependência e constituição no Quadro 1, a seguir.

Estrutura de dependência	Estrutura de constituição
<pre>[after [drinking [coffee]]]</pre>	<pre>[[after] [[drinking] [coffee]]]</pre>
Tradução: <i>depois de beber café</i>	
<pre>[[We] like [drinking [coffee]]].</pre>	<pre>[[We] [[like] [[drinking] [coffee]]]].</pre>
Tradução: <i>Nós gostamos de beber café.</i>	

QUADRO 1 - Representação das relações de dependência e de constituição pelo uso de colchetes. Adaptado de Osborne (2019, p. 368).

Fonte: Elaboração própria.

Uma outra maneira de representação da relação de dependência é pelo uso de setas (\rightarrow). Assim, para uma expressão como *livro caro*, tem-se a representação “livro \rightarrow caro”, da qual se pode entender que o adjetivo *caro* é dependente do substantivo *livro*. A representação das relações de dependência por setas é a norma nos modelos computacionais de PLN que têm por base a Gramática de Dependências, como ilustra a Figura 1. Esse é o caso do modelo das Dependências Universais (Seção 2).

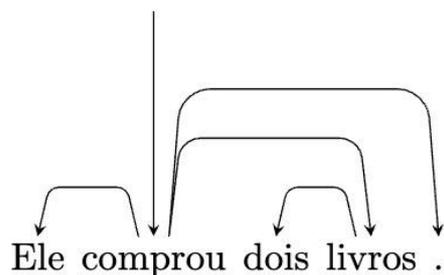


Figura 1 – Representação diagramática das relações de dependência da sentença "Ele comprou dois livros".

Fonte: Elaboração própria.⁵

Uma terceira maneira de representação é a árvore de dependência, ilustrada pela Figura 2. Esta última está mais próxima da representação originalmente desenvolvida por Tesnière (2015 [1959]).

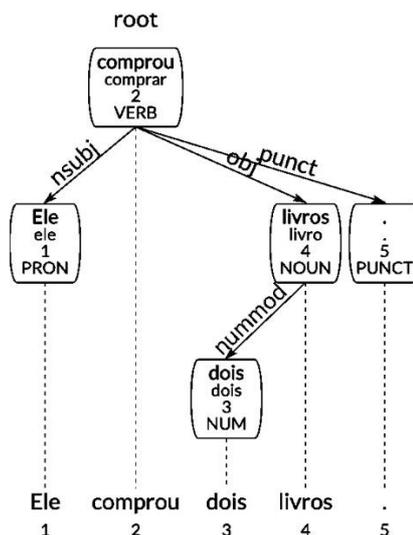


Figura 2 – Árvore de dependência da sentença "Ele comprou dois livros".

Fonte: Elaboração própria.

Como foi dito no início desta subseção, há pouco consenso entre linguistas a respeito da noção de dependência. É necessário fazer uma última indicação a esse respeito. Nas próximas sessões, discute-se a maneira pela qual o modelo das Dependências Universais aproveita a noção de dependência operacionalizando-a tipológica e computacionalmente.

⁵ As figuras foram criadas com o pacote tikz-dependency em um editor de LaTeX. Disponível em: <https://ctan.org/pkg/tikz-dependency>. Acesso em 12 de julho de 2023, às 11h39.

2 Uma visão geral das Dependências Universais

2.1 A organização geral do modelo

Conforme está na página introdutória do site das UDs⁶, as **Dependências Universais** são um projeto de desenvolvimento de bancos de dados anotados para as línguas do mundo, com finalidades práticas, como o desenvolvimento de *parsers* e outras tarefas de PLN.

No momento em que este artigo é escrito, o projeto apresenta: 245 bancos de dados anotados, para 141 línguas, de 30 famílias linguísticas diferentes (DE MARNEFFE; NIVRE; ZEMAN, 2023).

A anotação é feita em três domínios: o domínio morfossemântico, com a anotação das propriedades (*features*), o domínio morfossintático, com a anotação das relações de dependência (*deprels*), e o domínio lexical, com a anotação das classes de palavras (UPOS). A Figura 3 ilustra a anotação de uma sentença do português segundo o modelo.

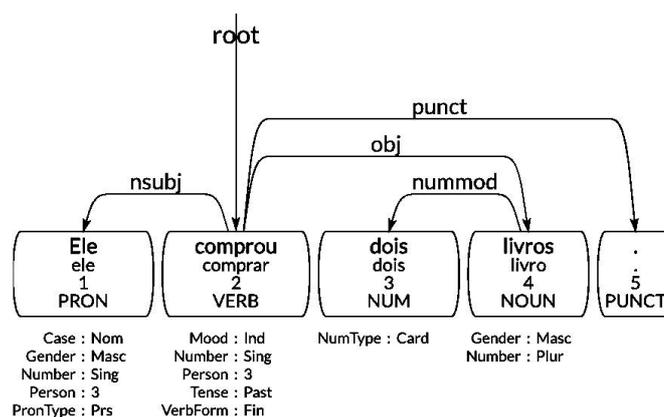


Figura 3 – Exemplo de anotação da sentença “Ele comprou dois livros” segundo o modelo das UDs.⁷

Fonte: Elaboração própria.

Na Figura, os arcos indicam as relações de dependência (anotação das *deprels*). Na primeira linha, está a sentença. Na linha abaixo, está a classificação das palavras segundo sua classe (anotação de UPOS), seguida de sua forma lematizada. Por fim, estão as propriedades morfossemânticas das palavras (anotação de *features*). A discussão a seguir se concentra nos domínios lexical e morfossintático.

O Quadro 2 apresenta as 17 etiquetas que fazem parte das classes de palavras universais, UPOS – *Universal Parts of Speech*.

⁶ Disponível em: www.universaldependencies.org. Acesso em: 9 de julho de 2023, às 17h25.

⁷ Esta figura foi gerada pelo programa *Arborator-Grew*. Disponível em: <https://arborator.icmc.usp.br/#/>. Acesso em 09 de julho de 2023, às 19h15.

Classes de palavras abertas		Classes de palavras fechadas		Outras	
ADJ	adjetivo	ADP	adposição	PUNCT	pontuação
ADV	advérbio	AUX	auxiliar	SYM	símbolo
INT	interjeição	CCONJ	conj. coordenada	X	outra
NOUN	substantivo	DET	determinante		
PROPN	subst. próprio	NUM	numeral		
VERB	verbo	PART	partícula		
		PRON	pronome		
		SCONJ	conj. subordinada		

Quadro 2 – UPOS nas UD. Adaptado de: *Universal POS tag*⁸.

Fonte: Elaboração própria.

Como se nota, o quadro categorial das UD é muito mais rico que a tradição gramatical, mas a distinção entre classes abertas e fechadas permanece. Até onde se sabe, não há uma explicação do porquê uma “interjeição” pertence à classe de palavras aberta e os “numerais”, à fechada. Note-se que há duas etiquetas, *punct* e *sym*, que são necessárias porque as UD são um modelo de PLN. A etiqueta *x* pode se explicar pelo fato de as UD serem um modelo tipológico e computacional: pode acontecer uma língua apresentar uma classe que não consta na lista disponível, ou de ocorrer uma palavra nonsense em uma sentença⁹, ou, ainda, de ocorrer problema de tokenização.

No total, as UD têm 37 relações de dependência (*deprels*) que se organizam pelo nível de funcionamento: nominal, oracional, palavra de modificação, palavra funcional, coordenação, expressão multi-palavra, *loose*, especial e outra. Para os quatro primeiros níveis, há os subníveis: argumentos centrais, dependentes não centrais, dependentes nominais. O Quadro 3 traz uma amostra da definição das *deprels* apenas no nível de funcionamento nominal.

Relação	Definição
EXPRESSÕES NOMINAIS EM FUNÇÃO ARGUMENTAL CENTRAL	
<i>nsubj</i>	SUJEITO NOMINAL (<i>nominal subject</i>) argumento nominal central que é o sujeito sintático (ou pivô) de um predicado
<i>obj</i>	OBJETO (<i>object</i>) argumento nominal central mais básico que não é o sujeito, em geral o participante mais diretamente afetado
<i>iobj</i>	OBJETO INDIRETO (<i>indirect object</i>) argumento nominal central de um verbo que não é sujeito ou objeto direto
EXPRESSÕES NOMINAIS EM FUNÇÃO DE DEPENDENTE NÃO CENTRAL	
<i>obl</i>	OBLÍQUO expressão nominal que funciona como modificador não central (obliquo) de um predicado
<i>vocative</i>	VOCATIVO (<i>vocative</i>) expressão nominal usada para se dirigir ao interlocutor

⁸ Tradução deste autor. Disponível em: <https://universaldependencies.org/u/pos/index.html>. Acesso em 09 de julho de 2023, às 19h.

⁹ O exemplo que se oferece no site é: “And then he just *xfg h pdl jklw*”.

dislocated	ELEMENTOS DESLOCADOS (<i>dislocated elements</i>) expressões nominais periféricas (inicial ou final) em uma oração que não preenchem um papel normal no predicado, mas desempenha função de tópico ou adendo
expl	EXPLETIVO (<i>expletive</i>) liga uma forma pronominal em posição argumental central à qual não é atribuído um papel semântico do predicado
EXPRESSÕES NOMINAIS EM FUNÇÃO DE DEPENDÊNCIA NOMINAL	
appos	MODIFICADOR APOSITIVO (<i>appositional modifier</i>) expressão nominal usada para definir, nomear, ou descrever um referente de um predicado
nmod	MODIFICADOR NOMINAL (<i>nominal modifier</i>) item nominal que modifica outro item nominal
nummod	MODIFICADOR NUMERAL (<i>numeral modifier</i>) expressão numeral que modifica um item nominal

Quadro 3 – *deprels* nominais nas UD. Adaptado de: *Universal Dependency Relations*.¹⁰
Fonte: Elaboração própria.

Na Seção 3 serão discutidos aspectos ligados à formulação das definições tanto das UPOS quanto das *deprels*.

2.2 As UD's como método com uma teoria que necessita explicitação

Neste ponto das reflexões é importante fazer algumas distinções conceituais. Dependências Universais são caracterizadas, na literatura, como *framework* (cf. DE MARNEFFE *et al.*, 2021). Em português não há uma palavra que traduza a noção expressa pelo termo em inglês. Desse modo, faz-se a opção, aqui, por tratar Dependências Universais como um “modelo”, isto é, como uma “metalinguagem sofisticada e complexa para fins de descrição linguística, a qual é pretendida para ser aplicada a qualquer língua”¹¹ (HASPELMATH, 2015, p. 288).

No que diz respeito às UD's, é importante manter distintos os significados de “modelo” e de “teoria”, que, em geral, se confundem. Haspelmath (2015, p. 289) documenta quatro sentidos possíveis para o termo “teoria”: (i) sinônimo de modelo descritivo, como em Teoria de Regência e Ligação, Teoria Linguística Básica etc.; (ii) descrição abstrata de algum domínio empírico, como em teoria sobre uso de linguagem, teoria dos gêneros discursivos; (iii) conjunto de hipóteses sobre um fenômeno, como em teoria sobre voz passiva; (iv) sinônimo de teorização, esforço científico teórico – na contraparte com esforço científico aplicado.

¹⁰ Tradução deste autor. Disponível em: <https://universaldependencies.org/u/dep/index.html>. Acesso em 09 de julho de 2023, às 19h13.

¹¹ Tradução deste autor. Texto original: “... a sophisticated and complex metalanguage for linguistic description that is intended to work for any language.”

Essa distinção decorre da proposta que se defende neste artigo: o modelo das UD's pode e deve ser enriquecido teoricamente do ponto de vista linguístico, sendo justamente a falta de uma base teórica unificada a causa de inconsistências tanto na proposição de categorias metalinguísticas quanto na conceituação dessas categorias (Seções 3 e 4).

Pelo andamento dessa discussão pode-se pressupor que seria possível estabelecer modelos, no sentido de uma metalinguagem descritiva, sem respaldo em uma base teórica. Acontece que, como ensina Saussure (2006 [1916]), é o ponto de vista que determina o objeto, ou seja, são as lentes teóricas que permitem ao analista estabelecer categorias, noções, bem como definir seu objeto de pesquisa. Como uma primeira problematização a esse respeito, considere-se o objetivo principal das UD's tal qual apresentado em de Marneffe *et al.* (2021, p. 256).

O objetivo das UD's é oferecer uma representação linguística útil para a pesquisa morfossintática, a interpretação semântica e para o processamento de língua natural prático para as diferentes línguas humanas. As UD's põem ênfase, portanto, nas representações simples de superfície que permitem traçar paralelos entre construções semelhantes nas diferentes línguas, apesar das diferenças de ordem de palavras, morfologia, e da presença ou ausência de palavras funcionais.¹²

O modelo das UD's já nasce com finalidades práticas, de representação computacional. Na base do modelo está implicada uma visão tipológica de linguagem (“para as diferentes línguas humanas”), a qual obviamente pressupõe a comparabilidade de estruturas gramaticais “nas diferentes línguas”. Pode parecer banal fazer uma explicitação dessa natureza, mas, como se discutirá adiante (Seções 3 e 4), tais pressupostos determinam a operacionalização prática do modelo, tanto no que diz respeito ao estabelecimento de categorias quanto na definição de noções.

Nessa direção, considerem-se, também, o conjunto de seis proposições que configuram as diretrizes do projeto das UD's.

O modelo deve ser satisfatório no que diz respeito à análise linguística de línguas individuais.
O modelo deve ser útil para a tipologia linguística, de modo a prover uma base adequada para a comparação entre línguas e famílias linguísticas.
O modelo deve ser adequado para anotação rápida e consistente por parte do anotador humano.
O modelo deve ser facilmente compreendido por não linguistas, seja ele um estudante de língua, ou um engenheiro com necessidades prosaicas de processamento de língua.
O modelo deve ser adequado para análise computacional com alta acurácia.
O modelo deve auxiliar tarefas de compreensão de língua natural que dependem de tarefas anteriores (extração de relações, tradução de máquina etc.)¹³ Zeman (2023)

¹² Tradução deste autor. Texto original: “The goal of UD is to offer a linguistic representation that is useful for morphosyntactic research, semantic interpretation, and for practical natural language processing across different human languages. It therefore puts an emphasis on simple surface representations that allow parallelism between similar constructions across different languages, despite differences of word order, morphology, and the presence or absence of function words.”

¹³ Tradução deste autor. Texto original: “1. UD must be satisfactory on linguistic analysis grounds for individual languages. 2. UD must be good for linguistic typology, i.e., providing a suitable basis for bringing out cross-linguistic parallelism across languages and

Essas seis proposições fazem a interface entre os campos da linguística, pela tipologia, e da computação, pelo PLN, naturalmente. Os propositores do projeto das UD's chegam até a dizer que facilmente seria possível pensar propostas que enriqueçam cada um desses objetivos propostos. O desafio, dizem eles, está em enriquecer o modelo em todas as suas dimensões – teórica, metodológica e prática (Zeman, 2023).

Nas próximas seções, discutem-se as duas primeiras proposições, com breve menção à quarta proposição. Ora, se está previsto no projeto tanto a análise de línguas individuais quanto a comparação interlinguística, o modelo deve contar com categorias e noções que devem tanto servir para a descrição e análise quanto para comparação. Esse requerimento, que também é um dos fundamentais da tipologia linguística, tem sido a 'pedra no caminho' dos tipólogos há décadas. A solução que pode ser útil às UD's é valer-se de "conceitos comparativos" (HASPELMATH, 2010), para o estabelecimento geral das diretrizes, aos quais se submetem "conceitos descritivos" (HASPELMATH, 2010), na proposição de diretrizes específicas para línguas individuais. De certa forma, as UD's já têm feito isso, como se mostrará nas próximas duas seções.

3 Dependências Universais na interface com a tipologia linguística: em foco, o estabelecimento dos universais

Nesta seção, discute-se a proposição 2 do projeto das UD's: "O modelo deve ser útil para a tipologia linguística, de modo a prover uma base adequada para a comparação entre línguas e famílias linguísticas.". Discute-se a proposição 2 antes da 1 porque, como ficará claro, é o modelo tipológico implicado no projeto das UD's (proposição 2) que garante a análise linguística de línguas particulares (proposição 1).

3.1 A adequação tipológica do modelo

A tipologia linguística é, inegavelmente, um campo consolidado na Linguística, cuja origem moderna pode ser traçada à figura de Joseph H. Greenberg (GREENBERG, 1963, 1966). No entanto, atualmente, há mais de uma proposta tipológica disponível: são alguns exemplos no campo da tipológica funcional, a Gramática Discursivo-Funcional (HENGEVELD; MACKENZIE, 2008), a Teoria Linguística Básica (DIXON, 2010a, 2010b, 2012), a Tipologia Distribucional (BICKEL, 2015), e a Tipologia "Framework-free" (HASPELMATH, 2015; CROFT, 2022). A diferença entre todas essas propostas é tanto teórica

language families. 3. UD must be suitable for rapid, consistent annotation by a human annotator. 4. UD must be easily comprehended and used by a non-linguist, whether a language learner or an engineer with prosaic needs for language processing. ... it leads us to favor traditional grammar notions and terminology. 5. UD must be suitable for computer parsing with high accuracy. 6. UD must support well downstream language understanding tasks (relation extraction, reading comprehension, machine translation, ...)."

quanto metodológica, embora todas convirjam no objetivo central da tipologia que é comparar as línguas do mundo e, a partir daí, estabelecer generalizações.

Talvez a Tipologia “Framework-free” seja a que melhor se alinha às Dependências Universais. A “comparação adequada”, referida na proposição 2, pode ser facilmente interpretada como sendo possibilitada pelo uso de “conceitos comparativos”, que são conceitos independentes de construções de línguas individuais, são criados especificamente com a finalidade de comparação interlinguística (HASPELMATH, 2010).

3.2 Uma discussão sobre a formulação de definições nas UDs

A terminologia linguística tende a ser pouco transparente, seja porque as formulações das definições não são claras o suficiente, seja porque um mesmo termo pode adquirir sentidos diferentes em diferentes teorias e propostas de análise. Não é simples a tarefa metateórica de formulação de definições de termos técnicos. Ela é, entretanto, necessária, se o que se objetiva é alcançar a consistência analítica.

No que diz respeito às UDs, duas questões devem ser apontadas. Em primeiro lugar, como se trata de um modelo computacional para a construção de bancos de dados anotados, as definições propostas funcionam como diretrizes de anotação ou, pelo menos, como um guia para o estabelecimento de diretrizes de anotação. Essa é a face “metacomputacional” do modelo. Em segundo lugar, as definições de categorias como *ADJETIVO* ou *nsubj*, quando parte da lista de UPOS ou *deprels* universais, devem ser feitas na forma de conceitos comparativos. É exatamente isso que se observa na definição da UPOS *ADJETIVO*, como se vê em (01).

- (01) **ADJETIVOS** são palavras que tipicamente modificam substantivos e especificam seus atributos ou suas propriedades.¹⁴

Note-se que a definição dos *ADJETIVOS* é formulada com base em dois critérios, um semântico – “... especificam seus atributos ou suas propriedades...” – e outro pragmático – “... são palavras que tipicamente modificam...”. Não há, aí, qualquer indicação de que o adjetivo deve apresentar tais morfemas, ou que deve ocorrer em tais construções gramaticais, porque essas indicações só podem ser feitas se se considerar a distribuição e o comportamento morfossintático dessas palavras em línguas específicas (veja-se a discussão na Seção 4.2). A caracterização de *ADJETIVO* em (01) permite que sejam identificadas todas as palavras que semanticamente denotam um conceito de **propriedade** (THOMPSON, 1988, 2000) e que pragmaticamente operam **modificação** (CROFT, 1991). Pode-se dizer que tal caracterização diz respeito ao adjetivo prototípico (nos termos de Croft (2022)).

¹⁴ Tradução deste autor. Texto original: “Adjectives are words that typically modify nouns and specify their properties or attributes”. Disponível em: <https://universaldependencies.org/u/pos/ADJ.html>. Acesso em: 8 de julho de 2023, às 22h06.

No domínio da morfossintaxe, uma definição semelhante pode ser verificada para a *deprel obj*, objeto, em (02).

(02) **obj**: argumento nominal central mais básico que não é o sujeito, em geral o participante mais diretamente afetado.

A formulação em (02) pode parecer circular ou tautológica: um argumento é objeto porque não é argumento sujeito. No entanto, a chave da interpretação dessa caracterização é “argumento central mais básico... o participante mais diretamente afetado”. Fica pressuposto que, numa predicação, argumentos tanto em posição de *sujeito* quanto de *objeto* são básicos; o que diferencia essas duas funções é a saliência do participante. Em uma predicação transitiva, se o participante agente é construído como mais saliente, então ele assume a posição de sujeito, e o participante afetado, a de objeto (cf. a discussão de Croft (2022, §6.2.1)).

Por meio da discussão do ADJETIVO e do *obj* nas UD, pode-se reconhecer sem muita dificuldade a importância de conhecer a teoria que está na base do estabelecimento de quadros categoriais do modelo. Caso contrário, é bastante provável que se incorra em interpretações rasas e confusas dessas categorias.

Acontece que a tipologia linguística, no estado atual de seu desenvolvimento teórico e descritivo, tem se transformado num campo do conhecimento extremamente complexo. E pode acontecer de não especialistas se interessarem pelas UD, o que não é incomum, já que é um modelo computacional. Neste ponto da discussão, é importante resgatar a proposição 4, “O modelo deve ser facilmente compreendido por não linguistas, seja ele um estudante de língua, ou um engenheiro com necessidades prosaicas de processamento de língua”. Ora, é necessário, então, que as definições que se apresentam para as categorias que compõem o modelo sejam acessíveis a não especialistas. A tarefa de formular definições acessíveis a não especialistas não é simples, porque muito facilmente pode-se incorrer em um ‘barateamento’ conceitual ou pode-se apresentar simplesmente uma definição inadequada. É o caso da definição que se apresenta para SUBSTANTIVO, em (03).

(03) **SUBSTANTIVO** é a classe de palavras que tipicamente denota uma pessoa, lugar, coisa, animal ou ideia.¹⁵

Essa definição é tipicamente nocional. Definições dessa natureza já estão ultrapassadas na tipologia linguística, porque reconhece-se que apenas a dimensão nocional não capta exatamente a natureza de uma classe de palavra (CROFT, 2000, 2001, 2022, 2023; HASPELMATH 2012, 2023; RAUH, 2010). Veja-se que, comparando-se a definição de SUBSTANTIVO com a de ADJETIVO, não se verifica um

¹⁵ Tradução deste autor. Texto original: “Nouns are a part of speech typically denoting a person, place, thing, animal or idea.”. Disponível em: <https://universaldependencies.org/u/pos/NOUN.html>. Acesso em 08 de julho de 2023, às 23h09.

paralelismo: a definição de ADJETIVO é tanto semântica quanto pragmática, ao passo que a definição de SUBSTANTIVO é puramente nocional.

Definições como a em (03) decorrem da tradição gramatical ocidental, a qual também está na base do modelo (DE MARNEFFE *et al.*, 2021, p. 260). O apego à tradição gramatical, pelo menos em parte, pode justificar-se pela necessidade de manter o modelo acessível a não especialistas. No entanto, é possível encontrar o ponto de equilíbrio entre tradição e tipologia: a nomenclatura pode ser a tradicional – obviamente não é necessário fazer um desmonte teórico –, mas as definições devem ser formuladas como conceitos comparativos, sempre que possível.

3.3 Uma amostra de comparação com base em uma amostra de línguas

Como foi dito na Introdução deste artigo, há poucos estudos de comparação interlinguística que se valem dos bancos de dados das UD. Um dos motivos para isso talvez seja o fato de que as 141 línguas que compõem o conjunto do projeto não se distribuem uniformemente de modo a constituir uma amostra linguística diversa o suficiente (MIESTANO; BAKKER; ARPPE, 2016; RIJKHOFF *et al.*, 1993; DRYER, 1989). De Marneffe *et al.* (2023) apresentam a seguinte distribuição das línguas em famílias na atual versão do projeto das UD: 48% indo-europeia, 13% outras, 9% urálica, 9% afro-asiática, 9% tupi, 4% turcomanas, 3% austronésia, 2% dravídica, 2% sino-tibetana, 2% níger-congo.

Nesta seção, esboça-se uma comparação entre as línguas de três famílias linguísticas: a afro-asiática, a indo-europeia e a turcomana. Os dados que se apresentam nos Quadros 4 a 6, a seguir, foram extraídos do site do projeto das Dependências Universais¹⁶.

Línguas afro-asiáticas	
Língua	UPOS
acadiano	13
amárico	16
árabe	16
assírio	13
beja	16
cóptico	17
hebraico	16
hebraico antigo	16
vietnamita	19
zaar	16

Quadro 4 – Número de UPOS usados nas línguas afro-asiáticas da amostra das UD.

Fonte: Elaboração própria.

¹⁶ Disponível em: www.universaldependencies.org. Acesso em 09 de julho de 2023, às 22h42.

Línguas indo-europeias					
Língua	UPOS	Língua	UPOS	Língua	UPOS
africâner	17	francês	17	lígure	17
albanês	14	francês antigo	s.i.	lituano	17
alemão	17	gaélico escocês	17	manquês	17
alemão suíço	17	galego	17	marathi	17
alto sorábio	17	galês	17	napolitano	17
antigo eslavo eclesiástico	14	gheg	15	nayini	16
armênio	17	gótico	14	norueguês	17
armênio ocidental	17	grego antigo	17	persa	16
baixo saxão	17	grego	16	polonês	17
bengali	17	hindi	17	pomak	17
bielorrusso	16	hitita	15	português	17
boiapuri	17	holandês	17	romeno	17
bretão	16	inglês	17	russo	17
búlgaro	15	irlandês	16	sânscrito	16
catalão	17	irlandês antigo	17	servo	17
croata	17	islandês	17	sinhala	13
dinamarquês	17	italiano	17	soi	16
eslavo oriental antigo	17	kanger	17	sueco	17
eslovaco	17	khunsari	17	tcheco	17
esloveno	17	kurmanji	17	ucraniano	17
espanhol	17	latim	17	umbrian	17
feroesa	17	letão	17	urdu	s.i.

Legenda: "s.i." = sem informação.

QUADRO 5 – Número de UPOS usados nas línguas indo-europeia da amostra das UDs.

Fonte: Elaboração própria.

Línguas turcomanas	
Língua	UPOS
cazaque	17
quirgiz	14
tártara	17
turco	16
turco antigo	13
uigur	16
yakut	17

QUADRO 6 – Número de UPOS usados nas línguas turcomanas da amostra das UDs.

Fonte: Elaboração própria.

Como se vê pelos Quadros acima, é unânime o uso de praticamente todas as 17 etiquetas de UPOS. As línguas que não fazem uso das 17, em geral não usam SYM, PUNCT e X, as três etiquetas que se ligam diretamente à natureza computacional do programa. Alguns casos interessantes a se mencionar são:

- a) a língua gheg (indo-europeia): não faz uso de SYM e PROPEN;

- b) a língua bretã (indo-europeia): não faz uso de PART;
- c) a língua acadiana (afro-asiática): não faz uso de AUX, INTJ, PUNCT e SYM;
- d) a língua assíria (afro-asiática): não faz uso de NUM, INTJ, SYMB e X;
- e) a língua vietnamita (afro-asiática): tem um conjunto de etiquetas absolutamente distinto.

Para que seja possível extrair qualquer generalização, no entanto, seria necessário examinar obras gramaticais dessas línguas. Além disso, deve-se considerar o tipo de texto contemplado nas amostras de anotação para cada língua. Por exemplo, nos bancos de dados de línguas como grego antigo e grego contemporâneo não foram encontradas INTJ, mas os anotadores justificam a não ocorrência devido ao tipo de texto da amostra.

Se o estudo ‘tradicional’ apresenta limitações no que diz respeito à composição da amostra de línguas e de material de análise, vê-se que as mesmas dificuldades são encontradas na tipologia de base computacional. De todo modo, os avanços são mais significativos do que as dificuldades.

4 Dependências Universais como um modelo de análise linguística: em foco, a documentação do particular

Nesta seção, discute-se a proposição 1 do projeto das UDs: “O modelo deve ser satisfatório no que diz respeito à análise linguística de línguas específicas”. Espera-se mostrar, nesta seção, que a análise de línguas específicas pode estar ancorada no modelo de tipologia discutido na seção anterior.

4.1 A adequação descritiva do modelo

Na linha da distinção que se fez entre “modelo” e “teoria” na seção 3.1, pode-se levantar a necessidade de se distinguir entre “descrição” e “análise”. Essa pode parecer uma distinção legítima na linguística, mas, como muitos autores já esclareceram (GIVÓN, 1995; DRYER, 2006; HASPELMATH, 2015, entre outros), ela é falsa, pois toda descrição implica análise – em diferentes níveis de generalização, obviamente.

Essa realidade é muito particularmente verificada na construção de bancos de dados anotados. Se do ponto de vista da descrição linguística para a produção de obras gramaticais de línguas individuais, descrição implica análise, pode-se facilmente argumentar que a prática de anotação para a construção de bancos de dados também implica análise. Como dizem Gerdes e Kahane (2016), todo projeto de desenvolvimento de banco de dados anotados se constrói sobre escolhas entre possíveis anotações, as quais são determinadas por “princípios gerais” (p. 139). Aí está, pois, o papel da teoria. Esses autores discutem uma série de possibilidades de anotações para diversos fenômenos

linguísticos segundo o modelo das UD_s, dentre eles, a coordenação, mostrando o modo pelo qual escolhas teóricas – no caso deles, alinhadas à gramática léxico-funcional (BRESNAN, 1981) e à teoria *meaning-text* (MEL'ČUK, 1988) – determinam o estabelecimento de critérios de anotação.

Na direção de propor a conciliação entre o exercício tipológico e a tarefa de descrição de línguas individuais, Mel'čuk (2006, p. 171) reconhece que “um problema central na relação entre tipologia e a escrita de gramáticas de línguas individuais é o desenvolvimento de um sistema conceptual viável do ponto de vista interlinguístico e um modelo terminológico correspondente”¹⁷. Essa mesma dificuldade se aplica à anotação de línguas individuais. No caso das UD_s, isso significa tomar categorias universais (UPOS e *deprels*) e modelá-las de acordo com as especificidades de cada sistema linguístico. A seguir discute-se essa questão examinando-se duas diretrizes de anotação para o português brasileiro (DURAN, 2021, 2022) em confronto com as diretrizes gerais das UD_s.

4.2 A elaboração de diretrizes para anotação de línguas específicas, em foco o português brasileiro

O português brasileiro é uma língua com cada vez mais recursos de PLN disponíveis. Há inúmeros grupos que trabalham com a preparação e a disponibilização de bancos de dados anotados morfosintática e semanticamente. Particularmente, a anotação de bancos de dados segundo o modelo das UD_s tem ganhado bastante espaço no cenário da linguística brasileira, sendo uma característica distintiva dessa empreitada a anotação de textos “não canônicos”, isto é, fora do domínio jornalístico (Souza *et al*, 2021; Pardo *et al*, 2021).

Nesta seção, direciona-se a atenção a dois manuais de anotação segundo o modelo das UD_s, um sobre anotação de classes de palavras (POS Tag – Duran (2021)) e outro sobre as relações de dependência (*deprels* – Duran (2022)). O exame vai na direção de explicitar o modo pelo qual as diretrizes gerais das UD_s podem (e devem) ser especificadas para acomodar os fatos sistêmicos da gramática de línguas particulares.

Considere-se a definição de Adjetivo¹⁸ apresentada no manual de Pos Tag, apresentada em (04).

(04) Adjetivos são palavras que **modificam** os substantivos, especificando suas **propriedades** e **atributos**. Os adjetivos constituem uma classe aberta de palavras. A maioria deles **sofre flexão** de gênero (feminino ou masculino), de número (singular ou plural), e de grau (diminutivo, aumentativo e superlativo). Ex: *bom, boa, bons, boas, bonzinho, bonzão, boníssimo*. Alguns adjetivos

¹⁷ Tradução deste autor. Texto original: “A central problem in the relationship between typology and the writing of individual grammars is that of developing a cross-linguistically viable conceptual system and a corresponding terminological framework.”

¹⁸ As categorias grafas com a primeira letra maiúscula correspondem a categorias descritivas de uma língua particular, não a um conceito comparativo (CROFT, 2001).

não variam em gênero (ex: *fácil, difícil*), outros não variam nem em gênero nem em número (ex: *simples*). (Duran, 2021, p. 7 – negritos e sublinhados deste autor, itálicos originais)

A primeira parte da definição é semelhante àquela oferecida pelas UD's, em (01) acima. Novamente, a definição afirma as propriedades semânticas – “especificando suas **propriedades e atributos**” – e pragmáticas – “são palavras que **modificam**” – que caracterizam a classe.

É interessante notar que na formulação da definição do Adjetivo em português, Duran (2021) apresenta uma série de propriedades morfosintáticas típicas dessa classe: a possibilidade de flexão em **gênero, número e grau**. Essas propriedades designam o “comportamento potencial” (CROFT, 2003, 2022) da classe. E é com base nelas que a autora estabelece duas classes de adjetivos em português: a classe dos que potencialmente flexionam e a classe dos que não flexionam. Essas são particularidades da gramática do português brasileiro. Em outras línguas, como a luó (Níger-congo), os adjetivos não apresentam flexão nominal (STASSEN, 1997, p. 161).

Duran (2021, p. 7) ainda apresenta as construções gramaticais básicas em que o Adjetivo pode ocorrer em português: (i) antes ou depois do substantivo – construção de admodificação; (ii) modificação por advérbio – construção de intensificação; (iii) o Adjetivo liga-se ao substantivo por verbo de ligação – construção de predicado nominal. A autora ainda fala da possibilidade do Adjetivo “ocorrer em grupo”, podendo aparecer em coordenação. Novamente, esses fatos são específicos da gramática do português. Em português, o Adjetivo em posição predicativa continua sendo Adjetivo, sem sofrer qualquer alteração morfológica. No entanto, em línguas como luó (Níger-congo) e oromo (Afro-asiática) o Adjetivo em função predicativa apresenta morfologia de nominalização e morfologia de tempo e aspecto verbal, respectivamente (STASSEN, 1997, p. 161-166).

Um outro caso que merece consideração é a anotação dos Pronomes Possessivos. Em Duran (2021, 2022), quanto à anotação de POS, está a proposta de que as palavras dessa categoria sejam anotadas em português como determinantes (DET). Quanto à anotação de *deprel*, as diretrizes gerais das UD's deixam aberta a possibilidade de se anotar essas palavras ou como estabelecendo a relação de determinação (*det*) ou a de modificação nominal (*nmod*).

A interpretação da categoria “pronomes possessivos” como determinante é uma herança da gramática gerativa (ABNEY, 1987) e não necessariamente corresponde à função que as palavras dessa classe desempenham na construção das sentenças. Acontece que, pelo menos em português, o pronome possessivo pode indicar posse propriamente, relação e papéis temáticos (NEVES, 2011), como se vê nas ocorrências de *corpus*¹⁹ (05) a (08).

- (05) Relação de posse: E assim que Petra desapareceu no elevador com suas **malas**...
- (06) Relação assimétrica: O meu **pai** era paulista.
- (07) Relação simétrica: Se não acredita, pega o telefone e liga para o meu **irmão**.

¹⁹ Corpus do Laboratório de Lexicografia da Unesp de Araraquara.

- (08) Papel temático de agente: Tudo somado, pode-se concluir que sua **aparição** não foi um bom negócio.

O pronome possessivo sempre opera uma relação entre duas pessoas do discurso: em (05), por exemplo, tem-se a relação entre uma 3ª pessoa (malas) e outra 3ª pessoa (Petra). Nesse sentido, o pronome possessivo sempre opera uma relação genitiva, que pode ser interpretada, no modelo das UDs, como estabelecendo a relação de *nmod*: “suas malas”, “as malas de Petra”. Qual é, portanto, a melhor anotação para o Pronome Possessivo em português – como estabelecendo a relação de dependência *det* ou *nmod*? Não há resposta correta. A escolha deve ser balizada, em primeiro lugar, pela manutenção da coerência teórica que está na base das diretrizes de anotação e, em segundo lugar, pela maximização da possibilidade de comparação com outros bancos de dados.

Esses dois pontos relativos à escolha de como anotar determinadas estruturas podem ser princípios gerais para o estabelecimento de diretrizes, juntamente com o princípio de praticidade na implementação computacional do modelo (cf. GERDES; KAHANE, 2016).

Considerações finais

Este artigo pretendeu apresentar o modelo das Dependências Universais, um modelo computacional de morfossintaxe, pela ótica da linguística teórica, particularmente de uma visão tipológica “framework-free”. Em última instância, espera-se ter podido delinear, pelo menos minimamente, a teoria que está na base do modelo. Espera-se, também, que as dificuldades enfrentadas pelo modelo das UDs não representam fraquezas, mas, sim, oportunidades para fortalecê-lo ainda mais.

O quadro categorial geral das UDs, isto é, tanto as etiquetas de POS quanto as *deprels* podem ser interpretadas como sendo conceitos comparativos, especialmente vistas pela proposição das diretrizes gerais do modelo, sendo possível transformá-las em categorias descritivas por meio da especificação das propriedades categoriais de cada língua, de acordo com o desenvolvimento de diretrizes de anotação individuais. Essa dupla visão sobre o quadro categorial das UDs pode garantir isonomia e equivalência teórico-metodológica, de modo a otimizar a realização de tarefas de análise e descrição linguísticas assistidas por modelos computacionais (proposições 1 e 2 do projeto das UDs – seção 2.2).

Muito significativamente, pode-se dizer que conhecer a teoria de base garante o avanço consistente do modelo, principalmente no que diz respeito ao estabelecimento de categorias analíticas (como é o caso das UPOS e das *deprels*), bem como no que diz respeito à proposição de diretrizes de anotação, como muito bem fez Duran (2021, 2022). Espera-se que as ideias apresentadas aqui sejam apenas o ponto de partida para discussões necessárias na comunidade de pesquisadores das Dependências Universais.

Informações complementares

Avaliação e resposta dos autores

Avaliação: <https://doi.org/10.25189/rabralin.v22i2.2176.R>

Editores

Jorge Manuel Evangelista Baptista

Afiliação: Universidade do Algarve

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

Adriana Silvina Pagano

Afiliação: Universidade Federal de Minas Gerais

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

Marta Deysiane Alves Faria Sousa

Afiliação: Universidade Federal de Sergipe

ORCID: <https://orcid.org/0000-0002-0480-0422>

RODADAS DE AVALIAÇÃO

Avaliador 1: Adriana Silvina Pagano

Afiliação: Universidade Federal de Minas Gerais

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

Avaliador 2: Juliano Desiderato Antonio

Afiliação: Universidade Estadual de Maringá

ORCID: <https://orcid.org/0000-0002-9816-5852>

AVALIADOR 1

O artigo indaga o modelo das UD's e argumenta em prol da discussão dos pressupostos teóricos subjacentes a esse modelo para podermos consolidar um arcabouço que garanta a comparabilidade entre línguas e dê sustentação às decisões analíticas tomadas pelas línguas de forma individual.

Trata-se de um artigo relevante no estado atual das discussões sobre gramática de dependências, não somente no contexto nacional, mas também no internacional, tendo-se em vista o impacto

do modelo das UDs, tanto no campo do Processamento de Linguagem Natural como no campo dos estudos da linguagem.

O artigo está bem estruturado e organizado e apresenta discussão inédita no escopo dos estudos gramaticais da linguagem. Atende plenamente as convenções de redação acadêmica.

Sugestões de aprimoramento foram inseridas no arquivo disponibilizado na plataforma de avaliações.

AVALIADOR 2

O trabalho está muito bem escrito e organizado e traz importantes contribuições para o modelo no qual se insere.

Fiz algumas observações pontuais no arquivo do manuscrito.

Conflito de Interesse

O autor não tem conflitos de interesse a declarar.

Protocolo e Pré-Registro de Pesquisa (obrigatório)

Avaliando os roteiros propostos pela Equator Network, considero que nenhum deles se mostra relevante para a pesquisa em tela. Também informamos que a pesquisa desenvolvida não foi pré-registrada em repositório institucional independente.

Declaração de Disponibilidade de Dados

O compartilhamento de dados não é aplicável a este artigo, pois nenhum dado novo foi criado ou analisado neste estudo.

Agradecimentos

Agradeço à Adriana Pagano, interlocutora arguta. As ideias presentes neste artigo são, em parte, fruto de nossas discussões. Obviamente, quaisquer erros e atalhos são inteiramente minha responsabilidade.

REFERÊNCIAS

ABNEY, Steven Paul. **The English noun phrase in its sentential aspect**. 363 ff. Tese (Doutorado em Linguística) – Massachusetts Institute Technology, MIT, Cambridge, MA, 1987.

ÁNGEL, Vilmos et al. (orgs.). **Dependency and valency: an international handbook of contemporary research**. 2 vols. Berlin: Mouton de Gruyter, 2003.

BICKEL, Balthasar. Distributional typology: statistical inquiries into the dynamics of linguistic diversity. In: HEINE, Bernd; NARROG, Heiko (orgs.). **The Oxford Handbook of Linguistic Analysis**. 2 ed. Oxford: Oxford University Press, 2015, p. 901-924.

BORBA, Francisco da Silva. (coord.) **Dicionário gramatical de verbos do português contemporâneo do Brasil**. São Paulo: Editora Ática, 1990.

BORBA, Francisco da Silva. **Uma gramática de valências para o português**. São Paulo: Editora Ática, 1996.

BRESNAN, Joan. An approach to Universal Grammar and the mental representation of language. **Cognition**, vol. 10, n. 1, p. 39-52, 1981.

CARNIE, Andrew et al. (orgs.). **The Routledge Handbook of Syntax**. New York: Routledge, 2014.

CHAFE, Wallace. **Significado e estrutura linguística**. Tradução de Maria Helena de Moura Neves et al. São Paulo: Livros Técnicos e Científicos Editora, 1979.

COLOMBAT, Bernard. The notion of dependency in Latin grammar in the Renaissance and the 17th century. In: IMRÉNYI, András; MAZZIOTTA, Nicolas (orgs.). **Chapters of dependency grammar: a historical survey from Antiquity to Tesnière**. Amsterdam: John Benjamins, 2020, p. 59-84.

CROFT, William. **Syntactic categories and grammatical relations: the cognitive organization of information**. Chicago: The University of Chicago Press, 1991.

CROFT, William. Parts of speech as typological universals and as language particular categories. In: VOGEL, Petra Maria; COMRIE, Bernard. **Approaches to the typology of word classes**. Berlin: Mouton de Gruyter, 2000, p. 65-102.

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

CROFT, William. **Typology and universals**. 2 ed. Cambridge, UK: Cambridge University Press, 2003.

CROFT, William. **Morphosyntax: constructions of the world's languages**. Cambridge, UK: Cambridge University Press, 2022.

CROFT, William. Word classes in Radical Construction Grammar. In: VAN LIER, Eva (org.). **The Oxford Handbook of Word Classes**. Oxford: Oxford University Press, 2023.

DE MARNEFFE, Marie-Catherine et al. Universal Stanford Dependencies: a cross-linguistic typology. In: Ninth International Conference of Language Resources and Evaluation (LREC), 2014. *Proceedings...* Reykjavik: ACL, p. 1585-1592.

DE MARNEFFE, Marie-Catherine et al. Universal dependencies. **Computational Linguistics**, vol. 47, n. 2, p. 255-308, 2021.

DE MARNEFFE, Marie-Catherine; NIVRE, Joakim; ZEMAN, Dan. **Tutorial on Universal Dependencies**. Apresentação em 1st Webinar UniDive Cost, 19 de junho de 2023. Disponível em: <https://github.com/UniDive/2023-uni-dive-webinar/blob/main/ud-tutorial/01-intro.pdf>. Acesso em 10 de julho de 2023, às 15h55.

DIXON, Robert M. W. **Basic Linguistic Theory**. vol. 1 – Methodology. Oxford: Oxford University Press, 2010a.

DIXON, Robert M. W. **Basic Linguistic Theory**. vol. 2 – Grammatical Topics. Oxford: Oxford University Press, 2010b.

DIXON, Robert M. W. **Basic Linguistic Theory**. vol. 3 – Further Grammatical Topics. Oxford: Oxford University Press, 2012.

DRYER, Matthew S. Large linguistic areas and language sampling. **Studies in Language**, vol. 13, n. 2, p. 257-292, 1989.

DRYER, Matthew S. Descriptive theories, explanatory theories, and Basic Linguistic Theory. In: AMEKA, Felix; DENCH, Alan; EVANS, Nicholas (orgs.). **Catching language: the standing challenge of grammar writing**. Berlin: Mouton de Gruyter, 2006, p. 207-234.

DURAN, Magali. **Manual de anotação de POS tags**: orientações para anotação de etiquetas morfossintáticas em língua portuguesa, segundo as diretrizes da abordagem Universal Dependencies. Relatório Técnico, n. 434, 2021. Disponível em: <https://sites.google.com/icmc.usp.br/poetisa/annotation-material>. Acesso em 10 de julho de 2023, às 15h50.

DURAN, Magali. **Manual de anotação de relações de dependência** – versão revisada e estendida. Relatório Técnico, n. 440, 2022. Disponível em: <https://sites.google.com/icmc.usp.br/poetisa/annotation-material>. Acesso em 10 de julho de 2023, às 15h50.

FREDDI, Maria. Corpus methods in syntax. In: LURAGHI, Silvia; PARODI, Claudia (orgs.). **The Bloomsbury Companion to Syntax**. Londres: Bloomsbury, 2013, p. 51-62.

GERDES, Kim; KAHANE, Sylvain. Dependency annotation choices: assessing theoretical and practical issues of Universal Dependencies. In: Proceedings of the 10th Linguistic Annotation Workshop, 2016. **Anais...** Berlin: ACL, p. 131-140.

GIVÓN, Talmy. **Functionalism and grammar**. Amsterdam: John Benjamins, 1995.

GREENBERG, Joseph. Some universals of grammar with particular reference to the order of meaningful elements. In: GREENBERG, Joseph (org.). **Universals of language**. Cambridge, MA: The MIT Press, 1963, p. 73-113.

GREENBERG, Joseph. **Language universals**. Berlin: Mouton de Gruyter, 1966.

HAIČOVÁ, Eva; SGALL, Petr. Dependency syntax in functional generative description. In: ÁNGEL, Vilmos et al. (orgs.) **Dependency and valency: an international handbook of contemporary research**. 2 vols. Berlin: Mouton de Gruyter, 2003, p. 570-593.

HASPELMATH, Martin. Comparative concepts and descriptive categories in crosslinguistic studies. **Language**, vol. 86, n. 3, p. 663-687, 2010.

HASPELMATH, Martin. How to compare major word-classes across the world's languages. **UCLA Working Papers in Linguistics**, vol. 17, p. 109-130, 2012.

HASPELMATH, Martin. Framework-free grammatical theory. In: HEINE, Bernd; NARROG, Heiko (orgs.). **The Oxford Handbook of Linguistic Analysis**. 2 ed. Oxford: Oxford University Press, 2015, p. 287-310.

HASPELMATH, Martin. Word class universals and language-particular analysis. In: VAN LIER, Eva (org.). **The Oxford Handbook of Word Classes**. Oxford: Oxford University Press, 2023.

HELBIG, Gerhard; SCHENKEL, Wolfgang. **Wörterbuch Zur Valenz Und Distribution Deutscher Verben**. Berlin: Mouton de Gruyter, 1991 [1983].

HENGEVELD, Kees; MACKENZIE, J. Lachlan. **Functional Discourse Grammar**. Oxford: Oxford University Press, 2008.

HERINGER, Hans Jürgen. Basic ideas and the classical model. In: STEGER, Hugo; WIEGAND, Herbert Ernst (orgs.). **Syntax: an international handbook of contemporary research**. vol. 1. Berlin: Mouton de Gruyter, 1993a, p. 298-316.

HERINGER, Hans Jürgen. Formalized models. In: STEGER, Hugo; WIEGAND, Herbert Ernst (orgs.). **Syntax: an international handbook of contemporary research**. vol. 1. Berlin: Mouton de Gruyter, 1993b, p. 316-329.

HUDSON, Richard. Word grammar. In: ÁNGEL, Vilmos et al. (orgs.). **Dependency and valency: an international handbook of contemporary research**. vol. 1. Berlin: Mouton de Gruyter, 2003, p. 508-526.

HUDSON, Richard. **An introduction to Word Grammar**. Cambridge, UK: Cambridge University Press, 2010.

JING, Yingqui; WIDMER, Paul; BICKEL, Balthasar. Word order variation is partially constrained by syntactic complexity. **Cognitive Science**, vol. 45, p. 1-25, 2021.

KERTÉSZ, András et al. (orgs.). **Current approaches to syntax: a comparative handbook**. Berlin: Mouton de Gruyter, 2019.

LANGACKER, Ronald. Structural syntax: the view from Cognitive Grammar. **Sémiotiques**, n. 6-7, p. 69-84, 1994.

LEVSHINA, Natalia. Cross-linguistic trade-offs and causal relationships between cues to grammatical subject and object, and the problem of efficiency-related explanations. **Frontiers in Psychology**, vol. 12, p. 1-20, 2021.

LEVSHINA, Natalia. Frequency, informativity, and word length: insights from typologically diverse corpora. **Entropy**, vol. 24, n. 280, p. 1-16, 2022.

LURAGHI, Silvia; PARODI, Claudia (orgs.). **The Bloomsbury Companion to Syntax**. Londres: Bloomsbury, 2013.

LUHTALA, Anneli. Syntactic relations in ancient and medieval grammatical theory. In: IMRÉNYI, András; MAZZIOTTA, Nicolas (orgs.). **Chapters of dependency grammar: a historical survey from Antiquity to Tesnière**. Amsterdam: John Benjamins, 2020, p. 23-58.

MAZZIOTTA, Nicolas; IMRÉNYI, András. Aspects of the theory and history of dependency grammar. In: IMRÉNYI, András; MAZZIOTTA, Nicolas (orgs.). **Chapters of dependency grammar: a historical survey from Antiquity to Tesnière**. Amsterdam: John Benjamins, 2020, p. 1-21.

MEL'ČUK, Igor. **Dependency syntax: theory and practice**. Albany: State University of New York Press, 1988.

MEL'ČUK, Igor. Calculus of possibilities as a technique in linguistic typology. In: AMEKA, Felix; DENCH, Alan; EVANS, Nicholas (orgs.). **Catching language: the standing challenge of grammar writing**. Berlin: Mouton de Gruyter, 2006, p. 171-206.

- MEL'ČUK, Igor. Dependency in natural language. In: POLGUÈRE, Alain; MEL'ČUK, Igor (orgs.). **Dependency in linguistic description**. Amsterdam: John Benjamins, 2009, p. 1-110.
- MIESTANO, Matti; BAKKER, Dik; ARPPE, Antti. Sampling for variety. **Linguistic Typology**, vol. 20, n. 2, p. 233-296, 2016.
- MORAVCSIK, Edith. Introduction. In: KERTÉSZ, András et al. (orgs.). **Current approaches to syntax: a comparative handbook**. Berlin: Mouton de Gruyter, 2019, p. 1-21.
- NEVES, Maria Helena de Moura. **Gramática de usos do português**. 2 ed. São Paulo: Editora Unesp, 2011.
- NIVRE, Joakim et al. Universal Dependencies v1: a multilingual treebank collection. In: 10th International Conference on Language Resources and Evaluation (LREC), 2016. **Proceedings...** Portorož: ACL, p. 1659-1666.
- NIVRE, Joakim et al. Universal Dependencies v2: an evergrowing multilingual Treebank collection. In: 12th Conference on Language Resources and Evaluation (LREC), 2020. **Proceedings...** Marseille: ACL, p. 4034-4043.
- OSBORNE, Timothy. Dependency Grammar. In: CARNIE, Andrew et al. (orgs.). **The Routledge Handbook of Syntax**. New York: Routledge, 2014, p. 604-626.
- OSBORNE, Timothy. **A dependency grammar of English: an introduction and beyond**. Amsterdam: John Benjamins, 2019a.
- OSBORNE, Timothy. Dependency grammar. In: KERTÉSZ, András et al. (orgs.). **Current approaches to syntax: a comparative handbook**. Berlin: Mouton de Gruyter, 2019b, p. 361-388.
- PARDO, Thiago et al. Porttinari – a Large Multi-genre Treebank for Brazilian Portuguese. In: Simpósio Brasileiro de Tecnologia de Informação e da Linguagem Humana. **Anais...** Porto Alegre: SBC, p. 1-10, 2021.
- PERCIVAL, W. Keith. Reflections on the history of dependency notions in linguistics. **Historiographia Linguistica**, vol. 17, n. 1-2, p. 29-47, 1990.
- RAUH, Gisa. **Syntactic categories: their identification and description in linguistic theories**. Oxford: Oxford University Press, 2010.
- RIJKHOFF, Jan et al. A method of language sampling. **Studies in Language**, vol. 17, n. 1, p. 169-203, 1993.
- SAUSSURE, Ferdinand de. **Curso de linguística geral**. 27. ed. Tradução de José Paulo Paes, Antônio Chelini, Izidoro Blikstein. São Paulo: Edifora Cultrix, 2006 [1916].
- SOUZA, Elvis et al. PetroGold – Corpus padrão ouro para o domínio do petróleo. In Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana. **Anais...** Porto Alegre: SBC, p. 29-38, 2021.
- STASSEN, Leon. **Intransitive predications**. Oxford: Oxford University Press, 1997.
- STEGER, Hugo; WIEGAND, Herbert Ernst (orgs.). **Syntax: an international handbook of contemporary research**. vol. 1. Berlin: Mouton de Gruyter, 1993.
- TESNIÈRE, Lucien. **Elements of structural syntax**. Tradução de Timothy Osborne e Sylvain Kahane. Amsterdam: John Benjamins, 2015 [1959].
- THOMPSON, Sandra. A discourse approach to the cross-linguistic category 'adjective'. In: HARKINS, John (org.). **Exploring language universals**. New York: Blackwell, 1988, p. 167-185.

THOMPSON, Sandra. Property concepts. In: BOOIJ, Geert et al. (orgs.). **Morphology**: an international handbook of inflection and word-formation. Berlin: Mouton de Gruyter, 2000, p. 1111-1117.

VAN LANGENDONCK, Williy. The dependency concept and its foundation. In: ÁNGEL, Vilmos et al. (orgs.). **Dependency and valency**: an international handbook of contemporary research. 2 vols. Berlin: Mouton de Gruyter, 2003, p. 170-188.

ZEMAN, Daniel. **Universal Dependencies**. Apresentação NPFL075 Dependency Grammars and Treebanks, 23 de março de 2023. Disponível em: <https://ufal.mff.cuni.cz/~zeman/vyuka/deptreebanks/ud1-intro-morpho.pdf>. Acesso em 10 de julho de 2023, às 15h48.