Estado da publicação: Não informado pelo autor submissor

Word2Vec: Um algoritmo saussuriano

Leonardo Giamarusti

https://doi.org/10.1590/SciELOPreprints.11678

Submetido em: 2025-04-07
Postado em: 2025-04-17 (versão 1)
(AAAA-MM-DD)

A moderação deste preprint recebeu o endosso de:
Eliane Silveira (ORCID: https://orcid.org/0000-0002-4862-4547)

WORD2VEC: UM ALGORITMO SAUSSURIANO

WORD2VEC: A SAUSSUREAN ALGORITHM

LEONARDO GIAMARUSTI
Universidade Federal de Uberlândia (UFU)
ORCID: https://orcid.org/0009-0001-9487-6892
<leonardogiamarusti@ufu.br>

RESUMO

Este artigo propõe uma leitura do funcionamento do Word2Vec, um algoritmo para geração de *embeddings* de palavras, à luz da Teoria do Valor (TdV) de Ferdinand de Saussure. O Word2Vec, nos últimos anos, tem sido bastante útil para diversas tarefas de PLN, tais como classificação de textos, análise de sentimentos e cálculos de probabilidade de ocorrências, devido ao manejo de vetores de alta dimensão. Defendo, portanto, que, por meio deste modelo de linguagem, é possível notarmos que algumas noções teóricas da linguística sausssuriana, a saber, o sistema, o signo e o valor, continuam sendo produtivos para refletir sobre aspectos teóricos e epistemológicos da determinação de significados nas línguas naturais; bem como de que forma esses sentidos parecem ser emulados por técnicas modernas de PLN, a exemplo do Word2Vec. Partimos de uma crítica às limitações do TF-IDF, passando pela influência da Semântica Distribucional e da Hipótese Distribucional em modelos vetoriais de linguagem modernos, para, enfim, propor que o Word2Vec apresenta indícios de poder operacionalizar, em níveis de semântica computacional, aquilo que Saussure já formulara conceitualmente no início do século XX, a saber: a ideia de que o significado de uma palavra não é fixo nem individual, mas relacional e determinado pelos valores semelhantes e dessemelhantes que a cercam. Nesse sentido, as fontes saussurianas mobilizadas, nesta pesquisa, para a delimitação dos conceitos abordados foram: o Curso de Linguística Geral; o conjunto de manuscritos *Notes pour le 3e Cours;* e o caderno de Émile Constantin, ouvinte do Terceiro Curso de Linguística Geral ministrado por Saussure em Genebra, entre 1910-1911. Nosso objetivo, assim, é propor que as noções saussurianas de *similia* e *dissimilia* podem ser percebidas nos bastidores teóricos do Word2Vec, promovendo uma aproximação entre o saussurianismo e o PLN contemporâneo. A hipótese que guia este trabalho, portanto, é a de que o Word2Vec pode ser lido como um *algoritmo saussuriano*, por aplicar computacionalmente a dinâmica dos valores linguísticos para emular a forma com que significados são determinados por meio da relação entre as palavras, conforme antecipara o mestre genebrino ainda no século passado.

Palavras-chave: Processamento de Linguagem Natural; Saussure; Teoria do Valor; Word2Vec.

ABSTRACT

This article proposes an interpretation of the functioning of Word2Vec, an algorithm for generating word embeddings, in light of Ferdinand de Saussure's Theory of Value (TdV). In recent years, Word2Vec has proven highly useful for various NLP tasks—such as text classification, sentiment analysis, and word occurrence probability estimation—due to its handling of high-dimensional vectors. I argue, therefore, that this language model allows us to recognize that certain theoretical notions from Saussurean linguistics—namely, system, sign, and value—remain productive for reflecting on the theoretical and epistemological aspects involved in meaning determination in natural languages, as well as on how such notions appear to be emulated by modern NLP techniques, such as Word2Vec. This study begins with a critique of the limitations of TF-IDF, proceeds through the influence of Distributional Semantics and the Distributional Hypothesis in modern vector-based language models, and ultimately suggests that Word2Vec shows signs of being capable of operationalizing, at the level of computational semantics, what Saussure had already formulated conceptually in the early twentieth century—namely: that the meaning of a word is neither fixed nor individual, but relational and determined by the similar and dissimilar values that surround it. In this sense, the Saussurean sources mobilized in this research to define the conceptual framework include: the *Course in General Linguistics*; the manuscript collection *Notes pour le 3e Cours*; and the notebook of Émile Constantin, a student in Saussure's Third Course in General Linguistics taught in Geneva between 1910 and 1911. Our objective, then, is to propose that the Saussurean notions of *similia* and *dissimilla* can be identified within the theoretical underpinnings of Word2Vec, promoting a convergence between Saussurean theory and contemporary NLP. The central hypothesis of this work is that Word2Vec can be read as a Saussurean algorithm, as it computationally applies the dynamics of linguistic values to emulate the way meanings are determined through the relations between words, as foreseen by the Genevan master more than a century ago.

Keywords: Natural Language Processing; Saussure; Theory of Value; Word2Vec.

Introdução

Nas últimas décadas, o Processamento de Linguagem Natural (PLN) tem desenvolvido técnicas cada vez mais sofisticadas para analisar e representar aspectos semânticos das línguas naturais. Os principais ganhos desses avanços, por consequência, residem no aumento da eficiência em tarefas de PLN baseadas em embeddings de palavras, como a classificação de texto (Sebastiani, 2002; Joachims, 1998), a recuperação de informação (Salton; Wong; Yang, 1975), a análise de sentimentos (Pang & Lee, 2008) e o cálculo de probabilidade de ocorrência de palavras (Masoni, 2021).

Dentre essas sofisticações, o TF-IDF (*Term Frequency-Inverse Document Frequency*) é uma técnica clássica no PLN para medidas estatísticas de frequência de termos (Aizawa, 2005). Por meio de cálculos via função logarítmica (Souza; Maia, 2019), a técnica "recompensa" palavras que são frequentes em um texto, mas raras no conjunto do *corpus*, conferindo ao termo maior ou menor grau de importância, o que é bastante útil para a busca semântica e para a classificação de textos. Porém, alguns desafios se impõem nessa abordagem, "como a alta dimensionalidade e a ambiguidade semântica" (Souza; Maia, 2019, p.717).

O grande desafio do TF-IDF, basicamente, reside no fato de que ele apenas "conta palavras". Isto é, o algoritmo não compreende que "cachorro" e "cão", por exemplo, designam o mesmo signo, nem entende que "banco" pode ser um lugar para sentar ou uma instituição financeira. Ou seja, o TF-IDF apresenta algumas limitações em grandes *corpora* de dados não estruturados, apresentando baixa eficiência frente a sinônimos, à ambiguidade e ao contexto de ocorrência.

Nesse cenário, técnicas mais modernas, como o Word2Vec (Mikolov et. al, 2013), oferecem alternativas para os limites enfrentados por algoritmos de TF-IDF. Isso porque, enquanto o TF-IDF enfoca na frequência de ocorrência apenas, ignorando as relações entre as palavras, o Word2Vec consegue aprender com *com quem esta palavra anda*[1]. Por exemplo, se "rei" aparece em contextos próximos de "trono", "reino", "coroa", e "rainha" aparece em contextos parecidos, o algoritmo entende que "rei" e "rainha" são próximos e, portanto, possivelmente relacionados semanticamente.

Nesse sentido, criado por Tomás Mikolov e por sua equipe no Google em 2013, o Word2Vec parte do princípio de que o sentido de uma palavra pode ser inferido a partir das palavras que a cercam, o que dialoga diretamente com a já conhecida Hipótese Distributiva (Firth, 1949; Harris, 1955) em PLN.

Como se sabe, o principal aporte linguístico-teórico por trás do Word2Vec – e de seus pares, como o BERT e o GloVe – advêm de uma abordagem linguística de forte herança estruturalista (Gastaldi, 2020), a saber, a Semântica Distribucional (SD). Na SD, que, neste trabalho, tratamos como

[1] Aqui faço uma analogia com a clássica citação de John R. Firth: "Reconhecerás uma palavra pela companhia que ela mantém" (Firth, 1955, p.108, tradução nossa).

uma teoria da significação (cf. Giamarusti, 2025), adota-se o entendimento de que palavras que ocorrem em contextos parecidos tendem a ter significados próximos. Em contrapartida, em nosso entendimento, pressupor que o significado de uma palavra está relacionado às palavras ao seu redor é o mesmo que afirmar, em termos de linguística saussuriana, que o significado de uma palavra é determinado pela rede de valores em *similia* (Saussure, 1910-1911, f. 27) e em *dissimilia* (ibidem) que cercam o signo.

Sendo assim, proponho-me a discutir que o Word2Vec pode nos auxiliar a demonstrar, computacionalmente, de que forma Saussure teoriza um sistema linguístico de valores relacionais, em que o significado não é dado por uma palavra isolada, mas "pelo concurso do que existe fora dela" (Saussure, 2012 [1916], p.162), a saber, as relações externas do signo e a Teoria do Valor (TdV).

A partir dessa discussão que aqui me dedico, espero possibilitar reflexões sobre a possível pertinência do arcabouço teórico de Ferdinand de Saussure para a construção de um novo ponto de vista, em PLN, sobre a forma como determinados modelos de linguagem entendem e processam significados em línguas naturais. O ponto de visto que trago, assim, tem como base as elaborações teóricas do mestre de Genebra contidas em sua obra póstuma, o Curso de Linguística Geral (CLG), no conjunto de manuscritos *Notes pour le cours III* (BGU Mr. 3951, 1910-1911, 56f)[2] e em anotações do caderno de Émile Constantin, um de seus alunos presentes no terceiro curso de linguística geral, ministrado por Saussure em Genebra entre 1910 e 1911.

Um passo anterior ao valor: os signos linguísticos e suas relações internas e externas

Antes de avançarmos sobre o que entendemos por valor linguístico e de que forma ele pode ser percebido como base teórica do Word2Vec para além da semântica distribucional, é fundamental voltarmos a nossa discussão para algumas questões anteriores à TdV que são bastante caras ao saussurianismo. Esse retorno a alguns conceitos prévios ao valor se faz importante, primeiro

[2] A referência "BPG Ms. 3951, 1910-1911, 56f" significa que este conjunto de manuscritos escrito por Saussure está arquivado na Biblioteca Pública de Genebra, sob a inscrição Ms. 3951, com data provável entre 1910 e 1911, contendo 56 folhas de material.

porque foi este o movimento de Saussure para chegar à exposição da Teoria do Valor, que ocorreu de forma mais amadurecida somente no Terceiro Curso de Linguística Geral[3], que ocorreu na Universidade de Genebra, com início em 29 de outubro de 1910 e término em 4 de julho de 1911 (cf. De Mauro, 1967, p.353).

Para nós, a chave para entender a complexidade da Teoria do Valor reside, em primeiro lugar, na compreensão de que a língua é um sistema: "a língua é um sistema que exprime ideias" (Saussure, 2012 [1916], p.34). Dessa afirmação, cabe perguntar-nos: o que seria, então, *sistema* para Saussure? Essa pergunta, por sua vez, já foi objeto de estudo do saussurianismo em diferentes fases da recepção das ideias do mestre genebrino (aqui, destacamos o trabalho de Silveira, 2016); e, a bem da verdade, ainda não possuímos evidências documentais suficientes que possam satisfazer a lacuna "A noção de sistema, para Saussure, é…".

Micaela P. Coelho, experiente pesquisadora brasileira nos estudos saussurianos, em sua dissertação de mestrado intitulada *A noção de sistema na fundação da Linguística Moderna* (2015), demonstra que, embora não haja, no saussurianismo, uma afirmação geral para o que se entende por *sistema*, pode-se notar que esta noção era um tema caro a Saussure mesmo antes da realização dos cursos de Linguística Geral em Genebra:

> O *Mémoire* – que é um livro publicado, embora não seja póstumo – também representa um elemento que contribui para a compreensão da teorização do linguista. Embora não se enquadre em um documento que apresenta reflexões sobre Linguística Geral, é possível notar que nele são apresentadas algumas noções cruciais para a contribuição original de Saussure. Todavia, essas noções – como é o caso da noção de sistema – não são apresentadas, no documento em questão, nem como conceitos definidos nem como princípios relacionados à língua enquanto um objeto sincrônico com funcionamento próprio. Apesar disso, assinalaram um caminho

[3] Como se sabe, entre 1907 e 1911, Saussure ministrou três grandes cursos intitulados como de "Linguística Geral", os quais serviram de base para a constituição de uma edição póstuma desses cursos, publicada em 1916 por Charles Bally e Albert Sechehaye: o conhecido "Curso de Linguística Geral". Para a elaboração do material, Bally indica, no prefácio à edição francesa, que foram utilizadas anotações dos cadernos de ouvintes dos três cursos, alguns manuscritos do próprio Saussure, além da experiência que Bally e Sechehaye tinham com a teoria saussuriana devido à proximidade com o próprio Saussure. Para compreender melhor a organização do CLG e a relação da edição com os cadernos dos alunos, sugerimos a leitura da edição crítica do *Cours* de Rudolf Engler, publicada em língua francesa em dois volumes, em 1967, pela editora Otto Harrasowitz e Wiesbaden.

> para que a língua fosse pensada dessa forma (Coelho, 2015, p.110).

Coelho (2015), no mesmo trabalho, demonstra que a noção de *sistema* no CLG[4], além de ser atravessada por flutuações terminológicas, não é clara e, em grande medida, faz referência à ideia de que a língua é um conjunto de estados de língua intrinsecamente conectados. Uma interpretação parecida, por sua vez, aparece em Ducrot e Todorov (1972), os quais argumentam que a noção saussuriana de sistema refere-se à organização interna da língua[5]. Os autores, ademais, afirmam que os sucessores de Saussure, posteriormente à publicação do CLG, notadamente os linguistas de Praga (cf. Altman, 2021), optaram por utilizar a nomenclatura "estrutura" em vez de "sistema" - de onde argumenta-se, inclusive, a ideia de que Saussure seria o precursor do acontecimento estruturalista na linguística ocidental:

> De uma forma positiva, agora, Saussure mostra que a linguagem, em qualquer momento de sua existência, deve se apresentar como uma organização. **Essa organização inerente a toda língua, Saussure a chama de sistema (seus sucessores frequentemente falam de estrutura)** (Ducrot; Todorov, 1972, p.31, tradução nossa, grifo nosso).[6]

Nesse sentido, neste trabalho, também trabalhamos com a ideia de que a noção de sistema refere-se à organização da língua e, por consequência, aos estados de língua que a compõem. Essa afirmação, consequentemente, leva-nos a uma outra teoria que, assim como o *sistema*, é fundamental para a linguística saussuriana e, por consequência, para a TdV: os signos linguísticos.

Os signos linguísticos, em linhas gerais, são as entidades linguísticas que representam os diferentes estados de língua que compõem o sistema linguístico. Embora Saussure não tenha sido o primeiro a refletir sobre a existência de signos em matéria de linguagem, o mestre genebrino, por outro lado, foi precursor em trazer maior notoriedade para essa discussão, tornando

---

[4] Como se sabe, o Curso de Linguística Geral (CLG) é uma obra póstuma de autoria atribuída a Saussure, a qual foi organizada por dois colegas do mestre genebrino, os linguistas Charles Bally e Albert Sechehaye, com base em manuscritos autógrafos de Saussure, como as notas para os cursos, bem como anotações de ouvintes dos cursos de Genebra, como E. Constatin, C. Patois, A. Riedlinger e outros.

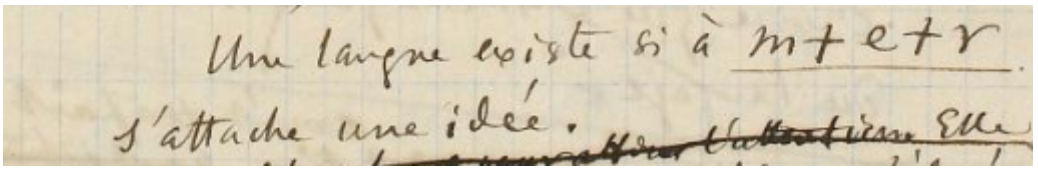
[5] Se consideramos que a noção de sistema refere-se à organização da língua, pode-se dizer que esta noção já permeia as discussões linguísticas desde a antiguidade (cf. Ducrot, 1968).

[6] *D'une façon positive, maintenant, Saussure montre que le langage, à tout moment de son existence, doit se présenter comme une organisation. Cette organisation inhérente à toute langue, Saussure l'appelle système (ses successeurs parlent souvent de structure).*

o signo linguístico um fator preditor de um dos conceitos de língua que aparece em seu escopo teórico[7].

Em outra ocasião (Giamarusti, 2025), apresentamos um trecho do manuscrito autógrafo de Saussure, intitulado *Le Double Essence du Langage* (Dupla Essência da Linguagem/EDL), arquivado na Biblioteca Pública de Genebra sob a classificação Arc. de Saussure 372, 1891, no qual pode-se encontrar uma espécie de reforço à ideia de que a língua é organizada por meio de relações internas entre os constituintes que compõem o signo linguístico, a saber, o significante - a imagem acústica; e o significado, o conceito. Nota-se, além disso, que essa relação interna entre significante e significado - arbitrária, por sinal - garante um possível conceito de língua. Esse ponto é particularmente evidente na folha 14 do EDL, como vemos:

Figura 1. Dupla Essência da Linguagem, BGU Arc. de Saussure, 372, 1891, 14f.



Une langue existe si à m + e + r j'attache un idée.
Uma língua existe se a m+e+r lhe corresponder uma ideia.[8]

Na folha 14 do manuscrito, Saussure faz uma reflexão sobre o aspecto sonoro do termo *mer* (mar), que o autor utiliza como exemplo para discutir o caráter acústico da língua (*le domaine de l'acoustique*). A partir desse pequeno fragmento de uma teoria ainda em formulação, pode-se dizer que a junção entre o significante a uma ideia é dos fatores constituidores do que, inclusive, Saussure entende por língua. Em outras palavras, uma língua "existe", na medida em que ela se constitui, sumariamente, de signos linguísticos que possuem relações internas entre suas duas faces. Essa união entre um significante e um significado - união arbitrária, por sinal (cf. Saussure, 2012 [1916]) - parece delimitar uma espécie de identidade para cada signo, o que, consequentemente, revela um conceito de língua que advém da relação entre

[7] De-Vitto, Oliveira e Souza (2022) demonstraram, a partir do CLG, a possibilidade de se pensar em diferentes conceitos de língua na teoria saussuriana, por isso, adotamos o plural "conceitos de língua".
[8] No original: "Un langue existe si à m+e+r l'attache une idée."

essas duas dimensões complementares — a imagem acústica e o conceito.

No Curso de Linguística Geral, a discussão em torno do signo é grande e ganha destaque em diferentes capítulos da obra. De forma geral, podemos anotar que “o signo linguístico é, pois, uma entidade psíquica de duas faces” (Saussure, 2012, p.106), formado pela união indissociável entre um significante e um significado: “esses dois elementos estão intimamente unidos e um reclama o outro” (ibidem). Um fato que não escapa à descrição dos signos, além do exposto, é a notável importância que a noção de *relação* (Marques, 2016) admite na teoria saussuriana, especialmente na teoria dos signos linguísticos e, consequentemente, na Teoria do Valor.

Apoiando-nos no trabalho saussurianista de Allana Marques, intitulado *A noção saussuriana de relação na teoria linguística de Ferdinand de Saussure* (2016), defendemos a ideia de que o arcabouço teórico de Saussure é atravessado por um princípio relacional que é materializado tanto a nível epistemológico, isto é, na relação complementar entre suas próprias teorias, quanto na natureza e na formação de seus pressupostos. No caso dos signos linguísticos, por exemplo, Saussure discute, como veremos a seguir, a existência de, pelo menos, dois tipos de relações principais, a saber, as relações internas e externas do signo. E é importante fazermos alguns comentários sobre essas relações, tendo em vista que o entendimento delas é fundamental para o estudo da Teoria do Valor.

Nas anotações do caderno de Émile Constantin, ouvinte do Terceiro Curso de Linguística Geral (TCLG), podemos acompanhar o percurso teórico que Saussure faz sobre a natureza relacional e arbitrária da(s) língua(s), partindo do pressuposto de que o signo é formado por relações interiores e exteriores, ambas arbitrárias. Quanto às “relações internas”, no excerto 307 da folha 53 do caderno de Constantin, pode-se ver uma delimitação desse termo por Saussure: “Há, por um lado, <a primeira> **relação interna**, que nada mais é do que uma associação entre a imagem acústica e o conceito.” (Constantin; Saussure, 2005 [1910-1911), p.235, tradução nossa, grifo nosso)[9].

Já as “relações externas” do signo, por sua vez, aparecem apenas mencionadas por Saussure neste curso, igualmente na folha 53. Nesta fase do

[9] *Il y a d’un côté <la première> une relation intérieure, qui n’est autre chose qu’une association entre l’image acoustique et le concept.*

terceiro curso, Saussure estava explicando sobre a natureza arbitrária do signo, isto é, a não motivação das relações internas entre o significante e o significado. Contudo, mesmo sem deixar muito claro exatamente do que se tratam as relações externas do signo, Saussure as menciona: “À primeira vista, parece que não há nada em comum entre essa relação interna e essa **relação externa** com um termo oposto” (ibidem).

Nesse sentido, qualquer teorização que fizermos sobre as relações externas do signo, na verdade, são apenas inferências. Nossa hipótese, assim, é que tais relações exteriores estão no plano das relações de um signo para com outro. Essa interpretação pode ser corroborada também por outras anotações de Constantin, notadamente por uma ilustração utilizada por Saussure para demonstrar as relações entre os termos do sistema:

Figura 2. Relações externas dos signos

Fonte: Adaptado de Constantin e Saussure (2005 [1910-1911]), p.235.

Dessa figura, podemos inferir que a delimitação de um signo se dá não apenas por suas relações interiores, entre o significante e o significado, mas também por suas relações exteriores, ou seja, pela posição relativa que cada termo, ou signo, ocupa no sistema à margem de outro signo.

Nessa perspectiva, essas relações externas entre os signos é o que consideramos, neste trabalho, como a principal fonte formadora dos valores linguísticos, tendo em vista que são elas as responsáveis por sugerir uma espacialidade para a(s) língua(s), conferindo a cada elemento do sistema um espaço e valores únicos. Sendo assim, na próxima seção, vamos aprofundar nessa discussão sobre a Teoria do Valor, a fim de, na sequência, hipotetizamos de que forma as relações externas do signo podem ser computacionalmente demonstradas por meio de técnicas de embeddings de palavras,

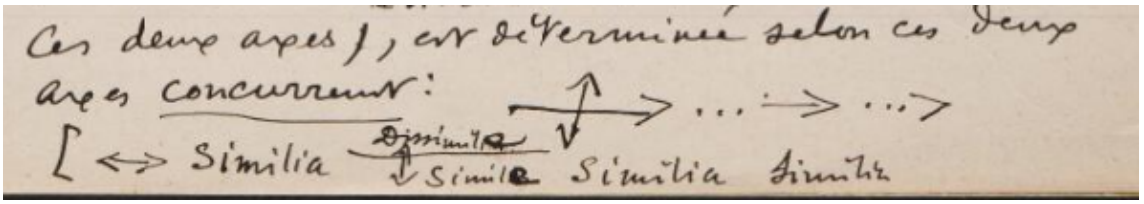
permitindo-nos visualizar as aproximações teóricas entre a Teoria do Valor de Ferdinand de Saussure e o funcionamento do Word2Vec.

A teoria do valor: entre semelhanças e diferenças

Em linhas gerais, a chave para entender a dinâmica dos valores reside na ideia de que a língua é um sistema de relações (cf. Saussure, 2012 [1916]). Em outra oportunidade, já explicamos que: "[...] a ideia básica construída por Saussure sobre a TdV, a qual podemos acompanhar no CLG nos capítulos dedicados ao valor, é de que a noção de relação é um princípio que atravessa a constituição dos valores linguísticos" (Giamarusti, 2024, p.499). Isso implica dizer que os valores fazem parte da própria natureza relacional do signo.

Para determinar então os valores linguísticos, Saussure indica, em pelo menos 3 ocasiões diferentes, a necessidade de dois eixos de relações, representados por segmentos de reta que retomam, em bastante grau, à noção de vetores[10]. A primeira evidência documental disso encontra-se no conjunto de manuscritos autógrafos intitulado "Notes pour le 3e cours de linguistique générale, 1910-1911" (*Notas para o Terceiro Curso de Linguística Geral, 1910-1911*), arquivado na Biblioteca Pública de Genebra (BGU) sob a inscrição Ms fr. 39551/23, como vemos:

Figura 3. Transcrição e tradução de trecho da folha 27 do conjunto de manuscritos Ms fr. 39551/23.



Ces deux aspects [les valeurs], il y a à déterminer selon ces deux axes concurrents

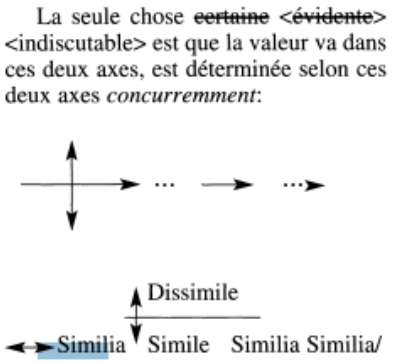
[10] Importante dizer que Saussure tem uma grande relação com as ciências exatas; e, portanto, afirmar que as "flechas" por ele utilizadas se assemelham a vetores não é uma super interpretação. Joseph (2012) relembra que, no período que permaneceu em Leipzig, na virada do século XIX para o XX, Saussure se dedicou a cursos de Química e Física, antes de engendrar na carreira linguística. Além disso, René de Saussure, seu irmão 6 anos mais novo, era um conhecido matemático, com quem Ferdinand trocava constantemente correspondências teóricas, já que René, para além da matemática, também era um grande entusiasta da morfologia, inclusive com textos publicados sobre processos de formação de palavra (cf. Anderson; De Saussure, 2018)

Esses dois aspectos [os valores] devem ser determinados segundo esses dois eixos concorrentes[11]

Fonte: Saussure, Notes pour le 3e cours de linguistique générale, 1910-1911

No caderno de Émile Constantin, encontramos uma explicação parecida e a mesma figura, o que sugere que os eixos de determinação dos valores se mantiveram das notas preparatórias à realização do Terceiro Curso de Linguística Geral:

Figura 4. Transcrição e tradução de trecho de folha do caderno de Émile Constantin sobre o TCLG.



La seule chose ~~certaine~~ <évidente> <indiscutable> est que la valeur va dans ces deux axes, est determinée selon ces deux axes concourrement:

A única coisa ~~certa~~ <evidente> <indiscutível> é que o valor segue nesses dois eixos, sendo determinado conforme esses dois eixos *simultaneamente.*

Fonte: Constantin; Saussure, 2005 [1910-1911), p.283, tradução nossa.

Porém, vale dizer, essas figuras não estão na edição do Curso de Linguística Geral. A provável hipótese para isso é o fato de que os editores do CLG não tiveram acesso ao caderno de Constantin durante a compilação das anotações dos alunos, conforme relembra De Mauro (1973)[12]. Por outro lado, no CLG, há uma explicação sobre a determinação dos valores via dois eixos, a saber, a semelhança e a dessemelhança :

[11]Vale ressaltar que o termo “valores” foi inserido por nós com base no conteúdo anterior a esse trecho do manuscrito, tendo em vista, que nesta folha, Saussure dedica-se à descrição do valor linguístico por eixos.

[12] Neste trabalho, não nos detivemos em comparações profícuas entre o CLG e os cadernos dos alunos. Sendo assim, caberia um trabalho secundário buscando analisar se essas imagens do caderno de Constantin aparecem nos cadernos de outros alunos.

> [...] verifiquemos inicialmente que, mesmo fora da língua, todos os valores parecem estar regidos por esse princípio paradoxal. Eles são sempre constituídos:
> 1º - por uma coisa *dessemelhante*, suscetível de ser *trocada* por outra cujo valor resta determinar:
> 2º - por coisas *semelhantes* que podem *comparar* com aquela cujo valor está em causa. (Saussure, 2012, p.162).

Em outras palavras, o valor de uma unidade linguística não pode ser definido de forma isolada, mas depende das oposições e semelhanças que ele mantém com as demais unidades do sistema, o que faz com que cada signo ocupe um espaço determinado na língua, marcado por diferenças e semelhanças. No próprio CLG, Saussure traz uma explicação mais clara sobre essa dinâmica dos valores. O mestre genebrino (Saussure, 2012 [1916]), no capítulo IV da Segunda parte do Curso de Linguística Geral, utiliza o exemplo de uma moeda de 5 francos para exemplificar essa dinâmica relacional do valor. Para determinar o valor do signo “5 francos”, Saussure indica a necessidade de trocar este signo por algo dessemelhante, como “pão”, e, em seguida, trocar por termos semelhantes, por exemplo, uma moeda de 1 franco. Dessa forma, Saussure demonstra que cada signo adquire valor ao se diferenciar dos outros e ao se aproximar dos signos em *similia*, configurando um jogo de distinções que sustenta toda a organização do sistema de relações.

Essa compreensão da língua como um sistema relacional leva-nos, portanto, à rejeição de qualquer abordagem que trate os signos como entidades estáveis e autônomas. Saussure propõe, ao contrário, que o valor de um signo é determinado não pelo seu conteúdo, isto é, apenas as relações internas e arbitrárias entre o significante e o significado, mas pelas diferenças que o signo mantém em relação aos outros signos, conferindo-lhe relações externas entre um signo e outros: “o valor [de uma palavra] não estará então fixado [...] falta ainda compará-la com os valores semelhantes, isto é, com as palavras que se lhe podem opor” (Saussure, 2012 [1916], p.131).

Posto isso, esse princípio é essencial para entender a dinâmica dos valores linguísticos, pois nos permite visualizar a língua como um sistema em constante equilíbrio, em que as relações externas do signo se deslocam e dão forma ao sistema, delimitando-lhe uma organização própria em função das relações dos termos que compõem esse sistema.

Nessa perspectiva, ao considerarmos o valor como produto dessas

relações externas entre os signos, é possível, enfim, refletir sobre a própria natureza da significação nas línguas. Da Teoria do Valor, portanto, pode-se chegar a uma interpretação sobre a própria constituição dos sentidos na(s) língua(s) naturais.

No CLG, no quarto capítulo da segunda parte, dedicado à TdV, Saussure deixa pistas de que a ideia de significado e de valor estão muito próximas, ainda que, terminologicamente, não se confundam. A dificuldade em delimitar a fronteira que separa o valor e o significado reside no fato de que, aparentemente, tanto os valores quanto os conceitos se dão por mecanismos de determinação muito parecidos. Isto é, Saussure rejeita a noção de que o significado das palavras deriva de uma conexão direta entre a imagem acústica e os conceitos que ela designa, isto é, as relações internas do signo. Em vez disso, ele parece sugerir que o significado de um signo, assim como valor, é sempre relacional e mutável, conforme as mudanças do sistema linguístico e os contextos semânticos em que o signo é usado. E, aqui, encontramos a grande semelhança entre Saussure e as bases teóricas do Word2Vec.

A Semântica Distribucional, fundamento central dos modelos de linguagem vetoriais, preconiza, tal como Saussure o faz por meio da teoria do valor, que o significado da palavra não é fixo, mas está contido nas relações do sistema, principalmente, por meio do eixo da semelhança:

> A semântica distribucional é ancorada na Hipótese Distribucional que preconiza que palavras que têm um contexto linguístico semelhante tendem a ter significado similar ou aproximado. É o caso, por exemplo, de palavras como “ensino” e “educação” que costumam aparecer no mesmo contexto de palavras como “aluno”, “escola” e “professor”, sugerindo que existe uma similaridade entre as duas palavras em certos contextos (Seno et. al, 2023, p.1).

Em seu clássico artigo, *Structural linguistics*, Firth (1955) propõe, então, o entendimento linguístico que sustenta a SD, a saber: “Reconhecerás uma palavra pela companhia que ela mantém” (Firth, 1955, p.108, tradução nossa). O mesmo já havia sido previsto por Harris, em 1954, por meio da obra *Distributional Structure*, em que o autor, como também já demonstramos (cf. Giamarusti, 2025), indica que termos com traços semânticos comuns estão próximos no campo semântico, enquanto que signos com traços distintivos estão mais distantes. Não é surpreendente, assim, que Saussure tenha

previsto essa mesma proposta semântica para o valor, delimitando-o como não fixado e interdependente dos signos, conforme vemos em outra anotação de Constantin:

> Não se pode tomar as palavras isoladamente. É assim que o sistema do qual procede o termo é uma das fontes do valor. É a soma dos termos comparáveis por oposição à ideia trocada. O valor de uma palavra [395] nunca será determinado senão pela concorrência dos termos coexistentes que o limitam; <ou, para reforçar o paradoxo levantado:> O que está dentro da palavra nunca é determinado senão pela concorrência do que existe ao seu redor (o que está na palavra é o valor) – ao seu redor sintagmaticamente ou ao seu redor associativamente. É preciso abordar <a palavra> de fora, partindo do sistema e dos termos coexistentes (Constantin; Saussure, 2005 [1910-1911], p.284, transcrição e tradução nossas)[13].

Assim, ao compararmos a proposta de Saussure sobre a teoria do valor com a Hipótese Distribucional, encontramos um ponto de convergência fundamental: em ambos os casos, os valores de uma palavra - e os significados por consequência - não são intrínsecos ao signo em si, mas emergem das relações externas que o signo estabelece no sistema linguístico. Saussure explica que o valor de um termo é determinado "pela concorrência dos termos coexistentes que o limitam", ou seja, o valor surge pela posição relativa de um termo em relação aos outros dentro do sistema.

A mesma lógica é aplicada pelos teóricos da semântica distribucional. A partir de Firth (1955) e Harris (1954), a Hipótese Distribucional sugere que o significado de uma palavra é construído com base nos contextos em que ela ocorre. Palavras que aparecem em contextos semelhantes tendem a ter significados semelhantes, enquanto palavras que ocorrem em contextos diferentes tendem a ter significados distintos. Essa perspectiva é exemplificada em modelos de representação vetorial, como o Word2Vec e o GloVe, que constroem vetores semânticos a partir de padrões de co-ocorrência em grandes *corpora* de textos. Tais modelos operam no mesmo princípio sugerido por Saussure: o valor de uma palavra - de um signo - é definido pelas relações de semelhança e dessemelhança que ela mantém com outras palavras/signos.

---

[13] *On ne peut prendre les mots isolément. C'est ainsi que le système d'où procède le terme est une des sources de la valeur. C'est la somme des termes comparables par opposition à l'idée échangée. La valeur d'un mot ne [395] sera jamais déterminée que par le concours des termes coexistants qui le limitent; <ou pour mieux appuyer sur le paradoxe relevé:> Ce qui est dans le mot n'est jamais déterminé que par le concours de ce qui existe autour de lui (ce qui est dans le mot, c'est la valeur) – autour de lui syntagmatique ou autour de lui associativement.*

Nesse sentido, podemos traçar um paralelo direto entre as explicações de Saussure e a matemática subjacente aos modelos vetoriais de linguagem, a saber: (i) em Saussure, o valor linguístico é compreendido como a soma das diferenças que uma palavra mantém em relação aos outros signos do sistema. (ii) Na Hipótese Distribucional, o significado de uma palavra é modelado também por outros signos, convertidos em vetores num espaço semântico no qual a proximidade vetorial indica similaridade semântica.

Sendo assim, argumentamos que o que Saussure previu em termos conceituais, a semântica distribucional de palavras parece demonstrar em termos computacionais: o significado não é um atributo fixo das palavras, mas uma entidade dinâmica, construído pelas relações e posições que os signos mantêm no sistema. Diante disso, consideramos que é o sistema, portanto, que atribui valor ao signo; e esse valor não é absoluto, mas relacional e dependente dos outros termos que compõem o conjunto.

Na prática, modelos como o Word2Vec ilustram essa ideia ao associar palavras como "ensino" e "educação", como explicado por Seno et al (2023), a contextos semelhantes, revelando que ambas compartilham traços semânticos. Da mesma forma, as palavras "água" e "fogo" tendem a estar distantes no espaço semântico, indicando que seus significados estão relacionados a contextos diferentes. Essa proposição, por sua vez, é central tanto para a teoria do valor saussuriana quanto para a Hipótese Distribucional.

Com base nessas discussões, na próxima seção, objetivamos demonstrar, de forma objetiva, de que maneira a teoria do valor pode ser convertida numa teoria de semântica distribucional, transformando valores em vetores num espaço vetorial guiado por eixos de semelhança e dessemelhança, os quais podem auxiliar diferentes modelos de linguagem na tarefa de modelagem semântica de itens lexicais, bem como possíveis análises de similaridade.

Word2Vec: um algoritmo saussuriano

Com base em nossa discussão, propomos que o Word2Vec emerge de um contexto de possível retorno à linguística formalista impulsionado por avanços no Processamento de Linguagem Natural (PLN). O artigo seminal que apresenta o algoritmo à comunidade científica é o *Efficient Estimation of Word*

Representations in Vector Space, publicado por Mikolov e colaboradores em 2013. Essa publicação inaugura oficialmente o Word2Vec e alavanca a aplicação de embeddings de palavras de alta dimensão para tarefas como classificação de texto, busca semântica e cálculo de probabilidade de ocorrência.

O modelo proposto por Mikolov representou um verdadeiro avanço no PLN. Isso porque o modelo utiliza redes neurais e técnicas de aprendizado profundo (deep learning) para capturar de forma mais eficiente as relações semânticas e sintáticas entre as palavras. Com isso, tornou-se possível que sistemas de inteligência artificial reconhecessem e manipulassem relações complexas entre signos, ajustando dinamicamente seus vetores em múltiplas camadas representacionais.

Apesar de sua importância, o modelo não está isento de desafios. Kenneth Church (2016), professor do Institute for Experiential AI (EAI), da Northeastern University (EUA), aponta limitações importantes, como a dificuldade do Word2Vec em lidar com palavras polissêmicas e com ambiguidades contextuais. Ainda assim, o pesquisador reconhece o valor do modelo, sobretudo por sua simplicidade técnica e ampla aplicabilidade:

> Word2vec não é o primeiro, o último ou o melhor [modelo de linguagem] para discutir espaços vetoriais, embeddings, analogias, métricas de similaridade etc. Mas o word2vec é simples e acessível. Qualquer um pode baixar o código e usá-lo em seu próximo artigo. E muitos o fazem (para o melhor e para o pior) (Church, 2016, p.156)[14]

Nessa perspectiva, o Word2Vec opera com base em dois algoritmos principais: Skip-Gram e CBOW (Continuous Bag of Words). Ambos têm como objetivo capturar regularidades estatísticas do uso das palavras em grandes corpora textuais, baseando-se, sobretudo, na hipótese distributiva, como vimos.

No modelo Skip-Gram, o objetivo é prever palavras de contexto a partir de uma palavra-alvo. Já no CBOW, faz-se o oposto: o modelo utiliza palavras do contexto para prever a palavra central. Em ambos os casos, o treinamento ocorre a partir de grandes quantidades de dados textuais não anotados, sendo

[14] No original: “Word2vec is not the first, last or best to discuss vector spaces, embeddings, analogies, similarity metrics, etc. But word2vec is simple and accessible. Anyone can download the code and use it in their next paper. Any many do (for better and for worse).”

que a eficiência e a capacidade de generalização decorrem do uso de redes neurais rasas e da vetorização de palavras em espaços contínuos de alta dimensionalidade.

Para ilustrar o funcionamento do Skip-Gram, consideremos a frase: “O gato caça o rato.” Vamos supor, por hipótese, a ausência do verbo “caça”. A tarefa do modelo seria prever a palavra ausente a partir das demais: “O gato _____ o rato.” Nesse exemplo, “caça” seria a palavra-alvo, e o contexto seria formado pelas palavras vizinhas. Com uma janela de tamanho 2, consideraríamos duas palavras à esquerda e duas à direita da palavra-alvo, totalizando até quatro palavras contextuais.

O objetivo do Skip-Gram é, portanto, maximizar a probabilidade de ocorrência das palavras de contexto (c), dado uma palavra-alvo (w). Isso pode ser expresso pela seguinte função de otimização: max $\sum \log p(c \mid w)$, para todas as palavras *w* e contextos *c* observados no corpus D.

Por exemplo, se a palavra-alvo for “caça” e o contexto for ["o", "gato", "o", "rato"], o modelo deverá atribuir maior probabilidade às co-ocorrências mais informativas, como p(“gato” | “caça”) e p(“rato” | “caça”), e menor probabilidade a termos funcionais de menor valor semântico, como p(“o” | “caça”). Durante o treinamento, os vetores são ajustados de modo a refletir essas probabilidades, de tal forma que palavras semanticamente próximas tendem a ocupar posições similares no espaço vetorial.

Nesse sentido, um ponto fundamental para este trabalho é o retorno à teoria saussuriana que tanto o método Skip-Gram quanto o CBOW promovem, a saber, a ideia de que nenhum signo pode ser determinado de forma isolada.

Para que a palavra “caça” seja reconhecida pelo computador, é preciso considerar a rede de signos com a qual ela contrai relações — seus semelhantes e dessemelhantes. Sob uma perspectiva saussuriana, poderíamos dizer que o valor de “caça” na frase “O gato _____ o rato” decorre de sua posição diferencial em relação a termos como “come”, “persegue” ou “brinca”, por exemplo. Assim, como afirma Saussure (2012 [1916]), o valor de um signo e, consequentemente o seu significado, emerge das relações diferenciais dos signos dentro do sistema. Essas relações

diferenciais, portanto, materializam-se em valores *similia* (semelhança) ou *dissimilia* (dessemelhança).

Sendo assim, defendemos que o que Word2Vec faz, na prática, é operacionalizar como a dinâmica de valores funciona e de que modo ela está articulada com a determinação de significados. Nas línguas naturais, Saussure teoriza que o significado necessariamente está intimamente relacionado com os valores que o cercam. No Word2Vec, de forma muito parecida, a semântica distribucional preconiza que o significado advém da companhia que a palavra mantém. Em ambos os casos, o fator relacional do signo é essencial para que, enfim, haja a produção de sentidos na língua; e, no caso da semântica distribucional, essa relação é fundamental para tarefas de PLN cuja anotação semântica seja uma etapa importante do projeto.

Considerações finais

Ao longo deste artigo, percorri um caminho teórico que partiu das limitações da técnica TF-IDF no Processamento de Linguagem Natural até chegar em técnicas de embeddings de palavras, como o Word2Vec, que diferentemente do TF-IDF, preconizam janelas de contexto e a relação entre as palavras para a modelagem semântica de itens lexicais.

Minha proposta, assim, foi ousada, admito, mas fundamentada numa recepção saussuriana contemporânea que permite uma análise do Word2Vec à luz da Teoria do Valor de Ferdinand de Saussure. Para tanto, revisitei os fundamentos teóricos do saussurianismo, sobretudo no que tange à noção de sistema, aos signos linguísticos e às suas relações internas e externas. Esse retorno foi essencial para compreendermos como a teoria saussuriana concebe o significado como algo relacional, dinâmico e constituído a partir das diferenças e semelhanças que os signos mantêm entre si no seio do sistema linguístico.

Sendo assim, considerei diferentes fontes saussurianas para compor de que forma o significado está articulado com a determinação dos valores linguísticos. Vimos que, no conjunto de manuscritos *Notes pour le 3e Cours*, Saussure indica a necessidade de, pelo menos, dois tipos de valores para a

determinação de um signo frente aos outros: o eixo da semelhança e da dessemelhança. Posto isso, demonstramos que esse mesmo entendimento presente em anotações de Saussure para o TCLG, manteve-se na execução do Terceiro Curso, entre 1910 e 1911. Essa conclusão, assim, advém da análise que realizamos de alguns trechos do caderno de Émile Constantin, demonstrando de que forma essa ideia de que o signo é perpassado por uma rede de valores é apresentada, posteriormente, no Curso de Linguística Geral.

Ao alinharmos essa teoria à Semântica Distribucional – corrente que sustenta matematicamente o funcionamento do Word2Vec –, identificamos uma convergência de princípios: tanto Saussure quanto os autores da Hipótese Distribucional compreendem que o sentido de um signo não reside nele próprio, mas nas relações que ele estabelece com os outros signos.

No Word2Vec, essa dinâmica se materializa por meio da vetorização de palavras, onde a proximidade vetorial corresponde à similaridade semântica, reiterando, sob uma nova linguagem, o que Saussure já anunciava no início do século XX: o valor e o significado de um signo decorre da concorrência dos termos que a limitam.

Com base nessa articulação, propomos que o Word2Vec pode ser compreendido, em última instância, como um algoritmo saussuriano. Ele não apenas compartilha com Saussure a compreensão relacional do significado, como também concretiza, em níveis computacionais, os eixos de semelhança e dessemelhança que fundamentam a Teoria do Valor. Tal proposta não busca reduzir a complexidade da teoria saussuriana a um modelo computacional, mas sim evidenciar o quanto os pressupostos teóricos do mestre genebrino seguem sendo atuais e produtivos para pensar em pontos de vista contemporâneas de representação e de processamento semântico nas línguas naturais.

Dessa forma, espero ter contribuído com um pequeno *start* para o reconhecimento da pertinência e da vitalidade do pensamento saussuriano na era digital, abrindo espaço para um diálogo fecundo entre a linguística saussuriana e as técnicas mais recentes de modelagem semântica em PLN.

REFERÊNCIAS

AIZAWA, A. An information-theoretic perspective of tf-idf measures. *Information Processing & Management*, v. 39, n. 1, p. 45–65, jan. 2003. DOI: https://doi.org/10.1016/S0306-4573(02)00021-3.

COELHO, M. P. Sistema e relação na Teoria do Valor de Ferdinand de Saussure. *Estudos Linguísticos*, São Paulo, v. 42, n. 1, p. 378-391, 2013a. Disponível em: https://revistas.gel.org.br/estudos-linguisticos/article/view/878/1179. Acesso em: 8 março. 2025.

_____. "Significação" em Saussure: os três cursos de linguística geral. *Anais do SILEL*, Uberlândia, v. 3, n. 1, p. 943-956, 2013b. Disponível em: https://www.ileel.ufu.br/anaisdosilel/wp-content/uploads/2014/04/silel2013_943.pdf. Acesso em: 8 março. 2025.

CONSTANTIN, E; SAUSSURE, F. Linguistique générale (Cours de M. le professeur de Saussure) semestre d'hiver 1910-1911. *Cahiers Ferdinand de Saussure*, n. 58, p. 82–290, 2005. Disponível em: http://www.jstor.org/stable/27758719. Acesso em: 4 abr. 2025.

DE MAURO, T. Notes. In: *Cours de linguistique générale.* Édition critique preparé par Tullio de Mauro. Paris: Payot, 1973.

FIRTH, J. P. The semantics of linguistic science. IN: *Lingua*, v. 1, p. 393–404, 1949. Disponível em: https://doi.org/10.1016/0024-3841(49)90085-6. Acesso em: 4 abril. 2025.

GASTALDI, J. L. Why Can Computers Understand Natural Language? IN: *Philosophy & Technology*, v. 34, p. 149–214, 2020. Disponível em: https://doi.org/10.1007/s13347-020-00393-9. Acesso em: 4 abril. 2025.

GIAMARUSTI, L. *A linguística saussuriana aplicada ao processamento de linguagem natural*. 2025. 96 f. Dissertação (Mestrado em Estudos Linguísticos) – Programa de Pós-Graduação em Estudos Linguísticos (PPGEL), Universidade Federal de Uberlândia, Uberlândia, 2025.

_____. Saussure na Era da IA: Modelagem semântica de Word Embeddings à luz da Teoria do Valor (TdV). *Revista Desenredo*, *[S. l.]*, v. 20, n. 3, 2024. DOI: 10.5335/rdes.v20i3.16461. Disponível em: https://seer.upf.br/index.php/rd/article/view/16461. Acesso em: 20 dez. 2024.

HARRIS, Z. S. Distributional Structure**.** IN: *WORD*, v. 10, n. 2-3, p. 146-162, 1954. Disponível em: https://doi.org/10.1080/00437956.1954.11659520. Acesso em: 4 abril. 2025.

JOSEPH, J. *Saussure*. Oxford: Oxford University Press, 2012. Traduzido por Bruno Turra, Campinas (SP): Editora Unicamp, 2023 [2012].

JOACHIMS, T. Text Categorization with Support Vector Machines. IN: *European Conference on Machine Learning (ECML)*, 1998. DOI: 10.1007/BFb0026683.

MARQUES, A. C. M. **A noção de relação na teoria linguística de Ferdinand de Saussure**. 2016. 117 f. Dissertação (Mestrado em Estudos Linguísticos) - Universidade Federal de Uberlândia, Uberlândia, 2016. DOI: http://doi.org/10.14393/ufu.di.2016.180. Acesso em 04 de abril de 2025.

MIKOLOV, T.; SUTSKEVER, I.; CHEN, K.; CORRADO, G.; DEAN, J. Distributed representations of words and phrases and their compositionality. In: *arXiv*, Nova York, Cornell University, v. 1, s/n, janeiro, 2013. Disponível em: https://arxiv.org/abs/1301.3781. Acessado em 25 de março de 2025.

MASONI, G. *Análise de textos por meio de processos estocásticos na representação word2vec*. 2021. 58 f. Dissertação (Mestrado em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2021.

PANG, B.; LEE, L. Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, v. 2, n. 1-2, p. 1-135, 2008. DOI: 10.1561/1500000011.

SAUSSURE, F. *Curso de linguística geral.* Trad. de A. Chelini; J. P. Paes e I. Blikstein. 27ª ed. São Paulo: Cultrix, 2012. Cours de linguistique générale. Charles Bally e Albert Sechehaye (orgs.), com a colaboração de Albert Riedlinger, [1916].

_____. Notes pour le cour III. In: *Papiers Ferdinand de Saussure*, 3951. Bibliothèque de Genève, 1910-1911. 56 f.

_____. Science du langage. De la double essence du langage et autres documents du ms. BGE Arch. De Saussure 372. *Éditions critique partielle mais raisonné et augmentée des Écrits de linguistique générale*, établie par René Amacker, Genéve, Droz. 2011.

SALTON, G.; WONG, A.; YANG, C. A vector space model for automatic indexing. *Communications of the ACM*, v. 18, p. 613-620, 1975.

SEBASTIANI, F. Machine Learning in Automated Text Categorization**.** *ACM Computing Surveys (CSUR)***,** v. 34, n. 1, p. 1-47, mar. 2002. DOI: 10.1145/505282.505283.

SENO, E. R. M.; CLARO, D.; MOTA, L.; RODRIGUES, J. Semântica Distribucional. IN: Caseli, H.M.; Nunes, M.G.V. (org.) *Processamento de Linguagem Natural: Conceitos, Técnicas e Aplicações em Português*. 2 ed. BPLN, 2024. Disponível em: https://brasileiraspln.com/livro-pln/2a-edicao. Acessado em 28 de janeiro de 2025.

SOUZA, A.; MAIA, J. E. B. Agente inteligente para classificação de notícias por assunto. *Anais do Congresso X Computer on the Beach*, v. 10, 2019. Disponível em: https://periodicos.univali.br/index.php/acotb/article/view/14373. Acesso em: 4 abr. 2025.

DECLARAÇÃO DE CONFLITO DE INTERESSE

Declaro não estar submetido a qualquer tipo de conflito de interesse junto aos participantes ou a qualquer outro colaborador, direto ou indireto, para o desenvolvimento do Projeto de Pesquisa intitulado Word2Vec: um algoritmo saussuriano, cujos pesquisadores envolvidos são: Leonardo Giamarusti dos Santos.

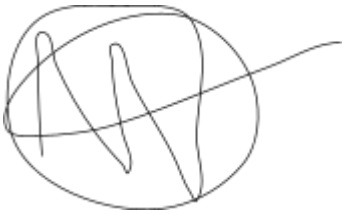
Declaro ainda que minha atuação como pesquisador é independente, autônomo e comprometida com o interesse precípuo de defesa de direitos e a segurança do(s) participante(s) de pesquisa nos termos da Resolução 466/12 e demais diretrizes éticas em pesquisas envolvendo seres humanos.

UBERLÂNDIA, 07 DE ABRIL DE 2025.

Nome completo do(a) Pesquisador(a)

Leonardo Giamarusti dos Santos

Assinatura do(a) Pesquisador(a)



Este preprint foi submetido sob as seguintes condições:

- Os autores declaram que estão cientes que são os únicos responsáveis pelo conteúdo do preprint e que o depósito no SciELO Preprints não significa nenhum compromisso de parte do SciELO, exceto sua preservação e disseminação.
- Os autores declaram que os necessários Termos de Consentimento Livre e Esclarecido de participantes ou pacientes na pesquisa foram obtidos e estão descritos no manuscrito, quando aplicável.
- Os autores declaram que a elaboração do manuscrito seguiu as normas éticas de comunicação científica.
- Os autores declaram que os dados, aplicativos e outros conteúdos subjacentes ao manuscrito estão referenciados.
- O manuscrito depositado está no formato PDF.
- Os autores declaram que a pesquisa que deu origem ao manuscrito seguiu as boas práticas éticas e que as necessárias aprovações de comitês de ética de pesquisa, quando aplicável, estão descritas no manuscrito.
- Os autores declaram que uma vez que um manuscrito é postado no servidor SciELO Preprints, o mesmo só poderá ser retirado mediante pedido à Secretaria Editorial do SciELO Preprints, que afixará um aviso de retratação no seu lugar.
- Os autores concordam que o manuscrito aprovado será disponibilizado sob licença Creative Commons CC-BY.
- O autor submissor declara que as contribuições de todos os autores e declaração de conflito de interesses estão incluídas de maneira explícita e em seções específicas do manuscrito.
- Os autores declaram que o manuscrito não foi depositado e/ou disponibilizado previamente em outro servidor de preprints ou publicado em um periódico.
- Caso o manuscrito esteja em processo de avaliação ou sendo preparado para publicação mas ainda não publicado por um periódico, os autores declaram que receberam autorização do periódico para realizar este depósito.
- O autor submissor declara que todos os autores do manuscrito concordam com a submissão ao SciELO Preprints.