

Análise da Anotação de Negação em Corpora sob as Diretrizes da Universal Dependencies

Isaac Souza de Miranda Jr.¹, Oto Araújo Vale¹

¹Departamento do Letras – Centro de Educação e Ciênicas Humanas
Universidade Federal de São Carlos (UFSCar)
Rodovia Washington Luís, km 235 – São Carlos – SP – Brazil

isc_jr@live.com, otovale@ufscar.br

Abstract. This paper analyzes negation annotation in Portuguese based on the Universal Dependencies (UD) guidelines. By reviewing negative operators and exclusion prepositions in the UD Portuguese Bosque, Porttinari-base corpora, and the PortiLexicon-UD tool, we identify inconsistencies in morphological features and syntactic relations. We propose adjustments to improve annotation accuracy and consistency, supporting NLP tasks such as sentiment analysis, information extraction, and machine translation.

Resumo. Este trabalho analisa a anotação da negação no português segundo as diretrizes da Universal Dependencies (UD). A partir da revisão de operadores negativos e preposições de exclusão nos corpora UD Portuguese Bosque, Porttinari-base e na ferramenta PortiLexicon-UD, identificamos inconsistências na marcação de atributos morfológicos e relações sintáticas. Propomos ajustes que tornam a anotação mais precisa e consistente, contribuindo para tarefas de PLN como análise de sentimento, extração de informação e tradução automática.

1. Introdução

A negação é um fenômeno linguístico universal, indispensável para a comunicação humana. Todas as línguas naturais dispõem de mecanismos para expressar oposição a um enunciado, por meio de partículas, advérbios, pronomes, morfemas ou construções sintáticas. Trata-se de uma operação cognitiva fundamental, associada à lógica e ao raciocínio, e amplamente discutida desde Aristóteles até a linguística contemporânea.

O Processamento de Língua Natural (PLN) é um campo multidisciplinar da Ciência da Computação, Inteligência Artificial e Linguística que estuda como computadores podem compreender e manipular a língua humana, seja em textos ou em fala [Li et al. 2021]. Considerando que a **negação** é um fenômeno linguístico universal e indispensável para a comunicação humana, sua análise deveria ocupar papel central nas descrições e aplicações de PLN. Contudo, isso não ocorre sempre [Jiménez-Zafra et al. 2020]. Segundo a autora, os trabalhos de PLN com enfoque no processamento da negação se iniciam em 2001, focados em identificar o escopo da negação em textos de saúde clínica [Chapman et al. 2001, Mutualik et al. 2001, Goldin and Chapman 2003] e vêm ganhando espaço na comunidade científica.

Apesar de sua centralidade, a negação apresenta alta complexidade. Ela pode envolver múltiplos elementos na sentença, especialmente no português, que apresenta

fenômenos como a *concordância negativa*. Além disso, operadores como *nem* e *não* podem assumir funções ambíguas, e preposições de exclusão — como *salvo*, *exceto* e *fora* — frequentemente não recebem tratamento formal adequado nos modelos de anotação.

Para o português brasileiro, a literatura descritiva sobre o funcionamento da negação é muito ampla, podemos citar trabalhos célebres como [Mioto 1992], [Mioto 1998] e [de Moura Neves 2000], como também trabalhos mais recentes, [Schwenter 2016], [Cyrino 2024] e [Moia 2024].

No português, a negação se manifesta tanto por meio de operadores sintáticos quanto por meio de itens lexicais. A *negação sintática* é realizada por elementos que ocupam posições específicas na sentença, como advérbios (*não*, *nunca*, *jamais*), pronomes (*nada*, *ninguém*, *nenhum*) e preposições accidentais de exclusão (*salvo*, *exceto*, *fora*, entre outras). Já a *negação lexical* é codificada por itens do léxico geral, como nomes (*falta*, *carência*), adjetivos (*ausente*, *impossível*) e verbos com sentido negativo (*negar*, *recusar*, *excluir*), além de processos morfológicos como prefixação negativa (*in-*, *des-*, *a-*).

A presença da *concordância negativa* é uma característica marcante do português, permitindo e até exigindo que múltiplos elementos da sentença carreguem traços de polaridade negativa. Estruturas como “*Não veio ninguém*” ou “*Não disse nada a ninguém*” são gramaticalmente preferíveis a versões com um único operador negativo. Por outro lado, esse fenômeno não se observa em línguas como o inglês, onde a dupla negação gera efeitos semânticos distintos.

Essas manifestações da negação no português podem ser organizadas conforme o escopo (oracional ou não-oracional) e o tipo de acionador (sintático ou lexical). O quadro 1 resume essa tipologia, destacando os tipos de operadores envolvidos e seus respectivos exemplos.

Quadro 1. Tipologia da negação no português brasileiro

Escopo	Tipo	Acionadores	Exemplo
2=Negação Oracional	Sintática	Advérbios de negação	“ Não existe outro jeito.”
	Lexical	Verbos, nomes e adjetivos com sentido negativo	“O médico negou que esteja tratando mais alguém.”
2=Negação Não-Oracional	Sintática	Advérbios, pronomes, preposições, flexão morfológica	“Erros não forçados.”
	Lexical	Adjetivos e nomes negativos	“ Falta de importância dos estaduais.”

No Processamento de Língua Natural (PLN), a anotação da negação permanece um desafio. A falta de uma modelagem consistente impacta diretamente tarefas como análise de sentimentos, extração de informações e tradução automática, além de gerar erros na identificação da polaridade e no escopo da negação, especialmente em línguas com padrões complexos como o português.

Diante desse cenário, este trabalho propõe uma análise dos operadores negativos do português, considerando advérbios, pronomes e preposições accidentais de exclusão,

dentro do framework da Universal Dependencies (UD) [Universal Dependencies 2021]. A análise toma como base os corpora UD Portuguese Bosque [Rademaker et al. 2017], Porttinari-base [Lopes et al. 2022b, Duran et al. 2023] e o léxico PortiLexicon-UD [Lopes et al. 2022a]. Buscamos identificar inconsistências na atribuição do atributo `Polarity=Neg` e nas relações sintáticas associadas à negação, propondo ajustes que tornam a anotação mais precisa e alinhada às necessidades da linguística computacional.

2. Metodologia

Para a obtenção dos dados deste trabalho, foram utilizados três corpora jornalísticos: UD Portuguese Bosque [Rademaker et al. 2017], Porttinari-base [Lopes et al. 2022a] e News of Brazilian Newspaper [Santana 2019], além do léxico PortiLexicon-UD. A metodologia dividiu-se em três etapas:

1. **Seleção dos corpora:** Os corpora foram escolhidos pelo gênero textual e pela quantidade disponível de informação. O Porttinari-Base é consistido de sentenças extraídas do News of Brazilian Newspaper, assim, ele foi utilizado para validação das análises linguísticas e extração das lexias não presentes no Porttinari-Base nem no UD Portuguese Bosque, (*tirante, excluso e exclusive*) mas presentes no PortiLexicon-UD.
2. **Extração dos dados:** Utilizou-se a ferramenta Unitex e scripts em Python para extrair sentenças contendo 20 lexias negativas previamente selecionadas, incluindo operadores canônicos (*não, nunca, nem*) e preposições accidentais (*exceto, salvo, menos*, entre outros). As lexias buscadas podem ser consultadas no quadro 1.
3. **Análise linguística:** Avaliaram-se aspectos morfológicos e sintáticos, com foco na presença ou ausência do atributo `Polarity=Neg` e na correção das relações de dependência sintática atribuídas às lexias.

3. Resultados e Discussão

Os resultados da análise dos corpora investigados indicaram a necessidade de aprimorar a anotação das lexias associadas à negação no português. Observou-se que, em diversos casos, falta a informação que sinaliza formalmente a presença de negação. A ausência dessa marcação dificulta a adequada representação do fenômeno e impacta diretamente o desempenho de ferramentas de Processamento de Língua Natural (PLN) que dependem desses dados para tarefas como análise de polaridade, recuperação de informações e tradução automática.

Em relação aos operadores adverbiais, observou-se que elementos como *nunca, jamais, sequer* e *tampouco* são frequentemente anotados apenas com sua classe gramatical (ADV) e a relação sintática `advmod`, sem que lhes seja atribuído o atributo morfológico `Polarity=Neg`. Essa ausência torna invisível para o processamento automático o fato de que esses operadores veiculam negação, o que compromete diretamente tarefas como análise de escopo negativo e detecção de polaridade.

Da mesma forma, lexias nominais negativas como *nada, ninguém* e *nenhum* frequentemente não apresentam uma distinção formal em relação a substantivos indefinidos neutros. Isso gera um problema significativo, uma vez que esses elementos têm função específica de negar a existência de entidades ou quantidades, e sua não diferenciação impede modelos de PLN de reconhecer corretamente estruturas negativas nominais.

A lexia *nem* ocorre no português desempenhando duas funções principais: como conjunção aditiva negativa e como operador negativo. Observa-se que, independentemente da função que exerce, não há a presença da informação de polaridade na anotação dos corpora analisados, o que impede que o traço de negação seja formalmente identificado nesses contextos.

Outro ponto crítico é a sub-representação das chamadas preposições accidentais de exclusão — como *exceto*, *salvo*, *fora*, *afora*, *menos*, *senão* e *tirante*. Apesar de semanticamente exercerem uma função claramente negativa, esses elementos não recebem nenhuma marcação de negação, nem no nível morfológico nem no sintático, sendo tratados apenas como preposições (ADP), o que empobrece a capacidade dos modelos de reconhecer estruturas excludentes como instâncias de negação.

Diante desses problemas, este trabalho propõe um conjunto de soluções para aprimorar a anotação da negação no português dentro das diretrizes da Universal Dependencies (UD):

- Inclusão sistemática do atributo morfológico `Polarity=Neg` em todas as lexias negativas, para às classes: ADV (advérbio), DET (determinante), ADP (preposição).
- Adição do atributo morfológico `PronType=Neg` para os pronomes negativos (*nada*, *nenhum* e *ninguém*); e
- Formalização da anotação das preposições accidentais de exclusão, garantindo tanto a atribuição do atributo `Polarity=Neg` quanto a definição de rótulos de dependência que refletem sua função excludente.

Todas as alterações e implementações podem ser consultadas no quadro 2.

Os resultados desta análise já estão sendo aplicados na preparação de uma nova versão do corpus Porttinari-Base [Lopes et al. 2022b, Duran et al. 2023], que servirá como base para o treinamento do parser LatinPipe [Straka et al. 2024]. No momento da escrita deste artigo, o processo de treinamento do parser encontra-se em andamento. Após sua finalização, tanto o modelo treinado quanto os dados derivados estarão disponíveis publicamente no repositório do <https://github.com/IscMr/PhDThesisGitHub>. Onde também será possível consultar a versão anotada semi-automaticamente, baseada na substituição dos padrões atuais por meio de scripts de pré-processamento e revisada manualmente pelo pesquisador, contendo todas as correções e refinamentos identificados durante esta pesquisa.

Esses recursos têm como objetivo não apenas validar empiricamente as propostas aqui apresentadas, mas também servir de base para futuros trabalhos que envolvam a detecção da negação, análise de escopo e desenvolvimento de modelos linguísticos mais robustos para o português.

4. Considerações Finais

A anotação precisa da negação é fundamental não apenas para a linguística computacional, mas também para uma compreensão mais aprofundada da estrutura formal da língua portuguesa. A ausência de uma representação consistente dos operadores negativos e das estruturas associadas à negação limita tanto os estudos linguísticos descritivos quanto o desenvolvimento de ferramentas robustas para Processamento de Língua Natural

Quadro 2. Lexias de polaridade negativa e seus atributos

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

(PLN). Este trabalho contribui diretamente para enriquecer os recursos linguísticos anotados do português, propondo soluções concretas para a marcação morfológica e sintática da negação dentro das diretrizes da Universal Dependencies (UD). Ao propor ajustes na marcação de atributos, relações sintáticas e no tratamento das preposições acidentais de exclusão, estabelece-se uma base sólida para o desenvolvimento de sistemas mais eficientes em PLN, especialmente em tarefas que dependem criticamente da detecção e interpretação correta da negação.

Além da modelagem da negação através da marcação de polaridade (Polarity=Neg), os resultados deste trabalho abrem caminho para uma investigação mais ampla de fenômenos relacionados, que não se restringem à simples presença de operadores negativos. A anotação da exclusão semântica, por exemplo, exige a identificação de construções que, embora não contenham um marcador morfológico negativo explícito, cumprem a função de negar, excluir ou restringir informações no enunciado. Da mesma forma, a diferenciação entre escopo estrito e escopo amplo da negação é uma dimensão ainda não formalizada na anotação dos corpora atuais, mas de extrema importância para

uma representação linguística mais precisa, tanto do ponto de vista teórico quanto computacional.

Adicionalmente, há um conjunto expressivo de palavras lexicais — como verbos do tipo *negar*, *recusar*, *excluir*, *dispensar*, entre outros — que também exprimem negações, mas que não são capturados pelo sistema atual de marcação baseado exclusivamente na polaridade. A anotação dessas relações exige uma modelagem que vá além da dependência direta de operadores morfológicos e sintáticos, incorporando informações de escopo semântico e estrutura argumental.

Nesse sentido, os próximos passos desta pesquisa contemplam a expansão da anotação para além das dependências básicas, com a adoção das *Enhanced Dependencies*, que permitem uma representação mais rica das relações sintáticas e semânticas presentes no enunciado. Com isso, será possível modelar não apenas os operadores negativos tradicionais, mas também as relações de exclusão, os contrastes semânticos, os diferentes níveis de escopo e as relações derivadas de palavras lexicais que expressam negação de maneira indireta. Essa etapa está prevista para as próximas versões do corpus Porttinari, que já incorporará essas relações aprimoradas como parte de sua evolução contínua.

Por fim, como trabalho futuro, propõe-se também estender esta análise para outras variedades do português — incluindo o português europeu e variedades regionais do português brasileiro —, bem como avançar na formalização da anotação de expressões multi-palavra que veiculam negação. Esses elementos, atualmente não contemplados, representam um desafio adicional, uma vez que exigem a identificação de unidades fraseológicas cujo sentido negativo não é recuperável pela simples análise das suas partes individuais. A integração dessas expressões e das relações de escopo semântico ao framework das *Enhanced Dependencies* constitui um avanço necessário para a construção de recursos linguísticos cada vez mais robustos e precisos para o português.

Agradecimentos

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.º 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.

Isaac Souza de Miranda Junior recebeu apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, processo n.º 23/01892-4).

Os autores agradecem o apoio financeiro concedido pelo PPGL/UFSCar de verba CAPES/PROAP para a participação neste evento.

Referências

- Chapman, W. W., Bridewell, W., Hanbury, P., Cooper, G. F., and Buchanan, B. G. (2001). A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of Biomedical Informatics*, 34(5):301–310.

- Cyrino, S. M. L. (2024). More on negation in brazilian portuguese. *Estudos linguísticos e literários*.
- de Moura Neves, M. H. (2000). *Gramática de usos do português*. Unesp.
- Duran, M., Lopes, L., Nunes, M. d. G., and Pardo, T. (2023). The dawn of the porttinari multigenre treebank: Introducing its journalistic portion. In *Anais do XIV Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana (STIL 2023)*, pages 115–124, Porto Alegre, RS, Brasil. Sociedade Brasileira de Computação.
- Goldin, I. and Chapman, W. W. (2003). Learning to detect negation with ‘not’ in medical texts. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2003) Workshop on Text Analysis and Search for Bioinformatics*.
- Jiménez-Zafra, S. M., Morante, R., Martín-Valdivia, M. T., and Ureña-López, L. A. (2020). Corpora annotated with negation: An overview. *Computational Linguistics*, 46(1):1–52.
- Li, Y., Thomas, M. A., and Liu, D. (2021). From semantics to pragmatics: where is can lead in natural language processing (nlp) research. *European Journal of Information Systems*, 30(5):569–590.
- Lopes, L., Duran, M. S., Fernandes, P., and Pardo, T. A. S. (2022a). Portilexicon-ud: a portuguese lexical resource according to universal dependencies model. In *Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022)*, Marseille, France.
- Lopes, L., Duran, M. S., Nunes, M. d. G. V., and Pardo, T. A. S. (2022b). Corpora building process according to the universal dependencies model: an experiment for portuguese. <https://repositorio.usp.br/item/003067405>.
- Mioto, C. (1992). *Negação sentencial no português brasileiro e teoria da gramática*. Tese de doutorado em linguística, Universidade Estadual de Campinas (UNICAMP), Campinas, SP.
- Mioto, C. (1998). Tipos de negação. *Cadernos de estudos linguísticos*, 34.
- Moia, T. (2024). A distribuição dos adjuntos temporais negativos no português contemporâneo: negação, concordância negativa e construções de grau. *Diacrítica*, 38(1):226–253.
- Mutalik, P. G., Deshpande, A., and Nadkarni, P. M. (2001). 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*, 8(6):598–609.
- Rademaker, A., Chalub, F., Real, L., Freitas, C., Bick, E., and Paiva, V. (2017). Universal dependencies for portuguese. In *Proceedings of the Fourth International Conference on Dependency Linguistics (Depling 2017)*, pages 197–206, Pisa, Italy.
- Santana, M. (2019). News of the brazilian newspaper.
- Schwenter, S. A. (2016). Some issues in negation in portuguese. In *The Handbook of Portuguese Linguistics*, pages 425–440. Wiley-Blackwell.

Straka, M., Straková, J., and Gamba, F. (2024). ÚFAL LatinPipe at EvaLatin 2024: Morphosyntactic analysis of Latin. In *Proceedings of the Third Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA) @ LREC-COLING-2024*, pages 207–214, Torino, Itália. ELRA and ICCL.

Universal Dependencies (2021). Universal dependencies. <https://universaldependencies.org/u/feat>. Acesso em: 03 junho 2025.