

ANÁLISE MÓRFICO-MORFÊMICA AUTOMÁTICA DE VERBOS REGULARES

Formação de Professores e Tecnologias Digitais (Tecnologias digitais na Educação Superior)

Lima, Wagner Ferreira; Universidade Estadual de Londrina¹

RESUMO

A implementação de um analisador mórfico-morfêmico automático para verbos regulares do português se deparou com a dificuldade dos estudantes em lidar com essa ferramenta; entre outras coisas, porque o analisador se mostrou disruptivo. Em vista disso, presumimos que um meio de minimizar essa dificuldade seria transformar esse analisador em um programa (app) web, de modo a torná-lo mais intuitivo ao usuário. O objetivo deste trabalho é, então, apresentar os resultados da pesquisa que visou confeccionar esse app. O trabalho de pesquisa consistiu em ajustar a ferramenta de análise mórfico-morfêmica, construída previamente, aos comandos do Flask e, através desse, transformá-lo em um app web. Os resultados sugerem que, mesmo aguardando por validação empírica, esse aplicativo pode proporcionar uma experiência mais amigável aos estudantes com análise morfológica, e ser um meio de divulgação *online* do conhecimento sobre morfologia dos verbos regulares.

Palavras-chave: Analisador mórfico-morfêmico automático. Processamento de linguagem natural. Ensino da morfologia portuguesa.

INTRODUÇÃO

Tecnologias de Processamento de Linguagem Natural (PLN), como as ferramentas e recursos para análise, síntese e tradução automáticas da linguagem verbal (JURAFSKY; MARTIN, 2014), podem em princípio auxiliar o ensino da morfologia linguística. Se considerarmos o ensino formal da morfologia, essa alegação soa muito atraente, haja vista que o formalismo subjacente às duas áreas – linguística estrutural e PLN – é em princípio basicamente o mesmo.

Diante disso, propomos um projeto de pesquisa em ensino que visou à criação

¹ Prof. Associado de Departamento de Letras Vernáculas e Clássicas da Universidade Estadual de Londrina. Coordenador do projeto “Contribuição de Tecnologias de Processamento de Linguagem Natural (PLN) para o Ensino da Gramática Formal do Português”. wflima@uel.br.

de uma função computacional especializada na análise morfológica dos verbos regulares do português. Essa função, denominada *analise_flexao_verbal()*, funcionaria como um gabarito de análise dos verbos e uma fonte de pesquisa sobre o tema (LIMA, 2021).

Entretanto, não obstante ela cumprir bem as funções analíticas para a qual foi designada, sua implementação em sala de aula não foi como era esperado. A ferramenta se mostrou disruptiva; entre outras coisas, por não dispor de uma interface gráfica amigável.

Por essa razão, propomos transformar a referida função em um aplicativo (app) web com dois objetivos, além daqueles de realizar a análise automática da morfologia verbal em termos do estruturalismo linguístico: (1) mitigar os efeitos disruptivos causados pela ferramenta anteriormente criada e, por extensão, (2) divulgar *online* o conhecimento sobre a morfologia dos verbos regulares do português.

A seguir, após apresentar o nível de abstração envolvido na abordagem estrutural da morfologia (“Análise mórfico-morfêmica”), descrevemos o app proposto. Em “Metodologia” apresentamos as ferramentas e recursos em PLN que nos auxiliaram na construção do app (“Resultados”). Em “Discussão” fazemos uma reflexão sobre as vantagens desse aplicativo e, finalmente, em “Considerações finais”, indicamos os passos futuros dessa pesquisa.

ANÁLISE MÓRFICO-MORFÊMICA

A abordagem científica da morfologia, como tudo o mais na ciência, é uma atividade pouco intuitiva. O estudo formal das palavras supõe a concepção de língua vista como uma estrutura cuja compreensão nem sempre é suave. Assim, uma palavra como (eu) *cantava* é explicada em termos de uma combinação de unidades menores, conhecidas como morfemas: *cant-*, *-a-*, *-va*. Um morfema é definido como “uma forma recorrente (significativa) que, por sua vez, não pode ser analisada em formas recorrentes menores (significativas)”² (BLOOMFIELD, 2023).

² “[...] is a recurrent (meaningful) form which cannot in turn be analyzed into smaller recurrent (meaningful) forms.”

Essa combinação apresenta uma distribuição arbitrária, a ordem em que essas unidades se seguem na cadeia da fala; disso resultando a estrutura da palavra. Câmara Jr. (2001, p. 104) encontrou para os verbos do português a seguinte estrutura: T (R + VT) + SF (SMT + SNP),³ que foi usada na descrição realizada pelo nosso app.

Ensinar os fundamentos da morfologia portuguesa seria sem problemas não fosse pelo fato de que uma língua natural apresenta variações estruturais, denominadas alomorfas. Um exemplo disso é o verbo analisado acima, o qual carece de sufixo número-pessoal na 1ª pessoa do singular em comparação com as demais pessoas verbais: (tu) *cantava(s)*, (nós) *cantáva(mos)* etc. (LAROCCA, 2003; CÂMARA JR., 2001).

Somemos a isso o complicador de que um mesmo morfema (p.ex., “1ª pessoa do singular”) pode apresentar expressões distintas: (eu) *cant(o)*, (eu) *cantava(Ø)* e (eu) *cante(i)* (LAROCCA, 2003). Um modelo de descrição foi proposto, então, para lidar com as alomorfas: o morfema passa a ser concebido como uma unidade mínima significativa abstrata que pode se manifestar nas palavras como segmentos fônicos, denominados alomorfes (LYONS, 1987). Para tanto, o morfema é notado entre chaves “{ }”; o alomorfe, pelos próprios segmentos fônicos ou ortográficos que formam a palavra. Assim, “-o”, “Ø” e “-i” são os alomorfes do morfema {1ª pessoa do singular}.

Essa lógica analítica, no entanto, coloca dificuldades aos estudantes, devido especialmente ao nível de abstração e formalização que ela requer. Assumimos, no início, que uma forma de tornar esse aprendizado menos demandante e mais atraente seria usar tecnologias digitais, como as ferramentas de PLN. Contudo, logo percebemos que tais ferramentas *per se* podem não surtir os efeitos esperados se elas não forem suficientemente interativas. Senão, vejamos.

METODOLOGIA

PLN dispõe de uma quantidade de trabalhos sobre morfologia do português brasileiro. Alguns dos quais incluem processamento baseado em regras; outros,

³T: tema do verbo; R: radical; VT: vogal temática; SF: Sufixo flexional; SMT: sufixo modo-temporal; SNP: sufixo número-pessoal.

processamento estatístico (BRANCO, SILVA, 2004). Esses últimos ainda usam estratégias de aprendizado de máquina (*machine learning*), como as redes neurais artificiais multicamadas (*deep learning*) (FONSECA; ROSA; ALUÍSIO, 2015); entre outros procedimentos.

As ferramentas de PLN, por possuírem uma base formal comum com o estruturalismo linguístico, poderiam ser empegadas na análise automática da linguagem. Dessa forma, criamos um aplicativo como parte do projeto de pesquisa em ensino, função *analise_flexao_verbal()*. A princípio, o objetivo era unicamente o desenvolvimento de uma ferramenta computacional para a análise automática dos verbos, e o emprego dela dentro da disciplina de morfologia⁴.

Assim, ela foi implementada em linguagem de programação Python (BIRD; KLEIN; LOPER, 2009), no *Google Colab*, um ambiente de edição de códigos oferecido gratuitamente pelo *Google*. A ferramenta criada cumpre a tarefa de análise verbal, conforme a lógica da morfologia estrutural descrita acima: segmentar os verbos, identificar os alomorfes constituintes, classificá-los segundo o morfema que eles representam e, finalmente, mostrar a estrutura do verbo através de um diagrama.

Todavia, os estudantes se mostraram relutantes em fazer uso dela, devido à dificuldade em usar o *Colab* e chamar a função. Isso nos obrigou a transformar essa ferramenta em um aplicativo web, por este ser mais interativo e menos disruptivo. Esse app está em fase de desenvolvimento, mas já é possível analisar suas funcionalidades. Vale lembrar, contudo, que uma avaliação mais contundente da efetividade de tais funcionalidades ainda aguarda por experimentação.

O desenvolvimento desse app contou com as tecnologias disponíveis para esse tipo de trabalho: *html* (*HiperText Markup Language*), *CSS* (*Cascading Style Sheets*) e *Flask*. As duas primeiras são usadas, nessa ordem, para diagramar e estilizar as páginas da web. Já o *Flask*, por sua vez, é um *framework* Python para desenvolvimento web (GHIMIRE, 2020). Ele oferece todos os recursos necessários

⁴ Vale destacar, não obstante, que, mesmo compartilhando o mesmo formalismo de base, a lógica analítica subjacente às pesquisas em PLN não é exatamente a mesma das dos estudos linguísticos. Enquanto estes procuram trazer à luz as regras de flexão e formação das palavras, aquelas buscam principalmente a etiquetagem automática das palavras.

para implementar funcionalidades, dentro das e entre as páginas, que não são possíveis unicamente com o *html*. Para o projeto em questão, ele foi fundamental para a execução de duas tarefas principais: quebrar o código da função *analise_flexao_verbal* () em suas funcionalidades constituintes e criar uma página web para cada uma delas.

Com isso, foi possível converter uma função que retornava, de uma só vez, uma série de informações metalinguísticas (alomorfes, morfemas e estruturas verbais) em uma sequência de páginas web – botões de informações metalinguísticas (Figura 1) –, cada qual especializada em uma dessas informações.

RESULTADO

O projeto deu lugar a um app constituído por duas páginas principais conectadas, uma de introdução, onde um panorama geral do projeto é oferecido, e uma de inserção de dados, um formulário, com campos em branco para preenchimento. Um botão de “link” na página inicial permite que o usuário navegue entre as páginas. Figura 1 mostra, no pé da página do formulário, o botão de link.

Figura 1 – Página do formulário

The image shows a web form with a light orange background and a green border. At the top, it asks the user to write the infinitive form of a regular verb. Below this, there are input fields for the infinitive and for conjugating the verb in various persons (Eu, Tu, Ele, Nós, Vós, Eles). At the bottom, there are eight buttons arranged in two rows, each representing a different linguistic analysis: 'Conjugação verbal', 'Verbos flexionados', 'Alomorfe de raiz', 'Alomorfe temático', 'Alomorfe modo-temporal', 'Alomorfes número-temporais', 'Morfemas do verbo', and 'Estrutura do verbo'. A 'Voltar à página inicial' button is located at the very bottom of the form area.

Fonte: própria.

Uma vez que o usuário entrar com os dados nos campos correspondentes, ele poderá acessar as informações metalinguísticas. Basta para isso clicar sobre o botão com o rótulo da informação desejada (base do formulário). Essa página oferece um campo contendo oito botões (Figura 1): os botões de informações metalinguísticas. Cada botão remete a uma página específica que traz os resultados de uma determinada funcionalidade da função *analise_flexao_verbal()*, a saber:

Botão “Conjugação verbal”: abre uma nova página com informações sobre o paradigma a que o verbo de entrada pertence: 1^a, 2^a ou 3^a conjugação; bem como porque ele é classificado dessa forma: se terminado em “ar”, “er” ou “ir”, respectivamente.

Botão “Verbos flexionados”: leva à página que mostra os verbos flexionados nas seis pessoas e explicita o tempo e o modo em que foram flexionados.

Botões “Alomorfes de raiz”, “Alomorfe temático”, “Alomorfe modo-temporal” e “Alomorfes número-pessoal”: abrem as páginas correspondentes a essas informações metalinguísticas, respectivamente. Os alomorfes são mostrados de acordo com a pessoa do verbo, uma vez que eles são analisados em relação a ela (LAROCCA, 2023).

Botão “Morfemas do verbo”: traz uma página contendo os morfemas de “radical”, “sufixo temático”, “sufixo modo-temporal” e “sufixos número-pessoais” dos verbos conjugados; cuja distribuição forma a estrutura proposta por Câmara Jr. (2021).

Botão “Estrutura do verbo”: está em via de construção, mas pretende ser uma página com diagramas mostrando a referida estrutura para cada flexão. O retorno à página inicial se dá por meio do botão de “retornar”, localizado no canto esquerdo superior do documento web.

DISCUSSÃO

O aplicativo aguarda por validação empírica, o que vai acontecer assim que ele entrar em modo de produção, ou seja, quando for disponibilizado na web na forma de um site. Contudo, o fato de ele ser uma aplicação web, de fácil acesso e interativa, já é uma vantagem em relação à ferramenta anterior.

Em tentativas prévias de usar a função *analise_flexao_verbal()* na disciplina

de morfologia, notamos que os estudantes se atrapalhavam com a ferramenta (LIMA, 2021). Entre outras coisas, porque, antes de entrar com os dados, era preciso que eles configurassem o *Google Colab*, ambiente onde a análise mórfico-morfêmica é implementada.

Além disso, mesmo oferecendo uma interface relativamente amigável em comparação com os editores congêneres (*Pycharm*, *VSCode* etc.), esse editor é bem menos intuitivo que as interfaces dos documentos web, com os quais os estudantes estão acostumados e que são designados para serem interativos.

Em suma, as experiências com o uso dessa tecnologia foram disruptivas e causaram desinteresse nos estudantes. Por outro lado, o aplicativo ora descrito, mesmo sem sua efetividade confirmada, tem o potencial de ser uma ferramenta pedagógica atraente. Aduzimos algumas razões para isso, a saber:

(1) Diferente da ferramenta anterior, que precisa ser aberta no drive do estudante, nosso app estará disponível na rede e poderá ser acessado como site.

(2) O estudante navegará pelas páginas através dos botões de informação metalinguística, ou seja, clicando sobre eles; o que já é uma prática intuitiva e corrente.

(3) O estudante não precisará escrever nada, exceto os verbos flexionados nas seis pessoas e de acordo com a norma culta do português.

(4) Uma vez feita a conjugação verbal, as informações metalinguísticas estão disponíveis para consultas, sem quaisquer esforços adicionais.

Em termos pedagógicos, o app web vai funcionar como um gabarito para as análises dos estudantes. Assim, correções poderão ser feitas e dúvidas esclarecidas em qualquer lugar e momento, em um dispositivo móvel. Finalmente, graças a essas facilidades, o conhecimento técnico e científico da análise mórfico-morfêmica dos verbos pode ser divulgado e alcançar mais pessoas.

CONSIDERAÇÕES FINAIS

Apresentamos um app web a ser posto em produção que realiza a análise mórfico-morfêmica dos verbos regulares do português através de uma interface interativa. Com isso, cumprimos os dois objetos propostos: minimizar os efeitos de

estranheza e divulgar *online* o conhecimento científico. O próximo passo desta pesquisa será validar esse instrumento junto a estudantes do curso de Letras. Com base nos resultados obtidos, revisões serão feitas a fim de melhorar a experiência do usuário com o aplicativo.

REFERÊNCIAS

BRANCO, António; SILVA, João R. 2004. Evaluating Solutions for the Rapid Development of State-of-the-Art POS Taggers for Portuguese. **Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC2004)**, Paris, 2004, p. 507-510.

BIRD, Steven; KLEIN, Ewan; LOPER, Edward. **Natural language processing with Python: analyzing text with the natural language toolkit**. O'Reilly Media, Inc., 2009.

BLOOMFIELD, Leonard. (2023 [1926]). "A set of postulates for the science of language".

[https://pure.mpg.de/rest/items/item_2282987/component/file_2282986/content]

CÂMARA JR, Joaquim Mattoso. **Estrutura da língua Portuguesa**. 34 ed. Petrópolis, RJ: Editora Vozes, 2001.

FONSECA, Erick R.; ROSA, João L. G.; ALUÍSIO, Sandra M. Evaluating word embeddings and a revised corpus for part-of-speech tagging in Portuguese. **Journal of Brazilian Computer Society**, v. 21, n.2, p. 1-14, 2015.

GHIMIRE, Devndra. Comparative study on Python web frameworks: Flask and Django. 2020. [Comparative study on Python web frameworks: Flask and Django - Theseus]

JURAFSKY, Daniel; MARTIN, James H. Martin. **Speech and Language Processing**. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. 2nd ed. New Jersey: Upper Saddle River, 2014.

LAROCA, Maria Nazaré de Carvalho. **Manual de morfologia do português**. 3 ed. Revisada e ampliada. Campinas: Pontes; Juiz de Fora: UFJF, 2003.

LYONS, John. **Linguagem e Lingüística**. Uma introdução. Trad. Marilda Winkler Averborg; Clarisse Sieckenius de Souza. Rio de Janeiro, RJ: LTC, 1987.

LIMA, Wagner Ferreira. Análise dos verbos regulares do português, no modo indicativo. **GitHub**, 2021.[https://github.com/wagner-ferreira/morfologia-do-portugues/blob/main/C%C3%B3pia_de_verbos_regulares_indicativo.ipynb]