RESEARCH REPORT

NEGATIVAS: A PROTOTYPE FOR **SEARCHING AND CLASSIFYING** SENTENTIAL NEGATION IN SPEECH DATA

Túlio Sousa de GOIS (iD) ✉
Laboratório Multiusuário de Informática e Documentação Linguística – Federal University of Sergipe (UFS)
São Cristóvão, Sergipe, Brazil

Paloma Batista CARDOSO (iD) ✉
Laboratório Multiusuário de Informática e Documentação Linguística – Federal University of Sergipe (UFS)
São Cristóvão, Sergipe, Brazil

CHECK FOR UPDATES

ABSTRACT

Negation is a universal feature of natural languages. In Brazilian Portuguese, the most commonly used negation particle is *não*, which can take scope over nouns or verbs. When it takes scope over a verb, "*não*"can occur in three positions: pre-verbal (NEG1), double negation (NEG2), or post-verbal (NEG3), e.g., "*não* gosto", "*não* gosto *não*", "gosto *não*"("I do not like it"). From a variationist perspective, these structures are different forms of expressing negation. Pragmatically, they serve distinct communicative functions, such as politeness and modal evaluation. Despite their grammatical acceptability, these forms differ in frequency. NEG1 dominates across Brazilian regions, while NEG2 and NEG3 appear more rarely, suggesting its use is contextually restricted. The low frequency of these structures challenges research, often resulting in subjective, non-generalizable interpretations of verbal negation with *não*. To address this, we developed *negativas*, a tool for automatically identifying NEG1, NEG2, and NEG3 in transcribed data. The tool's development involved four stages: i) analyzing a dataset of 22 interviews from the Falares Sergipanos database, annotated by three linguists, ii) developing the code using the Python language and Natural Language Processing (NLP) techniques, iii) running the tool, iv) evaluating accuracy. Inter-

annotator agreement, measured using Fleiss' Kappa, was moderate (0.57). The tool identified 3,338 instances of *não*, classifying 2,085 as NEG1, NEG2, or NEG3, achieving a 93% success rate. However, *negativas* has limitations. NEG1 accounted for 91.5% of identified structures, while NEG2 and NEG3 represented 7.2% and 1.2%, respectively. The tool struggled with NEG2, misclassifying instances as overlapping structures (NEG1/NEG2/NEG3). These challenges stem from the dataset's lack of punctuation, which in written texts, marks sentence boundaries. In spoken data, prosodic cues serve this purpose, recognized by speakers but not by the tool. This highlights the need for advancements in NLP to better handle the unique features of spoken language data.

RESUMO

A negação é uma característica universal das línguas naturais. No português brasileiro, a partícula de negação mais comum é o *não*, que pode incidir sobre nomes ou verbos. Quando incide sobre um verbo, o *não* pode ocorrer em três posições: pré-verbal (NEG1), dupla negação (NEG2) ou pós-verbal (NEG3), como em *não gosto*, *não gosto não* e *gosto não*. Sob uma perspectiva variacionista, essas estruturas são formas diferentes de expressar a negação. Pragmaticamente, elas desempenham funções comunicativas distintas, como polidez e avaliação modal. Apesar de sua aceitabilidade gramatical, essas formas apresentam frequência distinta. A NEG1 predomina em todas as regiões do Brasil, enquanto a NEG2 e a NEG3 ocorrem mais raramente, o que sugere que seu uso é contextualmente restrito. Essa baixa frequência impõe desafios à pesquisa, resultando, muitas vezes, em interpretações subjetivas e não generalizáveis sobre a negação verbal com *não*. Para lidar com essa questão, desenvolvemos o *negativas*, uma ferramenta para identificação automática de NEG1, NEG2 e NEG3 em dados de fala transcritos. O desenvolvimento da ferramenta ocorreu em quatro etapas: i) análise de um *corpus* de 22 entrevistas do banco de dados Falares Sergipanos, anotadas por três linguistas; ii) desenvolvimento do código utilizando a linguagem Python e técnicas de Processamento de Linguagem Natural (PLN); iii) execução da ferramenta; e iv) avaliação da acurácia. A concordância entre os anotadores, medida pelo Kappa de Fleiss, foi moderada (0,57). A ferramenta identificou 3.338 ocorrências de *não*, classificando 2.085 como NEG1, NEG2 ou NEG3 e alcançando uma taxa de acerto de 93%. Contudo, o negativas apresenta limitações. A NEG1 correspondeu a 91,5% das estruturas identificadas, enquanto a NEG2 e a NEG3 representaram 7,2% e 1,2%, respectivamente. A ferramenta apresentou

dificuldades com a NEG2, classificando erroneamente, em alguns casos, ocorrências como estruturas sobrepostas (NEG1/NEG2/NEG3). Esses desafios decorrem da ausência de pontuação no *corpus*, elemento que, no texto escrito, delimita as fronteiras sentenciais. Na fala, esse papel é desempenhado por pistas prosódicas, que são reconhecidas pelos falantes, mas não pela ferramenta. Isso evidencia a necessidade de avanços em PLN para que se possa lidar de modo mais eficaz com as particularidades dos dados de fala.

KEYWORDS

Negation; Natural Language Processing; Brazilian Portuguese.

PALAVRAS-CHAVE

Negação; Processamento de Linguagem Natural; Português Brasileiro.

INTRODUCTION

Negation is a common phenomenon in all natural languages (Dahl, 2010) and has been the focus of numerous descriptive linguistics studies, from different perspectives (Rocha, 2013; Goldnadel, 2016; Oliveira, 2022). In Brazilian Portuguese (BP), negative structures can be formed by *não* (no/not) in pre-verbal, double, and post-verbal positions (NEG1, NEG2, NEG3, respectively), as exemplified in (1):

(i) (Eu) *Não* gosto (NEG1)
(I) don't like it

(ii) (Eu) *não* gosto *não* (NEG2)
(I) don't like it no

(iii) (Eu) gosto *não* (NEG3)
(I don't) like it no

From a variationist perspective, these three negative structures are three ways of expressing opposition. From a pragmatic perspective, NEG1, NEG2, and NEG3 assume different communicative functions. These two approaches to negation with *não* in BP are commonly rooted in the analysis of sociolinguistic interviews.

These interviews are traditionally structured as dialogues between the informant and the interviewer, lasting an average of 50 minutes. During this time, a large volume of speech data is generated. Without the use of automatic tools to search for a specific element of interest, handling databases made up of dozens or hundreds of sociolinguistic interviews is a challenge. This is why, possibly, several studies about negative structures with *não* are carried out with a reduced sample of analysis. Consequently, many conclusions about the use of NEG1, NEG2, and NEG3 are based on subjective interpretations that cannot be generalized, which could be made through the use of sophisticated statistical analysis like models of conditional inference tree (Freitag *et al.*, 2020), that can give indications of the most relevant criteria for categorizing pre-verbal negation as an expression of opposition, for example. But, for this, it is necessary to have a large database of the phenomenon under analysis.

The lack of studies about the use of NEG1, NEG2, and NEG3 executed with large samples is also detrimental to the development of technologies based on the recognition of language patterns. Developing efficient Artificial Intelligence (AI) models to process Portuguese negative structures containing *não* requires accurate linguistic descriptions of these constructions. To contribute to studies about negation in BP and the processing of natural language, we present the *negativas* tool, which aims to make it possible to search, automatically, for negation structures formed with the adverb *não* in pre-verbal, double, and post-verbal positions. The tool was built using Python language and spaCy package (Honnibal *et al.*, 2020).

1. NEGATIVE STRUCTURES WITH NÃO IN BRAZILIAN PORTUGUESE (BP)

Negation is a property that can be expressed in different ways: by (i) morphemes or affixes (*infeliz*, *desiludido* – unhappy, disillusioned); (ii) negative particles (*não*, *nunca* – no, never); and (iii) negative verbs (*inviabilizar*, *impedir* – make it unviable, to stop).

In BP, one of the negative particles that can indicate negation is *não*. Traditionally, negation is understood as a strategy to indicate opposition to affirmation. From this perspective, NEG1, NEG2, and NEG3 have the same function and are, therefore, variants of the same variable. The approach of negation as opposition is heavily influenced by studies in logic. From this perspective, the propositional content of sentences is analyzed based on truth values, divided into true or false.

In the field of linguistic studies, numerous studies have focused on the description and analysis of negative structures with *não*. Variationist approaches assume that negative structures with *não* in pre-verbal, double, and post-verbal positions are three variants of the same variable. From this point of view, the use of NEG1, NEG2, and NEG3 are conditioned by linguistic (type of sentence in which *não* occurs, presence/absence of subject, type of verb, presence/absence of other negative adverbs) and social factors (level of education, age group, place of residence) (Rocha, 2013; Oliveira, 2022), suggesting that these uses may be a factor of regional identification since the occurrence of NEG3 is more present among speakers from the Northeast of Brazil.

Another approach to describing and analyzing negative structures with *não* in BP assumes that NEG1, NEG2, and NEG3 are not variants of the same variable. From this perspective, the three structures have specific pragmatic functions. Speakers can use a negative structure with *não* to indicate opposition and they can also use it to soften an information, to be polite to their listeners:

(1)   DOCLS: mas como os colegas do teu curso como que é a relação?
       But with your classmates, how is your relationship?

       GUI1MI: (…) assim eu go/ gosto não tenho problema com grande parte mas tem uns que assim não é que eu não tenha problema evito né (…)
       I mean, I li/ like, I don't have problems with most of them, but there are some of them that is not that I don't have problems with them, I even avoid right (…)

From a pragmatic point of view, negation has interactional properties and is therefore characterized as an act performed by the speaker in front of the listener. Approaches of this type also have a large place in studies describing Brazilian Portuguese (Goldnadel, 2016; Goldnadel; Petry; Lamberti, 2020), especially those derived from (Schwenter, 2005) assumptions. From this perspective, the uses of NEG1, NEG2, and NEG3 are conditioned by the informational status of the negated content, which can be new or old, expressed implicitly or explicitly.

To execute descriptive studies that can test different theories about the uses and functions assumed by NEG1, NEG2, and NEG3, it is necessary to make analyses based on a large volume of data, which is impossible to do by searching, manually, for each occurrence of negative structures with *não* in a set of sociolinguistic interviews. One possibility to deal with this problem is to use spaCy, a library for Natural Language Processing (NLP) which has models to recognize syntactic patterns of Brazilian Portuguese.

2. SPACY

spaCy (Honnibal *et al.*, 2020) is an open-source library for advanced natural language processing in Python. Its features and functionalities are related to linguistics, such as tokenization and data labeling, and also to machine learning (ML), such as model training. In addition to the variety of resources applied to NLP, spaCy also stands out for its pre-trained models for different languages, with 80 pipelines trained for 24 languages. For Brazilian Portuguese (BP), this library has 3 pre-trained models, differing in their size, training datasets, and functionalities. In this work, we use pt_core_news_lg, a pipeline recommended for tasks in which greater accuracy is required[1]. For *negativas*, three spaCy functionalities are central to its development and operation: tokenization, POS-tagging, and Rule-based Matching.

Tokenization is the process of segmenting texts into tokens (words, punctuation, etc.). In spaCy, this is a non-destructive process because, after segmentation, the tokens can be used to reconstruct the original sentence without any loss. This task is carried out according to the syntax of the language being processed. So, when we load the pre-trained model for BP, we are configuring specific syntactic specifications for processing the texts.

Through pre-trained pipelines, spaCy can predict the attributes of each token, depending on its context. These include part-of-speech tagging, which assigns a tag corresponding to the grammatical class of each word. The tags adopted by the library are the Universal POS tags[2] (Rademaker *et al.*, 2017).

The attributes of the tokens not only make it possible to understand more about the word and the syntactic context in which it is inserted but also help with tasks such as searching for certain phenomena in a large corpus. This process is carried out by assigning patterns to rule-based matching. Thus, it is possible to create search expressions using the grammatical class labels of the tokens (among other attributes, such as the text itself) and insert customized nomenclatures for each pattern created. This is how the search for structures with *não* is performed by *negativas*.

---

1 The pt_core_news_lg model achieved 0.97 accuracy in the POS Tags task in Portuguese. https://spacy.io/models/pt#pt_core_news_lg

2 https://universaldependencies.org/u/pos/#universal-pos-tags

3. METHODOLOGY

The process of developing and validating *negativas* took place in four main stages: i) understanding the sample Deslocamentos 2020 dataset, formed by sociolinguistic 22 interviews, and the NEG1, NEG2, and NEG3 structures; ii) building the code based on NLP techniques; iii) running the tool, and iv) analyzing accuracy if the results achieved.

3.1. THE SAMPLE DESLOCAMENTOS 2020

The Deslocamentos 2020 sample is part of the Falares Sergipanos database (Freitag, 2013). In total, this sample consists of 100 audio-recorded sociolinguistic interviews, structured in dialogues between informants and documenters. The informants are undergraduate students at the Federal University of Sergipe, and the documenters are researchers linked to the Grupo de Estudos em Linguagem, Interação e Sociedade – GELINS.

All the interviews that compose the Deslocamentos 2020 sample were audio-recorded and then saved in *.wav* format. They were then transcribed using ELAN software (Brugman; Russel; Nijmegen, 2004). The interviewee's and interviewer's speeches were recorded on individual tracks. Separately, on a third track, speech disfluencies, such as pauses and hesitations, were recorded. This procedure is justified by the possibility of generating .txt files containing the individual content of each track at a later date. In these files, we included a header containing information about the place where the interview was carried out, the speaker's gender and age, city of origin and residence, and period of undergraduate study. All the text files were used to create a single dataset with data from 22 interviews[3], which was then used for the search task with the proposed tool.

Negation structure	ID	Basic attributes
NEG1	pre-verbal	[{"TEXT": "*não*"}, {"POS": {"IN": ["VERB", "AUX"]}}]
NEG2	double negation	[{"TEXT": "*não*"}, {"POS": {"IN": ["VERB", "AUX"]}}, {"TEXT": "*não*"}]
NEG3	post-verbal negation	[{"POS": {"IN": ["VERB", "AUX"]}}, {"TEXT": "*não*"}]

Tabela 1. Search pattern attributes.

3 The Deslocamentos 2020 sample consists of 100 sociolinguistic interviews. Of these, 32 were recorded in audio and video format. The construction of the protocol using negatives was motivated by a Doctoral project whose goal is to describe the functions of negative structures with "não", considering, besides linguistic structure, and and face gestures. More information about it is presented by (Cardoso, 2023, 2025).

4. NEGATIVAS

The tool *negativas* was developed using Colab, a Google service that allows coding notebooks to be created. Besides that, we used *spaCy* libraries to apply NLP techniques, while pandas were used to structure and save the data.

In the dataset, each file has a header to identify information about the interview. Because of the transcription standards adopted by GELINS, the transcription files also have pauses and noise markings throughout the text (speech disfluences). Therefore, the data pre-processing tasks involved extracting the header information and cleaning the data by removing the aforementioned markings.

The search and classification of pre-verbal, double, and post-verbal negation structures started with the use of Matcher, from the spaCy library, a rule-based matching mechanism that allows searches to be carried out using token attributes as parameters. The search patterns are lists of dictionaries, and each dictionary represents a token. By adding them to the Matcher, it is possible to enter an ID (identifier), a structure used to classify each pattern found. For the phenomenon studied, we searched for forms corresponding to the negation structures presented (NEG1, NEG2, and NEG3). The patterns created used exact matching with the word "*não*"and POS tags to identify verbs and auxiliaries as attributes, and finally, corresponding IDs were then assigned to each structure. The attributes used as a basis for building the patterns, as well as their IDs, are shown in Table 1.

The occurrences found and their classifications are stored in spreadsheets, which contain the interview header data, the type of negation found and its identifier, in this case, the negation structure corresponding to the occurrence (NEG1, NEG2, NEG3).

4.1. CLASSIFICATION METRICS

In order to validate the classification performed by *negativas*, we generated a confusion matrix and calculated the main metrics based on the annotation of the data by three linguists. This stage was divided into two parts: the calculation of agreement between annotators and the evaluation of the tool's metrics.

Calculating agreement between annotators is necessary to assess the consistency of data annotations. In this work, the measure used was Fleiss' Kappa (Fleiss, 1971), which can be used with more than two annotators. It was applied using the *multi kappa function* from the NLTK library (Loper; Bird, 2002). In addition, Cohen's Kappa was used to assess the agreement between the annotators, using the *cohen_kappa_score* function from the *sci-kit learn library* (Pedregosa *et al*., 2011). After calculating the agreement, we unified the annotations by calculating the frequency of occurrence of negative structures. For each piece of annotated data, its final label was the one with the highest number of appearances in the annotations.

The classification was evaluated using the function *confusion_matrix* and *classification_report* also from sci-kit learn, to generate the confusion matrix and calculate precision, recall, F1 and accuracy, respectively. The Cohen's kappa coefficient was also calculated between human annotations and the classifications to determine the level of agreement.

5. RESULTS

The data annotation obtained 0.57 of agreement (Fleiss' Kappa), which is considered a moderate value. Cohen's Kappa was calculated for each pair of annotators and generated values between 0.52 and 0.65 (Figure 1). Human annotations resulted in 1893 (90.8%) occurrences of the pre-verbal structure, 92 (4.4%) occurrences of the post-verbal structure, and 100 (4.8%) of double negation.
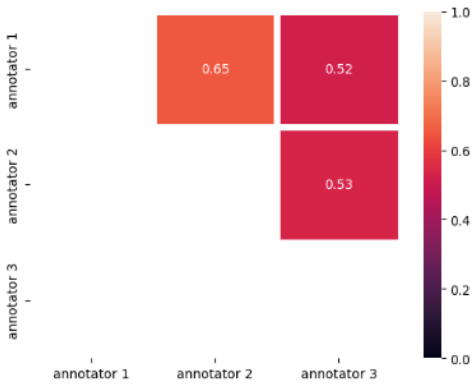

Figura 1. Inter-annotation agreement.

The application of *negativas* to the dataset resulted in the identification of 3338 occurrences of the word *não*, of which 2085 were classified as one of the three sentential negative structures. Pre-verbal negative structures (NEG1) were predominant, accounting for 91.5% of the recognized structures. Following this order, NEG3 (post-verbal) was identified in 7.2%, and finally, NEG2 (double negation), which occurred in only 1.2% of the negations.

At the stage of classifying negation structures, the *negativas* showed an accuracy of 93%. Besides this global metric, we also calculated precision, revocation, and F1 to each negation structure, as can be seen in Table 2. We calculated all the metrics from the elements present in the confusion matrix (see Figure 2).
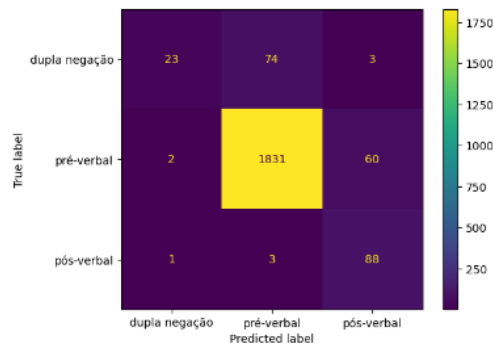
Figura 2. Negativas confusion matrix.

The agreement between human annotations and *negativas* classifications was calculated using Cohen's Kappa coefficient, yielding moderate agreement (k = 0.58).

6. DISCUSSION

Linguistic analyses based on speech data deal with a large volume of data. To deal with this amount of information, it is necessary to develop tools that make it possible to automate searches for the subsequent analysis of specific phenomena.

In the field of descriptive linguistics, studies are traditionally based on manual data handling. Transcription tools such as ELAN make it possible to search for specific items but in a generalized way. Searching for occurrences of the adverb *não* results in all the contexts in which this word is used. When one is interested in analyzing specific usage situations - such as *não* in pre-verbal, double, and post-verbal positions – filtering out specific syntactic structures becomes a long and expensive job. This is why the use of automatic tools is a latent need.

The different possibilities of negation with *não* do not occur randomly: they might be conditioned by linguistic, social, and pragmatic factors. Understanding these factors is important to describe how this universal property of natural languages works in BP. Understanding them also contributes to the development of efficient machine learning and AI models that adequately produce and process NEG1, NEG2 and NEG3. To do this, it is essential to work with a large volume of data, which is only feasible through the use of automatic tools whose efficiency is continually improved.

negativas has proved to be efficient in finding and classifying negative structures with *não*, but it does have its limitations. Such as the occurrence of double negation, which in some cases is not recognized as an individual syntactic structure, returning an occurrence that is simultaneously classified as NEG1, NEG2, and NEG3, and also the treatment of intervening material present in the data, which ends up making it difficult to model the search patterns. Another notable limitation

relates to the data imbalance, with NEG1 structures being significantly more frequent than other categories, which adversely affects precision when calculating classification metrics.

The tool's difficulty in recognizing the boundaries between NEG1, NEG2 and NEG3 is due to the type of data submitted. The interview transcripts in the Falares Sergipanos database have no punctuation marks, except the question mark. Elements which, in written data, delimit sentence boundaries and constituents (period, comma), are not present in the Displacements 2020 sample. In speech, these boundaries are delimited by prosodic parameters. Native speakers, due to the prosodic structure of Portuguese, perceive *não* fui *não* (I didn't went no) as a single block. This information is not accessible to the NLP algorithm. This fact indicates the limitations of the tool and also the need for improvements in natural language processing techniques so that they take into account the particularities of data submitted to automatic tools.

7. CONCLUSIONS

Given that *negativas* use a pre-trained model, the metrics calculated represent the efficiency of the patterns built using the pipeline in identifying negation structures. While the tool achieved a high accuracy of 93%, this metric primarily reflects the data imbalance rather than true performance. However, the Cohen's Kappa coefficient (k = 0.58) between *negativas* and human classifications indicates moderate agreement, demonstrating the tool's potential to automate the description of a large volume of data on negative structures with *não* in pre-verbal, double and post-verbal positions. However, to do this, it is necessary to overcome the limitations of speech data processing. Furthermore, the patterns built for the classification stage carried out by the tool for speech data also allow for the analysis of negation data from writing. However, for data from social networks, for example, changes to the structure would be necessary to accommodate different variations of *não* such as *n* and *ñ*.

Finally, the process of building *negativas* has shown that natural language processing models need to take into account the origin and characteristics of the data to be effectively efficient, thus contributing to advances in the fields of linguistics and AI.

ACKNOWLEDGMENTS

ADDITIONAL INFORMATION

CONFLICT OF INTEREST

The authors declare no competing interests.

CODE AND DATA AVAILABILITY

The code that supports the findings of this study are openly available in: (https://github.com/tuliosg/negativas). The interviews used to test negatives are available upon request to the Laboratório Multiusuário de Informática e Documentação Linguística, at the Federal University of Sergipe.

STATEMENT OF DATA AVAILABILITY

The Deslocamentos 2020 data sample is part of the Falares Sergipanos database (Call 02/2015 SENA- CON/MJ; Call CAPES/FAPITEC/SE 10/2016 PROMOB), approved by the Research Ethics Committee of the Federal University of Sergipe under process CAAE: 0386.0.107.000-11. The Deslocamentos 2020 sample consists of 100 interviews conducted with undergraduate students from the Federal University of Sergipe. The participants, by signing the informed consent form, authorized the use of the data generated for scientific research purposes.

AI USE STATEMENT

The authors did not use AI to produce this paper.

REVIEW AND AUTHORS' REPLY

Review: https://doi.org/10.25189/2675-4916.2025.V6.N4.ID861.R
Author's Reply: https://doi.org/10.25189/2675-4916.2025.V6.N4.ID861.A

REFERENCES

BRUGMAN, H.; RUSSEL, A; NIJMEGEN, X. A. Annotating Multi-media/Multi-modal resources with ELAN. *In*: LINO, M.; XAVIER, M.; FERREIRA, F.; COSTA, R.; SILVA, R. (Eds.). *Proceedings of the 4th International Conference on Language Resources and Language Evaluation (LREC 2004)*. Paris: European Language Resources Association, 2004. p. 2065-2068.

CARDOSO, P. B. Speech, hand, and facial gestures: a proposal of a multimodal approach to describe negative structures with não in Brazilian Portuguese. *Revista de Estudos da Linguagem*, v. 31, n. 2, p. 719-763, 2023.

CARDOSO, P. B. Entre palavras e gestos manuais: uma abordagem multimodal da negação no português brasileiro. Tese (Doutorado em Linguística) — Programa de Pós-Graduação em Letras, Universidade Federal de Sergipe, 2025.

DAHL, Ö. Typology of negation. *In*: HORN, L. R. (Ed.). The expression of negation. Berlim: De Gruyter Mouton, 2010. p. 9–38

FLEISS, J. L. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, v. 76, n. 5, p. 378-382, 1971.

FREITAG, R. M. K. Banco de dados falares sergipanos. *Working Papers em Linguística*, v. 14, n. 2, p. 156–164, 2013.

FREITAG, R. M. K.; PINHEIRO, B.; CARVALHO, C. d. S.; LOPES, N. d. S.; RODRIGUES, A. Modelo de árvore de inferência condicional para explicar usos linguísticos variáveis. *In*: CARVALHO, C.; LOPES, N.S.; RODRIGUES, A.T. (Orgs). Sociolinguística e Funcionalismo: vertentes e interfaces. Salvador: EdUNEB, 2020. p. 247–262.

GOLDNADEL, M. Funções pragmáticas de enunciados de dupla negação: análise de dados de Curitiba (PR). *ReVEL: Revista Virtual de Estudos da Linguagem*, v. 14, n. 13, p. 144-180, 2016.

GOLDNADEL, M.; PETRY, P.; LAMBERTI, L. Funções pragmáticas de enunciados de dupla negação em Florianópolis: um levantamento em entrevistas sociolinguísticas do Projeto VARSUL. *Cadernos de Estudos Linguísticos*, v. 62, p. e020017, 2020. Disponível em: https://periodicos.sbu.unicamp.br/ojs/index.php/cel/article/view/8658767.

HONNIBAL, M.; MONTANI, I.; LANDEGHEM, S. V.; BOYD, A. spaCy: Industrial-strength Natural Language Processing in Python. *Zenodo*, Honolulu, HI, USA, 2020.

LOPER, E.; BIRD, S. *NLTK*: The natural language toolkit. 2002. arXiv preprint cs/0205028.

OLIVEIRA, A. C. M. d. A variação sintática da negação na fala culta de Fortaleza-CE. Dissertação (Mestrado em Linguística) - Programa de Pós-graduação em Linguística, Centro de Humanidades, Universidade Federal do Ceará, Fortaleza, 2022.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, É. Scikit-learn: Machine Learning in Python. *The Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011.

RADEMAKER, A.; CHALUB, F.; REAL, L.; FREITAS, C.; BICK, E.; PAIVA, V. D. Universal dependencies for portuguese. *In*: MONTEMAGNI, S.; NIVRE, J. (Ed.). *Proceedings of the Fourth International Conference on Dependency Linguistics* (DepLing 2017). Linköping: Linköping University Electronic Press, 2017. p. 197–206.

ROCHA, R. S. A negação dupla no português paulistano. Dissertação (Mestrado em Linguística) - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2013.

SCHWENTER, S. A. The pragmatics of negation in Brazilian Portuguese. *Lingua*, v. 115, n. 10, p. 1427–1456, 2005.