



Programa de
Pós-Graduação em
Linguística

**NÃO É NADA NÃO: UMA ANÁLISE DAS LEXIAS NEGATIVAS DO PORTUGUÊS
PARA ANOTAÇÃO NAS UNIVERSAL DEPENDENCIES**

SÃO CARLOS

2022



Universidade Federal de São Carlos

ISAAC SOUZA DE MIRANDA JUNIOR

**UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE EDUCAÇÃO E CIÊNCIAS HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM LINGUÍSTICA**

**NADA NÃO, UMA ANÁLISE DO LÉXICO NEGATIVO DO PORTUGUÊS PARA
UNIVERSAL DEPENDENCIES**

ISAAC SOUZA DE MIRANDA JUNIOR
Bolsista: CAPES

Dissertação apresentada ao Programa de Pós-Graduação em Linguística da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do Título de Mestre em Linguística.

Orientador: Prof. Dr. Oto Vale

São Carlos - São Paulo - Brasil

2022

Miranda Junior, Isaac Souza de

Não é nada não: uma análise das lexias negativas do português para anotação nas Universal Dependencies / Isaac Souza de Miranda Junior -- 2022. 133f.

Dissertação (Mestrado) - Universidade Federal de São Carlos, campus São Carlos, São Carlos

Orientador (a): Oto Araújo Vale

Banca Examinadora: Adriana Silvana Pagano, Thiago

Alexandre Salgueiro Pardo

Bibliografia

1. Linguística Computacional. 2. Anotação de Corpus. 3. Negação. I. Miranda Junior, Isaac Souza de. II. Título.

Ficha catalográfica desenvolvida pela Secretaria Geral de Informática (SIn)

DADOS FORNECIDOS PELO AUTOR

Bibliotecário responsável: Ronildo Santos Prado - CRB/8 7325



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Educação e Ciências Humanas
Programa de Pós-Graduação em Linguística

Folha de Aprovação

Defesa de Dissertação de Mestrado do candidato Isaac Souza de Miranda Junior, realizada em 29/09/2022.

Comissão Julgadora:

Prof. Dr. Oto Araujo Vale (UFSCar)

Profa. Dra. Adriana Silvina Pagano (UFMG)

Prof. Dr. Thiago Alexandre Salgueiro Pardo (USP)

O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa de Pós-Graduação em Linguística.

AGRADECIMENTOS

Aos meus pais, Joelma e Isaac (o original), que fizeram o impossível para criar eu e meus irmãos. Agradeço imensamente todo o apoio e o amor incondicional que me ofertaram não apenas ao longo da pesquisa, mas da minha vida.

Aos meus irmãos, Rebeca, João e Emanuele, que, mesmo em períodos conturbados, sempre estiveram ao meu lado como suporte e apoio. Vocês são a chama que mantém meu coração aquecido.

À carérrima Nathália (com *h*), que, em pouco tempo, se tornou uma amiga indispensável. Obrigado pela companhia – mesmo que virtual –, suporte e ajuda incondicional.

À minha maior companheira de trabalho, Marcella (com dois *l*'s), que me auxiliou e aconselhou durante todo o período da escrita. Obrigado pelos debates, reflexões e discussões.

Às minhas amigas de pesquisa Gabriela, Francimeire e Roana. Nossos encontros do grupo de pesquisa tornaram os dois longos anos de pandemia menos angustiantes e mais produtivos. Obrigado por compartilharem as dificuldades, mazelas e as alegrias da vida acadêmica.

À querida Núbia, que aguentou dois longos anos de pandemia entre meus surtos de ansiedade e debates linguísticos enfadonhos. Muito obrigado por compartilhar, e tornar especial, cada dia dessa minha jornada.

Aos meus amigos Jonatas, Bruno e Renan, que, mesmo com uma índole duvidosa, foram parte essencial na manutenção da minha saúde mental.

Aos Professores Doutores Adriana Silvana Pagano e Thiago Alexandre Salgueiro Pardo que, como parte da banca, contribuíram com apontamentos importantíssimos para a finalização deste texto.

A meu orientador e amigo, Prof. Dr. Oto vale, que recepcionou um curioso de uma área distinta, me fornecendo apoio e orientação durante toda a pesquisa. Agradeço por todo conhecimento e suporte que recebi.

Por fim, à CAPES, pôr me oferecer subsídio durante os dois anos (e um pouquinho) do meu mestrado. O suporte à pesquisa é essencial para o bem-estar de cada pesquisado e para excelência das pesquisas.

Este Trabalho foi executado no Centro de Inteligência Artificial (C4AI-USP) com apoio da Fundação de Apoio à Pesquisa do Estado de São Paulo (Processo FAPESP 2019/07665-4) e da IBM Corporation.

O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 88887.499210/2020-00.

RESUMO

A negação é uma condição *sine qua non* das línguas naturais. Não há uma língua que não apresente algum meio de expressá-la dentro de seu sistema (HORN e WANSING, 2020). Mesmo sendo extremamente produtiva e indiscutivelmente indispensável para as línguas naturais, os estudos de processamento de língua natural que envolvem a negação são recentes (JIMÉNEZ-ZAFRA, *et al.*, 2020, p.190), o que torna o material sobre esse assunto pouco abundante. Em vista disso, este trabalho apresenta uma análise dos indicadores negativos (*não, nunca, jamais, sequer, tampouco, nada, ninguém, nenhum, sem, nem, exclusive, excluso, exceto, fora, afora, salvo, menos, senão e tirante*) do português em função da *Universal Dependencies* (UNIVERSAL DEPENDENCIES, 2021) a fim de melhorar o processamento automático da negação quando acionada por esses elementos, como também contribuir para a descrição linguística deles. A análise consistiu da revisão de literatura acerca de cada um dos indicadores e a busca dos mesmos indicadores nos *corpora* UD Portuguese Bosque (RADEMAKER *et al.*, 2017) e Porttinari-base (LOPES *et al.*, 2022a) como também no léxico/ferramenta lexical *PortiLexicon-UD*. Com o resultado das buscas, analisaram-se os atributos morfológicos e algumas relações sintáticas de cada um dos elementos, propondo possíveis correções e otimizações na anotação atual de cada *corpus* como propostas de melhoria para a ferramenta lexical. A revisão dos elementos negativos é de extrema importância para trabalhos futuros que pretendam realizar análises linguísticas automáticas que tenham a detecção da negação, tanto do escopo quanto do foco, como parte de sua tarefa.

Palavras-chave: Negação; Morfologia; Sintaxe; Universal Dependencies.

ABSTRACT

Negation is a sine qua non of natural languages, there is no language that does not present some way to express it in its system (HORN and WANSING, 2020). Despite being very productive and indispensable for natural languages, studies in natural language processing that involve negation is extremely recent (JIMÉNEZ-ZAFRA et al., 2020), which makes the material on this subject scarce. Therefore, this work presents an analysis of the Portuguese negative clues (não, nunca, jamais, sequer, tampouco, nada, ninguém, nenhum, sem, nem, exclusive, exclusivo, exceto, fora, afora, salvo, menos, senão e tirante) using the Universal Dependencies (UNIVERSAL DEPENDENCIES, 2021) methodology to increase the amount of information available of each element. The analysis consisted of a literature review of each clue and a search for them in the corpora UD Portuguese Bosque (RADEMAKER et al., 2017) and Porttinari-base (LOPES et al., 2022a) and the lexicon/lexical tool PortiLexicon-UD (LOPES et al., 2022b) as well. With the results, the morphological features and some syntactic relationships of the clues were analyzed, proposing possible corrections and optimizations in the current annotation of each corpus as proposals to improve the lexical tool. The review of negative clues is extremely important for future works that intend to perform automatic linguistic analyzes that have the detection of negation, both of scope and focus, as part of their task.

Keywords: Negation; Morphology; Syntax; Universal Dependencies.

LISTA DE FIGURAS

Figura 1 – Print do resultado de pesquisa no Google com a sequência ‘sem keanu reeves’ como palavra-chave	21
Figura 2 – Print dos primeiros quatro resultados da pesquisa no Google com a frase ‘filmes sem o keanu reeves’	21
Figura 3– Exemplo de anotação segundo as diretrizes da Universal Dependencies para a sentença: ‘Mulher morre em rio presa ao cinto do carro’	26
Figura 4– Exemplo de distintas anotações para a sentença: ‘um gato persegue ratos e camundongos’ em sueco, dinamarquês e inglês	39
Figura 5– Anotação sintática, simplificada, conforme as diretrizes da UD para a sentença “um gato persegue ratos e camundongos” em sueco, dinamarquês e inglês	40
Figura 6 – Comparação da anotação da sentença ‘o cachorro vai perseguir o gato fora do quarto’ em inglês, português e finlandês segundo as diretrizes da UD	44
Figura 7 – Anotação completa, segundo as diretrizes da UD, da sentença ‘um gato persegue ratos e camundongos’	46

LISTA DE QUADROS

Quadro 1 – Exemplo de extração dos indicadores negativos	29
Quadro 2 – Classes de PoS disponíveis na metodologia da UD	41
Quadro 3 – Características morfológicas relevantes	42
Quadro 4 – Relações de dependência da UD	45
Quadro 5 – Distribuição dos tipos de negação	61
Quadro 6 – Comportamento dos Indicadores Negativos.....	74
Quadro 7 – Ocorrências de JAMAIS.....	77
Quadro 8 – Ocorrências de NUNCA.....	77
Quadro 9 – Ocorrências de SEQUER.....	78
Quadro 10 – Ocorrências de TAMPOUCO.....	80
Quadro 11 – Ocorrências de NADA.....	82
Quadro 12 – Ocorrências de NINGUÉM	85
Quadro 13 – Ocorrências de NENHUM.....	88
Quadro 14 – Ocorrências de SEM.....	90
Quadro 15 – Ocorrências de NEM	92
Quadro 16 – Ocorrências de NÃO	94
Quadro 17 – Ocorrências de EXCLUSIVE.....	96
Quadro 18 – Ocorrências de EXCLUSO.....	97
Quadro 19 – Ocorrências de EXCETO	98
Quadro 20 – Ocorrências de FORA	98
Quadro 21 – Ocorrências de AFORA.....	99
Quadro 22 – Ocorrências de SALVO.....	100
Quadro 23 – Ocorrências de MENOS	101
Quadro 24 – Ocorrências de SENÃO.....	102
Quadro 25 – Ocorrências de TIRANTE.....	103
Quadro 26 – Propostas de anotação, na metodologia da UD para NUNCA	104
Quadro 27 – Propostas de anotação, na metodologia da UD para JAMAIS	104
Quadro 28 – Propostas de anotação, na metodologia da UD para SEQUER.....	105
Quadro 29 – Propostas de anotação, na metodologia da UD para TAMPOUCO.....	106
Quadro 30 – Propostas de anotação, na metodologia da UD para NADA.....	107
Quadro 31 – Propostas de anotação, na metodologia da UD para NINGUÉM.....	108
Quadro 32 – Propostas de anotação, na metodologia da UD para NENHUM.....	109

Quadro 33 – Propostas de anotação, na metodologia da UD para SEM	110
Quadro 34 – Propostas de anotação, na metodologia da UD para NEM.....	111
Quadro 35 – Propostas de anotação, na metodologia da UD para NÃO.....	112
Quadro 36 – Propostas de anotação, na metodologia da UD para a preposição acidental EXCLUSIVE	113
Quadro 37 Propostas de anotação, na metodologia da UD para a preposição acidental EXLCUSO.....	113
Quadro 38 – Propostas de anotação, na metodologia da UD para a preposição acidental EXCETO	114
Quadro 39 – Propostas de anotação, na metodologia da UD para a preposição acidental FORA	114
Quadro 40 – Propostas de anotação, na metodologia da UD para a preposição acidental AFORA.....	115
Quadro 41 – Propostas de anotação, na metodologia da UD para a preposição acidental SALVO	115
Quadro 42 – Propostas de anotação, na metodologia da UD para a preposição acidental MENOS	116
Quadro 43 – Propostas de anotação, na metodologia da UD para a preposição acidental SENÃO.....	116
Quadro 44 – Propostas de anotação, na metodologia da UD para a preposição acidental TIRANTE.....	117
Quadro 45 – PoS e Atributo(s) para as lexias investigadas	118

LISTA DE TABELAS

Tabela 1– Dados sobre os corpora/ferramenta	28
---	----

SÍMBOLOS E CONVENÇÕES

- * indica sequência inaceitável
- ? indica sequência com aceitabilidade duvidosa
- implicação lógica
- ↪ implicatura conversacional
- ¬ negação lógica
- PoS partes do discurso (Part of Speech)

Adota-se aqui como convenção as etiquetas correntemente utilizadas nos trabalhos de Universal Dependencies, a saber:

Partes do Discurso

Classe Aberta		Classe Fechada		Outras	
ADJ	Adjetivo	ADP	Preposições e posposições	PUNCT	Pontuação
ADV	Advérbio	AUX	Auxiliar	SYM	símbolo
INTJ	Interjeição	CCONJ	Conjunção coordenativa	X	PoS indefinida
NOUN	Nome	DET	Determinante		
PROPN	Nome próprio	NUM	Numeral		
VERB	Verbo	PART	Partícula		
		PRON	Pronome		
		SCONJ	Conjunção subordinativa		

Características morfológicas relevantes

Características		Valores	
Definite	Definitude	Def	Definido
		Ind	Indefindo
Gender	Gênero	Masc	Masculino
		Fem	Feminino
		Neut	Neutro
Number	Número	Sing	Singular
		Plur	Plural
Polarity	Polaridade	Neg	Negativa
PronType	Tipo de Pronome	Neg	Negativo
		Ind	Indefinido

Relações de dependência relevantes:

Relação de Dependência	
acl	Oração Adnominal
advmod	Modificador Adverbial
case	Caso
cc	Conjunção Coordenativa
conj	Conjunção
det	Determinante
mark	Marcador
nmod	Modificador Nominal
nsubj	Sujeito Nominal
obj	Objeto
nsubj:pass	Sujeito Nominal da Passiva
obl	Oblíquo

No que se refere aos exemplos, quando não são indicados como retomados de autores citados, serão anotados com as seguintes indicações de ocorrência:

FK ocorrência coletada no corpus Folha Kaggle – News Of Brazilian Newspaper

PTTB ocorrência coletada na Porrtinari base

UPB ocorrência coletada no UD Portuguese Bosque

Os exemplos que não tiverem indicação de origem são frases construídas por introspecção ou modificadas a partir das ocorrências.

SUMÁRIO

1.	INTRODUÇÃO E JUSTIFICATIVA	17
2.	MÉTODOS E MATERIAIS	25
2.1	OS CORPORA.....	25
2.2	A ANÁLISE.....	29
3	DEPENDÊNCIAS	31
3.1	O QUE É DEPENDÊNCIA	31
3.2	GRAMÁTICAS DE DEPENDÊNCIA.....	34
3.3	UNIVERSAL DEPENDENCIES	39
4	NEGAÇÃO NO PORTUGUÊS BRASILEIRO	50
4.1	OS TIPOS DE NEGAÇÃO	52
4.2	NEGAÇÃO SINTÁTICA.....	53
4.2.1	Negação Oracional	53
4.2.2	Negação Não-Oracional (Negação de Constituintes).....	57
4.3	NEGAÇÃO LEXICAL.....	58
4.4	CONCORDÂNCIA NEGATIVA.....	62
4.5	LEXIAS QUE REALIZAM A NEGAÇÃO SINTÁTICA.....	64
4.4.1	Nunca e Jamais	64
4.4.2	Sequer	65
4.4.3	Tampouco	66
4.4.4	Nada, Ninguém e Nenhum	68
4.4.5	Sem	70
4.4.6	Nem	71
4.5	PREPOSIÇÕES ACIDENTAIS	72
5.	ANÁLISES E RESULTADOS	76
5.1	NUNCA E JAMAIS	76

5.2	SEQUER.....	78
5.3	TAMPOUCO.....	79
5.4	NADA, NINGUÉM E NENHUM.....	80
5.4.1	Nada.....	81
5.4.2	Ninguém.....	84
5.4.3	Nenhum.....	87
5.5	SEM.....	89
5.6	NEM.....	91
5.7	NÃO.....	93
5.8	PREPOSIÇÕES ACIDENTAIS.....	96
5.8.1	Exclusive.....	96
5.8.2	Excluso.....	97
5.8.3	Exceto.....	97
5.8.4	Fora.....	98
5.8.5	Afora.....	99
5.8.6	Salvo.....	99
5.8.7	Menos.....	100
5.8.10	Senão.....	102
5.8.11	Tirante.....	103
5.9	SÍNTESE DOS RESULTADOS.....	104
5.9.1	Nunca e Jamais.....	104
5.9.2	Sequer.....	105
5.9.3	Tampouco.....	106
5.9.4	Nada, Ninguém e Nenhum.....	107
5.9.5	Sem.....	110
5.9.6	Nem.....	111
5.9.7	Não.....	112

5.9.8	Preposições acidentais	113
6.	CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS.....	119
	REFERÊNCIAS	122

1. INTRODUÇÃO E JUSTIFICATIVA

A capacidade de opor dois enunciados com relação ao seu valor de verdade (verdadeiro ou falso) é uma das principais distinções entre a faculdade de linguagem humana e os sistemas de comunicação dos animais (HORN, 2001, p.XIII). Ainda que existam estudos que comprovem que símios e algumas aves são capazes de serem treinados para compreenderem conceitos como *rejeição*, *recusa* e a *não existência* de algo, eles ainda apresentam dificuldade em, ou não são capazes de, gerar a oposição entre dois enunciados (HEINE e KUTEVA, 2007, p.141-143), fazendo com que a negação seja uma característica inerente à atividade de fala humana (HORN, 2001, p.XIII).

Por mais simples e intuitivo que pareça, a marcação de oposição entre duas sentenças gera diversas implicações, tanto lógicas, como problemas relacionados à implicação existencial e ambiguidade de escopo, quanto intencionais, envolvendo precisão e especificidade. O que torna o simples “ato de negar” um fenômeno de grande relevância para os estudos da filosofia e da linguagem. Ele é debatido desde a Antiguidade, por exemplo, nas Categorias de Aristóteles, até em textos mais modernos (FREGE, 1982; 2011; RUSSELL, 1905; WITTIGENSTEIN, 1922; GRICE 1989; entre muitos outros). Embora essa diversidade de problemas seja evidente, nota-se que a abordagem desse fenômeno ainda é pouco produtiva dentro do processamento de língua natural.

O Processamento de Língua Natural (PLN) é um campo multidisciplinar da Ciência da Computação, Inteligência Artificial e Linguística que estuda como computadores podem ser utilizados para compreender e manipular a língua humana em textos ou em fala (LI *et al.*, 2021, p.1). Considerando que a negação é uma condição *sine qua non* das línguas naturais, sua análise deveria ser incorporada na maior parte das análises e descrições para PLN. Contudo, esse não é o caso (JIMÉNEZ-ZAFRA, *et al.*, 2020, p,190).

Segundo Jiménez-Zafra (2020), os trabalhos de PLN voltados para o processamento da negação surgiram em meados de 2001 com foco na identificação do escopo negativo em sentenças em *corpora* de saúde clínica (CHAPMAN *et al.*, 2001; MUTALIK *et al.*, 2001; GOLDIN e CHAPMAN, 2003) e ganharam espaço na comunidade científica com o tempo. Mais recentemente, com o aumento de relevância da Análise de Sentimento e Mineração de Opinião, os trabalhos que envolvem o processamento da negação em PLN estão relacionados com a detecção, e reversão, da polaridade das sentenças (TABOADA *et al.*, 2011; JIMENEZ-ZAFRA *et al.*, 2017; MUKHERJEE *et al.*, 2021).

Existem quatro atividades normalmente realizadas em função da negação:

(i) detecção de indicadores negativos, a fim de encontrar as palavras que expressam negação; (ii) identificação do escopo, para descobrir quais partes da frase são afetadas pelos elementos negativos; (iii) reconhecimento de eventos negados, para determinar quais eventos são afetados pelos elementos negativos; e (iv) detecção de foco, para encontrar a parte do escopo que é mais proeminentemente negada.¹ (JIMÉNEZ-ZAFRA, *et al.*, 2020, p.190. Tradução do autor)

Segundo aquele trabalho, a maior parte dos esforços estão concentrados nas atividades (i) e (ii). Isso ocorre porque as tarefas de detecção de “escopo, evento e foco são mais complexas por dependerem da identificação dos indicadores negativos²” (JIMÉNEZ-ZAFRA, *et al.*, 2020, p.190. Tradução do autor).

As atividades (i) e (ii) são, em princípio, as bases para a detecção da negação. Elas consistem em identificar quais elementos funcionam como indicadores negativos, os elementos que acionam a negação, e quais elementos são modificados pelo indicador, o escopo:

1. *Você **não** consegue ter paz dentro de casa. (FK)*
2. ***Nem** todo mundo gostará que você assuma uma posição. (FK)*
3. *O corpo político parece uma geleia, **sem** ossos. (FK)*

Nos exemplos, os indicadores negativos estão destacados em negrito e o escopo está em itálico. Como é possível observar, existem casos em que o indicador irá atuar em toda a sentença, como em (1), mas também é possível que ele atue apenas sobre uma quantidade limitada de elementos, como em (2) e (3).

Contudo, existem alguns problemas com esse tipo de demarcação. Um deles é que uma simples listagem dos indicadores negativos não é suficiente para demarcar o seu funcionamento, uma vez que nem sempre eles funcionam com polaridade negativa (por exemplo, construções comparativas com *que nem* e marcadores discursivos como o *não é/né*):

4. *Cheio de memórias, cheio de componentes, **que nem** o carro. (FK)*
5. *Eu mereço pelo menos 10%, **né?** (FK)*

¹ “(i) negation cue detection, in order to find the words that express negation; (ii) scope identification, in order to find which parts of the sentence are affected by the negation cues; (iii) negated event recognition, to determine which events are affected by the negation cues; and (iv) focus detection, in order to find the part of the scope that is most prominently negated”, no original.

² “scope, event, and focus identification tasks are more complex because they depend on negation cue detection”, no original.

Isso faz com que a atividade de detecção da negação, apenas como “destaque” de operadores negativos, não seja suficiente. É necessária uma análise das diferentes formas que cada indicador pode ocorrer.

Uma outra dificuldade enfrentada pelo processamento da negação em trabalhos de PLN é a quantidade de sentenças com indicadores negativos dentro dos *corpora* anotados. Jiménez-Zafra (2020, p.191) apontam que para os *corpora* em inglês BioInfer (PYYSALO, *et al.*, 2007), Genia Event—a (KIM *et al.*, 2008), BioScope (VINCZE *et al.*, 2008), Product Review (COUNCILL *et al.*, 2010), PropBank FOCUS (BLANCO e MOLDOVAN, 2011), ConanDoyle-neg (MORANTE e DAELEMANS, 2012), SFU Review_{EN} (KONSTANTINOVA *et al.*, 2012), NEG-DrugDDI (BOKHARAEIA *et al.*, 2014), Deep Tutor Negation (BANJADE e RUS, 2016) e SOCC (KOLHATKAR *et al.*, 2019), a quantidade média de sentenças com negação varia entre 9,37% até 32,16% (valor atingido pelos *corpora* específicos para a negação). Dentro dos *corpora* utilizados nesta dissertação têm-se aproximadamente 16.36% de sentenças com elementos negativos no *corpus* UD Portuguese Bosque (RADEMAKER *et al.*, 2017), 18.59% para o Portinari-base (LOPES *et al.*, 2022a) e 18.86% para o *corpus* Folha Kaggle – News of Brazilian Newspaper (SANTANA, 2019).

Entretanto, mesmo com poucos trabalhos tratando sobre o seu processamento, a anotação e a detecção da negação é de grande relevância para processos como Recuperação de informação (LIDDY *et al.*, 2000), extração de informação (SAVOVA *et al.*, 2010), tradução automática (BAKER *et al.*, 2015) e Análise de Sentimento (LIU, 2020; MUKHERJEE *et al.*, 2021).

Como exemplo para sistemas de recuperação de informação, imagine duas entradas distintas em um sistema, a interface de uma biblioteca virtual:

6. Receitas de bolo chocolate.
7. Receitas de bolo **sem** chocolate.

Cada uma das solicitações deve retornar diferentes documentos para o usuário. Caso a negação não seja processada, é possível que o sistema apenas retorne para as duas entradas a mesma lista de livros/documentos contendo as mesmas receitas.

Para os sistemas de extração de informação, tomemos como exemplo um *corpus* clínico. A retirada dos sintomas apresentados, ou não, por um paciente é de extrema relevância para a extração de informação:

8. O paciente **não** apresenta febre.

9. O paciente apresenta febre de 38°C.

A oposição entre (8) e (9) pode ser uma distinção de vida ou morte em um quadro clínico.

A tradução automática de sentenças negativas entre línguas também é uma tarefa complicada. O *inglês* e o *português* são ótimos exemplos para a demonstração desse tipo de problema:

10. **Não** veio **ninguém** para escola

11. **No** one came to school

Nos exemplos (10) e (11) existe uma diferença na quantidade de elementos negativos presentes. A concordância negativa, com dois elementos, é a forma prototípica de realização da negação no português, enquanto a Dupla Negação (* *No one didn't come to school*) é inaceitável para o inglês.

Com relação a Análise de Sentimento, caso a negação não seja levada em consideração para o processamento, é possível que o sistema faça uma interpretação com sentido oposto ao pretendido pelo usuário:

12. O produto apresenta o resultado esperado

13. O produto **não** apresenta o resultado esperado

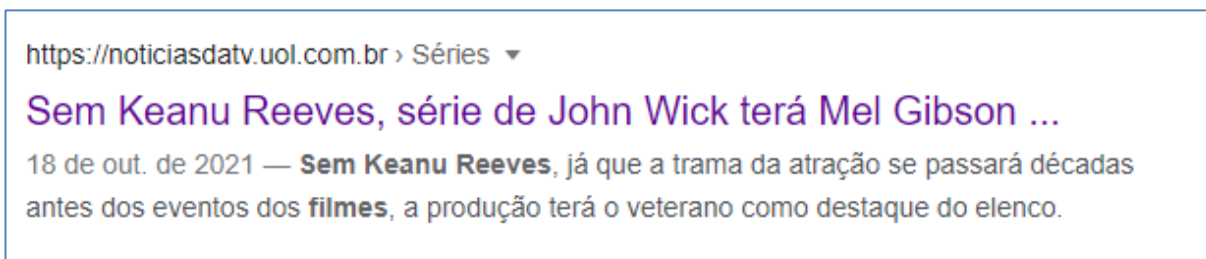
Sentenças como (13) podem ser interpretadas com o significado completamente oposto a interpretação presente em (12), pelo simples fato da negação não ter sido processada.

Os problemas com o processamento da negação podem ser visualizados em sistemas que as pessoas utilizam no seu dia a dia.

14. filmes **sem** o keanu reeves

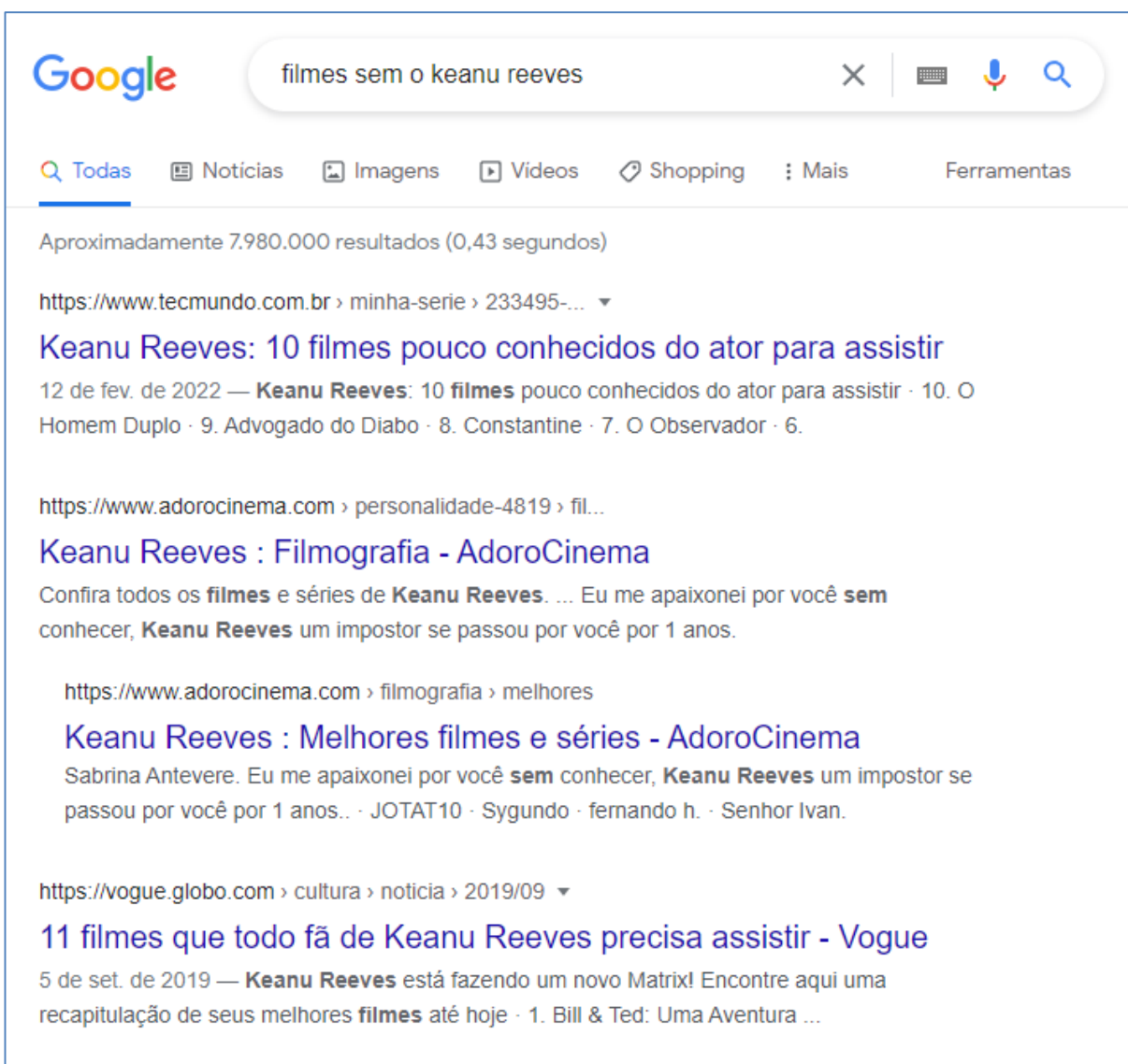
Caso seja buscado no Google, o exemplo (14) faz com que a ferramenta retorne em sua primeira página dez resultados em que apenas um deles têm a sequência *sem keanu reeves* como chave de busca, todos os outros resultados retornam listas de filmes e históricos acerca do Keanu Reeves, como observável nas figuras 1 e 2.

Figura 1 – Print do resultado de pesquisa no Google com a sequência ‘sem keanu reeves’ como palavra-chave



Fonte – O autor. Pesquisa realizada em 16/03/2022

Figura 2 – Print dos primeiros quatro resultados da pesquisa no Google com a frase ‘filmes sem o keanu reeves’



Fonte – O autor. Pesquisa realizada em 16/03/2022

Apesar de toda a relevância que o fenômeno da negação apresenta, há, até o presente momento, apenas um *corpus* anotado especificamente em função da negação³ para o português, presente no trabalho de Dalloux, *et al.* (2020). Os outros trabalhos que se relacionam com o assunto, ou preveem uma futura anotação (SCHNEIDER, *et al.*, 2020) ou discutem acerca da dificuldade enfrentada em realizar uma determinada atividade motivada pela negação (AVILA e MELO, 2013; MENDES, *et al.*, 2016).

Um outro problema encontrado no tratamento das negações são expressões multipalavras que veiculam a negação, mas que não contêm elementos negativos dentro de sua estrutura:

15. A Venezuela **está longe de ser** o único país latino-americano que sofre com a violência letal. (**FK**)
16. Os especialistas concordam que o Brasil **ainda precisa caminhar muito** para atingir os níveis de conformidade praticados nos Estados Unidos e na Europa. (**FK**)

Esses tipos específicos de construção não são facilmente detectados, uma vez que elas são dependentes de mais de um item lexical e a negação expressa por elas está vinculada à interpretação da expressão como um todo e não de suas partes individuais. Embora sejam uma parte desafiadora no trabalho de identificação da negação, essas expressões não serão tratadas no presente trabalho. Acredita-se que o fenômeno é bastante amplo e que será necessário um trabalho específico para tratar essas expressões.

Em vista do levantado, este trabalho busca realizar uma investigação na anotação dos *corpora* UD Portuguese Bosque (RADEMAKER *et al.*, 2017) e Portinari-Base (LOPES *et al.*, 2022a), como também na ferramenta lexical PortiLexicon-UD (LOPES *et al.*, 2022b). O intuito da investigação é analisar a anotação dos indicadores negativos que expressam a negação no português, sendo elas: *não, nunca, jamais, sequer, tampouco, nada, ninguém, nenhum, sem, nem*; e também das preposições acidentais⁴ de exclusão: *exclusive, excluso, exceto, fora, afora, salvo, menos, senão e tirante*. A finalidade da investigação é permitir que a anotação final do

³ No momento da escrita deste texto, não foi encontrado nenhum texto sobre o assunto nas bases Google Scholar (<https://scholar.google.com/>), SciELO (<https://www.scielo.br/>), Portal de Periódico CAPES (<https://www-periodicos-capes-gov-br.ez1.periodicos.capes.gov.br/>) e ACL Anthology (<https://aclanthology.org/>)

⁴ Utiliza-se aqui o termo *preposições acidentais* para nomear os “elementos que se estão gramaticalizando como **preposições** e que se empregam em contextos restritos” (NEVES, 2000, p.732, destaque original), ou seja, elementos que não são restritos a ocorrerem apenas como preposições, mas que, em casos específicos, exprimem comportamento prepositivo.

corpus Porttinari (PorTuguese Treebank) (PARDO *et al.*, 2021)⁵ contenha a identificação, como atributo morfológico, dos elementos gramaticais do português capazes de realizar a negação.

O *corpus* Porttinari (PorTuguese Treebank) é parte do POeTiSA⁶ (*Portuguese processing - Towards Syntactic Analysis and parsing*), “um projeto de longo prazo que visa o crescimento de recursos e o desenvolvimento de ferramentas e aplicativos baseados em sintaxe para a língua portuguesa brasileira, buscando alcançar o estado-da-arte mundial para esta área”⁷ (LOPES *et al.*, 2022a, p.2. Tradução do autor). O Porttinari (PARDO *et al.*, 2021) será um *corpus* multigênero anotado segundo as diretrizes do projeto *Universal Dependencies*, doravante UD, (Cf. seção 3.3) consistindo na união dos *corpora* Folha Kaggle (SANTANA, 2019), MAC-MORPHO (ALUÍSIO *et al.*, 2003), B2W-reviews01 (REAL *et al.*, 2019), DANTE (Dependency-ANalised corpora of TwEets) (SILVA *et al.*, 2020) e Book Reviews (BELISÁRIO *et al.*, 2020).

As duas questões que regem esta pesquisa são: (i) as anotações de *corpus* do português são satisfatórias em relação à negação? (ii) é possível realizar uma anotação da negação uniforme para o português? Nossa hipótese é que existe um conjunto de elementos que fazem parte de diferentes categorias gramaticais, capazes de indicar, morfológica e sintaticamente, a negação em uma sentença e que é possível efetuar uma anotação explícita desses elementos.

O objetivo deste trabalho é aumentar a quantidade de informação sobre a negação anotada em *corpus* dentro do português, em vista da pouca quantidade existente. Esta pesquisa também almeja não só facilitar, mas também contribuir com o possível desenvolvimento de ferramentas para a detecção de escopo e foco da negação, melhorando, assim, o desempenho de algoritmos de recuperação de informação, extração de informação, análise de sentimento/opinião e tradução automática para o português brasileiro.

Para isso, este trabalho se organiza da seguinte forma: apresentam-se os métodos e os materiais utilizados para a realização da pesquisa no capítulo 2, com a demonstração da seleção dos *corpora* e a justificativa pelo feito. No capítulo 3, é realizada uma revisão bibliográfica acerca das gramáticas de dependência e da *Universal Dependencies* (capítulo 3). O quarto capítulo promove uma revisão bibliográfica acerca das possíveis formas de realizar a negação dentro do português, dos acionadores negativos capazes de acionar/indicar a negação e das

⁵ <https://sites.google.com/icmc.usp.br/poetisa#h.arzw7f2o3g4d>

⁶ <https://sites.google.com/icmc.usp.br/poetisa>

⁷ “is a long term project that aims at growing syntax-based resources and developing related tools and applications for Brazilian Portuguese language, looking to achieve world state-of-the-art results in this area”, no original.

preposições acidentais que também o fazem no português. No capítulo 5, são apresentados os resultados juntamente com as investigações dentro dos *corpora* e da ferramenta lexical juntamente com as propostas de anotação realizadas pelo autor. Por fim, no capítulo 6 são apresentadas as considerações finais e as perspectivas para trabalhos futuros.

2. MÉTODOS E MATERIAIS

Neste capítulo estão organizadas as etapas referentes à metodologia e aos materiais utilizados para a realização dos objetivos desta dissertação. Assim, esta pesquisa foi recortada em três grandes etapas:

Etapa I – Seleção dos *corpora* de pesquisa: consistiu da seleção do *corpus* mais adequado para a realização da revisão e análise. Dessa forma, os exemplos da revisão bibliográfica (capítulo 4) são oriundos do *corpus* Folha kaggle (SANTANA, 2019), enquanto as análises das lexias são realizadas no *corpus* UD Portuguese Bosque (RADEMAKER *et al.*, 2017), no *corpus* Portinari-base (LOPES *et al.*, 2022a) e no léxico PortiLexicon-UD (LOPES *et al.*, 2022b).

Etapa II – Extração dos dados: para a extração dos exemplos para o texto e dos elementos analisados nos *corpora*, utilizou-se a ferramenta Unitex⁸, para a extração de padrões de concordância dentro dos *corpora*, assim como scripts em Python⁹.

Etapa III – Análise dos dados: com base nos dados extraídos na etapa anterior, foi realizada uma análise linguística dos resultados (apresentada no capítulo 5) para que fosse possível realizar propostas de solução para os problemas elencados durante a análise e propor possíveis anotações para as lexias investigadas.

2.1 OS CORPORA

Os três *corpora* aqui apresentados são da área jornalística. O recorte desse tipo de texto é dado em função da estrutura desse gênero textual. Textos jornalísticos tendem a ser escritos de acordo com a norma culta da língua, seguindo manuais específicos de redação, por exemplo o *Manual da Redação da Folha* (MANUAL, 2021). Outra característica importante desse gênero textual é que ele não é completamente narrativo, podendo ser alternado com discursos direto.

Um último reforço para a seleção dos *corpora* é a quantidade de sentenças com indicadores negativos presentes neles. Enquanto os três *corpora* aqui analisados têm em média 17.9% de sentenças com indicadores negativos – 16.36% para o UD Portuguese Bosque (RADEMAKER *et al.*, 2017); 18.59% para o Portinari-base (LOPES *et al.*, 2022a); e 18.86% para o *corpus* Folha kaggle (SANTANA, 2019) –, o *corpus* PetroGold (SOUZA *et al.*, 2021),

⁸ <https://unitexgramlab.org/pt/language-resources>

⁹ Tanto os scripts quanto os dados retirados dos *corpora* podem ser consultados em: <https://github.com/IscMr/Masterthesisdata>

um *corpus* de tamanho expressivo e também anotado em função da metodologia da UD, tem apenas 7.7% das suas sentenças com indicadores negativos (689 das 8949 sentenças) e apresenta apenas 8 dos 20 indicadores aqui analisados (*não, nunca, nem, nenhum, sem, exceto, fora e menos*), o que fez com que ele fosse descartado como objeto deste estudo.

Dos três *corpora* selecionados para este trabalho, apenas um não está anotado, o Folha kaggle (SANTANA, 2019). Ele é uma coletânea virtual de textos jornalísticos extraídos do jornal Folha de S. Paulo, disponível no site Kaggle¹⁰. Sua seleção se deu pelo fato de ser um *corpus* atual e com uma abrangência relevante: aproximadamente 160 mil notícias, 4 milhões de sentenças e 169 milhões de *tokens*. Devido à sua abrangência, ele foi utilizado para validar as informações presentes da revisão bibliográfica de cada uma das lexias gramaticais e para a obtenção dos exemplos presentes no capítulo 4. O *corpus* foi processado pelo UNITEX e foram extraídos os exemplos pela ferramenta de concordância do mesmo software.

Os outros dois *corpora* são textos processados. Ao contrário do Folha kaggle (SANTANA, 2019) que foi utilizado para a validação das ocorrências de cada lexia gramatical na revisão bibliográfica, eles são o objeto direto da investigação de anotação. Ambos foram selecionados por estarem anotados em função das diretrizes da UD, figura 3.

Figura 3– Exemplo de anotação segundo as diretrizes da Universal Dependencies para a sentença: ‘Mulher morre em rio presa ao cinto do carro’

# newdoc_id = CF877									
# text = Mulher morre em rio presa ao cinto do carro									
# sent_id = CF877-1									
# source = CETENFolha n=877 cad=Cotidiano sec=soc sem=94a									
1 Mulher	mulher	NOUN	_	Gender=Fem Number=Sing	2	nsubj	_	_	_
2 morre	morrer	VERB	_	Mood=Ind Number=Sing Person=3 Tense=Pres VerbForm=Fin	0	root	_	_	_
3 em	em	ADP	_	_	4	case	_	_	_
4 rio	rio	NOUN	_	Gender=Masc Number=Sing	2	obl	_	_	_
5 presa	prender	VERB	_	Gender=Fem Number=Sing VerbForm=Part	2	advcl	_	_	_
6-7 ao	_	_	_	_	_	_	_	_	_
6 a	a	ADP	_	_	8	case	_	_	_
7 o	o	DET	_	Definite=Def Gender=Masc Number=Sing PronType=Art	8	det	_	_	_
8 cinto	cinto	NOUN	_	Gender=Masc Number=Sing	5	obj	_	_	_
9-10 do	_	_	_	_	_	_	_	_	_
9 de	de	ADP	_	_	11	case	_	_	_
10 o	o	DET	_	Definite=Def Gender=Masc Number=Sing PronType=Art	11	det	_	_	_
11 carro	carro	NOUN	_	Gender=Masc Number=Sing	8	nmod	_	_	_

Fonte – UD Portuguese Bosque (RADEMAKER *et al.*, 2017)

¹⁰ <https://www.kaggle.com/marlesson/news-of-the-site-folhauol>

Na figura, é possível observar um exemplo de anotação em CoNLL-U, o formato adotado pela UD. Da esquerda para direita, as colunas representam: (i) índice do token, (ii) forma de ocorrência, (iii) lema, (iv) UPoS (parte do discurso universais), (v) XPoS (partes do discurso específicas da língua), (vi) atributo morfológico, (vii) HEAD (demarcação de dominância/dependência sintática), (viii) relação de dependência, (ix) relação de dependência melhorada e (x) outras anotações.

O Porttinari-base (LOPES *et al.*, 2022a) é um *corpus* pré-annotado nas diretrizes da UD (ver seção 3.3), contando apenas com a anotação de PoS (*Parts of Speech*) e dos atributos referentes à morfologia. A anotação foi dividida em duas etapas, uma manual e outra automática.

A anotação manual consistiu na revisão e na correção das etiquetas de PoS e dos atributos referentes à morfologia (LOPES *et al.*, 2022a, p.10-13). E a anotação automática consistiu na anotação do texto pelo *parser* UDPipe 2.0 (STRAKA, 2018) e uma revisão automática de padrões de erros comuns realizados pelo *parser* (LOPES *et al.*, 2022a, p.6-10). O Porttinari-base contém um total de 8.420 sentenças com um total de 168.399 *tokens*.

O Porttinari-base, assim como descrito por Lopes *et al.* (2022a, p.3-4), é uma seleção de sentenças do *corpus* Folha kaggle (SANTANA, 2019). Ainda que todas as sentenças presentes nele também estejam presentes no Folha kaggle (SANTANA, 2019), cada *corpus* tem uma função distinta nesta pesquisa. O Folha kaggle (SANTANA, 2019), como já descrito, é utilizado para a validação do comportamento dos indicadores negativos na revisão bibliográfica, enquanto o *Porttinari-base* é um dos *corpora* em que será feita a análise da anotação.

O *UD Portuguese Bosque* é uma adaptação do modelo PALAVRAS (BICK, 2014) para UD realizada por Bick (2016) no *corpus* Bosque. O *corpus* é parte do projeto Floresta Sintá(c)tica (AFONSO *et al.*, 2002) e consiste em uma coletânea de textos jornalísticos em português, brasileiro e português europeu, anotados sintaticamente pelo *parser* PALAVRAS (BICK, 2006) e revisado manualmente por linguistas. O UD portuguese Bosque tem um total de 227.825 *tokens* anotados em 9.364 sentenças.

A seleção do Bosque como *corpus* para pesquisa é devido ao fato de que ele foi utilizado como modelo de língua para o UDPipe (STRAKA, 2018) realizar a anotação automática do Porttinari-base (LOPES *et al.*, 2022a, p.7). Dessa forma, a anotação das lexias negativas presentes no *Porttinari-base* reflete a anotação do *corpus* *UD Portuguese Bosque*, tornando, assim, necessária sua revisão junto ao *Porttinari-base*. As questões referentes às anotações serão debatidas no capítulo 5.

Outro traço que fundamenta a necessidade de revisão do UD Portuguese Bosque é o fato de que ele é oriundo das transformações entre anotações distintas (AFONSO *et al.*, 2002; BICK, 2014; 2016). Mesmo que as conversões tenham baixíssimas perdas (BICK, 2016), ainda assim existem informações, como a marcação de polaridade, que existem na UD, mas não estavam presentes na anotação anterior.

Esta pesquisa também utilizará o léxico PortiLexicon-UD (LOPES *et al.*, 2022b) que é um recurso lexical com 858.922 palavras. Todas elas são compostas por uma 4-tupla que contém as seguintes informações: palavra (lexia), lema, classe de palavra (PoSTAG) e atributo morfológico (LOPES *et al.*, 2022b, p.6635). A ferramenta, assim como os *corpora* aqui analisados, segue as diretrizes da UD para a classificação das informações sobre as entradas.

O *PortiLexicon-UD* é originado da conversão dos recursos lexicais de outra ferramenta, UNITEX-PB (MUNIZ, 2004). Durante sua preparação, foram realizadas alterações e melhorias nas entradas, como também revisões humanas realizadas por linguistas; os dicionários do UNITEX-PB não contêm algumas informações que a metodologia da UD permite serem inseridas nas entradas, como por exemplo, as informações de polaridade negativa que são o foco desta pesquisa.

Em conclusão, as informações técnicas dos *corpora* e da ferramenta lexical estão dispostas na tabela 1:

Tabela 1– Dados sobre os corpora/ferramenta

Corpus	Sentenças	Sentenças ao menos um elemento negativo	Tokens
Folha kaggle – News of Brazilian Newspaper (SANTANA, 2019)	4.661.813	879.138	169.767.388
UD Portuguese Bosque (RADEMAKER <i>et al.</i> , 2017)	9.357	1.531	227.825
Porttinari-Base (LOPES <i>et al.</i> , 2022a)	8.420	1.563	168.399
Ferramenta	Palavras		
PortiLexicon-UD (LOPES <i>et al.</i> , 2022b)	858.922		

Fonte – O Autor

Assim, cada um dos *corpora* aqui descritos desempenhará funções distintas ao longo desta pesquisa.

2.2 A ANÁLISE

A análise deste trabalho foi realizada em duas etapas: uma automática e outra manual. A etapa automática consistiu na extração de todas as sentenças dentro dos *corpora* anotados (*UD Portuguese Bosque* e *Porttinari-Base*) que continham ao menos um dos 20 indicadores negativos investigados por esta pesquisa (*não, nunca, nem, jamais, sequer, tampouco, ninguém, nada, nenhum, sem, exclusive, excluso, exceto, fora, afora, salvo, menos, senão* e *tirante*).

Os *corpora* anotados seguem o formato CoNLL-U que é uma revisão e evolução do formato CONLL-X (BUCHHOLZ e MARSÍ, 2006). Os arquivos são organizados da seguinte maneira:

Cada arquivo pode conter múltiplas frases, separadas por uma linha em branco. Cada frase é iniciada por uma ou mais linhas contendo metadados, como identificador da frase e texto original. Em seguida, as palavras (tokens), são descritos um em cada linha. Também tokens multi-palavra (tokens ortográficos que foram quebrados em mais de uma palavra) recebem uma linha própria. Cada palavra ou tokens contém 10 campos separados por um único caractere de tabulação. São eles: numeração sequencial (ID); forma original da palavra no texto (FORM); lema, a forma canônica de uma palavra (LEMMA), classe gramatical universal (UPOSTAG), classe gramatical específica da linguagem (XPOSTAG), atributos morfológicos (FEATS), índice da palavra da qual depende (HEAD), a relação universal de dependência (DEPREL), relação de dependências enhanced (DEPS) e informações diversas (MISC). (MUNIZ *et al.*, 2017)

Assim, o arquivo resultante da extração automática foi um documento em formato *.txt* que contém para cada combinação possível das colunas LEMA, UPOSTAG, FEATS e DPREL a frequência desse padrão dentro do *corpus* e cada uma das sentenças que ocorreram com esse padrão, da seguinte forma para cada um dos indicadores negativos:

Quadro 1 – Exemplo de extração dos indicadores negativos

LEMA	UPOSTAG	FEATS	DPREL	ocorrências
nunca	ADV	_	advmod	62
sentenças				
Arnold Schwarzenegger deu entrevistas como ator principal, foi substituído por Bruce Willis há dois anos, mas o projeto nunca saiu do papel				
Alec Nunca usei um chapéu daqueles na vida!				
...				

Fonte – O Autor

Cada uma das lexias também foi buscada na interface de pesquisa do PortiLexicon-UD¹¹ (LOPES *et al.*, 2022b). Ainda que a ferramenta não tenha exemplos de ocorrência, ela traz as características de lema, classe de palavras (PoS) e atributos morfológico, as mesmas buscadas na extração automática dos *corpora*.

Após a pesquisa e extração automática, revisou-se cada uma das ocorrências e seu conjunto de sentenças, verificando possíveis inconsistências nas anotações. A característica analisada mais relevante foi a presença, ou ausência, dos atributos `Polarity=Neg` e `PronType=Neg` na coluna FEATS dos *corpora*, assim como na ferramenta. Por fim, os padrões encontrados foram comparados e categorizados, como é possível observar nos resultados expostos no capítulo 5.

¹¹ <https://portilexicon.icmc.usp.br/>

3 DEPENDÊNCIAS

Alice: Você poderia me dizer qual caminho eu devo tomar?

Gato: Bem, isso depende (...)

*Alice no País das Maravilhas –
Lewis Carroll*

O conceito de *Gramática de Dependências* segue de uma longa tradição linguística, podendo ser recuperada em textos das gramáticas de sânscrito de Panini até em trabalhos durante século XX, como Sgall, Hajičová e Panevová (1986), Hellwig (1986), Mel'čuk (1988), entre muitos outros. Como destaque, pode-se citar o célebre *Eléments de Syntaxe Structurale* de Lucien Tesnière, cuja publicação em 1959 é considerada como o marco inicial dos estudos contemporâneos da *Gramática de Dependências*.

Os estudos das gramáticas de dependência foram ofuscados pela principal teoria sintática de constituintes, a Gramática Gerativa (CHOMSKY, 1957). As motivações que levaram os trabalhos de Chomsky a se destacarem como teoria sintática estão para além da delimitação deste trabalho, todavia é inegável que a sintaxe de constituintes foi, durante muito tempo, a maneira “tradicional” de se abordar a sintaxe.

Recentemente a Gramática de Dependências ganhou grande destaque com o aumento exponencial de *parsers* que utilizam de sua metodologia para realizar análises linguísticas. Ela permite que uma ampla variedade de línguas, tipologicamente distintas, sejam processadas de uma maneira uniforme, confiável e facilmente replicável (KÜBLER *et al.*, 2009, p. XI).

O intuito deste capítulo é discorrer sobre o conceito de dependência nas línguas naturais e utilizar desse conceito para a discussão sobre a Universal Dependencies, uma metodologia de anotação de partes do discurso, características morfológicas e dependências sintáticas. Para isso, na primeira seção do capítulo será realizada uma revisão acerca do conceito de dependência, na segunda seção serão tratadas as gramáticas de dependência e na terceira seção serão abordadas as diretrizes da Universal Dependencies.

3.1 O QUE É DEPENDÊNCIA

O conceito de dependência pode ser definido de forma simples, assim como em Debusmann (2000, p. 2), como “em uma sentença, todas, exceto uma palavra, dependem de outra(s) palavra(s)”, porém essa definição não diz como identificar a relação de dependência entre as palavras. O próprio Tesnière (2015 [1959], p.3) tinha uma definição mentalista, e não

falseável, com relação à dependência: “a mente percebe a conexão entre uma palavra e seus vizinhos. A totalidade dessas conexões forma a base da frase”.

De Marneffe e Nivre (2019), trazem uma lista de critérios para a identificação de dependência com base em Zwicky (1993 [1985]) e Hudson (1990), em que critérios estão apenas em função da necessidade de processamento computacional, mas também como em função de diversas características linguísticas.

Segundo De Marneffe e Nivre (2019, p.203), a dependência entre a cabeça **H** (*head*) e o dependente **D** (*dependent*) em uma construção **C** (*construciton*), pode ser dada em função dos seguintes critérios:

- H determina a categoria sintática de C e normalmente pode substituir C;
- H determina a categoria semântica de C; D fornece a especificação semântica;
- H é obrigatória, D pode ser opcional;
- H seleciona D e determina quanto D é obrigatório ou opcional;
- A forma de D depende de H, em concordância ou governança;
- A posição linear de D é especificada com referência a H.

Contudo, essa lista, mesmo empregando diferentes critérios, tanto semânticos quanto sintáticos, se utiliza de termos das próprias gramáticas de dependências (*dependente, cabeça e governança*), que serão discutidos mais à frente, tornando-a endógena, ou seja, ela explica a relação de dependência em função de suas características próprias.

Uma definição não endógena para a relação de dependência entre elementos é encontrada em Robinson (1970, p.260), em que ela, com base nos trabalhos de Hays (1964) e Gaifman (1965) – obras que organizam e formalizam matematicamente os trabalhos de Tesnière (2015 [1959]) – postula quatro axiomas sobre estruturas de dependência bem formadas, sendo eles:

- I. Um, e apenas um, elemento é independente;
- II. Todos outros elementos dependem diretamente de algum outro elemento;
- III. Nenhum elemento depende diretamente de mais de um outro elemento;
- IV. Se um elemento p_1 depende diretamente do elemento p_2 e existe um elemento p_3 entre eles (na ordem linear da sequência de elementos), então p_3 depende diretamente de p_1 ou de p_2 ou de algum outro elemento que esteja entre eles.

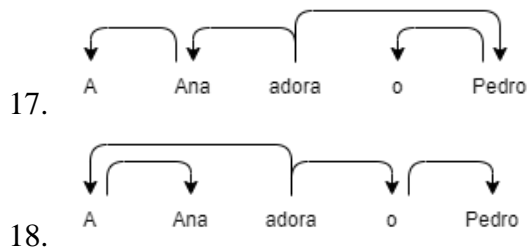
De maneira que a relação binária **p_1 depende de p_2** ($\langle p_1, p_2 \rangle \in R$) é transitiva, ou seja, se **p_1 depende de p_2** e **p_2 depende de p_3** , então **p_1 depende de p_3** ; assimétrica, ou seja, se **p_1 depende**

de p_2 , então p_2 não depende de p_1 ; e irreflexiva, ou seja, não existe nenhum elemento p_n de tal maneira que p_n depende de p_n .

Desses axiomas, é possível agora definir os termos – *dependente*, *cabeça* e *governança* – presentes na definição de De Marneffe e Nivre (2019) de maneira exógena.

Como exposto, a relação p_1 depende de p_2 é uma relação binária. Assim, é necessário que existam dois elementos que façam parte dela. O elemento que *depende* recebe o nome de *dependente*, enquanto o elemento que *domina* a relação recebe o nome de *governador* ou *cabeça*.

Partindo dessa nomenclatura pode-se definir a relação de *governança*, p_2 governa p_1 ($\langle p_2, p_1 \rangle \in G$), como a relação oposta à *dependência*, de maneira que ela também é transitiva, se p_3 governa p_2 e p_2 governa p_1 , então p_3 governa p_1 ; assimétrica, se p_2 governa p_1 , então p_1 não governa p_2 ; e irreflexiva, não existe nenhum elemento p_n , tal que p_n governa p_n . Porém, em contrapartida à *dependência*, a *governança* permite que um elemento governe mais de um dependente.



Os exemplos (17) e (18) trazem duas possíveis relações de dependências na sentença: *A Ana adora o Pedro*. Em (17), as palavras funcionais dependem das palavras de conteúdo, e, em (18), as palavras de conteúdo dependem das palavras funcionais. Nos exemplos, as setas indicam o dependente, logo, o elemento ao lado oposto do arco representa o governante. Assim, pode-se inferir que o conjunto R que descreve a relação de dependência é constituído pelas tuplas $\langle A, Ana \rangle$, $\langle Ana, adora \rangle$, $\langle Pedro, adora \rangle$ e $\langle o, Pedro \rangle$, para (17), e $\langle Ana, A \rangle$, $\langle A, adora \rangle$, $\langle o, adora \rangle$ e $\langle Pedro, o \rangle$, para (18), enquanto o conjunto G que descreve a relação de governança, é constituído pelas tuplas $\langle Ana, A \rangle$, $\langle adora, Ana \rangle$, $\langle adora, Pedro \rangle$ e $\langle Pedro, o \rangle$, para (17), e $\langle A, Ana \rangle$, $\langle adora, A \rangle$, $\langle adora, o \rangle$ e $\langle o, Pedro \rangle$, para (18).

Todavia, (17) e (18) exprimem apenas a relação $\langle p_m, p_n \rangle \in R$, ou seja, a dependência entre duas palavras. As árvores presentes nos exemplos não descrevem a nomenclatura das relações entre as palavras p_m e p_n , sendo linguisticamente pouco específicas. Isso ocorre porque

a anotação em (17) e (18) não leva em consideração uma gramática de dependência, apenas a dependência entre os elementos presentes em uma sequência de palavras.

Mesmo que seja essencial e necessário para as gramáticas de dependências, apenas o conceito de *dependência* não é suficiente para descrever linguisticamente os enunciados, fazendo-se necessários outros elementos. A próxima seção irá discorrer sobre os elementos que constituem uma gramática de dependência e como esses se correlaciona com o conceito de *dependência* para que seja possível ter uma análise mais uniforme e aprofundada do exemplo (17).

3.2 GRAMÁTICAS DE DEPENDÊNCIA

Como já dito, existem diversos trabalhos com base na gramática de dependência ao longo do século XX, o que acaba por gerar uma gama de teorias – a Gramática de Hays e Gaifman (HAYS, 1964; GAIFMAN, 1965), a Gramática de Palavras (Word Grammar) (HUDSON, 1990), a Descrição Gerativa Funcional (Functional Generative Description) (SGALL *et al.*, 1986), a Gramática de Unificação de Dependências (Dependency Unification Grammar) (HELLWIG, 1986), a Teoria do Sentido-Texto (Meaning-Text Theory) (MEL'ČUK, 1988), e a Gramática de Dependência Funcional (Functional Dependency Grammar) (JÄRVINEN e TAPANAINEN, 1998), e muitos outros. Entretanto, por mais numerosas e diversificadas, todas as teorias têm conceitos “básicos” compartilhados entre elas, além do conceito de dependência já estabelecido.

Em Polguère e Mel'čuk (2009, p.XIV) é possível encontrar uma lista de princípios básicos compartilhados, se não por todas, pela maior parte das teorias gramaticais fundamentadas no conceito de dependência, sendo eles:

- I. Conectividade da estrutura sintática;
- II. Direção das relações sintáticas;
- III. Organização hierárquica estrita da estrutura sintática;
- IV. “Significado” das relações sintáticas.

A *conectividade da estrutura sintática* faz referência à relação das palavras em um enunciado. “A estrutura sintática forma um todo organizado, isto é, um sistema contínuo de relações sintáticas” (POLGUÈRE e MEL'ČUK, 2009, p. XIV), ou seja, toda e qualquer palavra p_m que faz parte de um enunciado está relacionada com ao menos uma outra palavra p_n . Não existem palavras dentro de um enunciado que estejam fora da estrutura sintática. Cada par

relacional $R\langle p_1, p_2 \rangle$ representa um enunciado mínimo, como ocorre no exemplo (17): $R = \{\langle A, Ana \rangle, \langle Ana, adora \rangle, \langle Pedro, adora \rangle, \langle o, Pedro \rangle\}$.

É também visível no exemplo (17) que não existe nenhum elemento que faça parte do enunciado deixado de fora da estrutura, de maneira que cada uma das palavras estabelece ao menos uma conexão entre as outras palavras que a circundam. Segundo Polguère e Mel'čuk (2009, p. XIV), a propriedade formal que decorre desse princípio é que a estrutura sintática é um *grafo conectado*, ou seja, a estrutura de dependência é um conjunto de elementos que têm relações entre eles. O comportamento da estrutura de dependências como um grafo permite que todos os elementos da sentença sejam analisados como parte de um conjunto finito.

As relações sintáticas dos elementos de um enunciado são direcionadas. Assim como a relação de dependência é assimétrica, as relações sintáticas estabelecidas por essa relação também serão. Cada elemento dentro de um enunciado mínimo domina, ou é dominado, por um outro. Isso é demonstrável pela capacidade de um enunciado mínimo ser anexado a um outro elemento tendo o comportamento de sua *cabeça*. Tomemos os exemplos:

19.
 - a. Pedro tem olhos **muito belos**
 - b. Pedro tem olhos **castanhos**
 - c. Pedro corre **rápido**
 - d. (?) Pedro corre **muito belo**
 - e. Pedro corre **muito bem**

Muito belo(s) se comporta como um *adjetivo*, tal qual *castanho(s)*, como é visível pela oposição entre (19a) e (19b). Todavia, mesmo que haja um advérbio dentro do enunciado mínimo *muito belo(s)*, ele não se comporta como um *advérbio*, mas sim como um *adjetivo*, o governante do par, visível entre (19c) e (19d). Mesmo sendo possível uma interpretação em que (19d) seja gramatical (*Pedro corre e Pedro está/é muito belo*), ainda assim, o comportamento de *muito belo(s)* é completamente diferente do comportamento do advérbio *rápido* em (19c). Entretanto, *muito bem* em (19e) se comporta como um *advérbio*, assim como *rápido* em (19c), uma vez que o governante do enunciado mínimo é um *advérbio*.

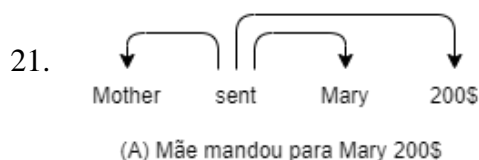
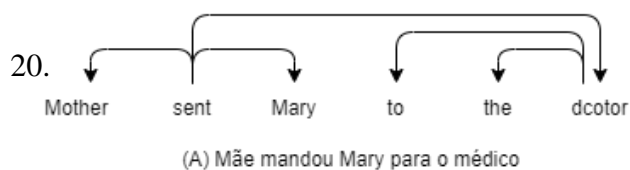
Segundo Polguère e Mel'čuk (2009, p. XIV), a propriedade formal que decorre do princípio de direção das relações sintáticas é que a estrutura sintática é um *grafo conectado e direcionado*. O direcionamento do grafo permite identificar quais elementos são mais relevantes para a estrutura da sentença analisada.

Assim como proposto pelos axiomas I e II de Robinson (1970, p.260), cada elemento dentro da estrutura sintática depende de um outro elemento, com exceção de um único elemento que governa todo enunciado. A organização hierárquica estrita se refere à presença de um, e apenas um, governador para cada elemento e um elemento máximo da estrutura de onde todos os outros elementos dependem, de maneira que o elemento que governa a sentença é o ponto máximo da estrutura de dependência.

Essa propriedade define que não existem ciclos dentro do grafo da estrutura sintática, ou seja, a estrutura parte do governador máximo, passa pelos pares intermediários e termina nos dependentes que não governam nenhum elemento. Em nenhum momento algum outro elemento irá dominar o governador máximo da sentença. Assim, no exemplo (17), o governador máximo do enunciado é o verbo *adorar* que não depende de nenhum outro elemento dentro da estrutura e os nomes *Ana* e *Pedro* governam respectivamente os artigos definidos *a* e *o*, que, por sua vez, não governam nenhum outro elemento.

A propriedade formal que decorre do princípio da organização hierárquica restrita do enunciado é que “a estrutura sintática é um *grafo direcionado, acíclico e conectado*, ou seja, uma árvore hierarquizada ou, mais brevemente, uma árvore” (POLGUÈRE e MEL'ČUK, 2009, p. XV, grifo nosso). Dessa forma, a estrutura de dependências relativa a uma construção é um sistema finito, ordenado e télico, ou seja, ele se inicia de um elemento e, necessariamente, termina em um outro.

Por último, a propriedade de “significado” das relações sintáticas. Pelos conceitos aqui estabelecidos, as relações sintáticas ainda são um reflexo direto da relação de dependência, porém, apenas esse tipo de relação não é suficiente para exprimir o comportamento das línguas naturais. Um exemplo disso é que se pode encontrar duas relações de dependência $R\langle p_1, p_2 \rangle$ que exprimam relações sintáticas distintas. Tomemos como exemplo as sentenças do inglês *Mother sent Mary to the doctor* e *Mother sent Mary 200\$* trazidas por Polguère e Mel'čuk (2009, p. XV):



Nesses exemplos, é possível identificar a relação de dependência $R(Mary, to\ send)$. Contudo, cada uma delas exprime uma relação sintática distinta. Em (20), *Mary* se relaciona com *to send* como objeto direto, enquanto em (21), *Mary* se relaciona com *to send* como objeto indireto. Mesmo que ambos os pares $R_{objD}(Mary, to\ send)$, em (20), e $R_{objI}(Mary, to\ send)$, em (21), sejam representados pela mesma distribuição na frase, *Mary* ocorre imediatamente após *to send*, eles representam relações de sentido distintas.

Assim, as relações sintáticas entre os elementos da frase não são apenas configurações estruturais, mas também carregam significado. Entretanto, deve-se notar que as relações sintáticas não carregam um sentido pré-definido, mas ele se altera em função da especificidade do governador da relação, como podemos observar nos seguintes exemplos:

22. A UFSCar irá contratar novos técnicos administrativos;

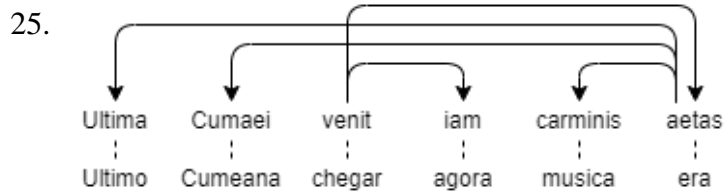
23. A UFSCar vem sofrendo com os cortes de gastos;

24. A UFSCar sediou a Reunião Anual da SBPC em 2015 um evento importante este ano.

A relação entre *UFSCar* e os verbos das sentenças é de sujeito ($R_{Suj}(p_1, p_2)$), mas em cada um dos exemplos ele desempenhará um papel semântico distinto: AGENTE, PACIENTE e LOCATIVO, respectivamente.

A propriedade formal que segue desse princípio é que a estrutura sintática é uma árvore em que os nós são rotulados e vinculados por arcos nomeados pelas relações sintáticas que representam (POLGUÈRE e MEL'ČUK, 2009, p. XVI).

Um último conceito muito relevante para as gramáticas de dependência é a *projetividade* (LECERF, 1960; HAYS; 1964; GAIFMAN, 1965), isto é, em uma árvore de dependências, caso ela seja projetiva, não pode haver interseções entre os galhos, ou seja, não existem cruzamentos entre as dependências em uma árvore. Todavia, por mais vantajosa que a *projetividade* seja para o processamento automático computacional (DE MARNEFFE e NIVRE; 2019, p.202), as línguas naturais nem sempre funcionam de acordo com essa propriedade. Línguas ricas em morfologia que marcam casos – como o latim, o russo e húngaro – aceitam ordenações distintas entre as palavras, o que pode acarretar problemas de projetividade. Isso pode ser observado no exemplo (25), em latim:



“É chegada agora a era da última canção cumeana”, (Virgílio, Bucólicas 4 *apud* POLGUÈRE e MEL’ČUK, 2009, p.87)

O dilema da projetividade é tratado de diferentes maneiras por diferentes teorias baseadas em dependência. A maior parte das soluções são multiestrato (SGALL *et al.*, 1986), em que a projetividade só existe nos níveis mais altos, sendo solucionada nos níveis mais profundos. Como solução para a projetividade há também propostas como a de Mel’čuk (1988) e Hudson (1990) que utilizam de transformações lógicas entre os multiestratos para solucionar a falta de projetividade.

Tomar-se-á aqui, considerando que a imensa maioria das sentenças das línguas naturais são projetivas (POLGUÈRE e MEL’ČUK, 2009, p.85) e as diretrizes da UD (UNIVERSAL DEPENDENCIES, 2021)¹², a projetividade como o padrão das línguas naturais.

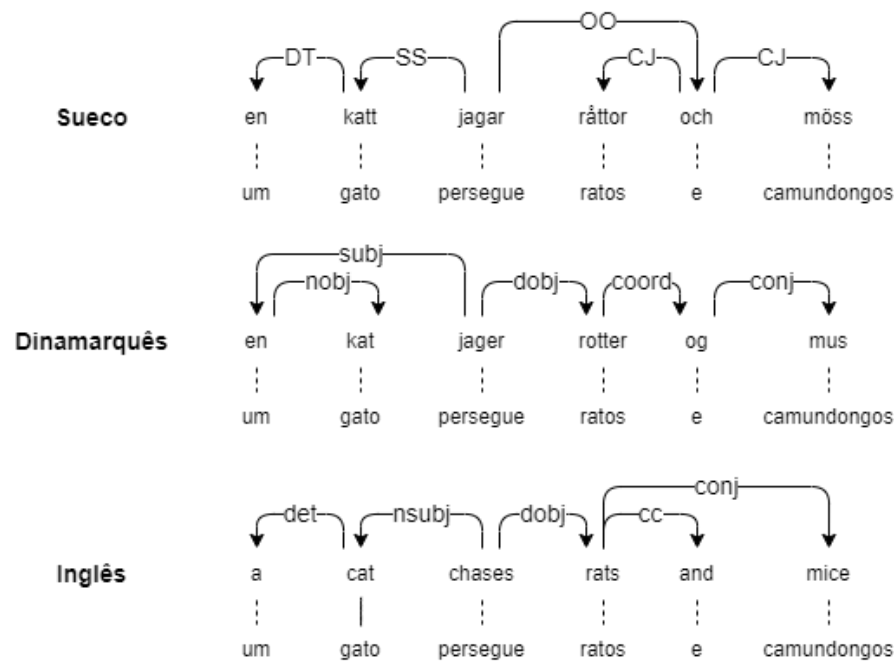
Antes de finalizar esta seção é necessário ressaltar que mesmo que todos os conceitos elencados aqui sobre *dependência* estejam diretamente relacionados à sintaxe das línguas naturais, *dependência* não está restrito a apenas esse nível de granularidade linguística, podendo ocorrer, como descrito por Enfield (2017), em outros níveis. Por motivos metodológicos, este trabalho discorrerá sobre *dependência* em função da sintaxe, todavia, essa não é a única abordagem para análises de dependência dentro das línguas naturais.

¹²<https://universaldependencies.org/guidelines.html>

3.3 UNIVERSAL DEPENDENCIES

O incrível aumento de anotações de *corpus* em Processamento de Língua Natural baseadas em dependência (KROMANN, 2003; DE MARNEFFE *et al.* 2006; NIVRE e MEGYESI, 2007; DE LHONEUX, 2018; HAN; 2018; ZHANG, 2020; entre muitos outros) é um fenômeno extremamente relevante na comunidade científica. Nos últimos anos, surgiram diferentes *corpora* em diferentes línguas. Em teoria, essa quantidade de *corpora* possibilitaria diversas atividades multilíngues, como a tradução automática e estudos comparativos de estruturas. Todavia, a aplicação desses *corpora* em ferramentas multilíngues esbarra em um grande problema: a compatibilidade entre as anotações. A figura 4 é uma adaptação de um exemplo trazido por Nivre *et al.* (2016) e de Marneffe e Nivre (2019).

Figura 4– Exemplo de distintas anotações para a sentença: ‘um gato persegue ratos e camundongos’ em sueco, dinamarquês e inglês



Fonte – (DE MARNEFFE E NIVRE, 2019, p.209). Adaptado pelo autor.

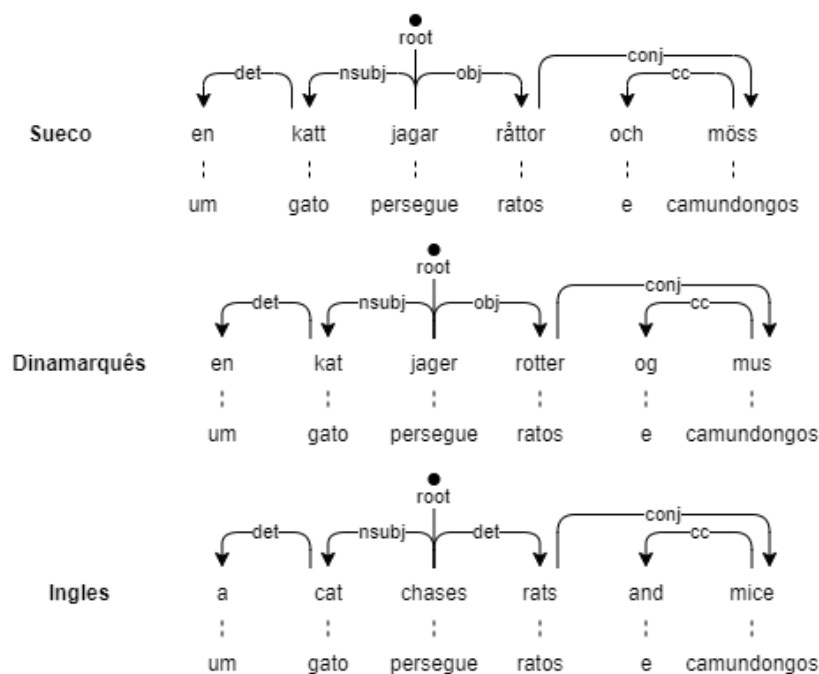
Nela é observável a comparação de três sentenças similares anotadas de acordo com as diretrizes de três diferentes propostas de anotação: em sueco, *Swedish Treebank* (NIVRE e MEGYESI, 2007); em dinamarquês, *Danish Dependency Treebank* (KROMANN, 2003); e inglês, *Stanford Typed Dependencies* (DE MARNEFFE *et al.* 2006), sobre as quais de Marneffe e Nivre (2019) afirmam que:

Embora as estruturas sintáticas sejam idênticas nas três línguas, a porcentagem de relações de dependência compartilhadas entre os pares de idiomas é de no máximo 40% (e 0% entre os três idiomas), tornando muito difícil comparar as árvores de dependência de maneira significativa¹³. (DE MARNEFFE E NIVRE, 2019, p.209. Tradução do autor)

Em suma, mesmo com estruturas sintáticas idênticas, a compatibilidade entre as anotações é muito baixa, tornando as atividades multilíngues muito mais complexas.

A UD (UNIVERSAL DEPENDENCIES, 2021)¹⁴ surgiu em 2014 como uma proposta de solução para este problema. Ela tem como objetivo uma anotação morfossintática multilíngue mais consistente (NIVRE *et al.*, 2016, p.1660), visando reduzir ao máximo o ruído gerado entre as diferentes anotações para aumentar a concordância e a quantidade de relações compartilhadas entre as línguas. Seguindo as diretrizes da UD, as sentenças da figura 4 têm uma mesma anotação entre as relações sintáticas de dependência, como é possível observar na figura 5.

Figura 5– Anotação sintática, simplificada, conforme as diretrizes da UD para a sentença “um gato persegue ratos e camundongos” em sueco, dinamarquês e inglês.



Fonte – (DE MARNEFFE E NIVRE, 2019, p.209). Adaptado pelo autor.

O modelo da UD é uma anotação gramatical para *corpus* que sistematiza as classes gramaticais ou PoS (*parts of speech*, partes do discurso), características morfológicas e estrutura

¹³ “While the syntactic structures are identical in the three languages, the percentage of shared dependency relations across pairs of languages is at most 40% (and 0% across all three languages), making it very hard to compare the dependency trees in a meaningful way”, no original.

¹⁴ Disponível em: <https://universaldependencies.org/>

sintática em função de dependências. Todo o modelo de anotação é baseado, e melhorado, nos modelos da *Stanford dependencies* (DE MARNEFFE *et al.*, 2006; DE MARNEFFE e MANNING, 2008; DE MARNEFFE *et al.*, 2014), para as dependências sintáticas; *Google universal part-of-speech tags* (PETROV *et al.*, 2012), para a anotação das classes gramaticais; e no *Internet Interlingua for Morphosyntactic Tagsets* (ZEMAN, 2008), para a classificação das características morfológicas.

A anotação morfológica consiste em três níveis de representação (Nivre *et al.*, 2020, p.4035):

- I. Um lema que representa a forma base do *token*¹⁵;
- II. Uma etiqueta de PoS que representa a classe gramatical do *token*; e
- III. Um conjunto opcional que representa as características morfológicas associadas ao *token*.

O lema é representado pela forma canônica e dicionarizada de cada língua. Para o conjunto de PoS existe um total de 17 etiquetas invariáveis distribuídas em três classes: *aberta*, os *tokens* da classe não são pré-definidos; *fechada*, os *tokens* da classe são pré-definidos; e *outras*, assim como no quadro 2:

Quadro 2 – Classes de PoS disponíveis na metodologia da UD

Classe Aberta		Classe Fechada		Outras	
ADJ	Adjetivo	ADP	Preposições e pósposições	PUNCT	Pontuação
ADV	Advérbio	AUX	Auxiliar	SYM	símbolo
INTJ	Interjeição	CCONJ	Conjunção coordenativa	X	PoS indefinida
NOUN	Nome	DET	Determinante		
PROPN	Nome próprio	NUM	Numeral		
VERB	Verbo	PART	Partícula		
		PRON	Pronome		
		SCONJ	Conjunção subordinativa		

Fonte – (NIVRE *et al.*, 2016, p.1661). Traduzido pelo autor.

As características morfológicas não são fixas, permitindo que seja possível adicionar as características necessárias para cada língua. Existe um total de 213 possíveis características disponíveis no site da UD¹⁶, sendo que 24 delas, presentes no quadro 3, são destacadas por

¹⁵ Considerando a ambiguidade envolvida com a palavra *palavra*, optamos por utilizar *token* para a significação de unidade de texto separada por espaços em branco.

¹⁶ <https://universaldependencies.org/u/feat/index.html>

Nivre *et al.* (2020, p. 4035) como “[...] um conjunto de características que foram atestadas em diversas línguas e precisam ser classificadas de forma uniforme [...]”¹⁷ (tradução do autor).

Quadro 3 – Características morfológicas relevantes

Flexionais		Lexicais	
Animacy	Animacidade	Abbr	Abreviação
Aspect	Aspecto	Foreign	Estrangeirismo
Case	Caso	NumType	Tipo de numeral
Clusivity	Inclusividade	Poss	Possessivo
Definite	Definitude ¹⁸	PronType	Tipo de Pronome
Degree	Grau	Reflex	Reflexivo
Evident	Evidencialidade	Typo	Erro de digitação
Gender	Gênero		
Mood	Modo		
NounClass	Classe de Nomes		
Number	Número		
Person	Pessoa		
Polarity	Polaridade		
Polite	Polidez		
Tense	Tempo		
VerbForm	Forma Verbal		
Voice	Voz		

Fonte – (NIVRE *et al.*, 2020, p.4036). Adaptado e traduzido pelo autor

Cada característica morfológica é anotada no formato *nome da feature=valor*, por exemplo *Gender=Masc/Fem* e *Number=Sing/Plur* e não há limites para a quantidade de características que um mesmo *token* pode receber, logo, o *token cachorros* recebe, nos conformes das diretrizes da UD, os atributos de gênero masculino e número plural:

26. **Cachorros** ► atributos: Gender=Masc | Number=Plur

¹⁷ “[...] an inventory of features that are attested in multiple languages and need to be encoded in a uniform way [...]”, no original

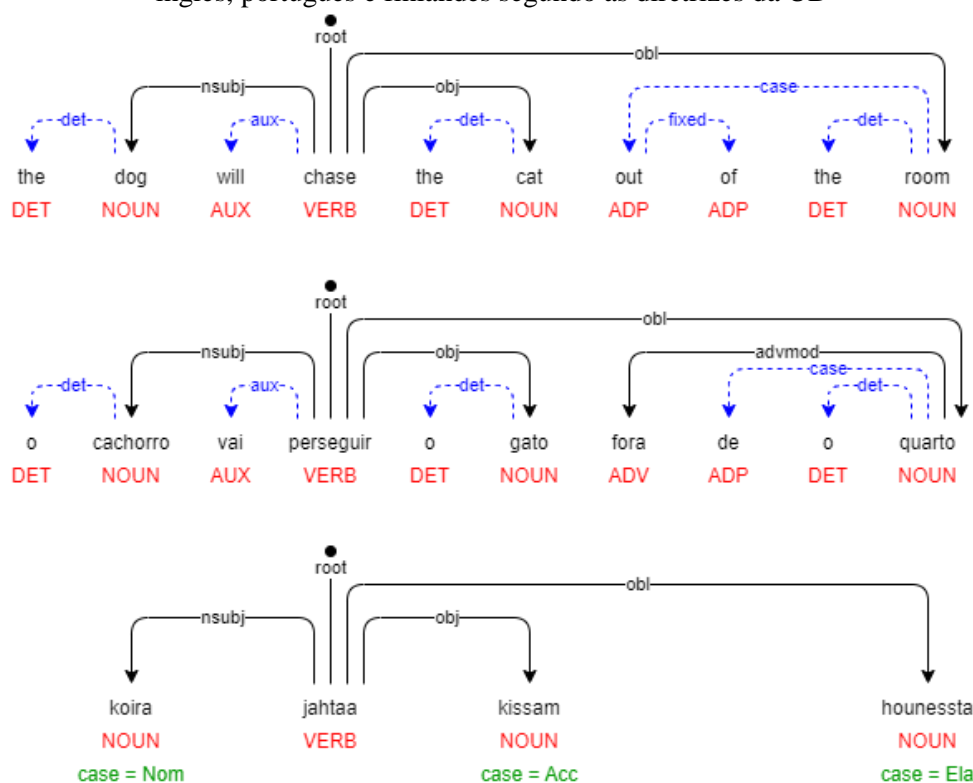
¹⁸ Tradução do termo *definite* do inglês. Optamos por *definitude* para categorizar a relação entre *definido* e *indefinido* a fim de evitar ambiguidade com o termo *definição*.

As relações sintáticas são baseadas em dependências, ou seja, cada *token* se associa de maneira binária com ao menos um outro *token* da sentença, de maneira que o *token* que governa a relação recebe o nome de *head* (cabeça) e o *token* que é governado recebe o nome de *dependent* (dependente).

Assim, como a proposta de Robinson (1970) para estruturas de dependência bem formadas, todos os *tokens* de uma sentença são dependentes de ao menos um outro *token*, com exceção do *token* que governa toda a sentença. Um mesmo *token* pode ser cabeça de mais de uma relação de dependência, mas não pode ser dependente de mais de uma relação. Da mesma forma, um *token* pode ser o *head* de uma relação de dependência e o *dependent* de uma outra relação. O *token* mais alto da sentença, aquele que não é dependente de nenhum outro, recebe o título de *root* (raiz).

Como critérios para a seleção de *head* e *dependent*, todo *head* de uma sentença sempre será uma palavra de conteúdo, pertencente à classe aberta, enquanto todas as palavras de funcionais, pertencentes à classe fechada, sempre serão dependentes. Essa organização entre os *dependents* e os *heads* ocorre por conta da proposta multilíngue. Línguas distintas expressam funções gramaticais com diferentes regras morfossintáticas. Algumas, como o inglês e o português, expressam essas relações por meio de palavras (preposições, artigos e auxiliares), enquanto outras expressam por meio de declinação morfológica de caso, como o finlandês e o russo.

Figura 6 – Comparação da anotação da sentença ‘o cachorro vai perseguir o gato fora do quarto’ em inglês, português e finlandês segundo as diretrizes da UD



Fonte – (DE MARNEFFE e NIVRE, 2019, p.207). Adaptado pelo autor

Na figura 6, é possível observar que as três sentenças apresentam a mesma organização sintática (um sujeito, um objeto e um modificador locativo), representados pelos arcos em preto *nsubj*, *obj* e *obl* respectivamente. Por sua vez, as palavras funcionais, representadas pelos arcos azuis, estão vinculadas às palavras de conteúdo que modificam, em inglês e em português¹⁹, enquanto em finlandês o que modifica as palavras de conteúdo são as declinações morfológicas (os casos nominativo, acusativo e elativo). Assim, o critério de manter todas as palavras de conteúdo como dependentes permite um paralelismo mais evidente em línguas com diferentes estruturas morfossintáticas.

A UD disponibiliza um total de 37 relações sintáticas possíveis para a anotação. Entretanto, nem todas as relações são dependências no sentido estrito, uma vez que existem relações que marcam disposições não hierárquicas como conjunções e expressões multipalavras.

¹⁹ Mesmo que a relação entre *fora* e *quarto* seja marcada como *advmod* (advérbio modificador) simbolizando uma relação entre duas palavras de conteúdo, ela desempenha a mesma função que a locução *out of* do inglês e o caso *elativo* do finlandês. A classificação dos dêiticos locativos como advérbios, ainda que seja reforçada pela gramática tradicional, ainda é alvo de debates sobre em quais classes esses dêiticos devem ser categorizados (por exemplo: XAVIER e KANTHACK, 2015; NEVES, 1993; 2000)

As relações estão distribuídas entre relações oracionais (vinculadas a orações), nominais (vinculadas a nomes), palavras modificadoras, palavras funcionais, coordenações, expressões multipalavras, livres, especiais e outras, como é possível observar no quadro 4.

Quadro 4 – Relações de dependência da UD

	Nominal	Oracional	Palavra Modificadora	Palavra Funcional
Argumento Essencial	nsubj obj iobj	csbj ccomp xcomp		
Dependente Não Essencial	obl vocative expl dislocated	advcl	advmod discourse	aux cop mark
Dependente Nominal	nmod appos nummod	acl	amod	det clf case
Coordenação	Expressão Multipalavra	Livres	Especial	Outras
conj cc	fixed flat compound	list parataxis	orphan goeswith reparandum	punct root dep

Fonte – <https://universaldependencies.org/u/dep/index.html>, acesso em: 18/02/2022.

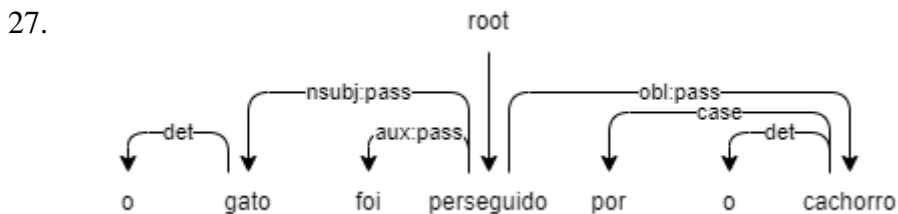
Traduzido pelo autor.

A classificação de argumentos essenciais (*core arguments*) e dependentes não essenciais (*non-core dependent*) está relacionada com a necessidade de um *token* ocorrer junto ao seu predicado. De maneira geral, os argumentos essenciais estão associados à capacidade do dependente em assumir os papéis semânticos de agente e dependente (NIVRE *et al.*, 2020, p.4036; DE MARNEFFE *et al.*, 2021, p.286) com relação a um predicado, enquanto os dependentes não essenciais estão vinculados aos dependentes que não se associam a esse papel temático.

Como já dito, ainda que a categorização das relações sintáticas seja feita em função de dependência, nem todas as relações representam uma dependência no sentido estrito. As categorias de conjunções, expressões multipalavras e de relações livres não dispõem de uma hierarquia vertical, ou seja, não existe uma relação de prioridade entre o *head* e o *dependent*, enquanto as categorias de pontuação e correções de erros (*goeswith* e *reparandum*) indicam uma marcação textual ou apenas que houve um erro de digitação. Dispor de todas essas relações como dependências é apenas uma metodologia para que dispositivos computacionais

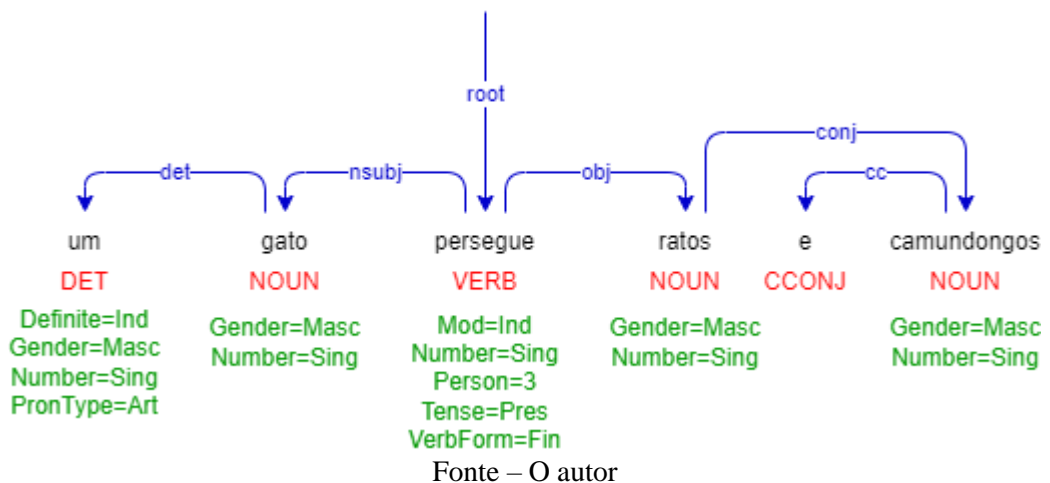
sejam capacitados a interpretar a estrutura linear das sentenças em uma árvore (NIVRE *et al.*, 2020, p.4035).

Todas as relações sintáticas são pré-definidas nas diretrizes da UD, todavia é possível gerar subtipos individuais para satisfazer as necessidades de cada língua, como é possível observar no exemplo (27), com a adição do subtipo `pass` para indicar que a relação se trata da organização de uma sentença na voz passiva, uma organização muito comum nas línguas latinas:



Assim, com todos os níveis da UD para a anotação textual, as sentenças devem ser anotadas, segundo as diretrizes, contendo sua representação de PoS, etiquetas morfológicas e relações de dependências. A figura 7 é o modelo completo, com todos os níveis de anotação, para a sentença “um gato persegue ratos e camundongos” (as relações sintáticas destacadas em azul, as classes gramaticais em vermelho e as características morfológicas em verde).

Figura 7 – Anotação completa, segundo as diretrizes da UD, da sentença ‘um gato persegue ratos e camundongos’



O português faz parte das 130 línguas já anotadas em função da UD (UNIVERSAL DEPENDENCIES, 2021). Dessa forma, o português dispõe de um total de três *corpora* disponíveis no site da UD: UD Portuguese Bosque (RADEMAKER *et al.*, 2017), UD Portuguese GSD (ZEMAN *et al.*, 2021) e PUD Portuguese Bosque (ZEMAN *et al.*, 2021). Todavia, eles não são os únicos trabalhos realizados na metodologia da UD em português.

Os trabalhos em português acompanharam a expansão da UD nos últimos anos, de maneira que já existem diversos trabalhos acerca dessa metodologia de anotação em função das especificidades do português. Os trabalhos se distribuem em três principais categorias: (i) anotação de *corpus* (PAIVA e REAL, 2016; REAL, 2018; SOUZA *et al.*, 2021; DA SILVA *et al.*, 2021; DI FELIPPO *et al.*, 2021; CONEGLIAN *et al.*, 2022; SOUZA e FREITAS, 2022), (ii) discussões morfossintáticas acerca da estrutura do português com relação às diretrizes da UD (DURAN *et al.*, 2021; SOUZA e FREITAS, 2021; DURAN, OLIVEIRA e SCANDAROLLI, 2022; DURAN *et al.*, 2022) e (iii) manuais de anotação específicos para o português (DURAN, 2021a; DURAN, 2021b; SOUZA *et al.*, 2021).

Discorreu-se no início desta seção sobre as vantagens de se adotar um modelo em função de dependências no lugar de um modelo de constituintes. A principal motivação para a adoção de um modelo de dependências é o comportamento bijetivo entre os elementos da sentença e a estrutura sintática, ou seja, cada elemento da sentença representa um, e apenas um, elemento dentro da estrutura sintática de dependência.

Para exemplificar como essa relação é produtiva, tomemos como exemplo a sentença em (28), retirada do Novo Manual de Sintaxe (MIOTO *et al.*, 2007, p.110):

28. A menina sentou desajeitada

No exemplo, existe uma complicação com relação à análise sintática referente ao termo *desajeitada*. Mesmo próximo ao verbo *sentar*, *desajeitada* não se comporta como um advérbio, o termo está relacionado com o substantivo *menina*, como é observável pela concordância de gênero entre os elementos. Caso o gênero do substantivo seja alterando, como em nossa modificação em (28), todos os elementos vinculados ao substantivo também terão seu gênero alterado motivados pela concordância:

29. O menino sentou desajeitado

Como *desajeitada* se comporta como um adjetivo em (28), pela estrutura de constituintes, é necessário que *desajeitada* e *menina* façam parte de um mesmo constituinte, mais precisamente, ambos precisam fazer parte de um *sintagma adjetival*. Entretanto, caso *desajeitada* seja movido para próximo de *menina*, a sentença resultante não terá a mesma interpretação de (30):

30. A menina desajeitada sentou.

Mioto et al (2007, p.110)

Mesmo *desajeitada* sendo um adjetivo e esteja vinculado a *menina*, a interpretação dos dois elementos como um único sintagma, *menina desajeitada*, não é suficiente para explicar a interpretação de (28), isso é motivado pela existência duas orações no exemplo:

- 31.
- a. [sv [SD A [SN menina]] sentou]
 - b. [sv [SD A [SN menina]] estava [SA desajeitada]] (quando sentou)

Por essa perspectiva, para que as duas orações de (31) estejam presentes em (28) é necessário que haja uma elipse no elemento conjuntivo *e*, e no verbo da segunda oração, *estar*:

32. [sc [sv [SD A [SN menina]] sentou] e [sv ~~estava~~ [SA desajeitada]]

Se tomarmos (32) como modelo para (28), teriam que ser adicionados dois elementos não presentes na sentença, para que seja possível realizar a análise sintática, além da adição de um novo sintagma conjuntivo para a conjunção *e*.

Outra forma de realizar a análise sintática de (28) em função de constituintes, derivada da gramática gerativa (CHOMSKY, 1957), é trazida por Mioto *et al.* (2007). O autor adota a existência de um constituinte denominado *small clause* que é “uma predicação que se estabelece entre um constituinte que é sujeito e um outro que é predicado sem que o núcleo desse predicado seja um verbo (ou uma flexão verbal)” (MIOTO *et al.*, 2007, p.107). Por essa proposta, pode-se assumir que exista um *substantivo* vazio (\emptyset) que se associa à *desajeitada* e que tem o mesmo referencial que *menina*²⁰:

33. [sv [SD A [SN menina_i]] sentou [small clause [SA [SN \emptyset _i desajeitada]]]]

A proposta presente em Mioto *et al.* (2007), não prescreve a adição de uma conjunção, mas adiciona a *small clause* e o substantivo \emptyset dentro do sintagma nominal, o que torna necessário a adição de mais elementos do que os presentes na sentença para realizar a análise sintática.

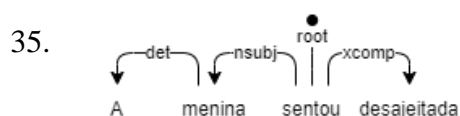
As duas propostas de análise em função de constituintes acabam por não serem computacionalmente muito eficiente. A necessidade de se inserir elementos e ou sintagmas abstratos geram passos de processamento muito mais longos que não são baseados nos dados analisados.

²⁰ O índice *i* indica que os elementos têm o mesmo referencial.

Em oposição às propostas em função de constituinte, pode-se realizar a análise sintática de (28) em função de dependência, de maneira simplista, partindo das relações entre governantes e dependentes a partir do verbo:



Entretanto, essa anotação exprime um problema, mesmo que em (34) os galhos não se cruzem, a relação entre *menina* e *desajeitada* impede que o verbo *sentar* seja o elemento mais alto da árvore. Não se pretende aqui discutir as implicações relativas à não projetividade de uma sentença. Assim, para solucionar esse problema, pode-se utilizar a metodologia da UD, construindo a dependência como:



Pela proposta da UD, a relação entre *sentar* e *desajeitada* é, afim de evitar a projetividade, denominada como uma oração complementar aberta (*xcomp*), que é a relação entre um *verbo* ou *adjetivo* que esteja relacionado a um elemento predicator, mas seus argumentos se encontram fora da dependência.

Mesmo que (34) não seja projetiva, ele e (35) compartilham a citada correspondência bijetiva entre os elementos da sentença e a estrutura sintática de dependência, ou seja, cada elemento da sentença é equivalente a cada elemento da estrutura sintática, não havendo a necessidade de serem inseridos novos elementos na sentença ou a criação de índices de correlação entre elementos.

Computacionalmente, essa forma de análise é mais produtiva, uma vez que torna mais rápido o processamento, por necessitar de menos passos, como também o torna mais simples, uma vez que não é necessário que cada elemento seja combinado em um constituinte, que por sua vez será combinado com outros constituintes de maneira recursiva até que se obtenha a sentença final.

Dito isto, este capítulo realizou uma revisão sobre o conceito de *dependência* dentro das línguas naturais para que, nesta subseção, fosse possível discorrer sobre as diretrizes da UD a fim de fundamentar as futuras análises realizadas no capítulo 5. O próximo capítulo será uma revisão acerca do funcionamento da negação dentro do português.

4 NEGAÇÃO NO PORTUGUÊS BRASILEIRO

Encantado: (...) me diga, boneco, onde está Shrek?

Pinóquio: Bem... Eu não sei onde ele não está.

Shrek Terceiro

Como elencado na introdução, a negação está presente em todas as línguas humanas já catalogadas. Não há uma língua que não apresente algum meio de expressar a negação dentro de seu sistema (HORN e WANSING, 2020). Mesmo assim, não existe uma tipologia única para a negação das línguas naturais. Existem célebres trabalhos propondo categorizações para a negação em línguas específicas (DAHL, 1979; PAYNE, 1985; MULLER, 1991; MIOTO; 1992), o que ressalta as diferentes formas que as diferentes línguas têm para expressá-la. Dentro desse paradigma, o português é uma língua muito produtiva quanto à negação.

A tradicional abordagem para a negação é sua equivalência com a negação lógica (\neg) (HORN, 2001; HORN e WANSING, 2020). Desse ponto de vista, ela não é nada mais que um mero operador que realiza oposição entre proposições. Se a proposição é verdadeira, a negação a torna falsa, se a proposição for falsa, a negação a torna verdadeira. Entretanto, essa simples oposição não é totalmente verificável nas línguas naturais. Mesmo existindo sentenças, como no exemplo (36), que representam essa oposição, as línguas naturais exprimem diversas formas de realizar a negação:

- 36.
- a. Existe outro jeito
 - b. **Não** existe outro jeito (*FK*)

O par opositivo <36.a, 36.b> é um claro exemplo do funcionamento da negação em língua natural ocorrendo em equivalência de sentido com a negação lógica. A única distinção entre (36a) e (36b) é a presença da negação, que é suficiente para gerar oposição entre as duas sentenças. Esse comportamento específico da negação recebe o nome de *negação padrão* (do inglês *standard negation*), uma vez que ela exprime comportamento idêntico à negação lógica.

Ainda assim, essa não é a única forma de realizar a negação em português. Existem outras palavras que são capazes, assim como o *não*, de realizar a negação dentro do português, como observável nos exemplos de (37) a (45):

37. **Nunca** me senti tão frágil e assustada na minha vida. (*FK*)
38. Já Palocci **jamais** teria tido qualquer compromisso com ela. (*FK*)

39. Eduardo Cunha **sequer** era filiado ao PMDB no momento da compra. (*FK*)
40. Para Naldinho, a música **não** é violenta **nem tampouco** sexual. (*FK*)
41. **Nada** indica que quebrará. (*FK*)
42. **Ninguém** falta ao trabalho, explica, e a empresa economiza com vale-transporte. (*FK*)
43. **Nenhum** dos serviços cobra taxa adicional dos clientes. (*FK*)
44. O longa é uma obra **sem** firulas vistosas para entreter o espectador. (*FK*)
45. **Nem** é preciso dizer que, depois dessa, o tatuador distraído levou mais umas bifas. (*FK*)

Dentre os elementos destacados nos exemplos, pela gramática tradicional (BECHARA, 2009; CUNHA e CINTRA, 2016; entre outros), os itens *nunca*, *jamais*, *sequer*, *tampouco* e *nem* são descritos como parte da categoria dos *advérbios de negação*. Todavia, alguns outros elementos, fora dessa categoria de advérbios, também são capazes de realizar a negação, como os pronomes *nada*, *ninguém* e *nenhum* e também a preposição *sem*.

Para além desses elementos, também é possível realizar a negação em sentenças com palavras lexicais, ou seja, palavras que não são tradicionalmente descritas como elementos negativos, mas que também são capazes de acionar a negação em um enunciado:

46. O médico **negou** que esteja tratando mais alguém (*UPB*)
47. Falamos sobre a **falta** de importância dos estaduais. (*FK*)
48. Outro tema **ausente** foi a análise de como o país prostrou-se perante o resultado (*FK*)

Nos exemplos (46), (47) e (48) a negação está sendo realizada por um verbo, um substantivo e um adjetivo, respectivamente. Todos os exemplos podem ser reescritos com o uso de um elemento negativo:

49. O médico afirmou que **não** está tratando mais ninguém.
50. Enfim, **não** existe/há uma parte importante da lição de casado projeto Centro Novo
51. Outro tema **não** presente foi a análise de como o país prostrou-se perante o resultado

Além das formas presentes nos exemplos citados, a negação no português brasileiro também pode ser realizada dentro do paradigma morfológico, de maneira que a presença de um morfema, mais especificamente um prefixo (*a-*, *des-*, *i(n)-*), é capaz de inventar a interpretação do significado do item associado.

50. O tênis brasileiro teve um dia **atípico** neste domingo [...] (*FK*)
51. A atitude de Graça **desagradou** à presidente Dilma Rousseff [...] (*FK*)

52. É ilógico alocar médicos nos quais se investiu tanto para desempenhar tarefas menos complexas [...] (**FK**)
53. Seu casamento é **infeliz**. (**FK**)

Dessa forma, o mapeamento da negação no português abrange uma vasta gama das categorias gramaticais.

Em função disso, o presente capítulo pretende realizar uma revisão de bibliografia acerca das diferentes formas de ocorrência da negação do português, a fim de propor uma tipologia para essa categoria. Além disso, busca-se também realizar uma revisão bibliográfica mais aprofundada voltada para cada uma das lexias que se pretende aqui investigar.

4.1 OS TIPOS DE NEGAÇÃO

Para a categorização da negação, seguir-se-á aqui a proposta realizada por Muller (1991, p. 55-60). O autor estabelece que a negação é capaz de ocorrer em duas formas distintas dentro da língua: (i) a **negação sintática** (*la négation syntaxique*) e (ii) a **negação no léxico/negação lexical** (*la négation dans le lexique*).

Os elementos pertencentes à **negação sintática** são lexias que desempenham uma função sintática específica para negação, ou seja, elas ocorrem em posições distintas e específicas para elementos que realizam a negação dentro de uma língua. Essas lexias fazem parte das classes de palavras denominadas como *fechadas* da gramática, ou seja, as palavras que têm seu significado em função do próprio sistema linguístico.

Já os elementos pertencentes à **negação lexical** são lexias que não ocupam a posição sintática reservada para a negação, mas a exprimem de alguma forma dentro de um enunciado. Essas lexias fazem parte das classes de palavras denominadas como *abertas*, ou seja, palavras que têm seu significado em função de um referente no mundo. Essas classes são denominadas *abertas*, pois é possível adicionar novos elementos a elas, não existe nenhuma limitação para a criação de novos nomes e adjetivos dentro da língua.

Por motivos de duração, uma vez que um estudo acerca das classes abertas necessitaria de mais tempo de investigação, esta pesquisa tem como objeto direto a negação sintática, ou seja, as lexias que assumem a posição sintática reservada para a negação dentro do português brasileiro – mais especificamente os advérbios de negação: *não, nunca, jamais, nem, sequer* e *tampouco*; os pronomes: *ninguém, nada, nenhum*; a preposição: *sem*; e as preposições acidentais *exclusive, excluso, exceto, fora, afora, salvo, menos, senão* e *tirante*. Cada uma das lexias será revista individualmente na seção 4.4. Entretanto, as primeiras seções deste capítulo

têm como objetivo realizar uma revisão das diferentes formas que a negação pode ocorrer no português brasileiro.

4.2 NEGAÇÃO SINTÁTICA

Como já dito, as lexias que fazem parte da **negação sintática** precisam ocorrer em uma posição específica para a realização da negação. Miotto (1992, p.99-104) traz que o *não*, o operador prototípico de negação, se comporta como um clítico no português e, assim como os outros clíticos, sempre ocorre anteposto ao verbo, ou ao elemento não verbal que ele irá negar (MIOTTO, 1998, p.110-116):

54. **Não** me deu um centavo para o tratamento. (**FK**)

55. Várias organizações **não** governamentais denunciaram a superlotação e desnutrição em centros de detenção venezuelanos. (**FK**)

Todavia, não é apenas o *não* que é capaz de assumir essa posição. Miotto (1992, p. 137-138) apresenta os outros advérbios de negação também são passíveis de ocorrerem em posição pré-verbal como acionadores negativos. Esse comportamento também é atestado em *corpus* nos exemplos:

56. **Nunca** me senti tão frágil e assustada na minha vida. (**FK**)

57. Já Palocci **jamais** teria tido qualquer compromisso com ela. (**FK**)

58. Eduardo Cunha **sequer** era filiado ao PMDB no momento da compra. (**FK**)

59. **Tampouco** na África ou na América Latina existe consenso sobre quais países deveriam representar essas regiões em uma sonhada atualização do conselho. (**FK**)

Em função dessa distribuição, é possível classificar a **negação sintática** em duas distintas ocorrências: (i) a *negação oracional*, quando o indicador negativo está associado a um verbo, e (ii) a *negação não-oracional*, quando o indicador negativo está associado a um elemento não verbal. Cada uma das ocorrências será melhor descrita nas próximas seções.

4.2.1 Negação Oracional

Na introdução do capítulo, utilizou-se o termo *negação padrão* para referenciar a ocorrência da negação nas línguas naturais com correspondência direta à negação lógica. Esse termo é originalmente proposto em Payne (1985), o autor a define como “é o tipo de negação que se pode aplicar nas sentenças mais mínimas e básicas”. Tais sentenças são

caracteristicamente orações principais e consistem um único predicado com a menor quantidade possível de sintagmas nominais e adverbiais²¹” (PAYNE, 1985 *apud* MIESTAMO, 2017, p. 408,. Tradução do autor).

Porém, a definição de Payne (1985) não especifica qual tipo de sentença se enquadra como “sentenças mais mínimas e básicas”. A fim de solucionar esse problema de ambiguidade de tipo de sentenças, Miestamo (2005) esboça uma definição mais precisa para a *negação padrão* em função das sentenças declarativas. Segundo ele, a negação padrão é uma negação “cuja função é modificar uma oração verbal declarativa principal que expressa uma proposição p de tal forma que a oração modificada expresse a proposição com o valor de verdade oposto a p, ou seja, $\neg p$ ” (MIESTAMO, 2005, p. 42. Tradução do autor). Dessa forma, apenas a negação presente em orações declarativas pode ser considerada equivalente à negação lógica.

Entretanto, a distinção entre a forma como a negação opera com os tipos de orações (declarativas, imperativas, exclamativas, interrogativas e vocativas) é debate de cunho totalmente semântico-cognitivo, o que está fora dos limites deste trabalho. O que se procura classificar aqui são as diferentes formas, em função da estrutura das sentenças, que a negação pode ocorrer dentro de um texto. Assim sendo, não se pretende utilizar do termo *oração padrão*, uma vez que ele está limitado apenas às orações declarativas.

Tomar-se-á aqui, em concordância com Miotto²² (1992; 1998), por *negação oracional* as ocorrências da negação em que ela tem escopo no verbo principal de uma oração, não se importando com o tipo específico de oração cuja negação está modificando. Dessa forma, a *negação padrão* é um tipo específico da *negação oracional*.

O escopo da negação é o elemento associado sintática e semanticamente à negação, ou seja, ele é o elemento no qual a negação atua. Existem duas possibilidades distintas para a ocorrência do escopo da negação: (i) escopo amplo, descrito nesta seção, que ocorre quando a negação se associa ao verbo principal de uma oração; e (ii) escopo estreito, descrito na próxima seção, que ocorre quando a negação se associa a um constituinte.

A forma mais prototípica de realizar a negação dentro do português é por meio da partícula *não* (NEVES, 2000, p.286). Todavia, o fenômeno da negação oracional não está

²¹ “that type of negation that can apply to the most minimal and basic sentences. Such sentences are characteristically main clauses, and consist of a single predicate with as few noun phrases and adverbial modifiers as possible”, no original.

²² Miotto (1992; 1998) adota a terminologia de “negação sentencial”, entretanto a sua proposta apresenta a mesma relação entre a negação e o verbo da oração, a “negação sentencial” é a forma de ocorrência da negação quando ela tem escopo sobre o verbo principal de uma oração. Assim, ela será tomada aqui como equivalente à “negação oracional”.

limitado apenas a este indicador negativo, podendo ser realizado pelos outros advérbios de negação (MIOTO, 1992, p.137-138) *nunca, jamais, sequer, tampouco* e *nem*:

60. **Nunca** me senti tão frágil e assustada na minha vida. (**FK**)
61. Já Palocci **jamais** teria tido qualquer compromisso com ela. (**FK**)
62. Eduardo Cunha **sequer** era filiado ao PMDB no momento da compra (**FK**)
63. **Tampouco** na África ou na América Latina existe consenso sobre quais países deveriam representar essas regiões em uma sonhada atualização do conselho. (**FK**)
64. **Nem** é preciso dizer que, depois dessa, o tatuador distraído levou mais umas bifas. (**FK**)

Essas ocorrências não estão limitadas apenas a realização da negação presente na oração principal, podendo também ocorrer em orações subordinadas:

65. Festeiros se esbaldavam até a invasão de uma turma que **não** estava na lista VIP (**FK**)
66. Lula e Dilma afirmam que **jamais** ouviram falar da tal conta. (**FK**)
67. Mas a presidente Dilma está tão fragilizada que **sequer** pode capitanear as medidas. (**FK**)
68. Chego a pensar que **tampouco** a classe operária sonhada por Marx era revolucionária (**FK**)
69. Espero que seja necessário menos sofrimento para que a Europa entenda que **nem tudo** deve ser tolerado (**FK**)

Um outro aspecto presente na negação oracional, dentro do português, é a Dupla Negação, mais precisamente, o português exprime 3 formas distintas de negação sentencial:

70. **Não** me deu um centavo para o tratamento (**FK**)
71. **Não** pode **não**, isto é crime (**FK**)
72. Pergunta ao Arnaldo Cezar Coelho: Arnaldo, isso pode? Ele vai dizer: pode **não**. (**FK**)

Enquanto em (70) tem-se a forma canônica da negação (Neg+Verbo), (71) e (71) exprimem formas distintas. A forma presente em (71) é chamada de Dupla Negação (MIOTO, 1992; NEVES, 2000; BARME, 2005), e a presente em (72) é chamada de negação pós-verbal (BARME, 2005).

Schwenter (2016) constata que as formas (71) e (72) se distinguem da forma canônica da negação por motivações pragmáticas de uso, sendo muito mais comuns em situações de

conversação do que em textos escritos. Segundo ele, as formas (71) e (72) estão vinculadas à negação de tópicos já inseridos dentro da conversação.

Uma última ocorrência da negação oracional que se deve ressaltar antes da finalização desta subseção é, como denominada por Horn (2001, capítulo 6), a *negação metalinguística*. Ela ocorre devido a uma *ambiguidade pragmática*, motivada pela capacidade da negação não necessariamente implicar, conversacionalmente²³, no oposto do predicado negado.

73. Eu **não** estou feliz com o plano, eu estou felicíssimo.

Em (73) temos exemplos de uma sentença que se utiliza da negação metalinguística. No exemplo, a negação da proposição da oração principal implica conversacionalmente na oposição direta de *feliz*, *triste*, ou seja, a maior parte das vezes em que alguém diz que *não está feliz*, a pessoa pretende dizer que *está triste*.

74. Eu não estou feliz \rightsquigarrow Eu estou triste²⁴

Entretanto, *não feliz*, não é necessariamente apenas *triste*, semanticamente *felicíssimo* também é um estado diferente de *feliz*, logo, *felicíssimo* implica, materialmente, *não feliz*.

75. Felicíssimo \rightarrow Não feliz

Assim, as interpretações de (74) e (75) permitem que em (73), a negação de *feliz* seja contraposta com *felicíssimo*. Entretanto, a existência das duas interpretações é necessária para que a *negação metalinguística* ocorra. Como Horn e Wansing (2020) salientam, caso não existam essas duas interpretações possíveis, a *negação metalinguística* falha, como é possível observar pela substituição de *não estou feliz* por *estou infeliz*, em (76):

76. (*) Eu estou **infeliz** com o plano, eu estou felicíssimo.

Como *infeliz* é por si só um item lexical que carrega o sentido de *não felicidade/tristeza*, ele remove a implicatura conversacional presente em (74), tornando a sentença em (76) inaceitável.

²³ Toma-se aqui por *implicatura conversacional* a mesma noção definida por Grice (1975).

²⁴ Optou-se por utilizar a notação “ \rightsquigarrow ” para a implicatura pragmática uma vez que ela é distinta da implicação material “ \rightarrow ”.

4.2.2 Negação Não-Oracional (Negação de Constituintes)

Em oposição à *negação oracional*, que tem escopo sobre o verbo principal da oração, pode-se classificar as ocorrências da negação que têm escopo em outros constituintes como *negação não-oracional* ou *negação de constituintes*. Essa forma de negação ocorre quando o escopo da negação fica limitado a apenas um constituinte da oração, não sendo aplicado a toda a proposição.

Podem-se destacar três formas distintas de *negação não-oracional* no português, sendo elas: (i) negação por meio da composição entre um elemento negativo e um constituinte; (ii) negação pronominal; e (iii) negação por derivação morfológica.

A primeira forma de negação ocorre de maneira similar à *negação oracional* anteposta ao verbo, mas o elemento negativo estará antecedendo a outro tipo de constituinte (MIOTO, 1998):

- 77. O **não** pagamento gerou queixas de consumidores em redes sociais e no Procon. (**FK**)
- 78. Paes em 2016 cometeu uma série de erros inteiramente **não** forçados. (**FK**)
- 79. [...] um quinto do resgate tinha sido fornecido **não** intencionalmente pela CIA [...]. (**FK**)

Nos exemplos, em (77) tem-se um caso de *negação não-oracional* com escopo sobre um nome, em (78) sobre um adjetivo e em (79) sobre um advérbio.

O segundo tipo de negação está limitado aos pronomes negativos indefinidos (MIESTAMO, 2017). Esses pronomes além de expressarem um referencial dêitico indefinido (pessoa, tempo e lugar), expressam a negação desses referenciais. O português expressa essa negação em três pronomes: *nada*, *ninguém* e *nenhum*, como nos exemplos:

- 80. **Nada** indica que quebrará. (**FK**)
- 81. **Ninguém** falta ao trabalho, explica, e a empresa economiza com vale-transporte. (**FK**)
- 82. **Nenhum** padre deve desfilar no sambódromo, apesar dos convites: eles dizem não querer desviar as atenções da homenageada. (**FK**)

Por último, a terceira forma da *negação não-oracional* ocorre por derivação morfológica. Em português, essa negação pode ocorrer por meio de prefixação por meio dos prefixos *a(n)-*, *des-* e *i(n)-*:

- 83. Normalidade x Anormalidade
- 84. Agradar x Desagradar
- 85. Lógico x Ilógico
- 86. Felizmente x Infelizmente

Essas formas de prefixação podem ocorrer nas quatro classes abertas, como nos substantivos, em (83); junto a verbos (84), a adjetivos, como em (85), e a advérbios, como em (86). Nota-se que a simples presença de um desses sufixos não é suficiente para exprimir a relação de oposição entre uma sentença positiva e outra negativa:

87. Ele pode **desfazer** as leis que ele mesmo fez e destituir instituições que se mostram corrompidas. (**FK**)

(?) *Ele pode **não fazer** as leis que ele mesmo fez e destituir instituições que se mostram corrompidas.*

88. Usando um feixe de laser para **inativar** temporariamente neurônios motores em camundongos geneticamente modificados (...) (**FK**)

(?) *Usando um feixe de laser para **não ativar** temporariamente neurônios motores em camundongos geneticamente modificados (...)*

Os sentidos dos verbos com *des-* e *in-*, presentes nos exemplos (87) e (88), não apresentam oposição negativa, mas sim o sentido de "contrariedade" (AZEREDO, 2008). Para que algo seja *inativado* ou *desfeito* é necessário que *ativo* e *feito* sejam ações prévias pressupostas ao enunciado.

Outro problema com verbos com morfologia negativa é a existência de verbos que não têm sua contraparte positiva, como *descascar* e *impedir*, que existem apenas com a morfologia negativa.

4.3 NEGAÇÃO LEXICAL

Como já exposto, toma-se aqui a divisão proposta por Muller (1991) para a categorização das lexias que podem exprimir a negação em duas categorias: (i) *Negação sintática*, discorrida na seção anterior, e (ii) a *negação lexical*.

A *negação lexical* não tem uma distribuição sintática estável, ou seja, os elementos que a realizam não ocorrem na posição sintática "padrão" para a negação. Esse tipo de negação ocorre nas categorias gramaticais abertas, como *verbos*, *nomes* e *adjetivos* (ADICIONAR ADVÉRBIOS):

87. O médico **negou** que esteja tratando mais alguém (**UPB**)

88. Falamos sobre a **falta** de importância dos estaduais. (**FK**)

89. Outro tema **ausente** foi a análise de como o país prostrou-se perante o resultado (**FK**)

Essas formas de ocorrência da negação ainda são pouco estudadas, tanto no português quanto em outras línguas, de maneira que apenas os *verbos* já foram discutidos como elementos capazes de realizar a negação nas línguas naturais²⁵.

Recentemente, Zeshan (2013) fez um estudo comparativo entre as línguas de sinais que revelou que existem domínios de significado que tendem a eximir o que ele chama de *negação irregular* (do inglês *irregular negative*). A autora define *negação irregular* como as formas de negação limitadas a um subconjunto de lexias não prototípicas para a negação, podendo variar entre algum tipo de derivação especial ou alguma forma supletiva.

Zeshan (2013) lista 6 domínios de significado que, nas línguas naturais, tendem a apresentar a *negação irregular* (tradução do autor):

- Cognição – por exemplo: “não+saber”, “não+entender”
- Atitudes emocionais – por exemplo: “não+querer”, “não+gostar”, “não+importar”
- Modais – por exemplo: “não+poder”, “não+precisar”, “não+dever”
- Posse/existencial – por exemplo: “não+ter”, “não+existir”, “não+obter”
- Tempo/aspecto – por exemplo: “*não+futuro*”, “não+passado”, “não+terminado”
- Julgamento de valor – por exemplo: “não+correto”, “não+possibilidade”, “não+suficiente”²⁶.

Para o português é possível observar a negação em formas derivadas com os morfemas citados na seção anterior, como também em formas supletivas de verbos, como por exemplo *excluir*, *faltar*, *negar*, *recusar* e *deixar*.

Ainda sobre verbos capazes de exprimir a negação, é possível encontrar para o espanhol descrições (HIDALGO, 2010; BOSQUE, 2019), com base em Mendikoetxea (1999), sobre *verbos de ausência* e *verbos de carência*. De maneira que os *verbos de ausência* representam predicados negativos estativos ou locativos:

89. Faltar (não+existir/não+estar presente)

90. Calar (não+falar)

E os *verbos de carência* exprimem predicados negativos de privação:

²⁵ No momento da escrita deste texto, não se encontrou nenhuma referência sobre o assunto nas bases Google Scholar (<https://scholar.google.com/>), SciELO (<https://www.scielo.br/>) e Portal de Periódico CAPES (<https://www-periodicos-capes-gov-br.ez1.periodicos.capes.gov.br/>).

²⁶ “Cognition: not know, not understand | emotional attitude: not want, not like, not care | modals: cannot, need not, must not | possession/existential: not have, not exist, not get | tense/aspect: will not, did not, not finished | evaluative judgement: not right, not possible, not enough”, no original.

91. Ignorar (não+saber)

92. Excluir (não+parte)

A proposta de Mendikoetxea (1999) também distingue entre as formas flexionais de realização da negação entre os verbos e as formas supletivas, de maneira que as formas flexionais também são classificadas em função dessas duas categorias.

Como dito anteriormente, não foram encontrados trabalhos acerca das formas nominais e adjetivais que também expressam a negação nas línguas naturais, tanto em português quanto em línguas estrangeiras²⁷. Entretanto, o mesmo comportamento com formas supletivas pode ser observado nos substantivos e adjetivos. Existem lexias que não indicam de forma nenhuma negação em seu significado, mas ainda assim a expressam, com nos exemplos:

93. A **ausência** de Lady Gaga, por motivos de saúde. (*FK*)

94. A **falta** de empregos é o maior desafio. (*FK*)

95. Outra **ausente** foi a deputada Soraya Santos. (*FK*)

Todas as sentenças, ainda que não expressem nenhuma característica gramatical, podem ser reescritas em função da negação:

96. A **não presença** de Lady Gaga, por motivos de saúde.

97. A **não existência** de empregos é o maior desafio.

98. Outra **não presente** foi a deputada Soraya Santos

Grande parte dos substantivos negativos no português ocorrem como nomes predicativos. Um nome predicativo é um “substantivo que, funcionando como núcleo da predicação, seleciona seus argumentos [...]” (MARTINEZ, 2019, p. 27-28). A definição de Martinez (2019) de nome predicativo é baseada em Gross (1981), na qual o autor define que um nome predicativo é um substantivo que requer, assim como um verbo, um argumento como complemento.

Como é possível observar nos exemplos (93) e (94), os substantivos destacados operam e modificam um outro constituinte: “A ausência **de Lady Gaga**”, em (93); e “A falta **de empregos**”, em (94). Tal especificidade é completamente compatível com a necessidade da

²⁷ No momento da escrita deste texto, não foi encontrado nenhum texto sobre o assunto nas bases Google Scholar (<https://scholar.google.com/>), SciELO (<https://www.scielo.br/>), Portal de Periódico CAPES (<https://www-periodicos-capes-gov-br.ez1.periodicos.capes.gov.br/>) e ACL Anthology (<https://aclanthology.org/>)

negação, que também opera sobre um elemento, seu escopo, ainda mais se comparado à negação *padrão* que também opera sobre predicados.

Por fim, assim como para os nomes capazes de exprimir a negação, não foram encontrados trabalhos específicos sobre esta categoria, tanto em português quanto em línguas estrangeiras²⁸. Os únicos trabalhos encontrados que relacionam adjetivos com a negação (KENNEDY, 2001; SASSOON, 2010, KOOSTER; 2019) são do inglês e têm como principal foco pares opositivos entre adjetivos: *bom x mal, claro x escuro, etc...*

As formas subjetivas de adjetivos que exprimem a negação tendem a ser formas derivadas morfologicamente dos verbos e nomes que exprimem a negação, como é possível observar em *ausente* (derivada de *ausência*), em (95).

Não se discorda que a descrição dos itens lexicais do português que acionem a negação seja de extrema relevância para a retenção de informação sobre a negação em um *corpus*, todavia, uma descrição desse tipo requer uma investigação aprofundada nas categorias lexicais (*substantivos, adjetivos, advérbios e verbos*), as maiores categorias de palavras. Dessa forma, em função do escopo e da duração desta pesquisa, tais formas de negação não serão aqui abordadas.

Assim, com as informações levantadas até esta seção, pode-se organizar, as formas de ocorrência da negação no português em função da seguinte tipologia:

Quadro 5 – Distribuição dos tipos de negação

Escopo	Tipo	Acionadores	Exemplo
Negação Oracional	Sintática	Advérbios de negação	Não existe outro jeito (<i>FK</i>)
	Lexical	Verbos negativos (<i>negar, faltar, recusar, etc.</i>) nomes predicativos negativos (<i>ausência, falta, carência, etc.</i>)	O médico negou que esteja tratando mais alguém (<i>UPB</i>)
Negação Não-Oracional	Sintática	Advérbios de negação Pronomes Negativos Flexão morfológica	Paes em 2016 cometeu uma série de erros inteiramente não forçados. (<i>FK</i>)
	Lexical	Adjetivos negativos (<i>ausente, recusado, negativo, etc.</i>). Nomes predicativos	Falamos sobre a falta de importância dos estaduais. (<i>FK</i>)

Fonte: O autor

²⁸ No momento da escrita deste texto, não foi encontrado nenhum texto sobre o assunto nas bases Google Scholar (<https://scholar.google.com/>), SciELO (<https://www.scielo.br/>), Portal de Periódico CAPES (<https://www-periodicos-capes-gov-br.ez1.periodicos.capes.gov.br/>) e ACL Anthology (<https://aclanthology.org/>)

4.4 CONCORDÂNCIA NEGATIVA

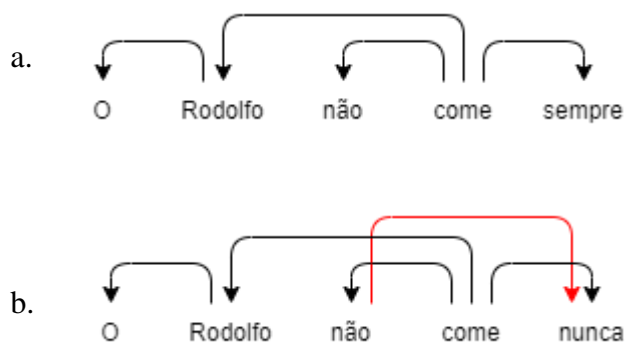
Um outro aspecto sintático importante da negação dentro do português é a concordância negativa (MIOTO, 1992; NEVES, 2000). A contrário de línguas germânicas que não aceitam a Dupla Negação, o português não apenas o faz, mas também exprime concordância em função da negação dentro dos elementos de um enunciado, como é possível observar em exemplo (99):

99.

- a. **Não** existe **ninguém** que Thoreau conheça tão bem como ele próprio. (**FK**)
- b. (?) *Não existe alguém que Thoreau conheça tão bem como ele próprio.*

Mesmo que a sentença em (99b) seja compreensível, ela é menos usual que a forma (99a). Isso ocorre devido ao fato de que o português concorda a polaridade do pronome em posição de objeto, assim, gerando preferência pela forma negativa (*ninguém*) no lugar da forma apenas indefinida (*alguém*)²⁹. Além da concordância com o pronome em posição de objeto, a concordância negativa também afeta os advérbios de negação, fazendo com que sentenças como “O Rodolfo não come sempre”, assim como (99b), sejam coesas, mas despreferidas quando comparadas a sua forma com concordância “O Rodolfo não come nunca”.

100.



Um problema enfrentado pela concordância negativa é uma possível quebra na projetividade das sentenças, uma vez que a concordância entre dois elementos dentro de um enunciado representa dependência entre o elemento que aciona a concordância e o elemento

²⁹ Não se discorda que existam contextos possíveis em que o *não* possa ocorrer junto de *alguém* ignorando a concordância negativa (Cf. *O novo ministro da Fazenda não é alguém que chega agora ao governo do PT*, FK), entretanto como dito, a forma com a concordância é a predileta.

que concorda (MEL'ČUK, 1988; HUDSON, 1990; POLGUÈRE e MEL'ČUK, 2009; DE MARNEFFE e NIVRE, 2019).

A relação, realçada em vermelho, entre *não* e *nunca* em (100b) não é só não projetiva, isto é, ela cruza os um ou mais arcos dentro da estrutura da árvore, como também viola o princípio de apenas um governante por dependente, *nunca* seria tanto dependente de *não* como de *comer*. Uma possível solução seria utilizar teorias multiestrato para justificar a projetividade apenas em níveis mais superficiais e separar as dependências entre *não e nunca* e *comer e nunca* em diferentes níveis de análise. Todavia, as análises aqui são voltadas para o modelo da UD.

A UD não suporta a relação realçada em vermelho, ou seja, não há em suas diretrizes nenhuma relação de dependência capaz de lidar com esse fenômeno. Porém, para tratar esse fenômeno, acredita-se aqui, em concordância com Rademaker *et al.* (2017), que a melhor forma de manter a informação da Dupla Negação nas sentenças, sem perder a estrutura sintática proposta pela UD, é realizar a marcação de traço de polaridade em todas as palavras negativas do enunciado:

“Para aqueles que estão interessados na dupla negação no português, a melhor forma de encontrá-la, na versão atual do UD_Portuguese corpus, será uma busca no atributo de polaridade (‘Polarity=Neg’) expressa pelas palavras que circundam os verbos. Nós esperamos que o uso consistente do atributo de polaridade em advérbios, pronomes e conjunções, como o *nem*, e outros possam prover uma análise completa deste fenômeno sem perder a análise sintática de superfície fornecida pelas relações da UD”³⁰ (RADEMAKER *et al.*, 2017. Tradução do Autor)

Assim, pretende-se aqui, com o uso do atributo de polaridade³¹, como descrita por Rademaker *et al.* (2017), e com o atributo de pronomes negativos³² (PronType=Neg) inserir informação o suficiente nos *corpora* para que seja possível identificar o fenômeno da Dupla Negação.

³⁰ “For those interested in double negations in Portuguese, the best way to look for them in the current UD_Portuguese corpus will be to check for the polarity feature (‘Polarity=Neg’) expressed in words that surround the verbs. We expect that the consistent use of the polarity feature in adverbs, pronouns, conjunctions, as *nem* (‘neither’), and others will provide us with a full analysis of this phenomenon without losing the surface syntactic analysis provided by the UD relations”, no original.

³¹ <https://universaldependencies.org/u/feat/Polarity.html#Neg>

³² <https://universaldependencies.org/u/feat/PronType.html#Neg>

4.5 LEXIAS QUE REALIZAM A NEGAÇÃO SINTÁTICA

Como já dito, além da forma tradicional de realizar a negação no português por meio do *não*, existem outras lexias que também são capazes de realizar a negação. Pretende-se aqui realizar uma revisão bibliográfica acerca dos itens: *não*, *nunca*, *jamais*, *sequer*, *tampouco*, *ninguém*, *nada*, *nenhum*, *sem* e *nem*; como também das preposições acidentais *exclusive*, *excluso*, *exceto*, *fora*, *afora*, *salvo*, *menos*, *senão* e *tirante*.

As lexias aqui analisadas não apresentam um comportamento único, dessa forma, elas estão organizadas em função de sua semelhança. As diferentes acepções trazidas pela literatura foram, quando possível, validadas no *corpus Folha kaggle* e estão disponíveis como exemplos. As acepções que não foram encontradas no *corpus* seguem com exemplos da obra que as descrevem.

4.4.1 Nunca e Jamais

Partindo-se de *nunca* e *jamais*, ambos são capazes de realizar a negação oracional, assim como o *não*, (MIOTO, 1992, p.204-212), eles também exprimem características aspectuais e temporais ao enunciado (NEVES, 2000, p.289-291):

101. **Nunca** me senti tão frágil e assustada na minha vida. (FK)
Não me senti tão frágil e assustada em nenhum momento na minha vida.
102. Já Palocci **jamais** teria tido qualquer compromisso com ela. (FK)
Já Palocci não teria tido qualquer compromisso com ela em nenhum momento

Note-se que, como consta em Lima (2011), as paráfrases de (101) e (102) são formas comutativas, ou seja, *nunca* e *jamais* podem ser substituídos por expressões adverbiais negativas de tempo, assim como as expressões podem ser substituídas pelos advérbios. Essas formas *tampouco* se limitam a ocorrer apenas na posição da negação pré-verbal, mas podem ocorrer junto ao *não* em um enunciado por meio de concordância negativa:

103. Durante mais ou menos um ano a hora mais feliz do meu dia era quando eu deitava na cama e fechava os olhos, com a esperança de que aquela noite de sono **não** terminasse **nunca**. (FK)
104. Em muitos deles, isso é suficiente para garantir ao ministro que pede a vista que ele **não** será derrotado **jamais**. (FK)

Todavia, mesmo sendo advérbios, tanto o *nunca* quanto o *jamais* não possuem liberdade em sua distribuição na oração quando acompanhados do *não*. Eles não podem anteceder a negação *não* dentro de um enunciado, sendo limitados a ocorrer após o sintagma verbal:

- 105.
- a. **Nunca** o João visitou a Bahia
 - b. O João **nunca** visitou a Bahia
 - c. O João **não** visitou **nunca** a Bahia
 - d. O João **não** visitou a Bahia **nunca**

(MIOTO, 1992, p.204)

Assim, tanto o *nunca* quanto o *jamais*, mesmo tendo um comportamento adverbial, são limitados em sua distribuição pela concordância negativa.

4.4.2 Sequer

Apesar das gramáticas tradicionais (BECHARA, 2009; LIMA, 2011; CUNHA e CINTRA, 2016) e dicionários (HOUAISS, 2009; MICHAELIS, 2022) apresentarem *sequer* como uma acepção ou sinônimo de *ainda*, esse advérbio sempre ocorre com polaridade negativa (DE SOUZA, 2008). De fato, *sequer* é capaz de realizar a negação oracional, estando presente sozinho em um enunciado, assim como os outros advérbios de negação:

106. Eduardo Cunha **sequer** era filiado ao PMDB no momento da compra. (**FK**)
Eduardo Cunha não era filiado ao PMDB no momento da compra
107. Uma situação estranha, pois ela **sequer** teve a chance de mostrar o resultado do seu trabalho em uma competição oficial. (**FK**)
Uma situação estranha, pois ela não teve a chance de mostrar o resultado do seu trabalho em uma competição oficial

As paráfrases dos exemplos (106) e (107), mesmo elidindo a interpretação escalar dos enunciados, ainda carregam a informação negativa das ocorrências originais, demonstrando o funcionamento de *sequer* em posição de negação sentencial.

Assim como os outros advérbios de negação, o *sequer* também ocorre em contextos de concordância negativa, junto com outros elementos em posição de negação oracional:

- 108.
- a. **Não** é o momento **sequer** de pensar isso. (**FK**)
 - b. **Nenhum** valor **sequer** foi retirado dos cofres públicos da cidade. (**FK**)
 - c. **Nem** Neymar chega **sequer** perto das indisciplinas cometidas pelos três. (**FK**)

Além disso, *sequer* também ocorre como elemento coesivo entre sentenças, introduzindo um enunciado negativo:

109. Desvinculados de qualquer constrangimento eleitoral, o governo e seus apoiadores estariam livres para adotar as medidas amargas. **Sequer** seriam medidas impopulares; (*FK*)
110. Não existia massa de trigo duro por aqui, nem arroz de risoto. **Sequer** azeite de oliva extravirgem [...] (*FK*)

Todavia, considerando que este trabalho está delimitado dentro da análise sintática dos enunciados, a relação de coesão entre as sentenças do texto não está dentro de seu escopo descritivo. As ocorrências de *sequer* como nos exemplos (109) e (110) serão tomadas como ocorrências de em posição de negação oracional.

Uma última ocorrência possível do *sequer* é como reforço negativo, podendo ocorrer logo em seguida de um elemento para enfatizar a negação presente no enunciado:

111. O Canadá **nem sequer** tem um sistema educacional nacional, pois a organização é baseada em províncias autônomas. (*FK*)
112. **Nunca sequer** falamos sobre pesca. (*FK*)
113. O outro "volpista" acredita que, por anos, Marieta esteja com uma versão falsificada **sem sequer** desconfiar de isso. (*FK*)
114. Nos EUA, **ninguém sequer** falou de isso. (*FK*)

Dessa forma, o *sequer* é capaz de realizar a negação sentencial, tem sua distribuição como advérbio limitada pela concordância negativa e funciona como reforço negativo.

4.4.3 Tampouco

Para além das gramáticas e dicionários, o *tampouco* é pouco explorado no PB³³, sua única acepção nas descrições tradicionais da língua é como “uso para reforçar uma negação”, como em “*não gosta de jabá, tampouco de jerimum*” (HOUAISS, 2009). Entretanto, o exemplo trazido no Houaiss não é apenas um “reforço da negação”, como descrito. *Tampouco* também

³³ No momento da escrita deste texto, não se encontrou nenhuma referência sobre o assunto nas bases Google Scholar (<https://scholar.google.com/>), SciELO (<https://www.scielo.br/>) e Portal de Periódico CAPES (<https://www-periodicos-capes-gov-br.ez1.periodicos.capes.gov.br/>).

está funcionando como uma conjunção no exemplo. Como é verificável pela sua reestruturação em (78):

115.

- a. **Não** gosta de jabá, **tampouco** de jerimum (HOUAISS, 2009)
- b. **Não** gosta de jabá **e também não gosta de** jerimum

Esse comportamento também é atestado em *corpus* em exemplos como:

116. **Não** se pode, porém, minimizar o peso simbólico desta viagem, **tampouco** o poder das imagens que produzirá. (*FK*)

117. A perspectiva da delação premiada já **não** intimida ninguém, **tampouco** convence a opinião pública. (*FK*)

118. Minha audiência **nunca** será imediata, **tampouco** grandiosa enquanto escrever for a minha teta de fora (...) (*FK*)

119. Minha audiência **nunca** será imediata e **nunca será grandiosa** enquanto escrever for a minha teta de fora (...) (*FK*)

Também podendo ocorrer precedido de um elemento de coordenação:

120. Segundo o TCU, porém, Fontes não havia apresentado o produto final **e tampouco** prestado contas dos R\$ 4,6 milhões que obteve. (*FK*)

121. Hoje vocês se aguentam – como aqueles casais velhos que não brigam, **mas tampouco** se falam. (*FK*)

122. Para Naldinho, a música não é violenta **nem tampouco** sexual. (*FK*)

No exemplo (122), mesmo que *tampouco*, esteja ocorrendo junto ao *nem* como elemento coordenativo, o *nem* também tem polaridade negativa, funcionando como um reforço negativo.

E por fim, *tampouco* é capaz de aparecer em uma sentença sem a necessidade de ser precedido por outro elemento negativo dentro da mesma sentença, realizando a negação oracional:

123. Segundo Libman, o projeto, apresentado na última edição da feira de tecnologia CES, em Las Vegas (EUA), **ainda não tem data para se tornar realidade**. (...). **Tampouco** há ideia de preço definido e tamanhos que serão disponibilizados a princípio, a tecnologia foi pensada para malas de mão, mas ela poderá ser implantada em outras bagagens e objetos, como carrinhos de compras em supermercados. (*FK*)

124. Os EUA manifestam abertamente seu apoio à Índia –e assim esperam relações privilegiadas com Nova Déli, mas o fazem sabendo que a China **não concorda de fato com o ingresso de um outro país asiático no quadro de membros permanentes**. (...) **Tampouco** na África ou na América Latina existe consenso sobre quais países deveriam representar essas regiões em uma sonhada atualização do conselho. (*FK*)

É válido salientar que, nos exemplos (123) e (124), *tampouco* ocorre em concordância negativa com outra sentença presente no texto, agindo como um elemento coesivo, da mesma forma como o *sequer*. Em vista disso, essas ocorrências também serão tomadas como ocorrências de *tampouco* em posição de negação oracional.

Assim, o *tampouco*, como os outros advérbios de negação, é capaz de realizar a negação sentencial e tem sua distribuição como advérbio limitado pela concordância negativa. Com adicional, ele também ocorrer como reforço negativo, assim como o *sequer*, e é também é capaz de ocorrer com a função de conjunção.

4.4.4 Nada, Ninguém e Nenhum

Não existe na tradição normativa do português a categoria de *pronomes negativos*, todos os pronomes que expressam a negação são categorizados com *pronomes indefinidos* (BECHARA, 2009; LIMA, 2011; CUNHA e CINTRA, 2016). O *nada*, o *ninguém* e o *nenhum* têm a propriedade de serem os únicos pronomes que alteram a polaridade da oração a qual estão vinculados.

125. **Nada** indica que quebrará. (*FK*)
 126. **Ninguém** falta ao trabalho, explica, e a empresa economiza com vale-transporte (*FK*)
 127. **Nenhum** dos serviços cobra taxa adicional dos clientes. (*FK*)

A única forma de reescrever os exemplos de (125), (126) e (127) com seus opostos não negativos seria adicionando algum outro elemento negativo na sentença para manter a mesma polaridade:

128. (?) **Não há algo** que indica que quebrará
 129. **Não há alguém que/Nenhuma pessoa** falta ao trabalho, explica, e a empresa economiza com vale-transporte
 130. **Todos** os serviços **não** cobram taxa adicional dos clientes.

Ainda assim, a reescrita em (128), sem a forma negativa *nada*, é estranha, pois, por concordância negativa, existe uma predileção em concordar o pronome com o elemento

negativo da frase:

131. Não há **nada** que indica que quebrará

Além do *nada*, os outros dois pronomes também são predileções em contexto de concordância negativa, tornando os mesmos enunciados com suas formas positivas menos usuais, como em (128):

132. Os salários **não** dão para **nada**. (*FK*)

133. Em 2011, Angelici **não** era **ninguém** e só venceu graças aos votos de Macri.
(*FK*)

134. A Folha **não** tinha **nenhum** indício de que teria havido tráfico de influência (...)
(*FK*)

Ainda com relação entre a distinção entre o *nada* e os outros pronomes, ele é o único capaz de ocorrer como advérbio dentro de uma oração sem exprimir funcionamento pronominal:

135. Não é **nada** extraordinário em relação ao que acontece no cotidiano das varas criminais. (*FK*)

No exemplo, *nada* pode ser substituído por qualquer advérbio de intensidade, o que não é possível em (125).

Um último comportamento assimétrico entre os pronomes é a capacidade de *nenhum* flexionar em gênero (*nenhum*, *nenhuma*) e também sua capacidade de ocorrer como determinante. Com relação ao gênero, o pronome pode ocorrer tanto em contexto feminino, quanto masculino:

136. **Nenhuma** de nós sabia bem o que isso significava. (*FK*)

137. **Nenhum** dos homens mostrou clemência. (*FK*)

enquanto os dois outros pronomes são invariáveis, sempre ocorrem com a mesma forma.

Com relação à capacidade de ocorrer como determinante, *nenhum(a)* ocorre tanto nas posições argumentais de uma oração (*sujeito* e *objeto*), como nos exemplos (138) e (139), assim como ele pode modificar, em posição de determinante um *nome*:

138. **Nenhum** grupo havia assumido a autoria do atentado até as 18h. (*FK*)

139. **Nenhuma** legenda obteve êxito parecido. (*FK*)

A distinção entre as duas ocorrências se dá pela presença da preposição *de* entre o *nenhum* e o *nome* que o segue, de maneira que é possível, na maioria das vezes, converter uma forma em outra:

140. **Nenhum** dos homens mostrou clemência. (**FK**)
Nenhum homem mostrou clemência.
141. **Nenhuma** legenda obteve êxito parecido. (**FK**)
Nenhuma das legendas obteve êxito parecido.

Assim, o pronome indefinido *ninguém* carrega além da negação a informação de pessoa (nenhuma pessoa), o pronome indefinido *nada* é capaz de negar um complemento ou adjunto negativo (*nenhuma coisa*) e o pronome indefinido *nenhum* é capaz de quantificar negativamente qualquer classe de elementos (*coisas, animais, pessoa, etc.*). (NEVES, 2000, p.290-291)

4.4.5 Sem

A preposição *sem* inicia sintagmas com o valor negativo em função da exclusão do elemento presente no sintagma (NEVES, 2000; p.291). Semanticamente, ela se comporta diretamente como a versão de polaridade negativa do item *com*:

142. O longa é uma obra **sem** firulas vistosas para entreter o espectador. (**FK**)
143. Ela venceu **sem** muitas dificuldades, ao aplicar um waza-ari e administrar o restante da luta. (**FK**)

O *sem* é limitado a realizar a negação não oracional, uma vez que o sintagma introduzido por ele, como já dito, é "excluído" da oração nuclear, mesmo que ele inicie o enunciado, não há negação total do predicado presente no enunciado:

144. **Sem** condição de contratar assessoria jurídica, ele conta que fez um mal negócio em um contrato e acabou perdendo R\$ 400 mil. (**FK**)

No exemplo, *sem* tem escopo apenas em *condição de contratar assessoria jurídica*, não interferindo no conteúdo da oração nuclear “*ele conta que fez mal negócio em um contrato e acabou perdendo R\$ 400 mil*”.

Por fim, o *sem* também aciona a concordância negativa, todavia, por não conseguir realizar a negação sentencial, a concordância fica limitada apenas ao sintagma em que *sem* tem escopo:

145. Dilma, por sua vez, chegou à Presidência da República **sem jamais** ter vivido o cotidiano de um Parlamento. (**FK**)

146. Lemmy foi um personagem único: cresceu nos escombros da Segunda Guerra, foi abandonado pelo pai e vagou pela Europa em estilo hippie, **sem nada** na cabeça ou nos bolsos. (**FK**)

147. Para os acadêmicos de plantão, o grafite deve permanecer na berlinda, como arte de periferia, **sem nenhum** reconhecimento. (**FK**)

Assim, o *sem* é limitado pelo sintagma em que opera, fazendo com que tanto seu escopo quanto a necessidade de concordância negativa sejam restritos.

4.4.6 Nem

O elemento *nem* é descrito, não apenas pelas gramáticas normativas (BECHARA, 2009; LIMA, 2011; CUNHA e CINTRA, 2016), como também em trabalhos descritivos (MATOS OLÍMPIO, 2005; CANCEIRO *et al.*, 2018; CAVAGUTI, 2018; MIRANDA JUNIOR, 2022) como um elemento negativo. As formas mais presentes em todas as descrições são seu comportamento como advérbio de negação em posição prototípica:

148. **Nem** é preciso dizer que, depois dessa, o tatuador distraído levou mais umas bifas. (**FK**)

E sua capacidade de coordenação em contexto de concordância negativa, sendo equivalente a um *e+não*:

149. O mais provável é que esse ancestral **não** tivesse veneno potente **nem** fosse capaz de esmagar suas vítimas, como fazem as sucuris e jiboias. (**FK**)

Também podendo ocorrer em concordância negativa com ele mesmo:

150. **Nem** a Apple **nem** as autoridades regulatórias explicaram (...) a medida. (**FK**)

Uma outra ocorrência do *nem* com polaridade negativa, mas pouco descrita, apenas encontrada em Cavaguti (2018) e em Miranda Jr (2022), é seu comportamento como acionador condicional concessivo em oração condicional:

151. **Nem** que Jesus Cristo desça à Terra e diga ser o Sport o campeão brasileiro de 1987, o Flamengo deixará de sê-lo. (**FK**)

E em oração condicional reduzida de gerúndio:

152. **Nem** quintuplicando o número de fiscais vai resolver o problema. (**FK**)

Em ambos os casos, *nem* não está realizando a negação da oração condicional, mas sim realizando a negação da oração nuclear em função de uma condição presente na oração subordinada:

153. **Mesmo** que Jesus Cristo desça à Terra e diga ser o Sport o campeão brasileiro de 1987, o Flamengo **não** deixará de sê-lo.
154. **Mesmo** quintuplicando o número de fiscais **não** vai resolver o problema.

A forma que opõe esse uso do *nem*, como é visível nos exemplos acima, é elemento *mesmo*. Sua presença faz com que a negação realizada pelo *nem* necessite ser demonstrada por outro elemento, neste caso o *não*, para que as sentenças (153) e (154) tenham, respectivamente, a mesma interpretação de (151) e (152).

Por fim, o *nem* não ocorre apenas com a interpretação de polaridade negativa, ele também faz parte da expressão comparativa cristalizada *que nem* (THOMPSON *et al.*, 2012), em que o elemento não modifica negativamente a sentença, mas tem significado similar a *como/igual a*:

155. Cheio de memórias, cheio de componentes, **que nem** o carro. (*FK*)
Cheio de memórias, cheio de componentes, como/igual a o carro.

4.5 PREPOSIÇÕES ACIDENTAIS

Todas as lexias descritas até aqui exprimem comportamento exclusivamente gramatical, ou seja, elas são termos que apenas se relacionam com o sentido dentro do sistema linguístico. Todavia, existem alguns elementos do léxico que são capazes de ocorrer dentro das classes abertas, mas que, em contextos específicos, são capazes de ocorrerem como preposições. Essas ocorrências específicas recebem o nome de *preposições acidentais* (NEVES, 2000, p.732).

Dentro desta classe de preposições, Neves (2000, p.735) lista uma sequência de elementos que, assim como o *sem*, estabelecem relação negativa de exclusão introduzindo um sintagma nominal ou uma oração infinitiva. Os elementos são: *exclusive, exclusivo, exceto, fora, afora, salvo, menos, senão* e *tirante*. Serão considerados aqui apenas as ocorrências dessas palavras como preposições, ou seja, suas ocorrências que introduzem uma relação semântica adverbial, mais especificamente, a relação adverbial de exceção (CASTILHO, 2014, p.575), ou seja, quando elas introduzem um sintagma nominal ou oração finita que é *excluído* da oração principal, assim como é observável nos exemplos extraídos do *corpus*:

156. Trabalhadores sem carteira de trabalho, **exclusivo** domésticos. (*FK*)
157. Acho que todos valem a leitura, **exceto** Reinaldo Azevedo. (*FK*)
158. É difícil encontrar no Brasil, **fora** a escravidão, um fenômeno social tão destrutivo quanto a motocicleta. (*FK*)
159. A operação não terá fim, **afora** um golpe ou acordão políticos ou erro, dada a extensão da pirataria contra o Estado. (*FK*)
160. A crise não irá acabar de uma vez, **salvo** eventos excepcionais. (*FK*)
161. Aqui você encontra comida árabe, americana, mexicana, **menos** a comida regional. (*FK*)
162. São quatro sócios, três deles sem ligações anteriores com a gastronomia, **senão** o gosto pela comida. (*FK*)
163. No episódio da Chapecoense, **tirante** meia dúzia de imbecis, o povo (de aqui e do mundo) tem tirado a razão de Néilson. (*FK*)

De todas as formas listadas por Neves (2000), apenas *excluso*, não obteve nenhuma ocorrência no *corpus* de validação, Folha kaggle. Todavia, esse item será mantido na categoria uma vez que, pela introspecção do autor, é possível substituir qualquer uma das proposições dos exemplos (156) a (163) mantendo a mesma interpretação.

164. Trabalhadores sem carteira de trabalho, **excluso** domésticos.
165. Acho que todos valem a leitura, **excluso** Reinaldo Azevedo.

Com o levantamento neste capítulo, de maneira reduzida, pode-se construir uma relação entre os elementos negativos aqui analisados e suas propriedades:

Quadro 6 – Comportamento dos Indicadores Negativos

Elemento	Negação Oracional	Negação Não-Oracional	Dupla Negação	Aciona Concordância	Concorda	Comportamento Adverbial	Comportamento Pronominal	Conjunção	Preposição	Reforço da Negação
não	X	X	X	X						
nunca	X				X					
jamais	X				X					
sequer	X				X					X
tampouco	X	X			X			X		X
nada	X	X			X	X	X			
ninguém	X	X			X		X			
nenhum	X	X			X		X			
nem	X			X	X			X		
sem				X					X	
exclusive									X	
excluso									X	
exceto									X	
fora									X	
afora									X	
salvo									X	
menos									X	
senão									X	
tirante									X	

Fonte – O Autor

No quadro 6, tem-se a distribuição de cada um dos indicadores negativos em função das características: capacidade de realizar negação oracional, capacidade de realizar negação não-oracional, capacidade de realizar dupla negação, acionador de concordância negativa, concordar negativamente, exprime comportamento adverbial, exprime comportamento pronominal, capacidade de ocorrer com conjunção, capacidade de ocorrer como preposição e reforço da negação. Essas características servirão como base para as análises, presentes no próximo capítulo, dos indicadores negativos dentro dos *corpora UD Portuguese Bosque, Portinari-base* e da ferramenta *PortiLexicon-UD*.

5. ANÁLISES E RESULTADOS

Neste capítulo serão apresentados os resultados e as análises da anotação das lexias dentro dos *corpora* aqui analisados; *UD Portuguese Bosque*, *Porttinari-base* e a ferramenta *PortiLexicon-UD*; assim como possíveis melhorias para a anotação em função das diretrizes da UD.

Deve-se salientar que todas as propostas aqui levantadas não são soluções únicas para a demarcação dos indicadores negativos no português. A UD é aberta em relação ao debate das anotações, de maneira que é possível acompanhar outras propostas e soluções por meio das *Issues* no Github³⁴.

Ademais, existem diversas anotações realizadas (PAIVA e REAL, 2016; REAL, 2018; SOUZA *et al.*, 2021; DA SILVA *et al.*, 2021; DI FELIPPO *et al.*, 2021; CONEGLIAN *et al.*, 2022; SOUZA e FREITAS, 2022) e algumas ainda em desenvolvimento para a descrição do português em função da UD, de maneira que nossas propostas aqui são apenas um recorte das diversas possibilidades de anotação.

5.1 NUNCA E JAMAIS

Em concordância com a revisão realizada (ver subseção 4.4.1) o comportamento do *nunca* e do *jamaís* é inteiramente adverbial com distribuições bem definidas. Nota-se que *nunca* é mais produtivo que *jamaís* com 62 ocorrências no UD Portuguese Bosque (RADEMAKER, 2017) e 53 ocorrências no *Porttinari-base* contra 7 e 11 ocorrências respectivamente. Ambos os advérbios constam com apenas uma entrada cada no *PortiLexicon-UD* sem nenhuma especificação de atributo morfológico.

Por sua baixa ocorrência, o *jamaís* não foi encontrado nos *corpora* revisados em concordância negativa, ocorrendo apenas na posição prototípica da negação sentencial, anteposto ao verbo, como advérbio:

³⁴ <https://github.com/universaldependencies/docs/issues>

Quadro 7 – Ocorrências de JAMAIS

		UD Portuguese Bosque	Porttinari-base	PortiLexicon-UD
jamais-1	PoS	ADV	ADV	ADV
	atributo	—	—	—
	dprel	advmod	—	—
	ocorrência	7	11	—

Fonte – O autor

Como nos exemplos:

166. Sei que **jamais** terei uma chance de trabalhar tão calmamente de novo. (*UPB*)
167. Prometo que aprenderei com isso e **jamais** tomarei qualquer atitude assim mais uma vez. (*PTTB*)

Em contrapartida, o *nunca* apresenta um total de três ocorrências em concordância negativa dentro do *corpus UD Portuguese Bosque*, duas junto ao *não* e uma junto ao *sem*:

Quadro 8 – Ocorrências de NUNCA

		UD Portuguese Bosque	Porttinari-base	PortiLexicon-UD
nunca-1	PoS	ADV	ADV	ADV
	atributo	—	—	—
	dprel	advmod	—	—
	ocorrência	62	53	—

Fonte – O autor

168. Está longe de ser o melhor de Tchaikovsky, mas o diabolismo da parte solista **não** deixa **nunca** de impressionar. (*UPB*)
169. Não há fotografias sem pessoas, ou sem os seus vestígios, o que **não** deixa **nunca** de remeter o trabalho deste fotógrafo (...) (*UPB*)
170. João Carlos se cansa com facilidade, pois também baixou os ombros e inclinou o queixo para o peito, **sem nunca** olhar para os lados das grandes janelas do tribunal. (*UPB*)

E com duas ocorrências junto a *sem* no *corpus Porttinari-base*:

171. Endiabrados ou não, baladeiros nova-iorquinos passaram anos dançando **sem nunca** saber da regra. (*PTTB*)
172. Poucos elementos são transmutados continuamente, mas **sem nunca** deixarem de ser eles mesmos. (*PTTB*)

Como já exposto, tanto o *jamais* quanto o *nunca* sempre ocorrem como advérbios, a única diferença entre a ocorrência com concordância negativa e a ocorrência sem ela é a presença de um item de polaridade negativa anteposto aos advérbios, tornando a distinção de notação entre as duas ocorrências desnecessária. Assim, há apenas a necessidade da adição do atributo de polaridade negativa nas ocorrências de ambos os *corpora*.

5.2 SEQUER

Todas as ocorrências de *sequer* apresentam polaridade negativa, como advérbio de negação, variando entre prototípica de negação oracional, concordância negativa e reforço da negação.

O quadro 9 traz as diferentes anotações ocorridas nos *corpora*. Assim como ocorre com *nunca* e *jamais*, *sequer* conta com apenas uma entrada no *PortiLexicon-UD* com nenhuma especificação de atributo morfológico:

Quadro 9 – Ocorrências de SEQUER

		UD Portuguese Bosque	Porttinari-base	PortiLexicon-UD
Sequer-1	PoS	ADV	ADV	
	atributo	–	–	
	dprel	advmod	–	
	ocorrência	14	2	
Sequer-2	PoS	NOUN		
	atributo	Gender=Masc Number=Sing		
	dprel	obl		
	ocorrência	9		
sequer-3	PoS	NOUN		
	atributo	Gender=Masc Number=Sing		
	dprel	nmod		
	ocorrência	1		

Fonte – O autor.

Como é possível observar no quadro 9, o *sequer* tem apenas duas ocorrências dentro do *Porttinari-base*. Para o *UD Portuguese Bosque*, existem 3 formas distintas de ocorrência

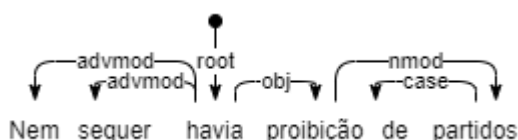
anotadas no *corpus*. Tanto as formas *sequer-2* e *sequer-3* de anotação ocorrem em exemplos em reforço da negação junto ao elemento *nem*:

173. Não havia cruzamento, não se tratou de uma tentativa de ultrapassagem, **nem sequer** de uma travagem brusca. (*UPB*)

174. **Nem sequer** havia proibição de partidos. (*UPB*)

Acredita-se que essa anotação seja oriunda da conversão automática do modelo do Palavras (BICK, 2014) para a UD, uma vez que tanto nas formas *sequer-1* quanto *sequer-2* a *deprel* associada ao elemento não é a relação adverbial (*advmod*), mas nominal oblíquo (*obl*) e modificador nominal (*nmod*), respectivamente. O que é contraintuitivo, considerando que *sequer* não é um substantivo.

Para os casos de *sequer-1* e *sequer-2*, em que *sequer* ocorre como reforço da negação, acredita-se que a anotação mais coerente seria, assim como em *sequer-3*, que o elemento receba a PoS ADV e o atributo de polaridade negativa com a relação de dependência *advmod* em que tanto *nem* quanto *sequer* são dependentes do verbo:



Assim, como com o *jamais* e o *nunca*, na ocorrência de *sequer-3*, a distinção entre sua ocorrência em concordância negativa quanto em posição prototípica de negação sentencial não diferem em anotação segundo os modelos da UD, todavia a marcação da polaridade negativa é essencial, tanto para a detecção da polaridade da sentença, como para a marcação da dupla negação dentro dos *corpora*.

5.3 TAMPOUCO

O *tampouco* é o elemento menos frequente dentro dos *corpora* com apenas 3 ocorrências dentro do *UD Portuguese Bosque* e 7 dentro do *Porttinati-base* (LOPES *et al.*, 2022a). Ele exprime os mesmos comportamentos que o *sequer* (ver subsecção 4.4.3) com a adição da possibilidade de relacionar enunciados como elemento de coordenação. Assim como os outros advérbios, ele conta com apenas uma entrada no *PortiLexicon-UD* sem nenhuma especificação de atributo morfológico, como pode ser observado no quadro 10:

Quadro 10 – Ocorrências de TAMPOUCO

		UD Portuguese Bosque	Porttinari-base	PortiLexicon-UD
tampouco-1	PoS	ADV	ADV	ADV
	atributo	—	—	—
	dprel	advmod	—	—
	ocorrência	2	6	—
tampouco-2	PoS	CCONJ		
	atributo	—		
	dprel	cc		
	ocorrência	1		

Fonte – O autor.

Em ambos os *corpora* a anotação é condizente com as ocorrências, como é possível verificar nos exemplos (176), para a *tampouco-1*, e (177), para a *tampouco-2*:

176. **Tampouco** havia grandes aglomerações, em uma campanha marcada pela discricção. (*PTTB*)
177. (...) enquanto os titulares de Telê privilegiam o meio-campo, com três, quatro, às vezes cinco volantes congestionando o setor, o que, na prática, não confere nem poder ofensivo ao time, **tampouco** proteção infalível à defesa, os garotos jogam com Mona, o único volante típico. (*UPB*)

Mesmo que não tenham sido encontradas ocorrências de *tampouco* como reforço negativo, a anotação dele como advérbio com polaridade negativa é suficiente para exprimir seu comportamento assim como ocorre com *sequer*, restando assim apenas a necessidade de adição do atributo de polaridade negativa para completar a anotação dos termos.

5.4 NADA, NINGUÉM E NENHUM

Como já descrito, a categoria de *pronomes negativos* não existe na classificação tradicional do português brasileiro (ver subseção 4.4.4). Cabe notar que não marcar a polaridade dos pronomes *nada*, *ninguém* e *nenhum* é uma perda tremenda de informação.

As diretrizes da UD descrevem que os pronomes negativos não recebem o atributo de polaridade negativa (*Polarity=Neg*)³⁵ mas têm uma anotação reservada a eles,

³⁵ <https://universaldependencies.org/u/feat/Polarity.html#Neg>

PronType=Neg³⁶. Entretanto, essa notação, mesmo sendo suficiente para demarcar a polaridade, não demarca a definitude do pronome, ocasionando também em uma perda de informação.

Propõe-se aqui, para manter tanto a polaridade quanto a definitude, que a anotação dos pronomes *nada*, *ninguém* e *nenhum* recebam os atributos de pronome negativo PronType=Neg e de definitude indefinida³⁷ Definite=Ind a fim de manter o máximo de informações possíveis na anotação.

Um problema identificado em todos os pronomes, com exceção do *nenhum*, é a anotação do atributo de gênero com a anotação de *gênero neutro* (Gender=Neut). Ainda que a existam trabalhos discutindo os novos usos do português com uma forma neutra de expressão (DE MOURA, 2021; DE SANTANA, 2021), a introdução desse atributo acarretaria algumas mudanças gerais no sistema, mas, sobretudo, não poderia ser aplicada especificamente no caso desses pronomes.

Palavras que podem ocorrer tanto em concordância com o gênero feminino quanto com o masculino, assim como os pronomes negativos, não apresentam flexão de gênero, ou seja, não têm gênero. Para que fosse possível a anotação do atributo de gênero como neutro, o português precisaria exprimir um, ou mais, morfema(s) específico(s) para o gênero neutro, assim como ocorre no alemão ou no romeno, por exemplo.

Dessa forma, propõem-se aqui que, com exceção do pronome *nenhum* que é capaz de flexionar em masculino (*nenhum_*) e feminino (*nenhuma*), os pronomes *negativos* não recebam a etiqueta de atributo de gênero (Gender=none)³⁸, em concordância com as diretrizes de anotação da UD, uma vez que não apresentam morfemas de gênero.

5.4.1 Nada

Dos três pronomes, o *nada* é relativamente mais produtivo com 72 ocorrências no UD Portuguese Bosque (RADEMAKER, 2017) e 71 ocorrências dentro do *Porttinari-base*. Assim como verificado pela literatura e pela verificação em *corpus* na subseção 4.4.4, o *nada* tem duas entradas dentro do *PortiLexicon-UD*, uma como *pronome* e outra como *advérbio*, como é possível observar no quadro 11:

³⁶ <https://universaldependencies.org/u/feat/PronType.html#Neg>

³⁷ <https://universaldependencies.org/u/feat/Definite.html#Ind>

³⁸ <https://universaldependencies.org/u/feat/Gender.html>

Quadro 11 – Ocorrências de NADA

		UD Portuguese Bosque	Portinari-base	PortiLexicon-UD	
nada-1	PoS	PRON	PRON	PRON	
	atributo	Gender=Masc Number=Sing PronType=Ind	Gender=Neut Number=Sing PronType=Ind	Gender=Masc Number=Sing PronType=Ind	
	dprel	obl	–	–	
	ocorrência	5	65	–	
	dprel	nsubj			
	ocorrência	9			
	dprel	obj			
	ocorrência	36			
	dprel	nmod			
	ocorrência	3			
	dprel	conj			
	ocorrência	3			
	dprel	root			
	ocorrência	1			
	dprel	xcomp			
	ocorrência	1			
dprel	nsubj:pass				
ocorrência	2				
nada-2	PoS	ADV			ADV
	atributo	Polarity=Neg	Polarity=Neg	–	
	dprel	advmod	–	–	
	ocorrência	8	5	–	
nada-3	PoS	DET			
	atributo	Definite=Ind Gender=Fem Number=Sing PronType=Art			
	dprel	det			
	ocorrência	1			
nada-4	PoS	NOUN	NOUN		
	atributo	Gender=Masc Number=Sing	Gender=Neut Number=Plur Typo=Yes		
	dprel	nmod	–		
	ocorrência	1	1		
	dprel	root			
	ocorrência	1			
	dprel	parataxis			
ocorrência	1				

Fonte – O autor

A ocorrência de diversas relações de dependência de *nada-1* no *UD Portuguese Bosque* é motivada pelo fato de que, com pronome, ele é capaz de ocorrer em diferentes posições sintáticas podendo ocorrer como sujeito de uma oração (*dprel nsubj*):

178. Mas também aqui ainda **nada** está definido. (*UPB*)

como objeto direto do verbo (*dprel obj*):

179. Durante o almoço familiar, por exemplo, fala-se muito sobre quase **nada** e ninguém mastiga. (*UPB*)

como argumento não nuclear, oblíquo, de um verbo (*dprel obl*):

180. Os onze remates do Sporting contra apenas um do Chaves, durante a metade inicial, dão bem a ideia da superioridade atacante dos homens de Alvalade e não abonam em **nada** a capacidade concretizadora dos seus avançados. (*UPB*)

como sujeito da passiva (*dprel nsubj*):

181. Os policiais foram a uma casa, indicada pelas duas mulheres, onde estariam outros integrantes da quadrilha, mas **nada** foi encontrado ali. (*UPB*)

ou como elemento participante de coordenação:

182. Estava sem trabalho, fazia uns biscates de vez em quando, mas **nada** de sério. (*UPB*)

Como o *Porttinari-base* ainda não foi anotado em função das relações de dependência, todos esses usos demarcados no *UD Portuguese Bosque* estão distribuídos dentro de *nada-1* apenas com a PoS PRON e os atributos Gender=Neutral, Number=Sing, PronType=Ind.

Como é verificável no quadro 11, a anotação do *Porttinari-base* corrigiu o atributo de gênero de Masc para Neut em apenas em *nada-4*. Todavia, como já discutido, o atributo Gender, segundo as diretrizes da UD, não recebe o valor de Neut em para línguas que não exprimem uma morfologia específica para tal declinação, fazendo-se necessária a alteração.

Apenas em *nada-3*, no *corpus UD Portuguese Bosque*, o elemento recebe a anotação de determinante (DET):

183. **Nada** mal para o primeiro contacto com um banco que está a gastar milhares de contos a promover a sua imagem. (*UPB*)

Entretanto, essa ocorrência ainda é *adverbial*, necessitando de uma alteração de DET para ADV, uma vez que *nada* não está quantificando nem alterando a definitude do adjetivo *mal*. Essa alteração/correção já foi realizada no *Porttinari-base* como é possível verificar em:

184. **Nada** mal para uma coleção bem tradicional de composições clássicas " animadas " (*PTTB*)

que ocorre em *nada-2*.

Considerando que *nada-2* já está demarcada com a polaridade negativa, resta apenas para as ocorrências com *nada* a alteração dos atributos de pronome de *nada-1* de *indefinido* para *negativo*, a adição do atributo de definitude indefinida (Definite=Ind), a alteração do valor do atributo de gênero e as alterações indicadas para todas as ocorrências.

5.4.2 Ninguém

Dos três pronomes analisados, o *ninguém* é o que exprime o comportamento prototípico da categoria, não podendo ocorrer com nenhuma outra função sintática dentro dos *corpora*. Contudo, no *PortiLexicon-UD* existe uma entrada com a PoS NOUN, *ninguém-2* no quadro 12. Essa entrada faz referência a ocorrência de *ninguém* anteposto por um determinante, os mais frequentes são *um* e *uma*:

185. Sou uma **ninguém** – sou só uma ex-trafficante. (*FK*)
 186. Eu era um **ninguém**. (*FK*)

Entretanto, por ser referente a uma classe de lexia de aberta, uma palavra de conteúdo, está fora do escopo da pesquisa. É válido ressaltar que esse tipo de ocorrência tem uma frequência muito baixa, com um total de cinco sentenças no *corpus* de validação.

As ocorrências como pronome podem ser observadas no quadro 12:

Quadro 12 – Ocorrências de NINGUÉM

		UD Portuguese Bosque	Portinari-base	PortiLexicon-UD		
ninguém-1	PoS	PRON	PRON	PRON		
	atributo	Gender=Masc Number=Sing PronType=Ind	Gender=Neut Number=Sing PronType=Ind	Number=Sing PronType=Ind		
	dprel	nsubj	–	–		
	ocorrência	30	49	–		
	dprel	obj				
	ocorrência	5				
	dprel	nmod				
	ocorrência	2				
	dprel	obl				
	ocorrência	2				
	dprel	conj				
	ocorrência	1				
	dprel	acl				
	ocorrência	1				
	dprel	nsubj:pass				
ocorrência	2					
ninguém-2	PoS					NOUN
	atributo					Gender=Masc Number=Sing
	dprel		–			
	ocorrência		–			

Fonte – O autor

Assim como em *nada-1*, as ocorrências como pronome, o *ninguém* consegue ocorrer com diversas anotações de dprel, no quadro 12, devido a sua capacidade de ocorrer com diferentes posições sintáticas, podendo ocorrer como sujeito de uma oração (*dprel nsubj*):

187. **Ninguém** força sua escalação porque há quem escale o time no São Paulo.
(UPB)

como objeto direto de um verbo (*dprel obj*):

188. O sismo de segunda-feira em Los Angeles não poupou **ninguém**, nem mesmo as estrelas. (*UPB*)

como argumento não nuclear, oblíquo, de um verbo (*dprel obl*):

189. Exigências que obrigam os grupos a montar peças a correr só para não serem penalizados monetariamente – o que não é saudável para **ninguém**, muito menos para o Gil Vicente. (*UPB*)

como sujeito da passiva (*dprel nsubj:pass*):

190. **Ninguém** foi encontrado ontem nas secretarias municipais e estaduais da Saúde para comentar o assunto. (*UPB*)

como modificador nominal (*dprel nmod*):

191. Tranquilidade absoluta, nenhuma hostilidade da parte de **ninguém**, uma paz realmente sem mácula. (*UPB*)

ou como elemento participante de coordenação (*dprel conj*):

192. A responsabilidade é a um só tempo de todos e de **ninguém**. (*UPB*)

Como é possível observar no quadro 12, a anotação do *Portinari-base* revisou o atributo de gênero para todas as ocorrências de *ninguém*, alterando o atributo *Masc* presente no *UD Portuguese Bosque* para *Neut*, todavia, como já descrito, não existe esta marcação morfológica no português. Assim, é necessária a alteração do atributo de pronome indefinido para negativo, de *Ind* para *Neg*, e a adição do atributo de definitude indefinida, *Definite=Ind* e a remoção do atributo de gênero (*Gender=none*).

5.4.3 Nenhum

Assim como o *nada*, o *nenhum* não exprime apenas o comportamento pronominal, porém, ele não exprime comportamento adverbial, mas, como quantificador (ver subseção 4.4.4), pode ocorrer na posição de determinante.

Como *nenhum*, ao contrário de *nada* e *ninguém*, flexiona em gênero (*nenhuma* e *nenhum*) não há necessidade da alteração em *nenhum* dos *corpora* do atributo de gênero. Ainda que exista mais de uma ocorrência no *corpus* o *nenhum* ocorre ou *determinante* ou como *pronome*, a única distinção entre as duas formas é a presença da preposição entre ele e o nome/pronome subsequente:

- 193. Mas não via **nenhuma** saída. (*UPB*)
- 194. A situação tende a se agravar, uma vez que **nenhuma** das partes parece mostrar disposição de recuar. (*UPB*)
- 195. Mas **nenhuma** delas (de+elas) tem medo de não encontrar o carro na saída do cinema ou de ser agredida pelo guardador, ou de parar no sinal de trânsito na volta para casa. (*UPB*)

Como é possível observar no quadro 13, houve a alteração do tipo de pronome de Negativo, no *UD Portuguese Bosque* para indefinido no *Portinari-base*. Porém, como já descrito, propõe-se aqui a anotação dos pronomes indefinidos e negativos com a anotação de pronome de tipo negativo (*PronType=Neg*) com a adição do atributo de definitude indefinida (*Definite=Ind*).

Quadro 13 – Ocorrências de NENHUM

		UD Portuguese Bosque	Porttinari-base	PortiLexicon-UD
nenhum-1	PoS		DET	DET
	atributo		Gender=Fem Number=Sing PronType=Ind	Gender=Fem Number=Sing PronType=Ind
	dprel		–	–
	ocorrência		18	–
nenhum-2	PoS	DET		DET
	atributo	Gender=Masc Number=Sing PronType=Neg		Gender=Masc Number=Sing PronType=Ind
	dprel	det		–
	ocorrência	16		–
	dprel	conj		
	ocorrência	1		
nenhum-3	PoS	PRON	PRON	PRON
	atributo	Gender=Fem Number=Sing PronType=Neg	Gender=Fem Number=Sing PronType=Ind	Gender=Fem Number=Sing PronType=Ind
	dprel	nsubj	–	–
	ocorrência	3	5	–
	dprel	obl		
	ocorrência	1		
nenhum-4	PoS			PRON
	atributo			Gender=Masc Number=Sing PronType=Ind
	dprel			–
	ocorrência			–

Fonte – O autor

5.5 SEM

O *sem* é o segundo item mais produtivo dos elementos negativos. Ele tem apenas um tipo de ocorrência dentro do *Porttinari-base* e um total de 5 tipos de ocorrências no UD Portuguese Bosque (RADEMAKER, 2017) como é possível observar no quadro 14.

O *sem* sempre inicia um sintagma (ver subseção 4.4.5), fazendo assim com que sua função seja sempre prepositiva, nunca introduzindo uma oração completa. Dessa forma, em concordância com Duran (2021a), propõe-se aqui que todas as ocorrências de *sem-1* como SCONJ, que não contém formas verbais flexionadas, sejam categorizadas, assim como feito no *Porttinari-base*, como preposições recebendo a PoS ADP.

A ocorrência de *sem-3*, referente a apenas uma frase, ocorre com relação a coordenação entre *sem* e *com*:

131. Os que olham para a sua câmara, com mais ou menos prazer, com ou **sem** indiferença, não mostram hostilidade ao fotógrafo. (**UPB**)

E as ocorrências de *sem-4* e *sem-5* são referentes a expressões multipalavras, duas em que *sem* ocorre como *head* e três que ele ocorre como dependente.

Por fim, a única ocorrência presente no *PortiLexicon-UD* a de *sem* como preposição (ADP) categorizado em *nem-2*.

Assim, para as ocorrências de *sem*, propomos aqui que *sem-1* seja anotada como *sem-2*, como já ocorreu no *Porttinari-base* e que *sem-1*, *sem-2* e *sem-3* recebam o atributo de polaridade negativa.

Quadro 14 – Ocorrências de SEM

		UD Portuguese Bosque	Porttinari-base	PortiLexicon-UD
sem-1	PoS	SCONJ		
	atributo	–		
	dprel	mark		
	ocorrência	47		
sem-2	PoS	ADP	ADP	ADP
	atributo	–	–	–
	dprel	case	–	–
	ocorrência	156	182	–
sem-3	PoS	ADP		
	atributo	–		
	dprel	conj		
	ocorrência	1		
sem-4	PoS	ADV		
	atributo	ExtPos=NOUN		
	dprel	nmod		
	ocorrência	2		
sem-5	PoS	PROPN		
	atributo	Number=Sing		
	dprel	flat:name		
	ocorrência	3		

Fonte – O autor

5.6 NEM

O *nem*, é um elemento negativo que ocorre com conjunção e em posição prototípica, sendo capaz de acionar a coordenação negativa e concordar com ela (ver subseção 4.6.6 e quadro 6). Todas essas ocorrências puderam ser verificadas em ambos os *corpora*, como observável no quadro 15.

Tanto para as ocorrências de *nem-1*, como conjunção, quanto para as de *nem-3*, como advérbio, a única necessidade é a adição do atributo de polaridade negativa (*Polarity=Neg*).

As ocorrências de *nem-2* e *nem-5*, que ocorrem apenas no UD Portuguese Bosque (RADEMAKER, 2017), têm uma característica interessante, ocorrem com as relações de dependência invertidas, *CCONJ* deveria ocorrer apenas com *cc* e *ADV* apenas com *advmod*. Acredita-se que essa inversão possa ter ocorrido no processo de conversão da anotação do Palavras para os conformes da UD (BICK, 2016). Sendo apenas necessário, além da adição do atributo de polaridade negativa, a correção da PoS e/ou *dprel*.

As ocorrências de *nem-6* são referentes a *nem sequer*, pela marcação de PoS de *sequer* como *NOUN*. Como já dito anteriormente, o que está ocorrendo é um reforço negativo. Para a solução desse problema já foi proposto a anotação tanto de *nem* quanto *sequer* como advérbios com polaridade negativa e dependentes do verbo que estão modificando, conforme a seção 5.2.

Para a ocorrências de *nem-7*, com apenas uma sentença, propomos aqui que, seguindo a proposta de Duran (2021a), apenas elementos que sejam capazes de introduzir uma oração com verbo flexionado recebam a marcação *SCONJ*, o que não é o caso desta ocorrência:

132. Tanto mais quanto, hoje em dia, o hemisfério esquerdo (dos dextrímanos) já não é considerado como o hemisfério cerebral “dominante”, **nem** como o único capaz de desempenhar altas funções mentais e de controlar movimentos precisos e sutis. (*UPB*)

Independentemente do tamanho do sintagma, o verbo presente nele não é o centro de uma oração, fazendo com que essa ocorrência da correção da PoS (de *SCONJ* para *CCONJ*) e da *dprel* (de *mark* para *cc*) com adição da polaridade negativa.

Quadro 15 – Ocorrências de NEM

		UD Portuguese Bosque	Porttinari-base	PortiLexicon-UD
nem-1	PoS	CCONJ	CCONJ	CCONJ
	atributo	–	–	–
	dprel	cc	–	–
	ocorrência	57	42	–
nem-2	PoS	CCONJ		
	atributo	–		
	dprel	advmod		
	ocorrência	2		
nem-3	PoS	ADV	ADV	ADV
	atributo	–	–	–
	dprel	advmod	–	–
	ocorrência	25	49	–
nem-4	PoS	ADV		
	atributo	–		
	dprel	root		
	ocorrência	1		
nem-5	PoS	ADV		
	atributo	–		
	dprel	cc		
	ocorrência	5		
nem-6	PoS	ADP		
	atributo	–		
	dprel	case		
	ocorrência	11		
nem-7	PoS	SCONJ		
	atributo	–		
	dprel	mark		
	ocorrência	1		

Fonte – O Autor

O *nem* também pode ocorrer como parte da locução *que nem* que não exprime polaridade negativa (conforme subseção 4.4.6). Essa locução funcionando como uma conjunção comparativa, assim como o *igual a* e *como* (THOMPSON *et al.*, 2012). Foram encontradas apenas duas ocorrências no *Porttinari-base*, classificadas como ADV, com essa locução:

133. E, mais do que nunca, concordei com a frase de que sorvete é **que nem** sexo. (PTTB)

134. A geração mais nova dificilmente tem acesso a outros tipos de música que conseguem unir corações, mentes e bundas - algumas boas **que nem** o Chacrinha, que, com suas chacetes, conseguia empolgar velhos e moços. (PTTB)

É válido ressaltar que esse tipo de ocorrência difere da sequência *que nem* quando o *que* é licenciado pelo verbo, como ocorre em:

135. A gente tem **visto que nem** sempre são demandas apropriadas. (PTTB)

em que o *que* introduz uma oração iniciada por *nem*.

Propomos aqui que a sequência presente em (198) e (199) seja anotada com a *dprel fixed*, uma vez que o significado da sequência *que nem* não é mais composicional, apenas gramaticalizado. De maneira que o *que* receba o atributo `ExtPos=CONJ` (ou `SCONJ` dependendo do complemento de *que nem*) e o *nem* mantenha a PoS de ADV, mas sem a polaridade negativa, em concordância com Duran (2021b).

5.7 NÃO

De todos os elementos negativos, a forma *não* é a mais produtiva para a realização da negação (ver capítulo 4). Em ambos os *corpora*, a soma da ocorrência de todos os outros elementos não ultrapassa 50% das ocorrências de não – 39,3% das ocorrências para o *UD Portuguese Bosque* e 35,72% para o *Porttinari-base*. A maior parte das ocorrências estão anotadas com a PoS ADV e com a marcação de polaridade negativa, como é possível observar no quadro 16.

Quadro 16 – Ocorrências de NÃO

		UD Portuguese Bosque	Porttinar-base	PortiLexicon-UD
não-1	PoS	ADV	ADV	ADV
	atributo	Polarity=Neg	Polarity=Neg	–
	dprel	advmod	advmod	–
	ocorrência	1347	1338	–
não-2	PoS	ADV		
	atributo	Polarity=Neg		
	dprel	fixed		
	ocorrência	5		
não-3	PoS	ADV		
	atributo	Polarity=Neg		
	dprel	discourse		
	ocorrência	4		
não-4	PoS	NOUN		NOUN
	atributo	Gender=Masc Number=Sing		Gender=Masc Number=Sing
	dprel	obj		–
	ocorrência	2		–
não-5	PoS	INTJ		
	atributo	–		
	dprel	parataxis		
	ocorrência	1		
não-6	PoS	INTJ		
	atributo	–		
	dprel	root		
	ocorrência	1		
não-7	PoS	INTJ		
	atributo	–		
	dprel	conj		
	ocorrência	2		

Fonte – O Autor

As ocorrências de *não-1* como ADV têm o atributo de polaridade negativa anotadas em ambos os *corpora*, não necessitando de revisão, sendo apenas necessário adicionar o atributo no *PortiLexicon-UD*. Todavia, o *Porttinari-base* por ainda não ter realizado anotação sintática não distribui o *não* em outras categorias como interjeição (INTJ), como observável nas duas únicas ocorrências do *corpus* UD Portuguese Bosque (RADEMAKER, *et al.*, 2017):

136. Eu falei: **não**, acho que isso está muito exagerado, afirmou Palocci. (**UPB**)
 137. **Não**, não consigo, desligue você (**UPB**)

Esses casos podem ser anotados com a PoS de interjeição (INTJ) com o atributo de polaridade negativa (Polarity=Neg), uma vez que não está associada a nenhum elemento para modificar como advérbio.

O caso das ocorrências de *não-2* é oriundo da decisão de anotação de anotar a sequência *a não ser* como um fixed para uma SCONJ. Concorde-se com a anotação, uma vez que “*a não ser*” é invariável com relação a sinônimos, não existindo a necessidade de alteração.

Para os casos das ocorrências de *não-3*, que são similares aos dos exemplos (201) e (202), propomos a anotação como a PoS de interjeição (INTJ) e o atributo de polaridade negativa (Polarity=Neg), mantendo a relação de dependência *discourse*.

As ocorrências de *não-4* apresentam o fenômeno, já descrito por Neves (2000, p.285-286), da capacidade do *não* ocorrer como equivalência a um enunciado negativo. Nas duas sentenças da ocorrência de *não-4*, ele está ocorrendo junto a um verbo de comunicação sendo referentes ao conteúdo da mensagem dita:

138. Ele a emprega quando tem que dizer **não** a uma pessoa sem desagradá-la no caso, a seus colegas de PFL que querem repartir os cargos com o PSDB antes do 1º turno. (**UPB**)
 139. Apesar de evitar dar um **não** definitivo, Marise deixou claro que deve recusar o convite de Brizola. (**UPB**)

Concorde-se com a anotação de *não* como NOUN neste caso, dessa forma, não é necessária nenhuma alteração, uma vez que essa classe está fora do escopo do trabalho.

A ocorrência de *não-5* é uma relação de parataxis dentro de um diálogo, como observável no exemplo:

140. R.(resposta) -- **Não**. (**UPB**)

Concorde-se com a anotação de interjeição para a PoS de *não*, mas propõe-se a adição do atributo de polaridade negativa.

Por fim, as ocorrências de *não-6* e *não-7* são referentes a uma mesma sentença:

141. **Não, não e não. (UPB)**

Em que todos os *não* recebem a marcação de interjeição sendo que dois eles ocorrem em conjunção com relação ao primeiro, sendo necessário apenas a adição do atributo de polaridade negativa (*Polarity=Neg*), concluindo esta subseção.

Acredita-se que as alterações propostas neste capítulo para a anotação do *Porttinari-base* sejam suficientes para uma melhor demarcação dos elementos negativos do português a fim de reter a maior quantidade de informação sobre eles dentro do *corpus*.

5.8 PREPOSIÇÕES ACIDENTAIS

Todas as nove preposições acidentais listadas por Neves (2000) constituem entrada dentro do *PortiLexicon-UD*, entretanto, apenas seis delas ocorrem nos *corpora*, como é possível observar nos quadros de 17 a 25. Das ocorrências encontradas, a mais relevante para a análise é a ocorrência das preposições como *ADP*, a POS específica para a marcação dessa classe de palavras.

5.8.1 Exclusive

Partindo de *exclusive*, a palavra está presente apenas no *PortiLexicon-UD* (LOPES e tal, 2022b). Todavia, ela não tem uma notação para a ocorrência como preposição, assim como verificado na subseção 4.1.7. A sua única entrada dentro do *PortiLexicon-UD* (LOPES e tal, 2022b) é como *ADV* (advérbio), como é verificável no quadro 17.

Quadro 17 – Ocorrências de EXCLUSIVE

		UD Portuguese Bosque	Porttinari-base	PortiLexicon-UD
Exclusive-1	PoS			ADV
	atributo			—
	dprel			—
	ocorrência			—

Fonte – O autor.

Todavia, é possível encontrar descrições, como em HOUAISS (2009), que trazem a função preposicional do advérbio: “o advérbio *exclusive* tem, modernamente, uso preposicional na língua, equivalente a 'tirante, salvo, exceto': traga para cá todos os livros, *exclusive* os não encapados”. Assim, tornando necessária sua marcação de POS como ADP.

5.8.2 Excluso

Assim como *exclusive*, *excluso* está presente apenas no PortiLexicon-UD (LOPES e tal, 2022b), entretanto, só existe uma entrada como ADJ, como verificável no quadro 18, não comportando suas ocorrências como ADP.

Quadro 18 – Ocorrências de EXCLUSO

		UD Portuguese Bosque	Portinari-base	PortiLexicon-UD
Excluso-1	PoS			ADJ
	atributo			Gender=Masc Number=Sing
	dprel			—
	ocorrência			—

Fonte – O autor

5.8.3 Exceto

Para o item *exceto*, também foi buscado nos *corpora* pela forma *excepto* (forma do português europeu), ele só está presente no UD Portuguese Bosque (RADEMAKER *et al.*, 2017). Como é possível verificar no quadro 19, todos os *corpora*, assim como a ferramenta lexical, têm ao menos uma ocorrência de *exceto* (ou *excepto*) como ADP.

Nenhuma das ocorrências possui a marcação de polaridade negativa, mesmo ocorrendo em contextos negativos de exclusão, assim como o *sem*:

131. Não temos uma política de seguro para o setor de frutas, **exceto** a maçã. (*PTTB*)
132. A margem de erro da pesquisa Datafolha é de 3,0 pontos percentuais, para mais ou para menos, **exceto** no Distrito Federal, que é de 4,0 pontos. (*UPB*)
133. Até à conclusão deste complicado processo, a CGD deverá continuar a suspender a execução das dívidas, **excepto** para as situações de abandono. (*UPB*)

Tornando necessária a adição do atributo de polaridade negativa (*Polarity=Neg*).

Quadro 19 – Ocorrências de EXCETO

		UD Portuguese Bosque	Portinari-base	PortiLexicon-UD
Exceto/ Excepto-1	PoS	ADP	ADP	ADP
	atributo	–	–	–
	dprel	case	–	–
	ocorrência	5	1	–
Exceto/ Excepto-2	PoS	–	–	NOUN
	atributo	–	–	Gender=Masc Number=Sing
	dprel	–	–	–
	ocorrência	–	–	–

Fonte – O autor

5.8.4 Fora

O *advérbio fora* ocorre nos dois *corpora* assim como na ferramenta lexical. Da mesma forma que o *exceto*, não há a marcação de polaridade negativa (*Polarity=Neg*) nas ocorrências, como é possível verificar no quadro 20, necessitando da adição.

Quadro 20 – Ocorrências de FORA

		UD Portuguese Bosque	Portinari-base	PortiLexicon-UD
Fora-1	PoS	ADP	ADP	ADP
	atributo	–	–	–
	dprel	nmod	–	–
	ocorrência	3	3	–
Fora-2	PoS	ADV	ADV	ADV
	atributo	–	–	–
	dprel	diversas	–	–
	ocorrência	61	51	–
Fora-3	PoS	–	–	NOUN
	atributo	–	–	Gender=Masc Number=Sing
	dprel	–	–	–
	ocorrência	–	–	–

Fonte – O Autor

5.8.5 Afora

O advérbio *afora* ocorre como preposição apenas no *PortiLexicon-UD*, dentro dos *corpora* ele têm ocorrência apenas como advérbio, como é possível verificar nos exemplos abaixo, e no quadro 21:

134. **Afora** historicismo, isso é menosprezar um fator interno à arte brasileira, que independe de contexto internacional. (*UPB*)
135. Che Guevara era conhecido pelo mundo **afora**, mas sua fama não lhe valera grande apreço entre os camponeses bolivianos. (*PTTB*)

Entretanto, o exemplo (210), retirado do UD Portuguese Boque (RADEMAKER *et al.*, 2017), é uma ocorrência prepositiva, uma vez que *afora* poderia ser substituído por *exceto* ou qualquer outra preposição acidental de exclusão.

Quadro 21 – Ocorrências de AFORA

		UD Portuguese Bosque	Portinari-base	PortiLexicon-UD
Afora-1	PoS	ADV	ADV	ADV
	atributo	–	–	–
	dprel	advmod	–	–
	ocorrência	1	1	–
Afora-2	PoS			ADP
	atributo			–
	dprel			–
	ocorrência			–

Fonte – O autor

5.8.6 Salvo

O adjetivo *salvo* está presente apenas no *UD Portuguese Bosque* e no *PortiLexicon-UD*. As ocorrências como preposição recebem a anotação de ADP, mas não têm a adição do atributo de polaridade negativa (*Polarity=Neg*), necessitando da adição. As ocorrências de *salvo* podem ser observadas no quadro 22:

Quadro 22 – Ocorrências de SALVO

		UD Portuguese Bosque	Porttinari-base	PortiLexicon-UD
Salvo-1	PoS	ADP		ADP
	atributo	–		–
	dprel	case/fixed		–
	ocorrência	2		–
Salvo-2	PoS			ADJ
	atributo			Gender=Masc Number=Sing VerbForm=Part
	dprel			–
	ocorrência			–

Fonte – O autor

5.8.7 Menos

O advérbio *menos* tem a segunda maior quantidade de ocorrências distintas entre as preposições acidentais. Ainda que a investigação tenha como intuito a análise das ocorrências prepositivas, não se pode deixar de verificar a marcação de número *Inv* nas ocorrências como determinante (*DET*) e nome (*NOUN*).

Não foi encontrada nenhuma descrição sobre qual a especificação da etiqueta *Inv*, tanto no manual de anotação de Duran (2021) quanto na descrição da ferramenta lexical *PortiLexicon-UD*. Todavia, tal etiqueta é listada na página do atributo de *Number* da UD (*NUMBER*, 2021) como *número inverso (inverse number)*, que a define como: “número inverso significa número não prototípico para aquele nome em particular. (Alguns nomes são assumidos por padrão como singular, alguns são duais ou plurais)”. O que não é o caso, uma vez que nas ocorrências do *Porttinari-base* a *lexia* se encontra em posição de determinante tanto de elementos singulares quanto plurais:

136. O trabalhador vai ter de escolher **menos** direitoS e emprego ou todos os direitos e desemprego", disse. (*PTTB*)
137. O problema é que, na prática, as mulheres acabam tendo **menos** sucesso_ nos tratamentos do que os homens. (*PTTB*)

Especula-se aqui que a etiqueta possa estar se referindo a palavras inflexíveis em número, ou invariáveis, uma vez que ela está presente em todas as categorias capazes de flexionar em número na descrição da ferramenta *Porttinari-base*. Caso seja este o caso, é necessário a alteração da etiqueta para evitar conflitos de concordância com as anotações em outras línguas.

Ademais, as ocorrências prepositivas da lexia, assim como a das anteriores, não exprime o atributo de polaridade negativa (`Polarity=Neg`), como é observável no quadro 23, fazendo-se necessária a adição.

Quadro 23 – Ocorrências de MENOS

		UD Portuguese Bosque	Porttinari-base	PortiLexicon-UD
Menos-1	PoS	ADV	ADV	ADV
	atributo	Gender=Masc Number=Plur	–	–
	dprel	nmod	–	–
	ocorrência	1	61	–
Menos-2	PoS	NOUN	NOUN	NOUN
	atributo	Gender=Masc Number=Sing	Gender=Masc Number=Sing	Gender=Masc Number=Sing
	dprel	nmod/obl	36	–
	ocorrência	38	51	–
Menos-3	PoS	DET	DET	DET
	atributo	Gender=Masc Number=Plur PronType=Ind	Gender=Neut Number=Inv PronType=Ind	Number=Inv PronType=Ind
	dprel	det	–	–
	ocorrência	1	11	–
Menos-4	PoS		ADP	ADP
	atributo		–	–
	dprel		–	–
	ocorrência		1	–

Fonte – O autor

5.8.10 Senão

O *senão* é a lexia com a maior quantidade de ocorrências distintas entre as preposições acidentais. Todavia, não há nenhuma ocorrência com que se destaque como as de *menos-3* (DET). Assim como as proposições anteriores, suas ocorrências como ADP não tem a marcação de polaridade negativa (*Polarity=Neg*) em nenhum dos *corpora* nem na ferramenta lexical, como é observável no quadro 24, fazendo-se necessário a adição do atributo.

138. Enquanto as universidades públicas ameaçam fechar as portas por falta de dinheiro para pagar a docentes e não docentes, é cada vez maior o número de estabelecimentos de ensino privados de qualidade duvidosa a ministrar somente cursos que outros recursos não exigem **senão** a caneta e o papel. (*UPB*)
139. Trabalho para milhares, **senão** milhões, de anos. (*PTTB*)

Quadro 24 – Ocorrências de SENÃO

		UD Portuguese Bosque	Portinari-base	PortiLexicon-UD
Senão-1	PoS	ADV		ADV
	atributo	–		–
	dprel	advmod		–
	ocorrência	6		–
Senão-2	PoS	SCONJ		
	atributo	–		
	dprel	mark		
	ocorrência	2		
Senão-3	PoS		CCONJ	CCONJ
	atributo		–	–
	dprel		–	–
	ocorrência		4	–
Senão-4	PoS	ADP	ADP	ADP
	atributo	–	–	–
	dprel	case	–	–
	ocorrência	2	2	–
Senão-5	PoS			NOUN
	atributo			Gender=Masc Number Sing
	dprel			–
	ocorrência			–

Fonte – O autor

5.8.11 Tirante

Por fim, *tirante* está presente apenas no *PortiLexicon-UD*. Ela não tem uma entrada para sua ocorrência como preposição (ADP), constando apenas com entradas para sua forma adjetiva (ADJ) e nominal (NOUN), como é observável no quadro 25.

Quadro 25 – Ocorrências de TIRANTE

		UD Portuguese Bosque	Portinari-base	PortiLexicon-UD
Tirante-1	PoS			ADJ
	atributo			Number=Sing
	dprel			–
	ocorrência			–
Tirante-2	PoS			NOUN
	atributo			Number=Sing
	dprel			–
	ocorrência			–

Fonte – O autor

Como verificado, todas as lexias que podem ocorrer como *preposições acidentais de exclusão* não têm a marcação de polaridade negativa ou até mesmo não têm registro em nenhum dos *corpora* ou na ferramenta lexical.

5.9 SÍNTESE DOS RESULTADOS

Com os dados levantados e discutidos neste capítulo, pretende-se realizar nesta seção uma condensação de todas as propostas de otimização sugeridas ao longo do texto. Os próximos quadros sumarizam cada uma das lexias aqui analisadas, com exemplos de ocorrências e formas de identificação para a anotação.

5.9.1 Nunca e Jamais

Quadro 26 – Propostas de anotação, na metodologia da UD para NUNCA

Nunca	
PoS	ADV
atributo	Polarity=Neg
<p>Exemplos:</p> <p>A banda nunca tocou melhor apenas veja aquelas 100 mil pessoas no Rock in Rio, diz o cantor (FK)</p> <p>Não há fotografias sem pessoas, ou sem os seus vestígios, o que não deixa nunca de remeter o trabalho deste fotógrafo (UPB)</p>	

Fonte – O autor

Quadro 27 – Propostas de anotação, na metodologia da UD para JAMAIS

Jamais	
PoS	ADV
atributo	Polarity=Neg
<p>Exemplos:</p> <p>Sei que jamais terei uma chance de trabalhar tão calmamente de novo (UPB)</p> <p>Deixava para trás o mundo explicado pelos orixás e não chegava jamais a alcançar o mundo explicado pela tabela periódica (FK)</p>	

Fonte – O autor

Tanto *nunca* quanto *jamais* são limitados a ocorrerem antepostos a o verbo, em contexto de negação oracional, ou pospostos, em contexto de concordância negativa.

5.9.2 Sequer

Quadro 28 – Propostas de anotação, na metodologia da UD para SEQUER

Sequer	
PoS	ADV
atributo	Polarity=Neg
Exemplos: Eduardo Cunha sequer era filiado ao PMDB no momento da compra (FK) Não é o momento sequer de pensar isso (FK) Nem sequer havia proibição de partidos (UPB)	

Fonte – O Autor

Assim como o *nunca* e o *jamais*, o *sequer* é limitado a ocorrer anteposto ao um verbo, em contexto de negação oracional, ou posposto, em contexto de concordância negativa. Para os casos de ocorrência como reforço negativo, assim como já proposto (ver seção 5.2), mantém-se a mesma notação de *PoS* e atributos, fazendo que tanto a *lexia* negativa, quanto o *sequer*, sejam dependentes do verbo modificado.

5.9.3 Tampouco

Quadro 29 – Propostas de anotação, na metodologia da UD para TAMPOUCO

Tampouco-1	
PoS	ADV
atributo	Polarity=Neg
<p>Exemplos:</p> <p>Tampouco havia grandes aglomerações, em uma campanha marcada pela discrição (<i>PTTB</i>)</p> <p>A empresa tampouco respondeu qual foi o percentual de espectadores que desistiram das séries antes de chegar ao "capítulo-gancho". (<i>FK</i>)</p>	
Tampouco-2	
PoS	CCONJ
atributo	Polarity=Neg
<p>Exemplos:</p> <p>Não se pode, porém, minimizar o peso simbólico desta viagem, tampouco o poder das imagens que produzirá (<i>FK</i>)</p> <p>A perspectiva da delação premiada já não intimida ninguém, tampouco convence a opinião pública. (<i>FK</i>)</p>	

Fonte – O autor

O *tampouco*, assim como os outros advérbios, ocorre anteposto ao verbo, mas com a possibilidade de ocorrer como *conjunção negativa*, ou seja, realizando coordenação entre duas orações, ou elementos, realizando negação, como em *tampouco-2*.

5.9.4 Nada, Ninguém e Nenhum

Quadro 30 – Propostas de anotação, na metodologia da UD para NADA

Nada-1	
PoS	ADV
atributo	Polarity=Neg
Exemplos: Nada mal para uma coleção bem tradicional de composições clássicas " animadas " (<i>PTTB</i>) Não estava nada à espera disto (<i>UPB</i>)	
Nada-2	
PoS	PRON
atributo	Gender=none Number=Sing PronType=Neg Definite=Ind
Exemplos: Durante o almoço familiar, por exemplo, fala-se muito sobre quase nada e ninguém mastiga. (<i>UPB</i>) Nada indica que quebrará. (<i>FK</i>)	

Fonte – O Autor

O *nada*, ocorre de duas formas: a primeira como advérbio sempre anteposto ao elemento que ele irá modificar, podendo ocorrer em negação oracional, anteposto a um verbo, ou como negação não oracional, anteposto a qualquer outro elemento; a sua segunda ocorrência é como pronome, podendo ocorrer em todos os contextos em que um *nome* é capaz de assumir (sujeito, objeto, complemento, etc.).

Quadro 31 – Propostas de anotação, na metodologia da UD para NINGUÉM

Ninguém-1	
PoS	PRON
atributo	Gender=none Number=Sing PronType=Neg Definite=Ind
<p>Exemplos:</p> <p>Ninguém força sua escalação porque há quem escale o time no São Paulo (UPB)</p> <p>Ninguém foi encontrado ontem nas secretarias municipais e estaduais da Saúde para comentar o assunto (UPB)</p>	

Fonte – O Autor

O *ninguém* é o único pronome negativo que exprime comportamento necessariamente pronominal, dessa forma ele é capaz de ocorrer em todos os contextos que um *nome* é capaz de assumir (sujeito, objeto, complemento, etc.).

Quadro 32 – Propostas de anotação, na metodologia da UD para NENHUM

Nenhum-1	
PoS	Det
atributo	Gender=Masc/Fem Number=Sing PronType=Neg Definite=Ind
Exemplos: Mas não via nenhuma saída (<i>UPB</i>) Nenhuma legenda obteve êxito parecido. (<i>FK</i>)	
Nenhum-2	
PoS	PRON
atributo	Gender=Masc/Fem Number=Sing PronType=Neg Definite=Ind
Exemplos: A situação tende a se agravar, uma vez que nenhuma das partes parece mostrar disposição de recuar (<i>UPB</i>) Nenhum dos serviços cobra taxa adicional dos clientes (<i>FK</i>)	

Fonte – O Autor

O *nenhum* exprime duas ocorrências, uma como *pronome* e outra como determinante. A única distinção das duas formas é a presença da preposição entre o *nenhum* e o elemento que ele irá modificar. Quando ela ocorre, ele tem comportamento pronominal, e quando não, tem comportamento determinante.

5.9.5 Sem

Quadro 33 – Propostas de anotação, na metodologia da UD para SEM

Sem-1	
PoS	ADP
atributo	Polarity=Neg
<p>Exemplos:</p> <p>O longa é uma obra sem firulas vistosas para entreter o espectador. (<i>UPB</i>)</p> <p>Ela venceu sem muitas dificuldades, ao aplicar um waza-ari e administrar o restante da luta. (<i>FK</i>)</p>	

Fonte – O Autor

O *sem* exprime comportamento estritamente prepositivo, sempre iniciando uma oração reduzida de infinitivo (ver subseção 4.4.5 e seção 5.7), dessa forma ele tem apenas uma forma como preposição, recebendo a etiqueta ADP, com polaridade negativa, recebendo a etiqueta Polarity=Neg.

5.9.6 Nem

Quadro 34 – Propostas de anotação, na metodologia da UD para NEM

Nem-1	
PoS	ADV
atributo	Polarity=Neg
<p>Exemplos:</p> <p>Nem é preciso dizer que, depois dessa, o tatuador distraído levou mais umas bifas. (<i>FK</i>)</p> <p>Mas Walter Feldman, o secretário menor da CBF, segue estimulando os clubes a sabotar a Medida Provisória do Futebol, embora nem tenha coragem de admitir. (<i>FK</i>)</p>	
Nem-2	
PoS	CCONJ
atributo	Polarity=Neg
<p>Exemplos:</p> <p>Nem o Corinthians está igual ao primeiro turno, nem o São Paulo dá confiança (Porttinari-base)</p> <p>JBS não é do BNDES , nem do governo (Porttinari-base)</p>	
Nem-3	
PoS	SCONJ
atributo	Polarity=Neg
<p>Exemplos:</p> <p>Não paira contra mim nenhuma suspeita de desvio de dinheiro público, não possuo conta no exterior, nem ocultei do conhecimento público a existência de bens pessoais (<i>FK</i>)</p> <p>Mas não respondeu às perguntas e nem retornou contatos posteriores. (<i>FK</i>)</p>	

Fonte – O Autor

O *nem*, para além do comportamento de advérbio de negação que ocorre anteposto ao *verbo*, também pode ocorrer realizando coordenação entre elementos internos a oração (CCONJ) ou realizando coordenação entre orações (SCONJ).

5.9.7 Não

Quadro 35 – Propostas de anotação, na metodologia da UD para NÃO

Não-1	
PoS	ADV
atributo	Polarity=Neg
<p>Exemplos:</p> <p>Não me deu um centavo para o tratamento (FK)</p> <p>Não pode não, isto é crime (FK)</p> <p>Pergunta ao Arnaldo Cezar Coelho: Arnaldo, isso pode? Ele vai dizer: pode não.</p> <p>Paes em 2016 cometeu uma série de erros inteiramente não forçados. (FK)</p>	
Não-2	
PoS	INTJ
atributo	Polarity=Neg
<p>Exemplos:</p> <p>Eu falei: não, acho que isso está muito exagerado, afirmou Palocci (UPB)</p> <p>Não, não consigo, desligue você (UPB)</p>	

Fonte – O Autor

O *não*, a forma prototípica de realizar a negação em português, ocorre como advérbio anteposto a um *verbo*, em contexto de negação oracional, ou outro elemento, em contexto de negação não oracional. Ele também pode ocorrer em dupla negação (NEG+VERB+NEG) como também apenas em negação posposta ao verbo. Para os usos em que o *não* substitui um enunciado negativo, propõe-se utilizar a notação de interjeição (INTJ), como feita por

Rademaker *et al.* (2017) no Portuguese UD Bosque, com a adição do atributo de polaridade negativa (Polarity=Neg)

5.9.8 Preposições acidentais

Todas as *preposições acidentais de exclusão* terão o mesmo funcionamento, assim todas terão a mesma anotação, como é possível verificar nos próximos quadros.

Quadro 36 – Propostas de anotação, na metodologia da UD para a preposição acidental EXCLUSIVE

Exclusive	
PoS	ADP
atributo	Polarity=Neg
Exemplos: Trabalhadores sem carteira de trabalho, exclusive domésticos. (FK) Para a América Latina, exclusive Brasil, tivemos 2,6%, 2,0%, 4,1% e 3,2%. (FK)	

Fonte – O Autor

Quadro 37 Propostas de anotação, na metodologia da UD para a preposição acidental EXCLUSO

Excluso	
PoS	ADP
atributo	Polarity=Neg
Exemplos: Trabalhadores sem carteira de trabalho, excluso domésticos. Acho que todos valem a leitura, excluso Reinaldo Azevedo.	

Fonte – O Autor

Quadro 38 – Propostas de anotação, na metodologia da UD para a preposição acidental EXCETO

Exceto	
PoS	ADP
atributo	Polarity=Neg
<p>Exemplos:</p> <p>Acho que todos valem a leitura, exceto Reinaldo Azevedo. (<i>FK</i>)</p> <p>Os PMs disseram aos estudantes que não vão permitir a entrada de outras pessoas na escola, exceto os funcionários. (<i>FK</i>)</p>	

Fonte – O Autor

Quadro 39 – Propostas de anotação, na metodologia da UD para a preposição acidental FORA

Fora	
PoS	ADP
atributo	Polarity=Neg
<p>Exemplos:</p> <p>É difícil encontrar no Brasil, fora a escravidão, um fenômeno social tão destrutivo quanto a motocicleta. (<i>FK</i>)</p> <p>Desenvolvido em conjunto com a Força Aérea Brasileira, que encomendou 28 unidades, o projeto absorveu mais de R\$ 4 bilhões em verbas oficiais de 2011 para cá, fora o investimento inicial (<i>FK</i>)</p>	

Fonte – O Autor

Quadro 40 – Propostas de anotação, na metodologia da UD para a preposição acidental AFORA

Afora	
PoS	ADP
atributo	Polarity=Neg
<p>Exemplos:</p> <p>A operação não terá fim, afora um golpe ou acordão políticos ou erro, dada a extensão da pirataria contra o Estado. (<i>FK</i>)</p> <p>Isto feito, é preciso esticar para oeste, dar uma volta pela Pensilvânia histórica e rural antes de ver Washington, seus museus magníficos de arte, história e ciência, afora monumento do poder mundial. (<i>FK</i>)</p>	

Fonte – O Autor

Quadro 41 – Propostas de anotação, na metodologia da UD para a preposição acidental SALVO

Salvo	
PoS	ADP
atributo	Polarity=Neg
<p>Exemplos:</p> <p>A crise não irá acabar de uma vez, salvo eventos excepcionais. (<i>FK</i>)</p> <p>O limite de gastos vigorará, no mínimo, até 2026, salvo em caso de mudança constitucional. (<i>FK</i>)</p>	

Fonte – O Autor

Quadro 42 – Propostas de anotação, na metodologia da UD para a preposição acidental MENOS

Menos	
PoS	ADP
atributo	Polarity=Neg
<p>Exemplos:</p> <p>Aqui você encontra comida árabe, americana, mexicana, menos a comida regional. (<i>FK</i>)</p> <p>E eu adoro o Japão porque tudo fala: porta do metrô fala, telão de LED fala, menos os japoneses, que não falam! (<i>FK</i>)</p>	

Fonte – O Autor

Quadro 43 – Propostas de anotação, na metodologia da UD para a preposição acidental SENÃO

Senão	
PoS	ADP
atributo	Polarity=Neg
<p>Exemplos:</p> <p>São quatro sócios, três deles sem ligações anteriores com a gastronomia, senão o gosto pela comida. (<i>FK</i>)</p> <p>A auto-estima nunca pode deixar de cair, depois que volta para seleção, a auto-estima tem que estar elevada, senão é melhor ficar em casa. (<i>FK</i>)</p>	

Fonte – O Autor

Quadro 44 – Propostas de anotação, na metodologia da UD para a preposição acidental TIRANTE

Tirante	
PoS	ADP
atributo	Polarity=Neg
<p>Exemplos:</p> <p>No episódio da Chapecoense, tirante meia dúzia de imbecis, o povo (de aqui e do mundo) tem tirado a razão de Néelson. (<i>FK</i>)</p> <p>Não faz tanto tempo assim, um time perdia três ou quatro partidas seguidas, tirante a primeira, contra um dos grandes, as demais eram equivalentes a jogos com Sampaio Correia, Figueirense e Avaí, as outras poderiam ser desculpáveis. (<i>FK</i>)</p>	

Fonte – O Autor

As *preposições acidentais de exclusão* sempre ocorrem em contexto prepositivo em que elas, assim como o *nem*, excluem o elemento a que elas se vinculam da oração principal. Mesmo que tenham sido investigadas nove preposições acidentais neste trabalho, acredita-se que é possível que existam outras lexias capazes de realizar a mesma função. Dessa forma, propõe-se aqui um teste para a identificação deste tipo de preposição. Sempre que um elemento de uma classe externa a preposição estiver exprimindo comportamento propositivo, ele será uma preposição acidental se, e somente se, ele pode ser substituído pela expressão *com exceção de* fazendo com que a oração tenha a mesma interpretação.

Com todos os resultados elencados, é possível unificar todos os indicadores negativos aqui investigados, e seus respectivos *PoS* e atributos, assim como na ferramenta *PortiLexicon-UD*:

Quadro 45 – PoS e Atributo(s) para as lexias investigadas

Lexia	Acepção	PoS	Lema	atributo(s)
Nunca	<i>Nunca-1</i>	ADV	Nunca	Polarity=Neg
Jamais	<i>Jamais-1</i>	ADV	Jamais	Polarity=Neg
Sequer	<i>Sequer-1</i>	ADV	Sequer	Polarity=Neg
Tampouco	<i>Tampouco-1</i>	ADV	tampouco	Polarity=Neg
	<i>Tampouco-2</i>	CCONJ		Polarity=Neg
Nada	<i>Nada-1</i>	ADV	nada	Polarity=Neg
	<i>Nada-2</i>	PRON		Gender=none Number=Sing PronType=Neg Definite=Ind
Ninguém	<i>Ninguém-1</i>	PRON	ninguém	Gender=none Number=Sing PronType=Neg Definite=Ind
Nenhum	<i>Nenhum-1</i>	DET	nenhum	Gender=Masc/Fem Number=Sing PronType=Neg Definite=Ind
	<i>Nenhum-2</i>	PRON		Gender=none Number=Sing PronType=Neg Definite=Ind
Sem	<i>Sem-1</i>	ADP	sem	Polarity=Neg
Nem	<i>Nem-1</i>	ADV	nem	Polarity=Neg
	<i>Nem-2</i>	CCONJ		Polarity=Neg
	<i>Nem-3</i>	SCONJ		Polarity=Neg
Não	<i>Não-1</i>	ADV	não	Polarity=Neg
	<i>Não-2</i>	INTJ		Polarity=Neg
Exclusive	<i>Exclusive-1</i>	ADP	exclusive	Polarity=Neg
Excluso	<i>Excluso-1</i>	ADP	excluso	Polarity=Neg
Exceto	<i>Exceto-1</i>	ADP	exceto	Polarity=Neg
Fora	<i>Fora-1</i>	ADP	fora	Polarity=Neg
Afora	<i>Afora-1</i>	ADP	afora	Polarity=Neg
Menos	<i>Menos-1</i>	ADP	menos	Polarity=Neg
Senão	<i>Senão-1</i>	ADP	senão	Polarity=Neg
Tirante	<i>Tirante-1</i>	ADP	tirante	Polarity=Neg

Fonte – O Autor

6. CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

A presente dissertação investigou e analisou o comportamento de itens capazes de acionar a negação em português brasileiro, mais especificamente, as lexias que atuam como itens gramaticais. O trabalho categorizou o comportamento das lexias *não, nunca, nem, jamais, sequer, tampouco, ninguém, nada, nenhum, sem* e das preposições acidentais *exclusive, excluso, exceto, fora, afora, salvo, menos, senão e tirante* em função das características morfossintáticas: capacidade de realizar negação oracional, capacidade de realizar negação não-oracional, capacidade de realizar dupla negação, acionador de concordância negativa, concordar negativamente, exprime comportamento adverbial, exprime comportamento pronominal, capacidade de ocorrer com conjunção, capacidade de ocorrer como preposição e reforço da negação.

Para que fosse possível realizar as análises, foi realizada uma revisão de literatura acerca das *gramáticas de dependência* e da *Universal Dependencies* (UNIVERSAL DEPENDENCIES, 2021) para que fosse possível comparar e verificar o comportamento da negação no português por meio dessas metodologias. Constatando assim, após as análises que as lexias aqui investigadas sempre exprimem comportamento de palavras gramaticais, ou seja, pela metodologia proposta pela UD, elas sempre serão dependentes dos elementos que elas modificam na estrutura da sentença. Contudo, os itens *não, nada, nenhum, nunca, jamais, ninguém e menos* são capacitados a ocorrerem com polaridade negativa como *root* de sentenças quando nos contextos em que ocorrem como resposta em segundo turno de conversação, como é possível verificar nos exemplos:

140. Teve crime de responsabilidade? **Não.** (FK)
141. E sabe como se chama a agência espacial norte-coreana? **NADA!** (FK)
142. E um amigo perguntou: "Com a mega acumulada em R\$ 130 milhões, daria pra comprar quantos deputados?". **Nenhum!** (FK)
143. Mas você já perdeu o controle sobre sua arte, como ocorre com o personagem? **Nunca.** (FK)
144. Isso quer dizer que o pessoal do cimento é inocente? **Jamais.** (FK)
145. O celular descarregou: olha que maravilha. **Ninguém.** (FK)
146. A festança chegou a ser comparada ao baile da Ilha Fiscal, promovido por dom Pedro 2º em 1889 e que passou para a História como o rega-bofe nonsense da derrocada do Império. **Menos.** (FK)

Outra configuração relevante entre a negação e as estruturas de dependência é a concordância negativa (MIOTO, 1992; NEVES, 200), verificada na seção 4.4. O comportamento do português em exprimir concordância entre os acionadores negativos e qualquer outro elemento negativo que ocorra após ele pode por gerar problemas de projetividade dentro da estrutura sintática de dependência, todavia, como levantado por Rademaker *et al.* (2017), a marcação dos atributos de polaridade negativa em todos os indicadores negativos é suficiente para que sejam evitados problemas de projetividade sem que seja perdida a informação da concordância negativa.

Verificou-se que, conforme descrito por Muller (1991), a negação pode ser acionada de duas distintas formas: (i) negação sintática, quando o indicador negativo é um item com função gramatical específica para o acionamento da negação, e (ii) negação lexical, quando o indicador negativo é um item de uma categoria aberta e não ocorre com a função gramatical específica para o acionamento da negação.

Para a primeira categoria, os acionadores negativos têm distribuição sintática limitada (MIOTO, 1992; 1998), ocorrendo, na maior parte das vezes como um clítico, mas também podendo ocorrer em contextos de concordância negativa, item preposicional e elemento conjuntivo. Esses elementos são limitados a contextos específicos e com distribuições sintáticas bem definidas, o que facilita sua detecção.

Para a segunda categoria, os acionadores negativos são itens lexicais com comportamento diversificado dentro das categorias abertas (nomes, verbos, advérbios e adjetivos). Mesmo que não tenham comportamento sintático distintos, a negação faz parte de seu significado, ou seja, ela está dentro da categoria semântica e não sintática. Tal fato torna a detecção e o mapeamento desses itens mais complexa, dificultando sua detecção.

Durante a análise dos dados verificou-se que o indicador *não*, a forma prototípica de realização da negação, e o indicador *nada* na forma adverbial (*nada-2*, cf. subseção 5.4.1) constam em ambos os *corpora* com a notação do atributo de polaridade negativa (Polarity=Neg), sendo necessário a adição dos atributos que indicam a polaridade negativa nos outros indicadores. Com relação aos pronomes (*nada*, *ninguém* e *nenhum*), propôs-se a adição do atributo de definitude indefinida (Definite=Ind) para que seja possível a marcação desses pronomes como indicadores negativos por meio do atributo PronType=Neg.

Ainda sobre as análises, verificou-se que existe um conjunto de lexias que não são limitadas a ocorrerem em contextos de elementos gramaticais, mas que em certos contextos são indicadores negativos, sendo elas as preposições acidentais de exclusão *exclusive*, *excluso*, *exceto*, *fora*, *afora*, *salvo*, *menos*, *senão* e *tirante*. É salientável que o comportamento como

preposição acidental capaz de exprimir negação não necessariamente está limitado aos itens aqui investigados, é possível que outras lexias também sejam capazes de realizar essa função.

Acredita-se que os objetivos propostos foram alcançados, uma vez que (i) foram investigados os trabalhos acerca da anotação/detecção da negação dentro do português, demonstrando que ainda há poucos trabalhos sobre o assunto; e (ii) demonstrou-se, por meio das metodologias da UD que é possível demarcar, utilizando um conjunto de etiquetas morfossintáticas, elementos gramaticais, ou que expressem função gramatical, que são indicadores negativos no português.

Pode-se considerar que a hipótese inicial foi confirmada, uma vez que, como verificado no capítulo 4, a negação ocorre em mais de uma classe gramatical fechada (advérbios de negação, preposições, conjunções e pronomes) e também ocorrer em itens lexicais (substantivos, adjetivos e verbos).

No capítulo 5 foram apresentadas as análises dos dados assim como os resultados obtidos. Cada uma das léxicas recebeu uma proposta de anotação segundo a metodologia da UD para suas diferentes ocorrências (advérbio, preposição, pronome, determinante e conjunção). Acredita-se que as propostas de anotação contribuem, para além da anotação final do Portinari (PARDO *et al.*, 2021), diretamente para as futuras tarefas de anotação em função da metodologia da UD, uma vez que trazem informações relevantes sobre o comportamento da negação no português.

Das propostas aqui levantadas, além da adição do atributo de polaridade negativa (`Polarity=Neg`), para os indicadores negativos que acionam a negação sintática, e o atributo de tipo de pronome negativo (`PronType=Neg`), para os pronomes capazes de acionar a negação, também se verificou a necessidade da adição do atributo de definitude indefinida (`Definite=Ind`) para que possa ser realizada a modificação dos pronomes negativos, também fazem parte da categoria dos pronomes indefinidos, sem que haja perda de informação relevante.

Para trabalhos futuros, acredita-se que a inserção das etiquetas aqui propostas para as lexias na anotação de *corpora* em função da metodologia UD permitirá a construção de sistemas capazes de lidar com o processamento da negação. Todavia, a descrição aqui realizada não é suficiente para o fenômeno da negação como um todo, fenômenos como *expressões multipalavra*, *verbos/nomes/adjetivos negativos*, *estruturas comparativas* e muitos outros são indicadores negativos que também contribuíram com sistemas de processamento e detecção da negação. Em função disso, é necessário e relevante o desenvolvimento de trabalhos que estudem a negação para aplicações em PLN.

REFERÊNCIAS

- AFONSO, S.; BICK, E.; HABER, R.; SANTOS, D. Floresta sintá(c)tica: a treebank for Portuguese. In: **Proceedings of the Third International Conference on Language Resources and Evaluation** Paris: ELRA. ELRA, 2002.
- ALMEIDA NETO, A. M. dos S; OSTI, B. A; AZEVEDO, C. A; NOBREGA, F. A. A; BRUM, H. B; PEINADO, L. H. O; BITTENCOURT, M. de M.; DA SILVA, N. L. P; CORTES, P. O. **SiDi-NLP-Team at IDPT2021: Irony Detection in Portuguese 2021**. IberLEF@SEPLN 2021: 933-939
- ALUÍSIO, S.; PELIZZONI, J. MARCHI, A. R.; DE OLIVEIRA, L.; MANETI, R.; MARQUIAFÁVEL, V. An Account of the Challenge of Tagging a Reference Corpus for Brazilian Portuguese. In: **Computational Processing of the Portuguese Language: 6th International Workshop, PROPOR 2003**, Faro, Portugal, June 26-27, 2003. Proceedings. Springer Science & Business Media, 2003. p. 110.
- AVILA, L. B.; MELLO, H. Challenges in modality annotation in a Brazilian Portuguese spontaneous speech corpus. In: **Proceedings of the International Conference on Computational Semantics (IWCS) 2013 Workshop on Annotation of Modal Meanings in Natural Language (WAMM)**. 2013. p. 1-6.
- AZEREDO, J. C. Gramática Houaiss de Língua Portuguesa. 2 ed. São Paulo: Publifolha, 2008.
- BAKER, K.; BLOODGOOD, M.; DORR, B. J.; CALLISON-BURCH, C.; FILARDO, N. W.; PIATKO, C.; LEVIN, L.; MILLER, S. Use of Modality and Negation in Semantically-Informed Syntactic MT. **Computational Linguistics**, [s. l.], v. 38, ed. 2, p. 411-438, 2015. DOI <https://doi.org/10.48550/arXiv.1502.01682>. Disponível em: <https://arxiv.org/abs/1502.01682>. Acesso em: 21/04/2022.
- BANJADE, R.; RUS, V.. Dt-neg: Tutorial dialogues annotated for negation scope and focus in context. In: **Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)**. 2016. p. 3768-3771.
- BANSAL, S.; SUHANE, A.; PATRO, J.; MUKHERJEE, A. Code-switching patterns can be an effective route to improve performance of downstream NLP applications: A case study of humour, sarcasm and hate speech detection. In: **Proceeding of the 58th Annual Meeting of the Association for Computational Linguistics (ACL) 2020**. 2020.
- BARME, S. A negação no brasileiro falado informal. **Zeitschrift für romanische Philologie**, v. 121, n. 3, p. 405-425, 2005.
- BECHARA, E. Moderna Gramática Portuguesa. 37^a. ed. Rio de Janeiro: Lucerna, 2009.
- BELISÁRIO, L. B.; FERREIRA, L.G.; PARDO, T.A.S. Evaluating richer features and varied machine learning models for subjectivity classification of book review sentences in portuguese. **Information**, v. 11, n. 9, p. 437, 2020.
- BICK, E.. Constraint grammar-based conversion of dependency treebanks. In: **Proceedings of the 13th International Conference on Natural Language Processing**. 2016. p. 109-114.
- BICK, E.. PALAVRAS, a constraint grammar-based parsing system for Portuguese. **Working with Portuguese corpora**, p. 279-302, 2014.

BLANCO, E.; MOLDOVAN, E.. Some issues on detecting negation from text. In: **Proceedings of the 24th International Florida Artificial Intelligence Research Society, FLAIRS-24**. 2011. p. 228-233.

BOKHARAEIAN, B.; DIAZ, A.; NEVEZ, M.; FRANCISCO, V. Exploring negation annotations in the DrugDDI Corpus. In: **Proceedings of the Fourth workshop on building and evaluating resources for health and biomedical text processing (BIOTxtM 2014)**. Citeseer, 2014. p. 1-8.

BONA, C. **Os prefixos de negação des- e in- no PB: Considerações morfossemânticas**. Dissertação de Mestrado - Universidade Federal do Rio Grande do Sul. 2014.

BOSQUE, I. Verbos de ausencia, carencia y omisión. In: **Lengua, cultura, discurso: estudios ofrecidos al profesor Manuel Casado Velarde**. EUNSA. Ediciones Universidad de Navarra, SA, 2019. p. 525-540.

BUCHHOLZ, S.; MARSI, E.. CoNLL-X shared task on multilingual dependency parsing. In: **Proceedings of the tenth conference on computational natural language learning (CoNLL-X)**. 2006. p. 149-164.

CANCEIRO, N.; MATOS, G.; COLAÇO, M. Especificidades de nem coordenativo em português europeu. **Revista da Associação Portuguesa de Linguística**, n. 4, p. 34-47, 2018.

CAVAGUTI, A.P.. **Uma análise semântico-pragmática das construções condicional-concessivas no português do Brasil**. 2018. Tese (Doutorado em Linguística) - Programa de Pós-graduação em Linguística, Universidade Federal de São Carlos, 2018.

CHAPMAN, W. W.; BIRDEWELL, W.; HANBURY, P.; COOPER, G. F; BUNCHANAN, G. B. A simple algorithm for identifying negated findings and diseases in discharge summaries. **Journal of biomedical informatics**, v. 34, n. 5, p. 301-310, 2001.

CHOMSKY, N.. Syntactic structures. De Gruyter Mouton, 1957.

CONEGLIAN, A. V.L.; GUIMARÃES, A.L.A.R.; FERREIRA, T.C.; PAGANO, A.S. Anotação de textos não canônicos: um estudo exploratório de Grande sertão: veredas pelas dependências universais. In: **Proceedings of the Universal Dependencies Brazilian Festival**. 2022. p. 1-11.

COUNCILL, I. G.; MCDONALD, R.; VELIKOVICH, L. What's Great and What's Not: Learning to Classify the Scope of Negation for Improved Sentiment Analysis. **Negation and Speculation in Natural Language Processing (NeSp-NLP 2010)**, p. 51, 2010.

CUNHA, C.; CINTRA, L. Nova Gramática do Português Contemporâneo. 7ª. ed. Rio de Janeiro: Lexicon, 2016.

DE LHONEUX, M.; BJERVA, J.; AUGENSTEIN, I; SØGAARD, A.; Parameter sharing between dependency parsers for related languages. In: **Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP) 2018**. 2018.

DE MARNEFFE, M.C.; MANNING, C. D.; NIVRE, J.; ZEMAN, D. Universal dependencies. **Computational linguistics**, v. 47, n. 2, p. 255-308, 2021.

DE MARNEFFE, M.C.; DOZAT, T.; SILVEIRA, N.; HAVERINEN, K.; GINTER, F.; NIVRE, J.; MANNING, C. Universal Stanford dependencies: A cross-linguistic typology. In:

Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC) 2014. 2014. p. 4585-4592.

DE MARNEFFE, M.C.; MACCARTNEY, B.; MANNING, C.D. Generating typed dependency parses from phrase structure parses. In: **Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC) 2006.** 2006. p. 449-454.

DE MARNEFFE, M.C.; MANNING, C. D. The Stanford typed dependencies representation. In: **Coling 2008: proceedings of the workshop on cross-framework and cross-domain parser evaluation.** 2008. p. 1-8.

DE MARNEFFE, M.C.; NIVRE, J. Dependency grammar. **Annual Review of Linguistics**, v. 5, p. 197-218, 2019. DOI: <https://doi.org/10.1146/annurev-linguistics-011718-011842>

DE MOURA, J.R.F. Língua (gem) e gênero neutro: uma perspectiva discursiva no português brasileiro. **Línguas e Instrumentos linguísticos**, v. 24, n. 47, 2021.

DE PAIVA, V.; REAL, L.. Universal POS tagging for Portuguese: Issues and Opportunities. In: **Proceedings of the 1th Workshop on Lexical Semantics for Lesser-Resourced Languages (LexSem) and 3rd Workshop on Logics and Ontologies (Logic) 2016.** 2016, p. 25-34.

DE SANTANA, M. S. O GÊNERO NEUTRO: PINTANDO O PORTUGUÊS PARA A LUTA. **Diversidade e Educação**, v. 9, n. 2, p. 695-710, 2021.

DE SOUZA, L.M; GRITTI, L.L.; DE OLIVEIRA, R.P. em **Linguística**, v. 9, n. 2, p. 23-40, 2008.

DEBUSMANN, R. An introduction to dependency grammar. **Hausarbeit fur das Hauptseminar Dependenzgrammatik SoSe**, v. 99, p. 1-16, 2000.

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, 55p. 2021a

DURAN, M.S. Manual de Anotação de Relações de Dependência: Orientações para anotação de relações de dependência sintática em Língua Portuguesa, seguindo as diretrizes da abordagem Universal Dependencies (UD). **Relatório Técnico do ICMC 435. Instituto de Ciências Matemáticas e de Computação**, Universidade de São Paulo. São Carlos-SP, Dezembro, 79p. 2021b

DURAN, M.S.; LOPES, L.; PARDO, T.A.S.. Descrição de numerais segundo modelo Universal Dependencies e sua anotação no português. In: SIMPÓSIO BRASILEIRO DE TECNOLOGIA DA INFORMAÇÃO E DA LINGUAGEM HUMANA (STIL), 13. , 2021, Evento Online. **Anais do Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana (STIL)**. Porto Alegre: Sociedade Brasileira de Computação, 2021 . p. 344-352. DOI: <https://doi.org/10.5753/stil.2021.17814>

DURAN, M.S.; OLIVEIRA, H.; SCANDAROLLI, C.. Que simples que nada: a anotação da palavra que em corpus de UD. In: **Proceedings of the Universal Dependencies Brazilian Festival.** 2022.

FREGE, G. Sobre o sentido e a referência. Tradução: Sérgio R. N. Miranda. **Fundamento**, n. 3, 2012.

GAIFMAN, H. Dependency systems and phrase-structure systems. **Information and control**, v. 8, n. 3, p. 304-337, 1965.

GERDES, K.; KAHANE, S. Defining dependencies (and constituents). **Frontiers in Artificial Intelligence and Applications**, v. 258, p. 1-25, 2013.

GOLDIN, I.; CHAPMAN, W. W. Learning to detect negation with 'not' in medical texts. In: **Proceedings of the 26th Advancing Computing as a Science and Profession Special Interest Group on Information Retrieval (ACM SIGIR) 2013 Workshop on Text Analysis and Search for Bioinformatics**. 2003.

GRICE, P.. *Studies in the Way of Words*. Harvard University Press, 1989.

HAN, Wenjuan; JIANG, Yong; TU, Kewei. Enhancing unsupervised generative dependency parser with contextual information. In: **Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics**. 2019. p. 5315-5325.

HAYS, D. G. Dependency theory: A formalism and some observations. **Language**, v. 40, n. 4, p. 511-525, 1964.

HEINE, B.; KUTEVA, T.. *The Genesis of Grammar: A Reconstruction*. OUP Oxford, 2007.

HELLWIG, P. Dependency unification grammar. In: **Coling 1986 Volume 1: The 11th International Conference on Computational Linguistics**. 1986.

HIDALGO, M.J. Verbos de estado y sus nominalizaciones. Una primera aproximación. **Hesperia: Anuario de Filología Hispánica**, v. 13, p. 101-116, 2010.

HORN, L. R.; WANSING, H. "Negation", *The Stanford Encyclopedia of Philosophy*, Spring. 2020. Disponível em: <https://plato.stanford.edu/entries/negation/#toc>. Acesso em: 18/02/2022

HORN, L. R. *A Natural History of Negation*. 2ª. ed. Califórnia: CSLI Publications, 2001.

HOUAISS, *Dicionário Eletrônico da Língua Portuguesa*: Editora Objetiva, 2009.

HUDSON, R. **English Word Grammar**. Oxford, Uk, 1990.

JÄRVINEN, T.; TAPANAINEN, P.. Towards an implementable dependency grammar. In: **Processing of Dependency-Based Grammars**. 1998.

JIMENEZ-ZAFRA, S. M.; MARTÍN-VALDIVIA M. T.; MARTÍNEZ-CÁMARA, E.; UREÑA-LOPEZ, L. A. Studying the scope of negation for Spanish sentiment analysis on Twitter. **IEEE Transactions on Affective Computing**, v. 10, n. 1, p. 129-141, 2017.

JIMÉNEZ-ZAFRA, S. M. MORANTE, R.; MARTÍN-VALDIVIA, S. T.; UREÑA-LOPEZ, L. A. Corpora annotated with negation: An overview. **Computational Linguistics**, v. 46, n. 1, p. 1-52, 2020.

KAHANE, B.. Bubble trees and syntactic representations. In: **Proceedings of mathematics of language (mol5) meeting**. DFKI, Saarbrücken, 1997. p. 70-76.

KENNEDY, C.. On the monotonicity of polar adjectives. **Perspectives on negation and polarity items**, p. 201-221, 2001.

KIM, J.-D.; OHTA, T.; TATEISE, Y.; TSUJII, J. GENIA corpus—a semantically annotated corpus for bio-textmining. **Bioinformatics**, v. 19, n. suppl_1, p. i180-i182, 2003.

KOLHATKAR, V.; WU, H.; CAVASSO, L.; FRANCIS, E.; SHUKLA, K.; TABOADA, M. The SFU Opinion and Comments Corpus: A Corpus for the Analysis of Online News Comments. **Corpus Pragmatics: International Journal of Corpus Linguistics and Pragmatics**, v. 4, n. 2, p. 155-190, 2019.

KONSTANTINOVA, N.; DE SOUZA, S. C. M.; CRUZ, N. P.; MAÑA, M. J.; TABOADA, M.; MITVOK, R. A review corpus annotated for negation, speculation and their scope. In: **Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)**. 2012. p. 3190-3195.

KOOSTER, W.. Negative adjectives. In: BAGRIACIK, M.; BREITBARTH, A.; DE CLERCQ, K. **Mapping Linguistic Data: Essays in honour of Liliane Haegeman**. [S. l.: s. n.], 2019. Disponível em: https://www.researchgate.net/profile/Karen-Clercq/publication/335422380_Mapping_Linguistic_Data_Essays_in_honour_of_Liliane_Haegeman/links/5dc2ce4a299bf1a47b1be6ff/Mapping-Linguistic-Data-Essays-in-honour-of-Liliane-Haegeman.pdf#page=146. Acesso em: 18 jul. 2022.

KROMANN, M. T.; MIKKELSEN, L.; LYNGE, Stine Kern. Danish dependency treebank. In: **Proceedings of the Second Workshop on Trekkans and Linguistic Theories (TLT) 2003**. 2003. p. 217-220.

KÜBLER, S.; MCDONALD, R.; NIVRE, J.. Dependency parsing. **Synthesis lectures on human language technologies**, v. 1, n. 1, p. 1-127, 2009.

LECERF, Y.. Programme des conflits, modele des conflits. **Bulletin bimestriel de l'ATALA**, v. 1, n. 4, p. 11-18, 1960.

LI, Y.; THOMAS, M. A.; LIU, D.. From semantics to pragmatics: where IS can lead In Natural Language Processing (NLP) research. **European Journal of Information Systems**, v. 30, n. 5, p. 569-590, 2021.

LIDDY, E. D.; PAINK, W.; MCKENNA, M. E.; WEINER, M. L.; YU E. S.; DIAMOND T. G.; BALAKRISHNAN, B.; SNYDER D. L. User interface and other enhancements for natural language information retrieval system and method. U.S. Patent n. 6,026,388, 15 fev. 2000. Disponível em <https://patents.google.com/patent/US6026388A/en>. Acesso em 16/03/2022.

LIMA, C. H. R. Gramática normativa da língua portuguesa. 49ª. ed. Rio de Janeiro: José Olympio, 2011.

LIU, B. Sentiment analysis: Mining opinions, sentiments, and emotions. Cambridge university press, 2020.

LOPES, L.; DURAN, M. S.; NUNES, M.G.V; PARDO, T.A.S.. Corpora building process according to the Universal Dependencies model: an experiment for Portuguese. 2022a. Disponível em: <https://repositorio.usp.br/item/003067405>. Acesso em: 18/03/2022.

LOPES, L.; DURAN, M. S.; FERNANDES, P.; PARDO, T.A.S.. PortiLexicon-UD: a portuguese lexical resource according to Universal Dependencies Model. In: **Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022)**. 2022b.

MANUAL de Redação: As normas de Escrita e Conduta do Principal Jornal do País. 22^a. ed. São Paulo: Publifolha, 2021. 512 p. ISBN 978-65-89548-00-3.

MATTOS OLÍMPIO, A.. O item nem no português brasileiro. **Estudos Linguísticos XXXIV**, 2005, p. 962-967.

MCDONALD, R.; Nivre J.; Quirmbach-Brundage, Y.; Goldberg, Y.; Das, D.; Ganchev, K.; Petrov, S.; Zhang, H.; Täckström O.; Bedini, C.; Castelló, N. B.; Lee, J. Universal dependency annotation for multilingual parsing. In: **Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)**. 2013. p. 92-97.

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

MENDES, A.; HENDRICKX, I.; ÁVILA, L.; QUARESMA, P.; GONÇALVES, T.; SEQUEIRA, J. Modality annotation for Portuguese: from manual annotation to automatic labeling. **Linguistic Issues in Language Technology**, v. 14, 2016.

MENDIKOETXEA, A. Construcciones inacusativas y pasivas. In: **Gramática descriptiva de la lengua española**. Espasa Calpe, 1999. p. 1575-1630.

MICHAELIS On-line: Editora Melhoramentos, 2022. Disponível em: <https://michaelis.uol.com.br/>. Acesso em: 14 fev. 2022.

MIESTAMO, M. Standard Negation: The Negation of Declarative Verbal Main Clauses in a Typological Perspective. Berlin. 2005.

MIESTAMO, M. Negation. Cambridge handbook of linguistic typology, 2017.

MIOTO, C. **Negação sentencial no português brasileiro e teoria da gramática**. Doutorado em Linguística-UNICAMP, 1992.

MIOTO, C. Tipos de negação. **Cadernos de Estudos Linguísticos**, v. 34, 1998.

MORANTE, R.; DAELEMANS, W. ConanDoyle-neg: Annotation of negation in Conan Doyle stories. In: **Proceedings of the eighth international conference on language resources and evaluation**, Istanbul. 2012. p. 1563-1568

MUKHERJEE, P.; BARD, Y.; DOPPALAPUDI, S; SRINIVASAN, S. M.; SANGWAN R. S.; SHARAMA R. Effect of negation in sentences on sentiment analysis and polarity detection. **Procedia Computer Science**, v. 185, p. 370-379, 2021.

MULLER, C. **La négation en français: syntaxe, sémantique et éléments de comparaison avec les autres langues romanes**. Librairie Droz, 1991.

MUNIZ, H.; CHALUB, F.; RADEMAKER, A. Cl-conllu: dependências universais em common lisp. In: **V Workshop de Iniciação Científica em Tecnologia da Informação e da Linguagem Humana (TILic)**. Uberlândia, MG, Brazil. 2017.

- MUTALIK, P. G.; DESHPANDE, A.; NADKARNI, P. M. Use of general-purpose negation detection to augment concept indexing of medical documents: a quantitative study using the UMLS. **Journal of the American Medical Informatics Association**, v. 8, n. 6, p. 598-609, 2001.
- NAMIUTI, C.; MIOTO, C.. Clíticos e negação em português: elementos para uma descrição gramatical. **Filologia e Linguística Portuguesa**, v. 16, n. esp., p. 95-123, 2014.
- NEVES, M. H. M. Gramática de usos do português. São Paulo: Unesp, 2000.
- NEVES, M. H. M. Os advérbios circunstanciais (de lugar e de tempo). In: ILARI, R. (Org.). Gramática do português falado, vol. II: Níveis de análise linguística. Campinas: UNICAMP, 1993, p. 249-284.
- NIVRE, J.; DE MARNEFFE, M.-C.; GINTER, F.; GOLDBERG, Y.; HAJIČ, J.; MANNING, C. D.; MCDONALD, R.; PETROV, S.; PYYSALO, S.; SILVEIRA, N.; TSARFATY, R.; ZEMAN, D. Universal dependencies v1: A multilingual treebank collection. In: **Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)**. 2016. p. 1659-1666.
- NIVRE, J.; DE MARNEFFE, M.-C.; GINTER, F.; HAJIČ, J.; MANNING, C. D.; PYYSALO, S.; SCHUSTER, S.; TYERS, F.; ZEMAN, D. Universal Dependencies v2: An Evergrowing Multilingual Treebank Collection. In: **Proceedings of the 12th Language Resources and Evaluation Conference. 2020**. p. 4034-4043.
- NIVRE, J. An efficient algorithm for projective dependency parsing. In: **Proceedings of the eighth international conference on parsing technologies**. 2003. p. 149-160.
- NIVRE, J.; MEGYESI, B. Bootstrapping a Swedish Treebank using cross-corpus harmonization and annotation projection. In: **Proceedings of the 6th international workshop on treebanks and linguistic theories**. Pennsylvania, PA: Association for Computational Linguistics, 2007. p. 97-102.
- PARDO, T. A. S.; DURAN, S. D.; LOPES, L. DI FELIPPO, A.; ROMAN, N. R.; NUNES, M. G. C. Porttinari-a Large Multi-genre Treebank for Brazilian Portuguese. In: **Anais do XIII Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana. SBC**, 2021. p. 1-10.
- PAYNE, J. R. Negation. **Language typology and syntactic description**, v. 1, p. 197-242, 1985.
- PETROV, S.; DAS, D.; MCDONALD, R.. A Universal Part-of-Speech Tagset. In: **Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC)**. 2012. p. 2089-2096.
- POLGUÈRE, A.; MEL'ČUK, I.A. (Ed.). Dependency in linguistic description. John Benjamins Publishing, 2009.
- PYYSALO, S.; GINTER, F.; HEIMONEN, J.; BJÖRNE, J.; BOBER, J.; JÄRVINEN, J.; SALAKOSKI T. BioInfer: a corpus for information extraction in the biomedical domain. **BMC bioinformatics**, v. 8, n. 1, p. 1-24, 2007.

- RADEMAKER, A.; CHALUB, F.; REAL, L.; FREITAS, C.; BICK, E.; PAIVA, V. Universal dependencies for Portuguese. In: **Proceedings of the Fourth International Conference on Dependency Linguistics (Depling 2017)**. 2017. p. 197-206.
- REAL, L.; RADEMAKER, A.; CHALUB, F.; PAIVA, V. Towards temporal reasoning in portuguese. In: **Proceedings of the 11th Internarnational Conferenve on Language Ressources and Evaluation (LREC2018) Workshop Linked Data in Linguistics**. 2018.
- REAL, L.; OSHIRO, M.; MAFRA, A.. B2W-Reviews01-An open product reviews corpus. In: **Proceedings of the XII Symposium in Information and Human Language Technology**. 2019. p. 200-208.
- ROBINSON, J. J. Dependency structures and transformational rules. **Language**, p. 259-285, 1970.
- RUSSELL, B.. On denoting. **Mind**, v. 14, n. 56, p. 479-493, 1905.
- SANTANA, M. R. O. News of the Brazilian Newspaper. 2019. Disponível em: <https://www.kaggle.com/datasets/marlesson/news-of-the-site-folhauol>. Acesso em: 20/01/2022.
- SARAIVA, M. E. F.. A distribuição dos advérbios de modo em "-mente". **Revista de Estudos de Língua Portuguesa**, v. 2, n. 2, p. 76-93, 1983.
- SASSOON, G. W. The degree functions of negative adjectives. **Natural language semantics**, v. 18, n. 2, p. 141-181, 2010.
- SAVOVA, G. K.; MASANZ, J. J.; OGREN, P. V.; ZHENG, J.; SOHN, S.; KIPPER-SCHULER, K. C.; CHUTE, C. G. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. **Journal of the American Medical Informatics Association**, v. 17, n. 5, p. 507-513, 2010.
- SCHER, A, P.; MONTEIRO, B, N.. O estatuto morfossintático dos prefixos negativos des-e in-em português. **Revista do GELNE**, v. 22, n. 2, p. 280-293, 2020.
- SCHNEIDER, E. T. R.; DE SOUZA, J. V. A.; KNAFOU, J.; COPARA, J. OLIVEIRA L. E. S.; GUMIEL, Y. B.; OLIVEIRA, L. F. A.; TEODORO, D.; PARAISO, E. C.; MORRO, C. BioBERT-pt-a portuguese neural language model for clinical named entity recognition. In: **Proceedings of the 3rd Clinical Natural Language Processing Workshop**. 2020. p. 65-72.
- SCHWENTER, S, A. Some issues in negation in Portuguese. **The Handbook of Portuguese Linguistics**, p. 425-440, 2016.
- SGALL, P.; HAJIČOVÁ, E.; PANEVOVÁ, J. **The Meaning of the Sentence in Its Semantic and Pragmatic Aspects**. Dordrecht: D. Reidel Publishing Company, 1986.
- SILVA, F, o J. V.; ROMAN, N, T.; CARVALHO, A, M. B. R. Stock market tweets annotated with emotions. **Corpora**, v. 15, n. 3, p. 343-354, 2020.
- SOUZA, E.; SILVEIRA, A.; CAVALCANTI, T.; CASTRO, M. C.; FREITAS, C.. PetroGold – Corpus padrão ouro para o domínio do petróleo. In: SIMPÓSIO BRASILEIRO DE TECNOLOGIA DA INFORMAÇÃO E DA LINGUAGEM HUMANA (STIL), 13., 2021, Evento Online. **Anais do Simpósio Brasileiro de Tecnologia da Informação e da Linguagem**

Humana (STIL). Porto Alegre: Sociedade Brasileira de Computação, 2021. p. 29-38. DOI: <https://doi.org/10.5753/stil.2021.17781>.

SOUZA, E.; FREITAS, C.. Polishing the gold—how much revision do we need in treebanks? In: **Proceedings of the Universal Dependencies Brazilian Festival**. 2022. p. 1-11.

SOUZA, E.; FREITAS, C.. Still on arguments and adjuncts: the status of the indirect object and the adverbial adjunct relations in Universal Dependencies for Portuguese. In: **Proceedings of the Universal Dependencies Brazilian Festival**. 2022. p. 1-10.

STRAKA, M. UDPipe 2.0 prototype at CoNLL 2018 UD shared task. In: **Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies**. 2018. p. 197-207.

TABOADA, M.; BROOKE, J.; TOFILOSKI, M.; VOLL, K.; STEDE, M. Lexicon-based methods for sentiment analysis. **Computational linguistics**, v. **37**, n. **2**, p. 267-307, 2011.

TEIXEIRA, Z. D. **Propriedades sintáticas e semânticas dos advérbios no português brasileiro**. 2015. Tese (Doutorado em Linguística) - Programa de Pós-graduação em Linguística, Universidade Federal de Brasília, 2015.

TESNIÈRE, L. **Elements of structural syntax**. John Benjamins Publishing Company, 2015. [1959].

THOMPSON, H. V. G.; DE OLIVEIRA TOTA, F.; RODRIGUES, V. V.. A trajetória de gramaticalização de QUE NEM. **JORNADA NACIONAL DO GRUPO DE ESTUDOS LINGUÍSTICOS DO NORDESTE**, v. **24**, 2012.

UNIVERSAL DEPENDENCIES. 2021. Disponível em: <https://universaldependencies.org/u/feat>. Acesso em: 01/02/2022.

VINCZE, V.; SZARVAS, G.; FARKAS, R.; MÓRA, G.; CSIRIK, J. The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes. **BMC bioinformatics**, v. **9**, n. **11**, p. 1-9, 2008.

WICK-PEDRO, G. **ComentCorpus: identificação e pistas linguísticas para detecção de ironia no português do Brasil**. 2018. Tese (Doutorado em Linguística) - Programa de Pós-graduação em Linguística - UFSCar, São Carlos, 2018.

WITTGENSTEIN, L. Tractatus logico-philosophicus by Ludwig Wittgenstein. *J. Hist. Ideas*, v. 59, p. 1-28, 1922.

XAVIER, A. C.; KANTHACK, G. S. O uso de advérbios locativos: uma abordagem morfossintática e semântica. **Revista Philologus: Anais da X CNLF**, Rio de Janeiro, ano 21, n. 63, 2015.

ZEMAN, D. Reusable Tagset Conversion Using Tagset Drivers. In: **Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC) 2008**. 2008. p. 28-30.

ZESHAN, U.. Irregular Negatives in Sign Languages. In: DRYER, Matthew S.; HASPELMATH, Martin. **The World Atlas of Language Structures Online**. [S. l.], 2013. Disponível em: <http://wals.info/chapter/139>. Acesso em: 2 ago. 2022.

ZHANG, S.; WANG, L.; SUN, K.; XIAO, X. A practical chinese dependency parser based on a large-scale dataset. **arXiv preprint arXiv:2009.00901**, 2020.

ZHENG, Z.; QIU, S.; FAN, L.; ZHU, Y.; ZHU, S.-C.. GRICE: A Grammar-based Dataset for Recovering Implicature and Conversational Reasoning. In: **Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021**. 2021. p. 2074-2085.

ZWICKY, A. M. 13 Heads, bases and functors. **Heads in grammatical theory**, p. 292, 1993. [1985].