

# FUNDAÇÃO CAED UFJF

ANA CAROLINA DOS SANTOS OVÍDIO

04/03/2022



## RELATÓRIO

Construção de um algoritmo ligado ao Processamento de Linguagem Natural

# 1. Introdução

Os textos, os manuscritos, as sentenças e as palavras são a materialização de uma comunicação entre humanos que vem sendo construída a anos e que, durante esse tempo, sofre diversas modificações em sua estrutura. Nesse viés, a compreensão de uma conversa não exige o profundo conhecimento em relação aos significados fonéticos, morfológicos e sintáticos - restritos à classificação das palavras em uma frase - ela apenas demanda o conhecimento semântico - ligado ao sentido e à capacidade de manter um canal de comunicação - visto que a fala é um processo natural intrínseco ao homem e às relações sociais. No entanto, acompanhar a evolução de uma língua é complexo e envolve análises gramaticais robustas, que vão além da compreensão cognitiva, exigindo outras ferramentas capazes de suprir mudanças abruptas as quais a língua sofre.

As linguagens artificiais, como linguagem de programação e notações matemáticas, auxiliam no Processamento de Linguagem Natural (PLN) - principal modelo para o estudo morfosintático de uma estrutura linguística. E, através dessas ferramentas, é possível unir os conceitos computacionais com a linguagem citadina vivenciada pelos seres humanos para que seja possível realizar diversos estudos sobre uma língua. Ademais, a computação sofreu um grande avanço com o advento tecnológico e, com isso, é possível criar, modificar, além de classificar, diversas frases em intervalos de tempo relativamente curtos, provando o benefício de realizar uma inteligência artificial e um aprendizado de máquina (*machine learning*) para fins didáticos.

Dessa forma, na tentativa de prever as diversas modificações que uma língua pode sofrer, criou-se um modelo preditivo com base em um algoritmo da linguagem *Python*. Nesse sentido, a modelagem computacional é vinculada a 6 tipos de transformação: trocar adjetivos por sinônimos ou antônimos; trocar o gênero de uma sentença; modificar as vozes (passiva e ativa); permutar as orações sem modificar o sentido delas e verificar a existência de palavras canônicas em uma frase. O algoritmo já possui reconhecimento das classes gramaticais de cada palavra (verbo, adjetivo, advérbio, pronome etc) juntamente com as características sintáticas, a exemplo de sujeito, predicado, adjunto adverbial etc. Notou-se que ele é otimizado, o que possibilita ser utilizado em computadores cuja velocidade de processamento é reduzida. Além disso, o código é ligado a uma interface web permitindo assim que pessoas que não são adeptas à computação possam inserir as frases e escolher as determinadas modificações de uma maneira simplificada.

## 2. Metodologia e Discussão

O projeto baseou-se na modelagem de um sistema web integrado a um banco de dados capaz de armazenar as sentenças que sofrerão modificações conforme o PLN. Dessa forma, escolheu-se utilizar o *Bootstrap* - conjunto das linguagens computacionais HTML, CSS e Java Script - para a formatação visual do site (FrontEnd) e o pacote *flask.py* da linguagem Python responsável pela estruturação interna desse sistema (BackEnd). Por outro lado, os algoritmos responsáveis pelo PLN foram feitos através de duas principais bibliotecas chamadas *nltk.py* e *spacy.py*, uma vez que elas possuem códigos já estruturados que auxiliam na otimização de mudanças no sentido morfológico, sintático e semântico de uma frase.

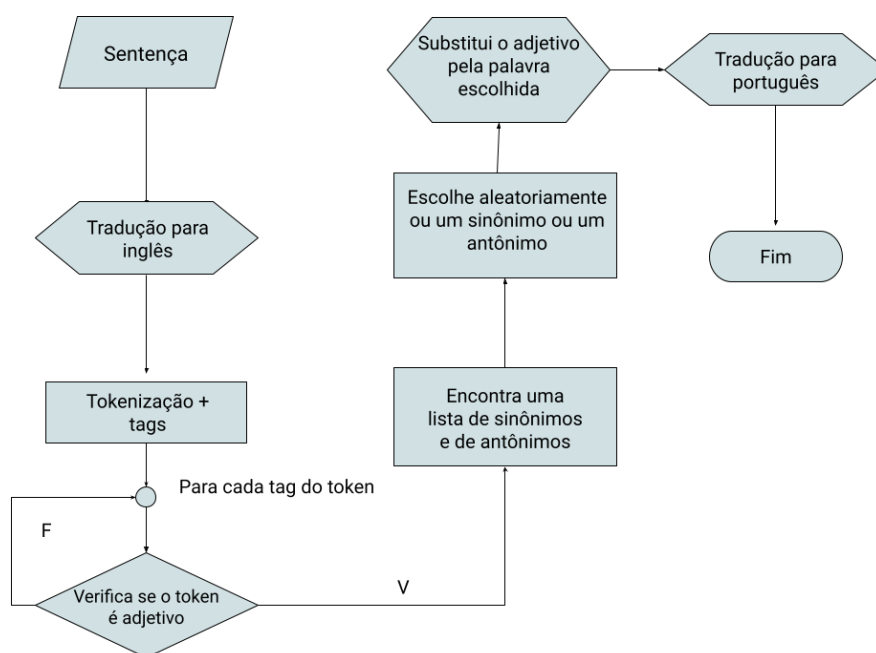
Embora existam diversas formas de reconhecimento de uma sentença e dos elementos que a constituem, o algoritmo de PLN foi desenvolvido em quatro estruturas, as quais serão descritas abaixo:

## 2.1 Substituição de termos

Na análise morfológica de uma frase, alguns termos, ao serem modificados, podem trazer um novo sentido para ela. Isso acontece quando substituímos um substantivo ou um adjetivo por sinônimos; antônimos e; palavras de gêneros opostos. Portanto, a construção da lógica computacional que seja capaz de realizar tais substituições mostra uma gama de possibilidades de sentenças ao usuário.

O primeiro passo foi realizar a troca de adjetivos por sinônimos e por antônimos. Para isso, é necessário que as frases sejam constituídas de substantivo, de um verbo e de um adjetivo, sendo o último elemento aquele que sofrerá modificação. Optou-se por não escolher outros termos para serem substituídos uma vez que se busca estabelecer um novo sentido para a frase e isso é possível através do elemento que dá característica.

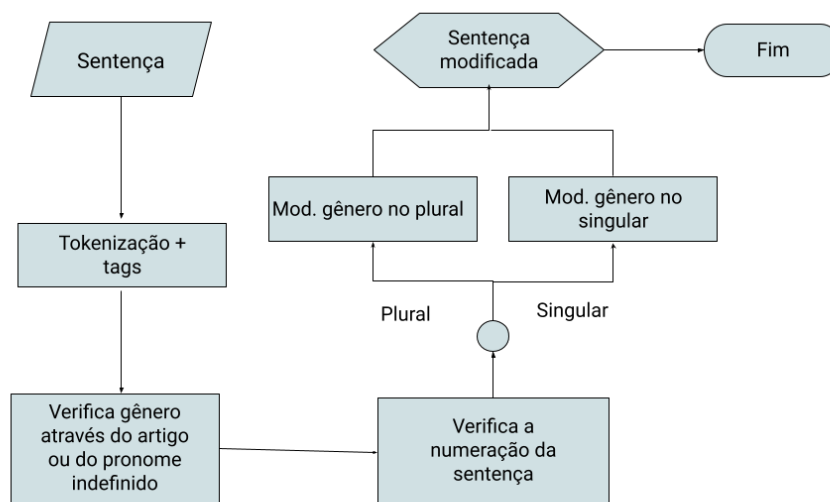
Como as bibliotecas computacionais são vinculadas à língua inglesa, o código traduz a sentença para o inglês, busca sinônimos ou antônimos associados ao adjetivo e, posteriormente, realiza a modificação passando para a linguagem original. Essa ação pode trazer algumas alterações no substantivo e, até mesmo, no verbo, porém isso não será prejudicial, uma vez que esses termos serão traduzidos de forma mais similar possível às palavras da sentença original (Fluxograma 1 e Tab. 1).



**Fluxograma1** Construção da lógica por trás do algoritmo que define a troca do adjetivo, de uma sentença, por um sinônimo ou por um antônimo.

Outra forma de trazer novos sentidos para as frases é por meio da transformação do gênero feminino para o masculino ou vice-versa. Essa estratégia baseia-se em identificar o gênero através de artigos ou de pronomes indefinidos e, posteriormente, troca-los para o gênero oposto juntamente com os substantivos e também adjetivos que são termos possíveis de serem flexionados (Fluxograma 2 e Tab. 1).

No entanto, a limitação do algoritmo é quando o substantivo possui palavras opostas que não possuem o mesmo radical. Ou seja, palavras cujo gênero não é restrito somente pelas últimas letras, a exemplo de cão e cadela ou homem e mulher. Uma solução para esse problema seria o uso de uma automação web que realizasse a busca na internet do gênero oposto do substantivo de uma sentença. O Python tem uma biblioteca especializada para tal caso -*selenium.py*- todavia ela depende do navegador padrão do computador e de um extensor para que a automação seja realizada, inviabilizando o sistema ser generalizado.



**Fluxograma 2** Construção da lógica por trás do algoritmo que define a troca de gênero de uma sentença.

	Sentença original	Modificação
Troca adjetivo	A menina é inteligente	A menina é burra
Troca gênero da sentença	Aqueles amigos eram engraçados	Aquelas amigas eram engraçadas.

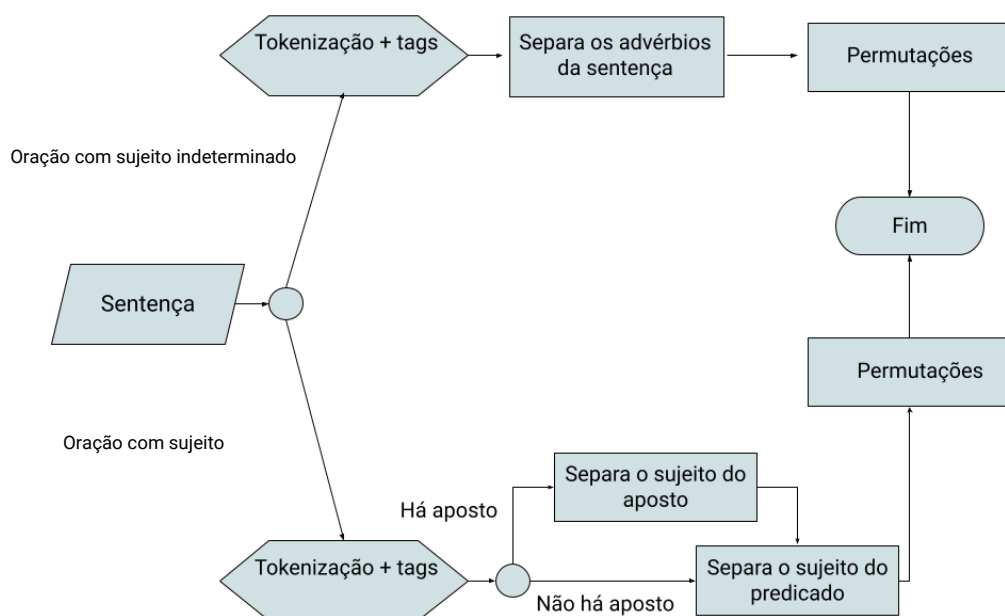
**Tab. 1** Comparação entre sentenças que o usuário pode inserir e os respectivos resultados do algoritmo de acordo com a modificação. Tudo que está supracitado foi realizado de forma computacional.

## 2.2 Permutação entre elementos de uma mesma oração

De acordo com a norma padrão da língua portuguesa, a ordem canônica de uma oração é composta por sujeito, verbo e complemento. Contudo, é possível inverter as posições entre os elementos sintáticos de uma sentença sem modificar o seu sentido e sem torná-la gramaticalmente incorreta. Para isso, dividiu-se entre duas linhas de oração: aquelas que possuem sujeito explícito e aquelas que possuem sujeito indeterminado.

Nas orações com sujeito, as posições podem ser permutadas desde que exista vírgula mostrando que houve inversão da ordem canônica; no entanto dependendo da ordem e do nível de ensino do leitor a sentença pode ficar muito rebuscada, dificultando, assim, o seu entendimento. Ademais, deve-se atentar também a presença de apostos, visto que eles dão características ao sujeito e, quando inseridos em uma sentença, devem anteceder ou sucedê-lo.

Por outro lado, nas orações cujo sujeito é indeterminado, deve-se atentar as locuções adverbiais ou as locuções verbais, pois elas não devem ser separadas no momento em que acontecesse permutação (Fluxograma 3 e Tab. 2).



**Fluxograma 3** Construção da lógica por trás do algoritmo que define a permutação entre elementos de uma oração, seja ela com sujeito explícito ou com sujeito indeterminado.

	Sentença original	Permutações
Oração com sujeito indeterminado	Trovejou nessa manhã	Trovejou de manhã Manhã trovejou
Oração com sujeito explícito	A Segunda Guerra Mundial, conflito militar entre 1939 a 1945, envolveu as maiores potências da época	A Segunda Guerra Mundial um conflito militar entre 1939 e 1945 envolveu as maiores potências da época.  Envolveu as maiores potências da época a Segunda Guerra Mundial um conflito militar entre 1939 e 1945.  Um conflito militar entre 1939 e 1945 a Segunda Guerra Mundial envolveu as maiores potências da época.  Envolveu as maiores potências da época um conflito militar entre 1939 e 1945 a Segunda Guerra Mundial.

**Tab. 2** Permutações entre as orações com sujeito indeterminado/oculto ou explícito Tudo que está supracitado foi realizado de forma computacional.

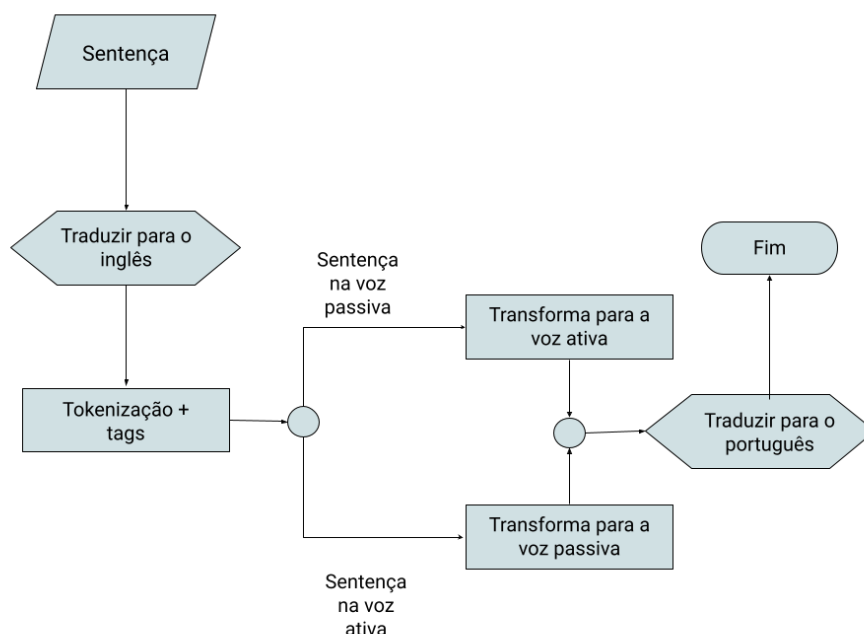
## 2.3 Voz ativa ou voz passiva

Tão obstante existisse complexidade na troca de vozes, sobretudo para a linguagem computacional, mostrar a possibilidade de trocar sujeito e objeto é excelente para o conhecimento lexical e cognitivo em uma criança. Sendo assim, o primeiro passo do algoritmo é identificar quem/ o que realiza a ação e quem/ o que sofre com a ação do verbo. Tal identificação acontece da diferenciação dos substantivos que vêm antes e depois do verbo principal, pois, na voz ativa, o sujeito antecede o verbo e na passiva o contrário. Para fins didáticos e computacionais, leva-se em conta que o que era sujeito passa ser objeto e vice-versa, porém sabe-se que os termos não mudam seus significados sintáticos gramaticalmente. (Fluxograma 4, Tab. 3)

Apenas algumas considerações devem ser observadas pois o verbo, na voz ativa, é flexionado no pretérito perfeito e, na voz passiva, ele muda para o pretérito mais que perfeito composto e existe o acréscimo da preposição "por" antes do sujeito. Além disso, mesmo que exista a tradução da sentença para o inglês antes das modificações das vozes, é possível identificar o que é passivo e ativo na outra língua sem modificar o sentido geral.

	Sentença original	Modificação
Voz ativa para a passiva	O carro atropelou o menino	O menino foi atropelado pelo carro
Voz passiva para a ativa	O objeto foi pego, no chão, pelo cachorro	O cachorro pegou, no chão, o objeto.

**Tab. 3** Transformação da voz passiva para a ativa ou vice-versa realizada pelo algoritmo. Tudo que está supracitado foi realizado de forma computacional.



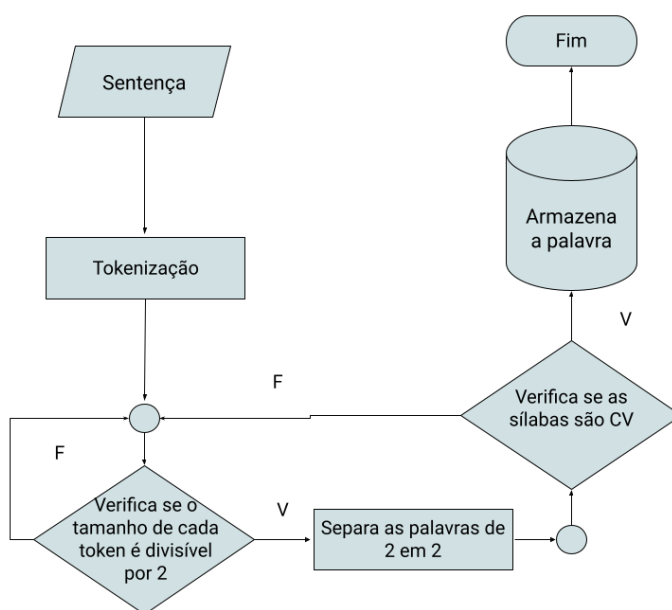
**Fluxograma 4** Construção da lógica por trás do algoritmo que define a troca da voz de uma sentença.

## 2.4 Verificação de palavras canônicas

Para este algoritmo não houve tradução da sentença original. O principal raciocínio empregado baseou-se em verificar se o tamanho da palavra era divisível por 2, uma vez que as palavras canônicas são compostas por sílabas constituídas de uma consoante e de uma vogal. Caso o tamanho não fosse, é nítido que não daria para constituir pares entre duas letras. Após a verificação, notou-se que se as palavras cujo tamanho era divisível por 2 teriam todas as sílabas respeitando a seguinte ordem: consoante-vogal (Fluxograma 5, Tab. 4).

	Sentença original	Palavras canônicas
Palavras canônicas	A menina estava com a cabeça na janela	['menina', 'cabeça', 'na', 'janela']

**Tab. 4** Verificação de palavras cujas sílabas são compostas por consoante+vogal. Tudo que está supracitado foi realizado de forma computacional.



**Fluxograma 5** Construção da lógica por trás do algoritmo que verifica a existência de palavras canônicas em uma sentença.

Além de realizar as modificações nas sentenças que o usuário insere no site, estima-se o grau de complexidade de acordo com as quantidades de verbo, com as quantidades de modificações escolhidas e com os tipos de modificações nas quais deseja-se fazer na sentença. Os parâmetros utilizados para definir o qual complexo é uma sentença estão expostos na Tab. 6. Os resultados resumem-se em nível Ensino Fundamental 1, Ensino Fundamental 2, Ensino Fundamental 3, Ensino Médio e Ensino Superior.

<b>A</b>		
	Níveis para as quantidades de modificações na sentença	Níveis de acordo com as quantidades de verbo
Quantidades		
0		Ensino Fundamental I, II ou III
1	Ensino Fundamental I	Ensino Fundamental I, II ou III
2	Ensino Fundamental II	Ensino Médio ou Superior
3	Ensino Fundamental III	Ensino Médio ou Superior
4	Ensino Fundamental III	Ensino Médio ou Superior
5	Ensino Médio	Ensino Médio ou Superior
6	Ensino Superior	Ensino Médio ou Superior

<b>B</b>	
	Tipo de modificação
Nível	
Ensino Fundamental I	Trocar adjetivos por sinônimos/antônimos
Ensino Fundamental II	Trocar de gênero
Ensino Fundamental III	Palavras canônicas
Ensino Médio	Paráfrase i e Paráfrase ii
Ensino Superior	Paráfrase iii

**Tab. 3** A tabela **A** representa os níveis de complexidade de acordo com as quantidades de modificações em que se deseja fazer nas sentenças e com a quantidade de verbo presente em cada sentença. Enquanto a tabela **B** indica os níveis de complexidade ligada a cada tipo de modificação.

Para cada etapa de medição da complexidade, estima-se um nível. Entre um resultado de mesmo nível sobressai o de menor grau, por exemplo o nível será Ensino Fundamental I caso a sentença tenha, no máximo, 1 verbo, embora a presença de 1 verbo também valha para Ensino Fundamental II e III. No fim, após ter estabelecido os níveis de complexidade para todas as etapas, observa qual nível, entre os escolhidos, é o maior para definir o grau da sentença. Ou seja, suponha que para a quantidade de modificações, a quantidade de verbo e o tipo de transformação os níveis sejam, respectivamente, Ensino Fundamental III, Ensino Médio e Ensino Fundamental III. Nesse sentido, para tal sentença o grau de complexidade é Ensino Médio, visto que, entre os três níveis, ele representa o maior.

### 3. Conclusão

O objetivo principal do projeto foi desenvolver um sistema web que recebe textos, frases ou apenas palavras de um usuário ligado ao sistema e realiza algumas modificações léxicas, semânticas, sintáticas e morfológicas de acordo com a norma culta da Língua Portuguesa, utilizando o *python* como principal construtor dessas mudanças. Nesse sentido, é notória a importância da aplicação computacional em conjunto com o Processamento de Linguagem Natural, uma vez que é possível propor projetos de ensinos de maneira dinâmica englobando diversas sentenças e suas transformações que, se fossem restritas apenas às atividades humanas, não existiriam tantas variabilidades. Assim, com auxílio do algoritmo descrito neste estudo, é possível que professores, estudantes e outros profissionais da área da educação utilizem a plataforma para introduzir textos e receberem de forma automatizada as variadas modificações que uma sentença pode apresentar. Os resultados foram consistentes ao esperado mesmo que os pacotes computacionais estejam vinculados especialmente a língua inglesa.