

Não diga nunca jamais: anotação de indicadores lexicais da negação no português

DOI: <http://dx.doi.org/10.21165/el.v55i1.3892>

– Seu tolo, nenhum homem pode me matar [...].
– Não sou um homem.
(*O senhor dos anéis: o retorno do rei*)

Isaac Souza de Miranda Jr.¹
Oto Araújo Vale²

Resumo

O presente artigo apresenta uma proposta de anotação de indicadores lexicais de negação no português no âmbito da metodologia Universal Dependencies (UD), visando ampliar a representatividade da informação negativa em *corpora* anotados. Tomam-se como base os *corpora* UD Portuguese Bosque (Rademaker *et al.*, 2017) e Porttinari-Base (Lopes *et al.*, 2022a), além do *corpus* News of Brazilian Newspaper (Santana, 2019), para exemplificação e validação adicional de ocorrências, e da ferramenta lexical PortiLexicon-UD (Lopes *et al.*, 2022b). Foram investigados dezenove indicadores negativos, organizados em nove categorias descritivo-funcionais (negação oracional, negação não-oracional, dupla negação, concordância negativa – acionamento e concordância, pronome negativo, conjunção, preposição e reforço da negação). A partir do levantamento e da inspeção das anotações existentes, propõem-se ajustes de PoS, lema e, sobretudo, a introdução sistemática das *features* morfológicas `Polarity=Neg` e `PronType=Neg`, de modo a capturar de forma mais consistente a contribuição desses itens para a polaridade negativa nas diretrizes UD.

Palavras-chave: Linguística Computacional; Processamento de Língua Natural; universal dependencies; negação; anotação linguística.

1 Universidade Federal de São Carlos (UFSCar), São Carlos, São Paulo, Brasil; Université Catholique de Louvain; isc_jr@live.com; <https://orcid.org/0000-0002-4004-3182>

2 Universidade Federal de São Carlos (UFSCar), São Carlos, São Paulo, Brasil; otovale@ufscar.br; <https://orcid.org/0000-0002-0091-8079>

Never Say Never Ever: annotation of lexical negation indicators in Portuguese

Abstract

This article proposes an annotation scheme for lexical negation indicators in Portuguese within the Universal Dependencies (UD) framework, aiming to improve the representation of negative information in UD-annotated *corpora*. The study draws on the UD Portuguese Bosque corpus (Rademaker *et al.*, 2017) and the Portinari-Base corpus (Lopes *et al.*, 2022a), uses the News of Brazilian Newspaper corpus (Santana, 2019) for additional validation and illustrative examples, and consults the PortiLexicon-UD lexical resource (Lopes *et al.*, 2022b). Nineteen negative indicators were examined and organized into nine descriptive-functional categories (sentential negation, non-sentential negation, double negation, negative concord—triggering vs. agreeing items, negative pronoun, conjunction, preposition, and negation reinforcement). Based on corpus evidence and the inspection of existing UD annotations, we propose adjustments to PoS, lemma, and, crucially, the systematic use of the morphological features `Polarity=Neg` and `PronType=Neg` to capture negative polarity information more consistently under UD guidelines.

Keywords: Computational Linguistics; Natural Language Processing; Universal Dependencies; Negation; Linguistic annotation.

Introdução

A habilidade de contrastar dois enunciados quanto ao seu valor de verdade (verdadeiro ou falso) é uma distinção fundamental entre a linguagem humana e os sistemas de comunicação animal (Horn, 2001, p. XIII). Estudos mostram que, embora símios e algumas aves possam ser treinados para entender conceitos como rejeição ou inexistência, eles têm dificuldade ou não conseguem criar oposição entre enunciados (Heine; Kuteva, 2007, p. 141-143), o que torna a negação uma característica única da fala humana.

A negação, embora seja parte inerente da atividade de fala, apresenta comportamentos linguísticos variados e consequências interpretativas relevantes, o que torna sua modelagem um desafio descritivo e computacional. Ainda assim, no âmbito do Processamento de Língua Natural (PLN), a representação explícita desse fenômeno em recursos linguísticos anotados permanece limitada, sobretudo no que diz respeito à codificação sistemática da polaridade negativa em *corpora* e ferramentas lexicais.

O PLN é um campo multidisciplinar da Ciência da Computação e da Linguística que investiga como computadores podem ser utilizados para compreender e manipular a língua humana em textos ou em fala (Li *et al.*, 2021, p. 1). Diversos fenômenos linguísticos,

como ironia (Wick-Pedro, 2018; Almeida Neto *et al.*, 2021), sarcasmo (Bansal, 2020) e inferências semântico-pragmáticas (Zheng *et al.*, 2021), continuam impondo dificuldades metodológicas e empíricas à área. Nesse contexto, a negação se destaca como um fenômeno particularmente sensível à anotação, pois sua interpretação depende de indicadores lexicais, de relações sintáticas e de aspectos de escopo e concordância, nem sempre capturados de maneira consistente pelos esquemas de anotação.

Jiménez-Zafra (2020) aponta que trabalhos em PLN voltados à negação se consolidaram a partir de meados de 2001, inicialmente com foco no processamento de textos clínicos (Chapman *et al.*, 2001; Mutalik *et al.*, 2001; Goldin; Chapman, 2003), e foram gradualmente incorporados a outras agendas de pesquisa. Estudos mais recentes têm explorado a negação em tarefas como detecção de polaridade e tratamento computacional de sentenças negativas (Taboada *et al.*, 2011; Jiménez-Zafra *et al.*, 2017; Mukherjee *et al.*, 2021), o que reforça a necessidade de recursos anotados que representem com maior sistematicidade os indicadores negativos e sua contribuição para a polaridade.

É possível separar as atividades de processamento da negação em quatro grandes distintas categorias:

- (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. (Jiménez-Zafra, *et al.*, 2020, p. 190, tradução própria³).

O presente trabalho é uma revisão e síntese dos resultados da pesquisa de mestrado de Miranda Jr. (2022). Aqui serão apresentadas as propostas de anotação para detecção de indicadores negativos do português no *corpus* Porttinari-Base (Lopes *et al.*, 2022).

Metodologia

Para a identificação dos indicadores negativos, este trabalho investigou a anotação dos *corpora* UD Portuguese Bosque (Rademaker *et al.*, 2017) e Porttinari-Base (Lopes *et al.*, 2022a), bem como na ferramenta lexical PortiLexicon-UD (Lopes *et al.*, 2022b). Assim, foram investigados os seguintes indicadores que expressam a negação no português:

3 “(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.

não, nunca, jamais, sequer, tampouco, nada, ninguém, nenhum, sem e nem; e também as preposições acidentais⁴ de exclusão: **exclusive, excluso, exceto, fora, afora, salvo, menos, senão e tirante.** Também se utilizou do *corpus* para News of Brazilian Newspaper (Santana, 2019) para a validação adicional das diferentes formas de ocorrências dos indicadores, todos os exemplos presentes neste trabalho serão oriundos dele. Os dados referentes a cada um dos *corpora*, como também da ferramenta utilizada, podem ser consultados na tabela a seguir:

Tabela 1. Dados sobre os *corpora*/ferramenta

Corpus	Sentenças	Sentenças com 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: Miranda Jr. (2022)

Tanto os *corpora* quanto a ferramenta estão organizados e anotados nas diretrizes da Universal Dependencies (Universal Dependencies, 2024), doravante UD, que é uma metodologia interlinguística de anotação sintática que leva em consideração partes do discurso, características morfológicas e dependência sintática.

A Universal Dependencies (UD) está enraizada no modelo de Gramática de Dependências, que constitui uma tradição de representação sintática em que as palavras de uma oração se organizam por meio de relações de dependência entre elementos, em vez de estruturas de constituintes hierárquicos complexos, como ocorre em abordagens convencionais de gramática de frase. Em uma análise de dependências, cada palavra em uma sentença está diretamente conectada a outra por meio de um vínculo de dependência, refletindo relações sintáticas básicas e evitando nós redundantes em estruturas sintáticas

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.

complexas (Miranda Jr.; Vale, 2025). No âmbito da UD, esse fundamento teórico permite construir *treebanks* multilíngues com categorias e relações comparáveis entre línguas, proporcionando uma base de anotação coerente mesmo diante da variação linguística, embora a aplicação sistemática de *features* específicas, como as de polaridade negativa, ainda dependa de decisões de anotação e curadoria nos *corpora* existentes.

Nas versões mais recentes do *framework*, Universal Dependencies v2 (Nivre *et al.*, 2020), já se encontram previstas *features* específicas para a marcação de polaridade negativa, como `Polarity=Neg` e `PronType=Neg`, o que demonstra que a representação da negação integra formalmente o modelo. Contudo, a efetiva implementação dessas marcações nos *corpora* depende de decisões de anotação, tempo de curadoria e esforço humano especializado. A ausência ou a aplicação restrita dessas *features* – frequentemente limitada aos indicadores mais prototípicos – não decorre de uma limitação estrutural do *framework*, mas antes evidencia uma assimetria entre a sofisticação das descrições teóricas da negação no português e sua operacionalização em recursos anotados de PLN. É precisamente nesse intervalo entre descrição gramatical e implementação computacional que se insere a presente proposta.

Foram retiradas dos *corpora*, de maneira automática, todas as sentenças que continham ao menos um dos indicadores. Em seguida, as sentenças foram agrupadas em função de quatro etiquetas dos indicadores: LEMA, a forma lematizada do indicador; UPOSTAG, anotação de partes do discurso; FEATS, anotação de características morfológicas; e DEPREL, relação de dependência. Ou seja, todas as sentenças que tinham anotações idênticas nas quatro etiquetas foram agrupadas.

Após o agrupamento, as sentenças foram analisadas e foi realizada uma verificação nas quatro etiquetas de cada grupo. A análise foi para verificar se seria possível, utilizando as diretrizes da UD, adicionar a informação lexical acerca da negação dentro dos *corpora*.

Para a análise dos agrupamentos das sentenças, foi realizado um levantamento de bibliografia acerca de cada um dos elementos (Mioto, 1991, 1998; Neves, 2000; Lima, 2011; Bechara, 2009; Cunha; Cintra, 2016; Houaiss, 2009; Michaelis, 2022; Mattos Olímpio, 2005; Canceiro *et al.*, 2018; Cavaguti, 2018; Miranda Jr., 2022; Thompson *et al.*, 2012) resultando em uma disposição dos elementos em nove distintas categorias: (1) *realiza negação oracional*, (2) *realiza negação não-oracional*, (3) *ocorre como dupla negação*, (4) *aciona concordância negativa*, (5) *concorda negativamente*, (6) *exprime comportamento pronominal*, (7) *realiza conjunção*, (8) *ocorre como preposição* e (9) *reforço da negação*.

Toma-se aqui como *negação oracional*, em concordância com Mioto (1991; 1998), as ocorrências da negação que têm escopo no verbo principal de uma oração, independentemente do tipo específico de oração cuja negação está modificando. Em outras palavras, a negação oracional ocorre quando o escopo da negação se aplica ao

verbo principal da oração, e não a um constituinte específico. Os indicadores capacitados a realizar esse tipo de negação são: *não, nunca, jamais, sequer, tampouco, nada e nem*, como se pode verificar nos seguintes exemplos:

1. **Não** me deu um centavo para o tratamento.
2. **Nunca** me senti tão frágil e assustada na minha vida.
3. Já Palocci **jamais** teria tido qualquer compromisso com ela.
4. Eduardo Cunha **sequer** era filiado ao PMDB no momento da compra.
5. Wesley **tampouco** afirma ter intenções de se tratar.
6. Afastado, a Câmara **nada** fez além de seguir a vida.
7. **Nem** é preciso dizer que, depois dessa, o tatuador distraído levou mais umas bifas.

A negação *não-oracional*, ou *negação de constituintes*, é as ocorrências da negação em que o escopo se aplica a um constituinte específico da oração, e não a toda a proposição. Isso significa que a negação atua sobre um elemento específico da frase, como um substantivo, adjetivo ou advérbio, e não sobre o verbo principal. Os indicadores capacitados a realizar esse tipo de negação são: *não, tampouco, nada, ninguém e nenhum*.

8. O **não** pagamento gerou queixas de consumidores em redes sociais e no Procon.
9. **Tampouco** na África ou na América Latina existe consenso sobre quais países deveriam representar essas regiões em uma sonhada atualização do conselho.
10. **Nada** indica que quebrará.
11. **Ninguém** falta ao trabalho, explica, e a empresa economiza com vale-transporte.
12. Nenhum padre deve desfilar no sambódromo, apesar dos convites: eles dizem não querer desviar as atenções da homenageada.

A dupla negação, em português, só pode ser realizada pelo *não*; ela ocorre quando há a repetição de um *não* no final da sentença (Mioto, 1991; Neves, 2000; Barme, 2005). Esse tipo de negação é bem mais frequente em situações de conversação do que em textos formais escritos. Segundo Schwenter (2016), essa forma de negação ocorre motivadamente por questões pragmáticas, de maneira que a dupla negação está vinculada à negação de tópicos inseridos anteriormente em uma conversação.

13. **Não** pode **não**, isto é crime.
14. **Não** confia **não**.
15. **Não** é fácil **não**, tá achando que é glamour?

Além da dupla negação, há outra forma de ocorrência em português de dois indicadores negativos em uma mesma sentença, a concordância negativa (Miotto, 1991; Neves, 2000). A concordância negativa é a necessidade da adição de outro indicador negativo, além do primeiro, na sentença. Caso isso não ocorra, a sentença acaba por se tornar marcada:

16. Não existe **ninguém** que Thoreau conheça tão bem como ele próprio.

17. Não existe alguém que Thoreau conheça tão bem como ele próprio.

Enquanto em 16 temos a forma preferencialmente usada, a falta de concordância negativa torna o exemplo 17 marcado, ou seja, mesmo que seja bem formado e compreensível, ele não é a forma esperada.

Assim, pode-se dividir os indicadores que ocorrem em concordância negativa em dois grupos: os que acionam a concordância e os que concordam. Os que acionam a concordância são: *não, nem e sem, nunca, jamais, nada, ninguém*; e os que concordam são: *nunca, jamais, nada, ninguém, nenhum, tampouco e nem*.

18. Não gosta de jabá, **tampouco** de jerimum.

19. 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**.

20. Em muitos deles, isso é suficiente para garantir ao ministro que pede a vista que ele **não** será derrotado **jamais**.

21. Os salários **não** dão para **nada**.

22. Me sinto perdida, **nunca** enterrei **ninguém**, nunca fui ao velório de familiares

23. [...] **jamais** apresentou **nada** de relevante para o futebol.

24. Evidentemente **nada** autoriza **nenhum** deles a nenhum tipo de abuso.

25. Eu sei, o mais provável é que **ninguém** faça **nada** disso no clima político envenenado em que vivemos.

26. Mesmo **sem** ver **nenhum** gol.

27. Em 2011, Angelici **não** era **ninguém** e só venceu graças aos votos de Macri.

28. A Folha **não** tinha **nenhum** indício de que teria havido tráfico de influência [...]

29. No Haiti, vi muita pobreza, muita tristeza, muitas famílias destruídas por causa do terremoto [de 2010], muitas pessoas **sem** o que comer **nem** beber.

30. Nem a Changi **nem** a Odebrecht querem essa solução.

Na categoria dos pronomes negativos, estão os indicadores *nada*, *ninguém* e *nenhum*. Quando ocorrem com distribuição sintática de pronomes, são capazes de alterar a polaridade de uma oração. Ou seja, diferentemente de outros pronomes, a presença de *nada*, *ninguém* ou *nenhum* em uma frase é o suficiente para torná-la negativa.

31. Nada deu certo.

32. Ninguém quer ir preso.

33. Nenhum sinal da fatura.

Os elementos que realizam coordenação negativa operam como elementos coordenativos tradicionais; entretanto, eles adicionam uma nova sentença/oração dentro de uma sequência negativa. Os indicadores que realizam essa função são o *nem* e o *tampouco*:

34. Eu **não** acelero **nem** desacelero, eu vou fazendo na medida que tem para fazer.

35. Passado esse período, sobretudo após a manifestação das defesas, **não** deveria ser possível acrescentar fatos novos e, como consequência lógica, **tampouco** deveriam ser aceitas provas não indicadas no pedido inicial.

Alguns dos indicadores também ocorrem como preposições, ou seja, introduzem uma oração relativa adverbial, mais especificamente, no caso das negativas, uma adverbial de exclusão/exceção (Miranda Jr., 2022). Os indicadores capacitados a realizar essa função são: *sem*, *exclusive*, *excluso*, *exceto*, *fora*, *afora*, *salvo*, *menos*, *senão*, *tirante*⁵.

36. Trabalhadores **sem** carteira de trabalho, **exclusive** domésticos.

37. Acho que todos valem a leitura, **exceto** Reinaldo Azevedo.

38. É difícil encontrar no Brasil, **fora** a escravidão, um fenômeno social tão destrutivo quanto a motocicleta.

39. A operação não terá fim, **afora** um golpe ou acordo político ou erro, dada a extensão da pirataria contra o Estado.

40. A crise não irá acabar de uma vez, **salvo** eventos excepcionais.

41. Aqui você encontra comida árabe, americana, mexicana, **menos** a comida regional.

42. São quatro sócios, três deles sem ligações anteriores com a gastronomia, **senão** o gosto pela comida.

5 De todas as formas listadas, a única que não foi identificada no *corpus* de verificação foi a *excluso*, todavia, por introspeção, é possível substituir qualquer uma das proposições dos exemplos 36 a 43 mantendo a mesma interpretação.

43. No episódio da Chapecoense, **tirante** meia dúzia de imbecis, o povo (de aqui e do mundo) tem tirado a razão de Nélon.

Por fim, tem-se a categoria de indicadores que ocorrem como reforçadores da negação. Esses indicadores ocorrem após um indicador negativo para enfatizar/reforçar a negação. Sendo esses elementos o *sequer* e o *tampouco*.

44. O Canadá **nem sequer** tem um sistema educacional nacional, pois a organização é baseada em províncias autônomas.

45. Nunca sequer falamos sobre pesca.

46. Para Naldinho, a música não é violenta **nem tampouco** sexual.

Após a análise de todas as sentenças elencadas no *corpus*, foram propostas anotações, mais bem descritas na próxima seção, para melhorar a quantidade de informação acerca dos indicadores negativos nos *corpora*.

Propostas de anotação

Para cada um dos indicadores elencados, foram levantadas suas ocorrências dentro dos *corpora*, como também a presença da etiqueta indicando a polaridade negativa nos *corpora* e na ferramenta lexical. Constatou-se que no portilexicon UD nenhum deles recebeu a marcação de polaridade, enquanto nos *corpora* apenas três indicadores receberam a marcação de polaridade negativa dentro do UD Portuguese Bosque e um no Portinari-base. A tabela a seguir exprime a quantidade de ocorrências e a quantidade de ocorrências com anotação relativa à negação em que cada indicador ocorreu nos *corpora*.

Para cada indicador elencado, foram contabilizadas suas ocorrências nos *corpora*, bem como examinada a presença da etiqueta `Polarity=Neg` tanto nos *corpora* quanto na ferramenta lexical. Constatou-se que, no PortiLexicon UD, nenhum dos indicadores recebeu marcação de polaridade, ao passo que, nos *corpora*, apenas três indicadores apresentaram a marcação de polaridade negativa no UD Portuguese Bosque e um no UD Portuguese Portinari-base. A Tabela 2 apresenta o número total de ocorrências de cada indicador nos *corpora*, bem como o número de ocorrências anotadas com marcação de negação.

Tabela 2. Ocorrências dos indicadores negativos em cada *corpus*

Elemento	Total de sentenças		Sentenças com etiqueta de indicação de polaridade			
	UD Portuguese Bosque	Porttinari-base	UD Portuguese Bosque	% rel	Porttinari-base	% rel
Não	1362	1338	1356	99,5%	1338	100%
Jamais	7	11	0	0%	0	0%
Nunca	62	53	0	0%	0	0%
Sequer	24	2	0	0%	0	0%
Tampouco	3	6	0	0%	0	0%
Nada	72	71	8	11,1%	5	7%
Ninguém	43	49	0	0%	0	0%
Nenhum	21	23	21	100%	0	0%
Sem	209	182	0	0%	0	0%
Nem	102	91	0	0%	0	0%
Exclusive	-	-	-	-	-	-
Excluso	0	0	0	0%	0	0%
Exceto	5	1	0	0%	0	0%
Fora	64	54	0	0%	0	0%
Afora	1	1	0	0%	0	0%
Salvo	2	0	0	0%	0	0%
Menos	40	124	0	0%	0	0%
Senão	10	6	0	0%	0	0%
Tirante	0	0	0	0%	0	0%

Fonte: Elaboração própria

Com o levantamento realizado para os indicadores, realizou-se uma revisão nas anotações e algumas propostas de alterações para que seja possível capturar a informação negativa presente nos elementos estudados. Para isso, utilizando a metodologia da UD, propõe-se aqui a adição das etiquetas de polaridade negativa.

A UD dispõe de duas etiquetas referentes às *features* morfológicas para a demarcação da polaridade negativa, sendo elas a etiqueta `Polarity=Neg`⁶ e a etiqueta `PronType=Neg`⁷. A distinção entre elas é simples: todo elemento com polaridade negativa que não seja um pronome recebe `Polarity=Neg`, enquanto os pronomes que são indicadores recebem a etiqueta `PronType=Neg`.

Em vista das duas etiquetas disponíveis, foram avaliadas as ocorrências dos indicadores em cada um dos *corpora* e as diferentes etiquetas que cada um poderia assumir, agrupando-as em acepções em função da marcação de PoS e de atributos morfológicos. Verificando assim que os indicadores *nunca, jamais, sequer, tampouco, nada, sem, nem e não* ocorrem com a PoS de ADV (advérbio); *nada, nenhum e ninguém* ocorre com a PoS de PRON (pronome); *excluso, exceto, fora, afora, menos e tirante* ocorrem com a PoS de ADP (adposição, *tag* da UD para preposições); *tampouco* e *nem* também ocorre como CCONJ (coordenação entre constituinte); o *nem* também ocorre como SCONJ (coordenação entre sentenças); e, por fim, o *não* também ocorre como INTJ (interjeição).

Com base nas ocorrências, foram propostas as etiquetas finais de POS, LEM e atributos morfológicos disponíveis na Tabela 3, em que pode ser consultado o indicador negativo, o número da acepção, a PoS da acepção, seu lema e os atributos morfológicos indicando a polaridade negativa:

Tabela 3. Propostas de anotação

Indicador	Acepção	PoS	Lema	Atributo(s)
Nunca	Nunca-01	ADV	nunca	<code>Polarity=Neg</code>
Jamais	Jamais-01	ADV	jamais	<code>Polarity=Neg</code>
Sequer	Sequer-01	ADV	sequer	<code>Polarity=Neg</code>
Tampouco	Tampouco-01	ADV	tampouco	<code>Polarity=Neg</code>
Tampouco	Tampouco-02	CCONJ	tampouco	<code>Polarity=Neg</code>
Nada	Nada-01	ADV	nada	<code>Polarity=Neg</code>
Nada	Nada-02	PRON	nada	<code>Gender=none,</code> <code>Number=Sing, PronType=Neg,</code> <code>Definite=Ind</code>
Ninguém	Ninguém-01	PRON	nenhum	<code>Gender=none, Number=Sing,</code> <code>PronType=Neg, Definite=Ind</code>
Nenhum	Nenhum-01	DET	nenhum	<code>Gender=Masc/Fem, Number=Sing,</code> <code>PronType=Neg, Definite=Ind</code>

⁶ Disponível em: <https://universaldependencies.org/u/feat/Polarity.html#Neg>. Acesso em: 13 nov. 2024.

⁷ Disponível em: <https://universaldependencies.org/u/feat/Polarity.html#Neg>. Acesso em: 13 nov. 2024. Acesso em: 13 nov. 2024.

Nenhum	Nenhum-02	PRON	nenhum	Gender=none, Number=Sing, PronType=Neg, Definite=Ind
Sem	Sem-01	ADV	sem	Polarity=Neg
Nem	Nem-01	ADV	nem	Polarity=Neg
Nem	Nem-02	CCONJ	nem	Polarity=Neg
Nem	Nem-03	SCONJ	nem	Polarity=Neg
Não	Não-01	ADV	não	Polarity=Neg
Não	Não-01	INTJ	não	Polarity=Neg
Exclusive	Exclusive-01	ADP	exclusive	Polarity=Neg
Excluso	Exclso-01	ADP	excluso	Polarity=Neg
Exceto	Exceto-01	ADP	exceto	Polarity=Neg
Fora	Fora-01	ADP	fora	Polarity=Neg
Afora	Afora-01	ADP	afora	Polarity=Neg
Menos	Menos-01	ADP	menos	Polarity=Neg
Tirante	Tirante-01	ADP	tirante	Polarity=Neg

Fonte: Miranda Jr. (2022)

Considerações finais e trabalhos futuros

Com base na análise dos *corpora* UD Portuguese Bosque, Porttinari-Base e News of Brazilian Newspaper, e da ferramenta lexical PortiLexicon-UD, este trabalho propôs um sistema de anotação para indicadores de negação em português, utilizando a metodologia da Universal Dependencies (UD).

A pesquisa identificou a necessidade de aprimorar a representação da informação negativa nos *corpora*, visto que a maioria dos indicadores não possuía marcação de polaridade. Foram analisadas as diferentes ocorrências dos indicadores, categorizando-os em nove classes: negação oracional, negação não oracional, dupla negação, concordância negativa (acionando ou concordando), pronome negativo, conjunção, preposição e reforço da negação.

Com base nessa análise, propõe-se a adição de etiquetas de polaridade negativa (*Polarity=Neg* e *PronType=Neg*) às *features* morfológicas dos indicadores, de acordo com a metodologia da UD. A tabela 3 apresenta as etiquetas propostas para cada indicador, considerando sua classe gramatical (PoS), lema e atributos morfológicos.

Visando o futuro, pretende-se aplicar as propostas de anotação nos *corpora* estudados, como também em outros *corpora* de português. Reconhecemos que este trabalho representa apenas uma pequena parcela das diferentes formas em que a negação

pode ocorrer no português e que há necessidade de se expandir a lista dos indicadores negativos para que se possa aprimorar trabalhos futuros que dependam da identificação da negação como parte de sua tarefa.

Espera-se que a implementação dessas propostas de anotação contribua para: uma melhor qualidade da informação sobre indicadores de negação em *corpora* de português; o desenvolvimento de ferramentas de PLN mais precisas na identificação e interpretação da negação; e avanços em áreas como análise de sentimentos e mineração de opinião, que dependem da correta identificação da polaridade das sentenças.

Agradecimentos

Isaac Souza de Miranda Jr. é doutorando em Linguística no PPGL/UFSCar e é apoiado pela FAPESP como beneficiário de uma Bolsa de Doutorado (proc. 2023/01892-4) e de uma Bolsa de Estágio de Pesquisa no Exterior FAPESP (proc. 2025/20010-8).

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

Referências

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, p. 933-939. Disponível em: https://ceur-ws.org/Vol-2943/idpt_paper6.pdf. Acesso em: 02 mar. 2023.

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: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL) 2020*. 2020. DOI: <https://doi.org/10.18653/v1/2020.acl-main.96>. Disponível em: <https://aclanthology.org/2020.acl-main.96/>. Acesso em: 19 nov. 2024.

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. ed. Rio de Janeiro: Lucerna, 2009.

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. DOI: <https://doi.org/10.26334/2183-9077/rapln4ano2018a29>. Disponível em: <https://ojs.apl.pt/index.php/rapl/article/view/30>. Acesso em: 19 nov. 2024.

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, São Carlos, 2018. Disponível em: <https://repositorio.ufscar.br/handle/ufscar/10080?show=full>. Acesso em: 19 nov. 2024.

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. DOI: <https://doi.org/10.1006/jbin.2001.1029>. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/12123149/>. Acesso em: 19 nov. 2024.

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

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. Disponível em: https://www.researchgate.net/publication/2906352_Learning_to_Detect_Negation_with_Not%27_in_Medical_Texts. Acesso em: 19 nov. 2024.

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

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

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

JIMENEZ-ZAFRA, S M.; MARTIÍN-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. Disponível em: <https://ruja.ujaen.es/server/api/core/bitstreams/2db57c6b-53af-4a1b-9267-24c66461edb2/content>. Acesso em: 17 nov. 2024.

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. DOI: https://doi.org/10.1162/coli_a_00371. Disponível em: <https://aclanthology.org/2020.cl-1.5/>. Acesso em: 19 nov. 2024.

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. DOI: <https://doi.org/10.1080/0960085X.2020.1816145>.

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

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: 13 nov. 2024.

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. Disponível em: <https://aclanthology.org/2022.lrec-1.715/>. Acesso em 19 nov. 2024.

MATTOS OLÍMPIO, A. O item nem no português brasileiro. *Estudos Linguísticos XXXIV*, p. 962-967, 2005. Disponível em: <http://www.gel.hospedagemdesites.ws/estudoslinguisticos/edicoesanteriores/4publica-estudos-2005/4publica-estudos-2005-sumario.htm>. Acesso em: 19 nov. 2024.

MICHAELIS On-line: Editora Melhoramentos, 2022. Disponível em: <https://michaelis.uol.com.br/>. Acesso em: 13 nov. 2024.

MIOTO, C. *Negação sentencial no português brasileiro e teoria da gramática*. 1991. Tese (Doutorado em Linguística) – Instituto de Estudos da Linguagem, UNICAMP, 1991. Disponível em: <https://repositorio.unicamp.br/acervo/detalhe/41980>. Acesso em: 19 nov. 2024.

MIOTO, C. Tipos de negação. *Cadernos de Estudos Linguísticos*, v. 34, 1998. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/cel/article/view/8637055>. Acesso em: 19 nov. 2024.

MIRANDA JR., I. S. de. *Não é nada não*: uma análise das lexias negativas do português para anotação nas Universal Dependencies. 2022. Dissertação (Mestrado em Linguística) – Centro de Educação e Ciências Humanas, Universidade Federal de São Carlos, São Carlos, 2022. Disponível em: <https://repositorio.ufscar.br/handle/20.500.14289/17087>. Acesso em: 19 nov. 2024.

MIRANDA JR., I. S.; VALE, O. A. Dependência: o conceito e as gramáticas. *Linguamática*, [S. l.], v. 17, n. 2, p. 71-84, 2025. DOI: 10.21814/lm.17.2.467. Disponível em: <https://www.linguamatica.com/index.php/linguamatica/article/view/467>. Acesso em: 25 fev. 2026.

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. DOI: <https://doi.org/10.1016/j.procs.2021.05.038>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1877050921011236?via%3Dihub>. Acesso em: 13 nov. 2024.

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. DOI: <https://doi.org/10.1136/jamia.2001.0080598>. Disponível em: <https://academic.oup.com/jamia/article/8/6/598/781176?login=true>. Acesso em: 02 mar. 2026.

NEVES, M. H. de M. *Gramática de usos do português*. São Paulo: Editora UNESP, 2000.

NIVRE, J.; de MARNEFFE, M.-C.; GINTER, F.; HAJIČ, J.; MANNING, C. D.; PYYSSALO, S.; SCHUSTER, S.; TYERS, F.; ZEMAN, D. Universal Dependencies v2: An Evergrowing Multilingual Treebank Collection. In: *Proceedings of the Twelfth Language Resources and Evaluation Conference (LREC 2020)*, Marseille, France, 2020. p. 4034-4043. European Language Resources Association. Disponível em: <https://aclanthology.org/2020.lrec-1.497/>. Acesso em: 13 nov. 2024.

O senhor dos anéis: o retorno do rei. Direção: Peter Jackson. Produção: Barrie M. Osborne; Fran Walsh; Peter Jackson. Estados Unidos: New Line Cinema, 2003. Filme (201 min.)

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. Disponível em: <https://aclanthology.org/W17-6523/>. Acesso em: 19 nov. 2024.

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 jan. 2022.

SCHWENTER, S. A. Some issues in negation in Portuguese. *The Handbook of Portuguese Linguistics*, p. 425-440, 2016. Disponível em: https://www.academia.edu/10062480/Some_Issues_in_Negation_in_Portuguese_REVISED. Acesso em: 19 nov. 2024.

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. DOI: https://doi.org/10.1162/COLI_a_00049. Disponível em: <https://aclanthology.org/J11-2001/>. Acesso em: 19 nov. 2024.

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

UNIVERSAL DEPENDENCIES. Disponível em: <https://universaldependencies.org> Acesso em: 13 nov. 2024.

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, Universidade Federal de São Carlos, São Carlos, 2018. Disponível em: <https://repositorio.ufscar.br/handle/ufscar/10710?show=full>. Acesso em: 19 nov. 2024.

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. DOI: <https://doi.org/10.18653/v1/2021.findings-acl.182>. Disponível em: <https://aclanthology.org/2021.findings-acl.182/>. Acesso em: 19 nov. 2024.