



**UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE LETRAS E LINGUÍSTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS LINGUÍSTICOS**

LEONARDO GIAMARUSTI DOS SANTOS

**A LINGUÍSTICA SAUSSURIANA APLICADA AO PROCESSAMENTO DE
LINGUAGEM NATURAL**

**UBERLÂNDIA – MG
2025**

LEONARDO GIAMARUSTI DOS SANTOS

A LINGUÍSTICA SAUSSURIANA APLICADA AO PROCESSAMENTO DE
LINGUAGEM NATURAL

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Estudos Linguísticos do Instituto de Letras e Linguística da Universidade Federal de Uberlândia como exigência parcial para a obtenção do título de Mestre em Estudos Linguísticos.

Área de Concentração: Estudos em Linguística e Linguística Aplicada

Linha de Pesquisa: Linguagem, sujeito e discurso Orientadora:
Prof.^a Dra. Eliane Silveira

UBERLÂNDIA – MG

2025



ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Estudos Linguísticos				
Defesa de:	Dissertação de Mestrado - PPGEL				
Data:	Vinte e cinco de fevereiro de dois mil e vinte e cinco	Hora de início:	10:00	Hora de encerramento:	12:30
Matrícula do Discente:	12312ELI012				
Nome do Discente:	Leonardo Giamarusti dos Santos				
Título do Trabalho:	A linguística saussuriana aplicada ao Processamento de Linguagem Natural				
Área de concentração:	Estudos em Linguística e Linguística Aplicada				
Linha de pesquisa:	Linguagem, sujeito e discurso				
Projeto de Pesquisa de vinculação:	Ferdinand de Saussure e a Linguística Geral: da elaboração dos seus conceitos aos seus efeitos				

Reuniu-se, na sala 213, bloco U, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Estudos Linguísticos, assim composta: Professores Doutores: Eliane Silveira - UFU, orientadora da Dissertação; Allana Cristina Moreira Marques - UFRGS; Antônia Coutinho - Universidade Nova de Lisboa.

Iniciando os trabalhos a presidente da mesa, Dra. Eliane Silveira, apresentou a Comissão Examinadora e o candidato, agradeceu a presença do público, e concedeu ao Discente a palavra para a exposição do seu trabalho. A duração da apresentação do Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir, a senhora presidente concedeu a palavra, pela ordem sucessivamente, aos examinadores, que passaram a arguir a candidata. Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando a candidata:

Aprovado.

Esta defesa faz parte dos requisitos necessários à obtenção do título de Mestre.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar, foram encerrados os trabalhos. Foi lavrada a presente ata que, após lida e achada conforme, foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **Maria Antónia Diniz Caetano Coutinho, Usuário Externo**, em 25/02/2025, às 20:25, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Eliane Mara Silveira, Professor(a) do Magistério Superior**, em 26/02/2025, às 14:14, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Allana Cristina Moreira Marques, Usuário Externo**, em 24/03/2025, às 11:26, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **6124518** e o código CRC **F4AC34A7**.

Referência: Processo nº 23117.011729/2025-66

SEI nº 6124518

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU
com dados informados pelo(a) próprio(a) autor(a).

S237 Santos, Leonardo Giamarusti dos, 1998-
2025 A Linguística Saussuriana Aplicada ao Processamento de
Linguagem Natural [recurso eletrônico] / Leonardo
Giamarusti dos Santos. - 2025.

Orientadora: Eliane Silveira.
Dissertação (Mestrado) - Universidade Federal de
Uberlândia, Pós-graduação em Estudos Linguísticos.
Modo de acesso: Internet.
Disponível em: <http://doi.org/10.14393/ufu.di.2025.270>
Inclui bibliografia.

1. Linguística. I. Silveira, Eliane ,1965-, (Orient.).
II. Universidade Federal de Uberlândia. Pós-graduação em
Estudos Linguísticos. III. Título.

CDU: 801

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:

Gizele Cristine Nunes do Couto - CRB6/2091
Nelson Marcos Ferreira - CRB6/3074

Sumário

LISTA DE ABREVIACÕES E SIGLAS	4
LISTA DE FIGURAS	5
LISTA DE TABELAS	6
AGRADECIMENTOS	7
RESUMO.....	8
APRESENTAÇÃO.....	10
PARTE 1 DA NOÇÃO DE RELAÇÃO À SIGNIFICAÇÃO NA LÍNGUA	14
1. Uma rápida discussão sobre o sentido em Saussure	15
1.1. Breve percurso teórico de Saussure: do <i>Ensaio</i> aos <i>Cursos</i>	22
2. O princípio relacional e a noção de relação na teoria saussuriana.....	28
3. O signo linguístico e suas relações internas e externas	41
4. Delimitando “sistema” em Saussure	45
6. A Teoria do Valor: entre semelhanças e dessemelhanças	49
7. Considerações parciais.....	59
PARTE 2 DO SIGNO AOS VETORES DE PALAVRAS	62
1. A Semântica Distribucional sob a égide do saussurianismo.....	63
2. Word Embeddings	75
4. Funcionamento do Word2Vec e a determinação de palavras com base nos valores semelhantes e dessemelhantes	79
5. Considerações finais	86
REFERÊNCIAS.....	89

LISTA DE ABREVIAÇÕES E SIGLAS

- AM:** Aprendizado de Máquina Automático (Machine Learning);
BoW: Bag of Words;
CFS: Cahier Ferdinand de Saussure;
CLG: Curso de Linguística Geral;
EDL: Dupla Essência da Linguagem;
ELG: Escritos de Linguística Geral;
GPT: Generative Pre-trained Transformer;
HD: Hipótese distributiva;
IA: Inteligência Artificial;
Ms. Fr.: Manuscrit Français (Manuscrito Francês);
PCLG: Primeiro Curso de Linguística Geral;
PLN: Processamento de Linguagem Natural;
SBR: Sistemas Baseados em Regras;
SCLG: Segundo Curso de Linguística Geral;
SD: Semântica Distribucional;
TCLG: Terceiro Curso de Linguística Geral;
TF-IDF: Term Frequency-Inverse Document Frequency (Frequência do termo-inverso da Frequência em Documentos);
TdV: Teoria do Valor;
WE: Word Embeddings.

LISTA DE FIGURAS

Figura 1: Da noção de relação à significação na língua.

Figura 2: *Notes pour un livre sur la linguistique générale* (Ms. Fr. 3951/9.f. s/n), excerto 4, transcrito e traduzido por Marques, 2016, p.50.

Figura 3: *Notes pour un livre sur la linguistique générale* (Ms. Fr. 3951/9.f. s/n), excerto 5a, transcrito por Marques, 2016, p.51.

Figura 4: *Notes pour un livre sur la linguistique générale* (Ms. Fr. 3951/9.f. s/n), excerto 5b, transcrito e traduzido por Marques, 2016, p.51.

Figura 5. *Notes pour un livre sur la linguistique générale* (Ms. Fr. 3951/9.f. s/n), excerto 6, transcrito por Marques, 2016, p.53.

Figura 6. *Notes pour un livre sur la linguistique générale* (Ms. Fr. 3951/9.f. s/n), excerto 6, transcrito e traduzido por Marques, 2016, p.53.

Figura 7. *Notes pour un livre sur la linguistique générale* 19f (Ms. Fr. 3951/11.f. 4), excerto 12, transcrito por Marques, 2016, p.63.

Figura 8: *Notes pour un livre sur la linguistique générale* 19f (Ms. Fr. 3951/11.f. 4), excerto 12, transcrito e traduzido por Marques, 2016, p.63.

Figura 9: Reprodução EDL Arch.372/1891 14f, disponível em formato digital

Figura 10: Esquema de valores entre os signos, conforme apresentado no CLG, p. 161.

Figura 11: Exemplo de representação vetorial criada pelo word2vec.

Figura 12: Representação vetorial utilizando Word2Vec e t-NFE.

LISTA DE TABELAS

Tabela 1. Organização do conteúdo do CLG de acordo com os cursos proferidos em Genebra (1907-1911).

Tabela 2. Recorte de algumas ocorrências do termo “relação”, e seus sinônimos, no Curso de Linguística Geral (2012), adaptado de Marques, 2016.

Tabela 3. Ocorrências do termo “sistema” no Curso de Linguística Geral (Saussure, 2012 [1916]), adaptado de Flores (2023, p.101-102).

AGRADECIMENTOS

Ao final desta jornada, percebo que um trabalho acadêmico nunca é uma construção solitária. Ele é tecido por vozes, diálogos e encontros que moldam não apenas o percurso da pesquisa, mas também o pesquisador que se torna. À minha orientadora, professora doutora Eliane Silveira, minha mais profunda gratidão. Com sua dedicação, paciência e imenso conhecimento, guiou-me pelos caminhos intrincados do pensamento saussuriano, ajudando-me a compreender sua complexidade e a reconhecer sua relevância para a linguística brasileira.

Ao Grupo de Pesquisa Ferdinand de Saussure (CNPq/UFU) e aos colegas membros, agradeço pelo ambiente de troca, aprendizado e crescimento. As discussões, os debates e o compartilhamento de saberes foram essenciais para minha formação, ampliando horizontes e fortalecendo minha paixão pelos estudos da linguagem.

Aos meus pais, Maria Teresa Giamarusti e Mario Sergio dos Santos, que sempre me incentivaram a seguir o caminho do conhecimento e nunca duvidaram das minhas escolhas, ofereço este trabalho como um reflexo do amor, do apoio e dos valores que me transmitiram.

Por fim, agradeço a todas as pessoas que, de alguma forma, fizeram parte da minha trajetória acadêmica e pessoal. Cada palavra de encorajamento, cada gesto de apoio e cada momento de partilha foram preciosos e permanecerão comigo para além destas páginas.

A todos, meu mais sincero obrigado.

RESUMO

Esta dissertação propõe um diálogo entre a teoria linguística de Ferdinand de Saussure e o campo do Processamento de Linguagem Natural (PLN), explorando as possíveis interseções teóricas entre a Teoria do Valor (TdV), bem como os pressupostos saussurianos que a circunscrevem, e a Semântica Distribucional (SD), esta última amplamente aplicada no Processamento de Linguagem Natural (PLN) para tarefas semânticas. Na primeira parte, delineamos uma teoria da significação a partir de Saussure que apresenta evidentes elementos articuláveis com a SD, de modo que sustentamos a hipótese de que tanto a Teoria do Valor quanto a Semântica Distribucional possuem um cerne comum, isto é, a ideia de que a língua e seus significados são produtos de um sistema de relações diferenciais. Dito isso, examinamos como a noção de *relação*, e o que dela deriva, a saber, o sistema, o signo, o valor e o significado, pode contribuir para a constituição de uma sólida teoria da significação possivelmente útil ao PLN para tarefas de modelagem semântica, análise de similaridades e cálculo de probabilidade de ocorrência de palavra. A partir da análise das relações que compõem o sistema linguístico, investigamos como os valores linguísticos emergem e geram significação por meio de diferenças de *similia* e *dissimilia*. Na segunda parte, examinamos as conexões epistemológicas entre a linguística saussuriana e a SD, fundamentada nas ideias de Harris e Firth, além de análises à luz da linguística saussuriana de embeddings de palavras gerados pelo modelo Word2Vec, buscando por evidências de articulações possíveis entre Saussure e o PLN. Nesse sentido, nossos resultados indicam que a construção de significados em modelos de linguagem baseados na SD parece espelhar os princípios diferenciais e relacionais que Saussure desenvolveu para o funcionamento da língua no início do século XX. Adicionalmente, discutimos uma aplicação da SD em um projeto da Universidade de Genebra, destacando a utilidade dos conceitos saussurianos para a modelagem semântica de textos históricos. Em síntese, os resultados indicam que a linguística saussuriana oferece um aporte teórico significativo para o PLN, possibilitando uma compreensão mais aprofundada dos processos de significação em modelos vetoriais de linguagem. Tendo em vista isso, argumentamos a possível validade de nossa principal hipótese investigativa, a saber, a ideia de que princípios saussurianos permanecem relevantes na linguística contemporânea e, ao mesmo tempo, proporcionam uma base sólida para o avanço de novas tecnologias de linguagem, reafirmando a atualidade da linguística saussuriana no contexto do Processamento de Linguagem Natural no século XXI.

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

ABSTRACT

This dissertation proposes a dialogue between Ferdinand de Saussure's linguistic theory and the field of Natural Language Processing (NLP), exploring potential theoretical intersections between the Theory of Value (ToV), along with the Saussurean assumptions that frame it, and Distributional Semantics (DS), the latter widely applied in NLP for semantic tasks. In the first part, we outline a theory of signification based on Saussure that presents evident articulable elements with DS, thereby supporting the hypothesis that both the Theory of Value and Distributional Semantics share a common core—namely, the idea that language and meaning are products of a system of differential relations. Accordingly, we examine how the notion of relation—and the concepts derived from it, such as system, sign, value, and meaning—may contribute to the construction of a robust theory of signification potentially useful for semantic modeling, similarity analysis, and probability calculations in NLP tasks. Based on the analysis of the relations that comprise the linguistic system, we investigate how linguistic values emerge and generate meaning through differences of similia and dissimilia. In the second part, we examine the epistemological connections between Saussurean linguistics and DS, grounded in the ideas of Harris and Firth, and analyze word embeddings generated by the Word2Vec model from a Saussurean perspective, seeking evidence of possible articulations between Saussure and NLP. In this regard, our findings indicate that the construction of meaning in DS-based language models appears to mirror the differential and relational principles that Saussure developed for the functioning of language in the early 20th century. Additionally, we discuss a DS application in a project at the University of Geneva, highlighting the usefulness of Saussurean concepts for the semantic modeling of historical texts. In summary, the results indicate that Saussurean linguistics offers significant theoretical support for NLP, enabling a deeper understanding of the processes of signification in vector-based language models. In light of this, we argue for the possible validity of our main investigative hypothesis—namely, the idea that Saussurean principles remain relevant in contemporary linguistics while also providing a solid foundation for the advancement of new language technologies, reaffirming the contemporary relevance of Saussurean linguistics in the context of 21st-century Natural Language Processing.

Key-words: Saussure; Natural Language Processing; Theory of Value; AI.

APRESENTAÇÃO

A obra do linguista suíço Ferdinand de Saussure (1857-1913) continua a ocupar um lugar central nos estudos linguísticos e em outras áreas das ciências humanas. Reconhecido como o fundador da Linguística Moderna e autor póstumo do influente *Curso de Linguística Geral* (CLG), Saussure pôde estabelecer bases teóricas que, ainda hoje, alimentam pesquisas na busca por uma maior compreensão sobre a natureza e o funcionamento da(s) língua(s) humana(s).

No século XXI, o alcance de Saussure adquire novas dimensões, especialmente no campo do Processamento de Linguagem Natural (PLN). Isso porque princípios da linguística saussuriana, a exemplo da Teoria do Valor (TdV), parecem oferecer bases teóricas sólidas para se compreender o modo como determinados modelos de linguagem, notadamente aqueles baseados na Semântica Distribucional, conseguem compreender os sentidos da língua, transformando signos linguísticos em vetores de palavras capazes de representar significados possivelmente aplicáveis a diferentes tarefas de PLN, como a busca e a indexação semânticas, bem como análises de similaridades.

Portanto, nesta pesquisa, nosso principal objetivo é demonstrar a possível pertinência da linguística saussuriana para o Processamento de Linguagem Natural. Nossa hipótese investigativa, assim, é que a Teoria do Valor de Saussure e os pressupostos que a cercam podem constituir uma teoria da significação que encontra articulações com as bases linguísticas da Semântica Distribucional, amplamente aplicada hoje em dia para a modelagem semântica de embeddings de palavras.

A Semântica Distribucional (SD), por sua vez, é uma teoria da significação que admite que as relações na língua, especialmente entre os signos, são fundamentais para a delimitação de um valor semântico para determinada entidade linguística. Tal teoria, desenvolvida a partir dos ideais linguísticos distribucionais de John R. Firth (1890-1960) e Zelling Z. Harris (1909-1992), no Processamento de Linguagem Natural¹, é comumente associada a

¹ Segundo Freitas (2023), considera-se o Processamento de Linguagem Natural (PLN) como o lado prático da Linguística Computacional. Em outras palavras, o PLN se dedica, em linhas gerais, a entender e a problematizar a forma como computadores entendem a língua humana. Além disso, o PLN também se dedica ao desenvolvimento de técnicas e de modelos de

modelos de linguagem baseados em vetores de palavras, como o Word2Vec (Mikolov et. al, 2013), o Doc2Vec e o próprio modelo GPT, da OpenAi. Nesse sentido, assim como Ferdinand de Saussure o parece fazer ao longo de seu percurso teórico, notadamente por meio da Teoria do Valor (TdV), a SD também coloca em evidência o entendimento de que a língua é um sistema de relações (Saussure, 2012 [1916]) capaz de produzir sentidos por meio da rede de relações de semelhança e dessemelhança entre os signos.

Nessa perspectiva, propomos que a Teoria do Valor, quando lida sob o ponto de vista de uma teoria semântica, pode ser utilizada como base teórica para o desenvolvimento de modelos de linguagem baseados na Semântica Distribucional. Isso porque pretendemos demonstrar que há um aparente cerne comum entre a TdV e a SV, que reside na ideia de que a(s) língua(s) constituem sistemas relacionais que geram valores e engendram significados por meio da relação entre os signos.

Em níveis formais, nossa pesquisa divide-se em duas partes. A primeira é dedicada a esclarecer de que modo a TdV pode ser interpretada, também, como uma teoria da significação. Para esse entendimento, propomos uma semântica saussuriana baseada nas relações internas e externas do signo. Essas relações, por sua vez, dão origem ao sistema linguístico - considerado, por nós, como um sistema de relações. O estudo do sistema, então, possibilita uma maior compreensão sobre a natureza da língua humana, a qual é constituída de valores diferenciais entre os signos. Os valores, por fim, representados pelas relações externas do signo, representam o principal mecanismo que leva à significação na(s) língua(s), haja vista que os significados, para nós, só podem ser determinados por meio da análise de dois tipos de relações: os valores semelhantes e os valores dessemelhantes. Na figura a seguir, ilustramos o mecanismo pelo qual se chega à significação que nos propusemos a investigar.

linguagem que auxiliem a máquina a processar a língua e fala em diferentes níveis linguísticos. As principais aplicações do PLN atual concentram-se em: tradução automática; levantamento de dados linguísticos empíricos; conversa com Chatbots; produção de texto, imagem e som; catalogação de texto; análise de similaridade etc.

Figura 1. Da noção de relação à significação na língua.



Fonte: autoral

Já na Parte 2 desta pesquisa, buscamos investigar se esta leitura da Teoria do Valor como uma teoria da significação pode ser útil para o PLN propriamente. Nossa hipótese, assim, é de que a TdV e os pressupostos que a fundamentam podem ser uma alternativa teórica para a compreensão da Semântica Vetorial² e podem, portanto, possivelmente ser aplicados para a modelagem semântica de embeddings de palavras. Dito isso, nossa investigação buscou por evidências capazes de sugerir que a TdV e a SV possuem pontos em comum tanto em níveis teóricos, com a análise de filiações epistemológicas entre Firth, Harris e Saussure; quanto em níveis práticos, por meio de uma análise saussurianista de embeddings de palavras (Gastaldi, 2020) gerados pelo modelo Word2Vec.

Em suma, o presente estudo visa estabelecer um diálogo entre a teoria linguística de Saussure e os avanços tecnológicos na área de PLN, especialmente no que diz respeito ao desenvolvimento de modelos e de técnicas baseados na Semântica Distribucional. Ao final, esperamos demonstrar como algumas noções saussurianas clássicas, como o valor, o sistema e a relação podem oferecer subsídios teóricos para interpretar e expandir as possibilidades

² Ao longo desta pesquisa, utilizaremos a Semântica Distribucional (SD) e Semântica Vetorial (SV) como sinônimas, haja vista que, na literatura especializada em PLN, ambos os termos são equivalentes.

práticas da semântica computacional. Nesse sentido, essa reflexão sobre tal interação interdisciplinar pode evidenciar a atualidade do saussurianismo, como também lançar luzes sobre sua contribuição para a compreensão e o desenvolvimento de Inteligências Artificiais na era digital.

PARTE 1

DA NOÇÃO DE RELAÇÃO À SIGNIFICAÇÃO NA LÍNGUA

É possível uma teoria saussuriana da significação?

(Normand, 1992, p.50)

1. Uma rápida discussão sobre o sentido em Saussure

O estudo da significação a partir de Saussure, durante muito tempo, foi de grande interesse dos linguistas que se dedicam a pesquisas em torno do arcabouço teórico do mestre de Genebra. Nesta pesquisa, partimos da noção de relação e da TdV como principais bases epistemológicas para uma possível semântica saussuriana. Contudo, não podemos negar a existência de outras teorias sobre o sentido em Saussure.

É verdade, assim, que não poderíamos iniciar esta seção sobre possíveis teorias acerca do sentido em Saussure se não fizéssemos um breve retorno a uma teoria semântica anterior ao próprio mestre Genebra, a saber, a semântica histórica de Michel Bréal; não porque a semântica de Saussure seja igual à semântica de Bréal, mas porque, durante determinado período, associou-se uma possível semântica saussuriana como produto, ou efeito, de uma semântica brealiana, conforme discutido por Rudolf Engler, em artigo intitulado *Rôle et place d'une sémantique dans une linguistique saussurienneno*, publicado nos Cahier Ferdinand de Saussure³, n.28, 1973.

Dito isso, muito se sabe sobre os (des)encontros entre a linguística geral de Michel Bréal e a linguística geral de Saussure. Aarsleff (1981), Cruz (2009) e Dall'ortivino (2017) demonstraram, com base em evidências históricas e documentais, que alguns pressupostos de Bréal faziam parte do arcabouço teórico de Saussure, inclusive porque Saussure fora aluno de Bréal no *College de France*, em Paris, como mostra Cruz (2009, p.111).

Dall'ortivino (2017) aponta também outros vários episódios em que Bréal demonstrou afinidade com Saussure, como o fato de Michel convidar seu então ex-aluno para ser seu substituto na *École de Hautes Études* entre 1881 e 1891; assim como “as correspondências entre Henri de Saussure – pai de Saussure – e Bréal, para quem escreveu a fim de comunicar a presença do filho em Paris” (Dall'ortivino, 2017, p.1967).

Além disso, Marc Décimo, em artigo intitulado *De Quelques Candidatures*

³ Neste trabalho, optamos também por usar a notação CFS para nos referirmos aos *Cahier Ferdinand de Saussure* (Cadernos de Ferdinand de Saussure), coletânea de trabalhos acadêmicos que reúne artigos, resenhas e ensaios de respeitados pesquisadores da fortuna saussuriana. Os CFS são organizados e publicados pelo Cercle Ferdinand de Saussure desde 1941 até os dias de hoje.

Et Affinités Électives De 1904 À 1908, À Travers Un Fragment De Correspondance: Le Fonds Michel Bréal, publicado no CFS, n.47. 1993, resgata um texto de Antonie Meillet datado de 1881, intitulado “Nécrologie de Michel Bréal”, publicado na revista *Annuaire* entre 1916-1917, em que Meillet é enfático ao dizer que, se Saussure pôde ensinar em Paris, foi porque Bréal lhe fizera o convite para que o mestre se mudara de Genebra:

Se Ferdinand de Saussure ensinou entre nós, se formou entre nós tantos alunos brilhantes, foi porque Michel Bréal teve a ideia de trazê-lo de Genebra para Paris, de instalá-lo em seu próprio lugar e em sua própria cadeira, e de confiar a ele mesmo a tarefa de reconstruir a linguística, com carta branca para destruir antes de tudo (Meillet, 1916-1917 [1881], p. 41-42, citado por Décimo, 1993, p.38, tradução nossa)⁴.

Ademais, Dall’ortivino (2017) e Aarsleff (1981) demonstraram que termos caros ao saussurianismo, como *valor*, *língua*, *sincronia* e *diacronia*, já faziam parte do arcabouço teórico de Bréal, notadamente no capítulo intitulado “Leis intelectuais da linguagem”, do clássico *Ensaio de Semântica* (1992 [1891]). Contudo, Correia (2020) ressalta que Bréal e Saussure diferem na forma e na importância com que os referidos conceitos são abordados em suas teorias⁵.

Nesse sentido, a partir de possíveis filiações teóricas entre a linguística geral de Saussure e de Bréal, os editores do Curso, Bally e Sechehaye, propõem que uma semântica saussuriana estaria, portanto, atrelada à diacronia, conforme indica Engler (1973), afirmando que Bally e Sechehaye chegaram a propor que a semântica, em Saussure, se referiria ao “[...] estudo das mudanças de significação” ao longo do tempo (Engler, 1973, p. 36, tradução nossa)⁶.

A proposta dos editores do CLG à que se refere Engler se assemelha, como se sabe, à semântica histórica de Bréal (1992 [1891]) - e muito se pode hipotetizar sobre a influência deste importante linguista francês para com a formação dos editores do CLG, sobretudo com Bally, tendo em vista que ambos

⁴ No original: “Si Ferdinand de Saussure a enseigné chez nous, s'il a formé chez nous tant de brillants élèves, c'est que Michel Bréal a inventé de le faire venir de Genève à Paris, de l'installer à sa propre place et dans son propre fauteuil, et de lui confier lui-même le soin de reconstruire la linguistique, avec carte blanche pour détruire d'abord.”

⁵ Para um aprofundamento nas diferenças e aproximações teóricas entre Bréal e Saussure, sugerimos conferir na íntegra a dissertação *O Lugar do Sentido em Saussure*, de Tibério T. T. S. Correia (2020).

⁶ No original: “La définition que, dans leur note, Bally et Sechehaye donnent de la sémantique est conventionnelle (Wunderli le démontre) : pour eux, « la sémantique [...] étudie les changements de signification »”.

eram contemporâneos, partilhavam interesses teóricos – como a semântica e estilística - possuíam conhecidos em comum, como o próprio Saussure etc.

Em seu *Ensaio sobre semântica* (1992 [1891]), Bréal deixa claro o lugar do estudo dos sentidos no domínio da historicidade da língua. Em contrapartida, aplicar essa lógica da significação como produto histórico dos estados de língua à teoria saussuriana parece apresentar algumas lacunas epistemológicas, haja vista que o aparente lugar do sentido, em Saussure, é o que supõe divergências significativas entre a linguística saussuriana e a brealina.

Correia (2020), nesse sentido, explica que, mesmo que Bréal utilize de uma noção de sistema que sutilmente lembre à saussuriana, ainda faltariam argumentos que demonstrassem maior filiação teórica em matéria de semântica entre o mestre francês e o mestre genebrino:

A colocação de Bréal [sobre a língua como um sistema] nos lembra a noção de Valor linguístico de Saussure, no entanto, e **aí está o grande abismo entre os dois, Bréal propõe um estudo histórico do Sentido e isto está diretamente ligado ao que ele considera o principal fator da mudança linguística: A vontade do homem de significar.** Assim, o homem se inscreve na história através da língua e o estudo semântico, por óbvio diacrônico, visa o evoluir dos Sentidos marcados na história. Para Saussure, uma vez que haja uma mudança de Sentido, nós não temos como recuperar o que o precedeu, pois não temos acesso ao Sistema de um estado de língua que se modificou (Correia, 2020, p. 11, grifo nosso).

Engler (1973), então, nega a hipótese de Bally e Sechehaye de que o lugar da semântica em Saussure estaria associado à diacronia, tal qual Bréal o faz, trazendo a discussão sobre o sentido para o plano da sincronia: “Concluo, portanto, que a definição saussuriana de *semântica* não apenas não exclui a sincronia, como também certamente lhe confere um lugar preponderante” (Engler, 1973, p.39, tradução nossa). Isso porque a hipótese de Engler para a semântica saussuriana advém da oposição que Saussure faz entre o estudo dos sentidos e as formas da língua no CLG: “Em sua definição, Saussure opõe a ciência dos sentidos das palavras à das formas, a semântica e a morfologia” (*ibidem*, p.37).

Tullio de Mauro, em artigo publicado no CFS, n.45,1991, intitulado *Ancora Saussure e la Semantica*, também propõe que os estudos saussurianos sobre o sentido promovem um retorno à ideia de que a língua é forma e não substância.

De igual forma, o autor sugere uma interpretação sobre o sentido que recai sobre a natureza relacional do signo:

A teoria semântica de Saussure tem um centro: a distinção entre *forma* e *substância*. Se olharmos para os signos de uma língua, para os já existentes e para aqueles regularmente previsíveis, os vemos constituídos por um lado externo, feito de sons ou grafias, e por um lado interno, os sentidos [...]. Uns e outros fazem parte de um conjunto de entidades infinitas concretas, profundamente diferentes entre si. Se nos compreendemos, se podemos de alguma forma superar essa diversidade e entender o que outros dizem ou o que nós mesmos já dissemos em outro momento, e se conseguimos nos fazer entender por outros, isso só é possível sob a condição de admitir que a infinita e concreta diversidade de sentidos e das emissões fonéticas, constitutivas de atos de fala individuais e sempre diversos, se organiza e ordena em um sistema de equivalências, em uma série de esquemas ou formas que unificam ou conectam as diferentes *paroles* (De Mauro, 1991, p.105, tradução nossa)⁷.

Em contrapartida, embora haja diferentes teorias que apontem para caminhos distintos para se chegar a uma teoria da significação em Saussure, é preciso enfatizar que partimos do pressuposto de que a linguística saussuriana é, em primeiro nível, uma reflexão sobre a língua e as línguas, sendo o sentido um efeito de um sistema de relações arbitrárias. Ou seja, até o momento, são escassas as evidências documentais que sustentem a hipótese de que a significação era o principal ponto de partida teórica de Saussure, ainda que haja trabalhos em direcionamentos opostos, como sugere a análise do manuscrito saussuriano *Notes Item*, apresentado por Sortica (2021).

Por outro lado, filiando-nos a Engler (1973) e a De Mauro (1991), sustentamos a hipótese de que, embora a linguística saussuriana seja, em maior grau, uma teoria geral sobre a língua e as línguas, ela não exclui a possibilidade de, a partir do CLG e de materiais autógrafos do mestre genebrino, refletir como se dá a produção do sentido na(s) língua(s).

⁷ No original: "La teoria semantica di Saussure ha un centro: la distinzione di forma e sostanza. Se guardiamo ai segni di una lingua, a quelli già esistenti e a quelli regolarmente prevedibili, li vediamo costituiti da un versante esterno, fatto di suoni o grafie, e da un versante interno, i sensi. Le singole emissioni foniche e i singoli concreti sensi li incontriamo quando produciamo actes de parole. Le une e gli altri fanno parte di un insieme di infinite entità concrete, profondamente diverse ciascuna da ogni altra. Se ci capiamo, se possiamo in qualche modo scavalcare questa diversità e intendere quello che altri dicono o che noi stessi altra volta abbiamo detto, e se possiamo farci intendere da altri, ciò è possibile soltanto a condizione di ammettere che la infinita concreta diversità dei sensi e delle emissioni foniche, costitutive di individuali e sempre diversi atti di parole, si disciplini e ordini in un sistema di equivalenze, in una serie di schemi o forme unificanti o congiungenti le diverse paroles."

Dos referidos autores, emprestamos os termos “relações internas” e “relações externas” do signo, ainda que nossa leitura sobre esses conceitos seja diferente - já que os autores (Engler, 1973; De Mauro, 1991) privilegiam, para a constituição da significação, as relações internas, e nós, as externas. Além disso, nos apoiamos também na ideia trazida por eles de que as relações entre os signos constituem valores, e estes, por sua vez, atravessam e compõem os sentidos pelo jogo de diferenças da língua.

Não obstante, a nossa proposta para o lugar do sentido em Saussure reside na mesma base epistemológica da Semântica Distribucional, a saber, o entendimento de que a língua é um sistema de signos e, por consequência, de relações (cf. Saussure, 2012 [1916]). Dessa afirmação, extraímos o que consideramos o princípio mais importante para o entendimento da complexidade de uma possível semântica saussuriana, a saber, a noção de *relação* (Marques, 2016).

Em linhas gerais, partimos do pressuposto de que a significação está intimamente relacionada à noção de *relação* (*ibidem*) que, ora e sempre, aparece na obra de Saussure, isto é, a ideia de que, na língua, há um aparente *princípio relacional* que preconiza que os sentidos residem, primeiro, no jogo de relações estabelecido entre as entidades⁸ do sistema. Nesse sentido, este “jogo de relações” nos revela um segundo princípio, também igualmente, para nós, fundamental para a construção da significação na(s) língua(s): os valores. Os valores, por sua vez, possibilitam que os signos possam contrair relações externas (entre si) de semelhança e dessemelhança, formando, por consequência, valores positivos e valores negativos.

Esta nossa proposta para a significação, baseada na noção de *relação* e do que dela deriva, a saber, o signo, o sistema e o valor, conforme já demonstramos na Apresentação, possibilita, também, retomarmos uma discussão já iniciada por Bréal sobre a linguagem ser inerente e exclusiva à inteligência humana: “A linguagem é um ato do homem: não tem realidade fora

⁸ Nesta dissertação, não nos deteremos na concepção de “entidade” na linguística saussuriana do ponto de vista terminológico. Sendo assim, alhures, vamos utilizar o termo como um possível substituto de “signo”. Sabemos, contudo, que ambos os termos não são estritamente equivalentes. Para aprofundar nessa discussão, sugerimos a leitura de *Les Problème de la Définition des entités linguistiques chez Ferdinand de Saussure*, publicado no CFS, n.62, 2009, por Estanislao Sofia.

da inteligência humana” (Bréal, 1991, citado por Aarsleff, 1981, p.25, tradução nossa)⁹.

A afirmação supracitada de Bréal, talvez, atendesse aos anseios da linguística francesa daquele período, bastante marcada pela oposição - ora rejeição - aos ideais organicistas e darwinistas da gramática comparada e histórica. Nos dias de hoje, no entanto, o constante avanço de modelos computacionais de linguagem, como os inspirados na Semântica Distribucional, notadamente o Word2Vec (Mikolov et. al, 2013), o BERT e o GPT, possibilita repensarmos se a afirmação de Bréal ainda é válida para o nosso tempo¹⁰.

Os referidos modelos de linguagem, como veremos nesta dissertação - com ênfase para o Word2Vec - parecem emular a produção de sentidos de forma bem próxima ao mecanismo da significação na(s) língua(s) que enxergamos em Saussure, a saber, o entendimento de que o sistema linguístico é, também, um sistema de relações que gera valores e sentidos por meio do jogo de diferenças entre os signos.

O fato é que, nos últimos anos, as inteligências artificiais transformaram radicalmente o modo como interagimos com o mundo e processamos informações. O modelo GPT (*Generative Pre-Trained Model*), por exemplo, têm demonstrado potencial para colocar em discussão, na Linguística, possíveis novas interpretações para o modo como a(s) língua(s) funciona(m); bem como incide um papel importante para a atualização de conceitos que, antes, pareciam estritos às ciências da linguagem, como a noção de *valor linguístico* (Saussure, 2012 [1916]) e o entendimento sobre a construção da significação na língua.

Nesse sentido, enquanto as máquinas têm sido desenvolvidas para serem cada vez mais capazes de gerar textos complexos e emular diálogos que se assemelham aos produzidos por humanos, surge uma questão fundamental: em

⁹ No original: “Il lui manque pour cela une condition capitale : c'est que l'objet dont elle traite n'existe pas dans la nature. Le langage est un acte de l'homme : il n'a pas de réalité en dehors de l'intelligence humaine”.

¹⁰ Para deixar claro: não estamos afirmando que Inteligências Artificiais produzem língua ou a entendem tal como humanos. O que estamos dizendo é que, se admitirmos que a linguagem é um conjunto de signos interconectados, cujo lado social reside na coletividade, não seria insano pensar que, cada vez mais, podemos estar próximos de uma realidade na qual as IAs poderão, de maneira **semelhante** à inteligência humana, produzir linguagem. Por outro lado, ainda é cedo para dizermos que um modelo de linguagem será capaz de produzir “língua” no sentido de *la langue* (Saussure, 2012 [1916]), já que o grande desafio do PLN atual é emular relações linguísticas complexas que, a princípio, continuam estritamente humanas, como a ambiguidade e a polissemia, a ironia, o humor, a sensibilidade poética etc.

que medida inteligências artificiais compreendem os sentidos na língua? Ou ainda: será que os sistemas modernos de IA já entendem os sentidos tal como humanos o parecem fazer?

Para responder a esses questionamentos, precisamos nos colocar diante de uma questão epistemológica que envolve, em primeiro lugar, refletirmos sobre como, afinal, a língua funciona. Essa pergunta, assim, indubitavelmente, levam-nos a outra, a saber: qual a natureza da língua humana?

Ao longo da história das ideias linguísticas, vários foram os momentos em que pesquisadores e filósofos da linguagem se debruçaram sobre essas questões. É verdade, porém, que encontramos, nas teorias de Ferdinand de Saussure (1857-1913), uma produção teórica efervescente que pode nos ajudar a compreender melhor algumas leis de funcionamento universal para a língua e, também, para as línguas, as quais podem nos fornecer possíveis interpretações para compreendermos o que constitui e o que faz da língua humana um fato diferenciado entre outros objetos das ciências humanas.

Nessa perspectiva, argumentamos que algumas teorias do mestre genebrino têm apresentado indícios de que os estudos linguísticos dos séculos XIX e XX podem auxiliar o linguista na busca pela natureza da língua e, ao mesmo tempo, contribuir para os cientistas da computação e linguistas computacionais com aportes teóricos para o desenvolvimento de novas técnicas de Processamento de Linguagem Natural (PLN) no século XXI. Isso se deve ao fato de que, ao analisarmos o funcionamento de alguns modelos de linguagem, como os *word embeddings*, temos encontrado algumas possíveis semelhanças - e também sutis diferenças - entre o modo como Saussure parece conceber a produção de significação e a modelagem de significados por modelos vetoriais de linguagem.

De maneira ampla, a forma com que os referidos modelos de linguagem parecem conceber uma ideia de “significação” reside, em primeiro lugar, no comportamento dos signos linguísticos - transformados em vetores numéricos - os quais estão em associação uns com os outros por semelhança e dessemelhança. Em nível teórico, para a modelagem do sentido de vetores de palavras, tradicionalmente, tem-se utilizado, no PLN, a Hipótese Distributiva de Zelling Harris (1955), a qual prediz que a significação nas línguas está, na verdade, relacionada às ocorrências de outros signos. Ou seja, quanto mais

similar o contexto de uso de um signo em relação ao outro, maior a probabilidade de ocorrência de ambos numa mesma oração. Em contrapartida, há evidências de que a hipótese distributiva, anteriormente discutida por John R. Firth (1934; 1950), já fora antecipada por Saussure de uma forma muito particular, haja vista que, no próprio CLG, o mestre genebrino possibilita pensarmos, tal qual propõe a linguística distribucional, que a significação não está contida na palavra em si, mas no “concurso do que existe fora dela” (Saussure, 2012 [1916], p.162).

Em síntese, na próxima seção, então, propomos uma breve discussão sobre o percurso teórico de Saussure, a fim de analisarmos o quanto a noção de relação e de sistema eram constantes inquietações do professor de Genebra. Na sequência, discutiremos com mais afinco o que, afinal, *relação*, nos estudos saussurianos, significa; e como essa noção, *a posteriori*, desdobra-se em outros conceitos saussurianos, os quais, quando juntos e organizados hierarquicamente (rever Figura 1, p.10), constroem o percurso pelo qual se chega à significação na(s) língua(s).

1.1. Breve o percurso teórico de Saussure: do *Ensaio* aos *Cursos*

Pontuar que Saussure foi apenas um linguista dos séculos XIX e XX não parece corresponder aos efeitos duradouros do seu pensamento. Ao mesmo tempo, não se deve também limitar o alcance de sua obra aos mitos fundadores que a cercam. Começamos, então, esta seção¹¹ deixando claro ao(à) nosso(a) leitor(a) que a história de Saussure não se confunde com a história da linguística. Já existia linguística antes do linguista genebrino - na verdade, bem antes.

Assim, se não é Saussure o “criador” da Linguística, mas o precursor da Linguística Moderna, o que nos caberia relatar sobre seu percurso teórico? Do ponto de vista histórico, de fato, pouco temos a contribuir. Quijano (2008) e Joseph (2023 [2012]) já têm extensos trabalhos dedicados a contar sobre a vida de Ferdinand de Saussure, os quais recomendamos aos entusiastas de bibliografias de personalidades históricas. Porém, há um aspecto particular em toda a vida do mestre genebrino que merece nossa atenção: a sua curiosidade

¹¹ Ao longo desta dissertação, optamos por utilizar “seção” em vez de “capítulo” para nos referirmos aos níveis hierárquicos deste texto, tendo em vista que, embora o trabalho necessite das divisões didáticas que aqui propomos, estas, por sua vez, por uma convenção de número de páginas, não poderiam ser chamadas de capítulos.

recorrente sobre o funcionamento das línguas e da língua.

Em carta endereçada a Adolphe Pictet em 1872, o jovem Saussure pede, aos impressionantes 15 anos de idade, que Pictet leia o seu *Ensaio para reduzir as palavras do grego, do latim e do alemão a um pequeno número de raízes* (1872)¹², como mostra Joseph (2012, p.148). Nesse trabalho, Saussure afirma ter descoberto algo extremamente curioso, isto é, a evidência de um “sistema” nas línguas:

Eu não teria escrito nada se, pelo hábito, eu não tivesse constatado como evidente um sistema que me intriga desde o ano passado; eu sempre tenho o hábito de fazer sistemas antes de estudar as coisas nos mínimos detalhes (Saussure, 1872, citado e traduzido por Coelho, 2018, p. 397).

Tratava-se, naquele momento, de sua descoberta sobre a nasal sonante *v*, a partir de um texto de Heródoto, já se valendo de ideais comparatistas para criar suas próprias teorias sobre as transformações das línguas, baseando-se, desde esse momento, numa possível noção de sistema para sustentar suas hipóteses.

Nesse sentido, o contínuo interesse de Saussure pelo funcionamento e mudança das línguas fez com que, em 1876, o jovem suíço fosse aceito na Sociedade de Linguística de Paris aos 20 anos (Joseph, 2023 [2012]). Nesse período de 1876 a 1880, Saussure frequentou alguns cursos de linguística histórica e de gramática comparada na Universidade de Leipzig, sendo aluno de influentes neogramáticos, como August Leskien, Karl Brugmann, Hermann Osthoff, Berthold Delbrück e Hermann Paull (cf. Altman, 2024).

Em dezembro de 1878, Saussure defende a sua dissertação: *Mémoire sur le système primitif des voyelles dans les langues indo-européennes*, um estudo de caso sobre o sistema primitivo das vogais nas línguas indo-europeias, estabelecendo-se como um nome importante, na Europa, para o fortalecimento da gramática comparada.

Ao longo de sua dissertação, Saussure se dedica a resgatar evidências sobre diferenças fonéticas e fonológicas do indo-europeu antigo e do indo-europeu de seu tempo. Isso porque, naquele período, havia um “problema

¹² Quando nos referirmos a esse trabalho, utilizaremos apenas a notação *Ensaio*.

descritivo” (cf. Altman, 2024) relacionado às vogais indo-europeias. O mestre genebrino, então, retoma o linguista Franz Bopp (1791-1867), quem Saussure argumenta ser o responsável por ter demonstrado a existência de 3 vogais no antigo indo-europeu, a saber: a; e; o.

Bopp e aqueles que seguiram imediatamente o ilustre autor da *Grammaire Comparée* se limitaram a constatar que, em relação às três vogais a, e, o das línguas europeias, o ariano mostrava uniformemente a. O a, e, o passaram, a partir de então, a ser considerados como enfraquecimentos próprios aos idiomas do Ocidente e relativamente recentes do a único indo-europeu (Saussure, 1879, p.2, tradução nossa)¹³

Na sequência de suas discussões no *Memóire*, Saussure faz menção, também, a diferentes hipóteses sobre o surgimento de determinadas vogais indo-europeias, como o “sistema de M. Curtius”, retomado por M. Fick; o ‘sistema de Schleicher’, que ele menciona por quatro vezes nas páginas 5 e 6, que seria retomado por Amelung, e o de Karl Brugmann (cf. nota 2)” (Sofia, 2017, p.52). Nesse momento, o mestre genebrino parece retomar a algumas discussões que o próprio já havia previsto no *Ensaio*, a saber, a sua “descoberta” sobre a existência de uma noção de “sistema” nas línguas.

Sofia (2017), em trabalho intitulado *A noção de Sistema no Memóire*, faz um levantamento da palavra “sistema” na referida obra. O autor pontua que foram encontradas pouco mais de 20 ocorrências do termo, contudo “nenhuma delas revela uma conceptualização, uma problematização, ou uma reflexão sobre o alcance teórico daquela noção [sistema]” (*ibidem*, p.51). Esse esclarecimento teórico, porém, aparece de forma um pouco mais amadurecida somente mais tarde, nas notas para os cursos de Genebra, como veremos em seções adiante dedicadas exclusivamente à *noção de sistema* em Saussure.

Nos anos posteriores à defesa de sua dissertação, Saussure começa a ser lido e discutido em cursos de Linguística na Europa durante o final do século XIX. A publicação do *Memóire* e a aproximação de Saussure com expoentes da Gramática Comparada fizeram com que, em 1891, o mestre genebrino fosse

¹³ No original: “Bopp et ceux qui suivirent immédiatement l'illustre auteur de la *Grammaire Comparée* se bornèrent à constater qu'en regard des trois voyelles a e o des langues européennes, l'arien montrait uniformément a. L'e et l'o passèrent dès lors pour des affaiblissements propres aux idiomes de l'Occident et relativement récents de l'a unique indo-européen.”

admitido, na Universidade de Genebra, para a cadeira de história e comparação das línguas indo-europeias, onde ministrou as três clássicas conferências de Linguística Geral entre 1907 e 1911.

Essas conferências, como se sabe, serviram de base para a sua célebre obra póstuma, o *Curso de Linguística Geral* (1916) - editado por dois colegas de Saussure - os também linguistas Charles Bally e Albert Sechehaye. Para a composição do CLG, foram utilizadas diferentes fontes: manuscritos autógrafos, isto é, de autoria do próprio Saussure, com os quais Bally e Sechehaye tiveram contato, como as notas preparatórias aos cursos; e os cadernos de alguns estudantes, com destaque para as anotações de A. Riedlinger.

Uma questão interessante era a proximidade e afinidade que Saussure tinha com Bally e Sechehaye, principalmente com o primeiro. O próprio Saussure chegou a reforçar, em carta endereçada a Charles Bally, em 1905, o nome de Bally, a quem Saussure se referia como “discípulo”, para suceder à sua cadeira na Universidade de Genebra. A evidência documental desse momento se encontra no conjunto de manuscritos intitulado “Manuscrit à l'intention de Charles Bally, son disciple et successeur à l'Université de Genève” (1905-1908), arquivado sob a inscrição Ms. Fr. 5134/208f na biblioteca pública de Genebra, disponível também em formato digital.

Em vista disso, muito diferente do que se afirma no prefácio à tradução do CLG de Marcos Bagno (2021), em “Não temos, portanto, nenhuma garantia de que o texto [do CLG] corresponde, de fato, às lições do mestre genebrino” (Bagno, 2021, p.19), temos, ao contrário, várias evidências de que Bally e Sechehaye conheciam de perto a produção teórica de Saussure, tanto como alunos quanto como amigos. Isto é, a crítica do prefácio pode levar o leitor ao erro de desconsiderar os fatos históricos que entrelaçam as experiências que Bally e Sechehaye tiveram com Saussure em vida, dando a entender que as escolhas editoriais dos autores do CLG não poderiam ser cientificamente confiáveis e, por isso, o texto do CLG não refletiria às lições de Saussure - o que discordamos em gênero, número e grau, diga-se.

Nesse sentido, Flores (2023) retoma os cursos de Genebra, detalhando os conteúdos abordados em cada um deles¹⁴, os quais, depois, foram

¹⁴ Embora seja uma informação importante, não iremos explorar com detalhes, neste momento, o que foi abordado em cada um dos cursos de linguística geral proferidos por Saussure. Para

distribuídos no CLG da seguinte forma:

Tabela 1. Organização do conteúdo do CLG de acordo com os cursos proferidos em Genebra (1907-1911):

CURSO 1 (1907)	Base da <i>terceira</i> parte do CLG, dos Apêndices A e B e do Capítulo III da quinta parte.
CURSO 2 (1908-1909)	Fonte complementar de todo o CLG, além de base principal do Capítulo V da Introdução; dos Capítulos III, VI e VIII da Segunda parte; do Capítulo VIII da terceira parte e dos capítulos I e II da quinta parte.
CURSO 3 (1910-1911)	Fonte da Introdução do CLG (menos o Capítulo V e o Apêndice Princípios de Fonologia); do restante a Primeira parte, da Segunda Parte, da Quarta parte e dos dois últimos capítulos da Quinta Par

Fonte: adaptado de Flores, p.57.

Como se pode notar, há uma complexidade na organização do CLG em níveis temporais, o que, naturalmente, confere uma “flutuação terminológica no livro, e consequentemente, flutuação conceitual” (Flores, 2023, p.57). Em contrapartida, essas dissonâncias foram incapazes de apagar o impacto que essa obra demarcou na fundação da Linguística Moderna.

Nessa perspectiva, o CLG pode ser considerado um conjunto de teorias e de reflexões de Saussure - ainda que sob a voz de terceiros - que conduz o leitor a pensar sobre a possibilidade de princípios universais e particulares para as línguas, as quais são organizadas em sistemas que possuem leis próprias de funcionamento.

A ênfase dada pelo autor ao sistema de valores, a qual podemos acompanhar no CLG, entre os signos linguísticos, e a relação desses valores com a construção de significados numa língua, também confere determinada originalidade às suas reflexões. E, a partir dessas possíveis inovações apontadas, inclusive, é que podemos ver Saussure no Processamento de

isso, recomendamos a leitura de Flores (2023).

Linguagem Natural, haja vista que a forma com que um computador entende um “significado”, hoje, já possui proximidades com o que Saussure antecipou no início do século XX¹⁵.

Feito esse pequeno recorte do percurso teórico de Saussure, nas próximas seções, iremos, então, nos dedicar a discutir algumas noções e conceitos que compõem a fortuna intelectual herdada de Saussure e de que forma esses pressupostos podem se desdobrar em uma teoria da significação na língua; teoria essa que, como veremos na Parte II, apresenta indícios de ser fundamental para a compreensão da semântica distribucional e, por consequência, para o desenvolvimento de modelos vetoriais de linguagem.

É importante frisar que, embora iremos nos dedicar, isoladamente, a alguns pressupostos de Saussure para o funcionamento da(s) língua(s) com uma finalidade didática, como as noções de relação, sistema, valor e significação, sabemos que não é possível apartar esses conceitos, de fato, de outros princípios postulados pelo mestre genebrino. Isso porque defendemos que a linguística saussuriana é, de antemão, uma *linguística integradora*; ou seja, uma forma de se ver a língua e seus fenômenos em que cada lei está relacionada com outra que lhe é similar ou diferente e, ao mesmo tempo, com todas os demais pressupostos (cf. Marques, 2016).

Em outras palavras, para compreendermos o que Saussure entende por significação, é necessário retomarmos ao valor linguístico, o qual, por exemplo, é fundamental para o entendimento da noção de sistema. Para entender o que é sistema, porém, é necessário retomar à natureza da língua e de seus constituintes - os signos linguísticos. Para compreendermos a ideia de signo, em contrapartida, faz-se preciso esclarecer também a sua face dual: o significante e o significado; e assim por diante.

Essa característica relacional do saussurianismo, portanto, parece estar presente ao longo de todo o percurso teórico do mestre genebrino, como bem relembra Marques (2016). Este fundamento, assim, será tratado, neste trabalho, como o *princípio relacional de Saussure*, tema este que pode servir, como veremos, para a determinação dos sentidos na(s) língua(s) e também para o Processamento de Linguagem Natural atual explicar vários mecanismos

¹⁵ Ainda que já tenhamos antecipado um pouco essa discussão, a articulação entre valor, significado e modelos de linguagem será melhor desenvolvida na parte 2 deste trabalho.

linguísticos, como a modelagem - ou a identificação - de significados por IAs a partir da dinâmica de relações inerentes ao sistema linguístico e seus constituintes.

2. O princípio relacional e a noção de relação na teoria saussuriana

Em primeiro lugar, precisamos responder a um questionamento que talvez nosso leitor ainda esteja com dúvida: mas, afinal, por que a noção de relação é tão importante para este trabalho que articula Saussure e o PLN? Primeiramente, porque a base epistemológica da TdV e da SV reside na mesma ideia, a saber, a de que os signos estão intimamente conectados, ou seja, *relacionados*, tanto em sua constituição - por meio das relações internas entre o significante e o significado; quanto entre os próprios signos, por meio de relações externas materializadas em valores semelhantes e dessemelhantes.

Contudo, precisamos dizer que esta noção de *relação* não é de tão fácil entendimento, principalmente porque esta ideia relacional perpassa a produção teórica de Saussure em diferentes momentos de suas reflexões, de modo que ela parece ser uma chave para se desvendar muitas das encruzilhadas epistemológicas contidas no Curso de Linguística Geral e em alguns manuscritos.

O trabalho seminal que discute o que se entende por *relação* na linguística saussuriana, com o qual assumimos filiação teórica, é de autoria de Marques (2016). A dissertação intitulada *A noção de relação na teoria linguística de Ferdinand de Saussure*, para nós, marca um importante ponto de reflexão acerca da necessidade de se evidenciar, no saussurianismo, uma retomada a esse princípio epistemológico que, rotineiramente, aparece em fontes saussurianas.

Nesse sentido, o estudo proposto por Marques divide-se em três eixos principais:

- (i) Comparação com a Teoria Neogramática;
- (ii) Análise de Manuscritos de Saussure;
- (iii) A Relação como Base da Linguística Moderna.

No primeiro capítulo, intitulado "Reestabelecendo as relações fundamentais: da Linguística Histórica à Linguística Geral", Allana Marques examina como Saussure se afasta da abordagem da escola neogramática,

representada por Hermann Paul. Para os neogramáticos, a investigação da mudança linguística era central, e a noção de "relação" era abordada em função dessa mudança, sendo tratada de maneira secundária por eles.

Paul (1966 [1880]), por exemplo, enfatiza as relações no tempo e a ideia de que a linguagem evolui de acordo com essas relações, o que é importante para os estudos históricos, como vemos: "como é possível, na hipótese de forças e relações constantes, haver uma evolução histórica, um progresso das formas mais simples e primitivas para as mais completas?" (Paul ,1966[1880], p. 14).

No entanto, em determinado momento, Paul começa a se distanciar do que, tradicionalmente, era amplamente aceito pela gramática comparada para dar espaço ao seu método advindo da gramática histórica. Isto é, para o autor, os estudos comparatistas buscam as relações entre famílias de línguas, enquanto que a abordagem histórica "investiga a continuação da evolução baseando-se num ponto de partida que lhe é transmitido pela tradição" (Paul, 1966 [1880], p.33).

Entretanto, Saussure rompe com essa nova perspectiva adotada por Paul. Sobre isso, poderíamos inferir, então, que o mestre genebrino, principalmente no CLG, propõe um avanço que, de certo modo, era inovador à sua época. Em outras palavras, enquanto alguns neogramáticos preconizavam uma gramática histórica baseada nas relações entre línguas para identificar possíveis transformações, Saussure propõe um modelo de análise linguística ao focar nas relações sincrônica e nas leis internas do sistema.

No Curso de Linguística Geral, porém, vale dizer, há uma relação conflituosa estabelecida entre Saussure e a gramática histórica. O professor de Genebra, ainda que busque formas de contornar a sua insatisfação com a linguística de seu tempo, como se lê em cartas endereçadas a Antoine Meillet em 1984 (cf. Marques, 2016, p.11), não despreza totalmente os avanços científicos obtidos por meio dos estudos históricos. Para ele, essa escola foi responsável por "colocar em perspectiva histórica todos os resultados da comparação" (Saussure, 2012 [1916], p.35). Por outro lado, essas relações notadamente comparatistas estabelecidas pelos neogramáticos, para Saussure, "não correspondem a nada na realidade e [...] são estranhas às verdadeiras condições de toda linguagem" (ibidem, p.34).

Sendo assim, não se pode afirmar que Saussure rejeitou totalmente os

pressupostos da Gramática Histórica e da Gramática Comparada, de modo que foi nas lacunas epistemológicas deixadas por essas escolas que o mestre genebrino encontrou argumentos para propor uma análise linguística cujo objeto resida, em primeiro lugar, nas relações entre os fatos sincrônicos, notadamente as conexões entre diferentes estados de língua. Em outras palavras, o linguista suíço admite uma nova interpretação para a noção de relação, distanciando-se do que esse termo aparentemente significava para as escolas anteriores. Sobre isso, Marques (2016) esclarece:

Diferentemente do que acontece na reflexão saussuriana, na qual ela [a relação] ocupa lugar de destaque, Saussure se distancia dos estudos propostos pela escola que se formava no final do século XIX, sobretudo, a partir do modo de compreender **as relações em linguística.**" (Marques, 2016 p.11)

Nesse sentido, no conjunto de manuscritos *Notes pour un livre sur la linguistique générale*^{16;17} (1893-1894), atualmente arquivado na biblioteca pública de Genebra sob a inscrição Ms. Fr. 3951/11, Saussure retoma essa importância da noção de relação na língua ao torná-la a coluna vertebral de alguns de seus pressupostos. Ao longo de sua escrita, marcada por rasuras e reformulações (cf. Silveira, 2007), a primeira menção a uma ideia primitiva de "relação" aparece no trecho que Marques (2016, p.50) trata como "excerto 4", como lemos:

Figura 1. *Notes pour un livre sur la linguistique générale* (Ms. Fr. 3951/9.f. s/n) , excerto 4, transscrito por Marques, 2016, p.50.

¹⁶Segundo as notas catalográficas do acervo digital de Genebra, os manuscritos que integram o conjunto denominado *Notes de linguistique* foram doados por Jacques e Raymond de Saussure em três ocasiões: em 1955, 1958 e 1967.

¹⁷ Marques (2016, p. 38) indica que as *Notes pour un livre sur la linguistique générale* possam ser rascunhos de um livro que Saussure havia prometido a Meillet em carta de 1894. A autora considera três argumentos para isso: i) a possível coincidência entre a data da carta e dos manuscritos; ii) o conteúdo do manuscrito, que reflete claramente a mesma preocupação expressa na carta, ou seja, a necessidade de reformular a terminologia inadequada da linguística da época; e iii) os aspectos formais do manuscrito, que indicam uma preocupação com o público leitor.

Elle³¹ existe, parce que nous la déclarons identique à elle-même.
 Mais nous ne pouvons pas la déclarer identique à elle-même sans invocation tacite d'un point de vue : autrement nous pourrions tout aussi bien déclares identique à lui-même cantāre : chanter. Nous faison donc tacitement appel, pour proclamer l'existence de nū, au jugement d'identité prononcé par l'oreille, de même que nous faison appel pour affirmer l'identité^{existence unie} de cantāre et chanter à une autre [x] espèce d'identité, découlant d'un autre ordre de jugements ; mais dans aucun cas nous ne cessons de recourir à une opération de l'esprit : l'illusion des choses que seraient naturellement données dans le langage est profonde.³²

très positive

Figura 2. *Notes pour un livre sur la linguistique générale* (Ms. Fr. 3951/9.f. s/n) , excerto 4, transcrito e traduzido por Marques, 2016, p.50.

³² Ela existe porque nós a declaramos idêntica a si mesma. Mas nós não podemos declará-la idêntica a si mesma sem invocação ^{tacita} de um ponto de vista: caso contrário, poderíamos declarar também idêntico a si mesmo cantāre : chanter. Apelamos, então, tacitamente, para proclamar a existência de nū, ao julgamento de identidade pronunciado de ouvido, do mesmo modo que apelamos, para afirmar a identidade^{existencia unida} de cantāre e chanter, a uma outra [x] espécie de identidade, decorrente de uma outra ordem de julgamentos, mas, em nenhum caso deixamos de recorrer a uma operação ^{muito positiva} do espírito: a ilusão das coisas que seriam naturalmente dadas na linguagem é profunda.

Neste momento, a noção de relação, ainda que em estado de formulação, começa a adquirir um estatuto mais teórico na elaboração de Saussure. No trecho acima, Marques (*ibidem*) aduz ao entendimento de que a noção de relação pode estar contida, nesse momento, na ideia que Saussure constrói sobre as “identidades” linguísticas, ainda que não seja muito claro o que o autor pretendeu ao afirmar com a expressão “idêntica a si mesma”. É no excerto 5 (Marques, 2016, p.51), porém, que o termo “relação” aparece, pela primeira vez neste manuscrito, de forma explícita:

Figura 3. *Notes pour un livre sur la linguistique générale* (Ms. Fr. 3951/9.f. s/n) , excerto 5a, transcrito por Marques, 2016, p.51.

- Il y a différents genres d'identité.
 C'est ce qui crée ^{differents ordres} de faits linguistiques,
 de différents ordres. Hors d'une relation
 quelqu'un ^{un fait linguistique} [x] n'existe pas. Mais
 d'identité, la relation d'identité dépend d'un
 point de vue ^{variable} qu'on décide d'adopter ;
 il n'y a donc aucun rudiment
 de fait linguistique hors du point
 de vue défini que préside aux distinc-
 -tions³⁴

Figura 4. Notes pour un livre sur la linguistique générale (Ms. Fr. 3951/9.f. s/n) , excerto 5b, transcrito e traduzido por Marques, 2016, p.51.

³⁴- Há diferentes gêneros de identidade. É isso que cria ^{diferentes ordens} de fatos linguísticos **de diferentes ordens**. Fora de uma relação ^{qualquer} de identidade, ^{um fato linguístico} [x] não existe. Mas a relação de identidade depende de um ponto de vista ^{variável} que se decide adotar; não há, portanto, nenhum rudimento de fato linguístico fora do ponto de vista definido que preside às distinções.

Em nossas leituras dos excertos apresentados por Marques, os diferentes tipos, ou graus, de identidade (ou gêneros de identidade) correspondem a diversas formas de identidade (o que também consideramos como uma antecipação de “valor”) que podem ser estabelecidas pelos “signos” - embora essa última palavra não tenha aparecido, por enquanto. Esses graus, por consequência, geram diferentes categorias de fatos linguísticos, os quais só existem devido às relações de identidade formadas com base em determinado ponto de vista, como mencionado por Saussure em folhas anteriores às analisadas. Ao avançar nessa discussão, Marques (2016, p.52) argumenta que:

Esse entendimento nos permite afirmar que a relação de identidade criada pelo linguista e o ponto de vista adotado por ele são, então, simultâneos e dependentes. Dessa forma, bem entendido, o ponto de vista cria a relação ou, como veremos adiante em análise ao manuscrito seguinte, o próprio fato linguístico.

O manuscrito seguinte ao qual a autora se refere é o excerto 6 (Ms. Fr. 3951/9.f. s/n). Nesse excerto, Saussure destaca alguns termos, como "tempo" e "ligação", inclusive deixando-os em caixa alta, como veremos nas figuras a seguir. Ele sugere que o tempo é o fator que cria a **ligação** entre *alka* e *ôk*¹⁸, transformando *alka* em *ôk*.

¹⁸ Os exemplos *Alka* e *Ôk*, utilizados por Saussure, referem-se a diferentes estados de língua de um mesmo signo utilizado no indo-europeu antigo.

Figura 5. Notes pour un livre sur la linguistique générale (Ms. Fr. 3951/9.f. s/n) , excerto 6, transcrito por Marques, 2016, p.53.

lui-même

comme fait primordial

Done alka, moyennant le facteur TEMPS,
le trouve être ôk. Au fond, où est le LIEN entre
alka et ôk? Si nous entrons dans cette voie,
et il est inflexiblement nécessaire d'y entres, nous
verrons bientôt qu'il faudra de demander où est
le LIEN entre alka et alka, et à ce moment
nous comprendrons qu'il n'y a point des choses
précisément par autre chose que le liens ou les genres
de rapport que nous établissons nulle part d'abord
une chose qui sait alka^{ni aucun chose}; mais qu'il y a d'bord
un genre de rapport que nous établissons, par
exemple le rapport que entre alka et ôk[...]³⁷

Figura 6. Notes pour un livre sur la linguistique générale (Ms. Fr. 3951/9.f. s/n) , excerto 6, transcrito e traduzido por Marques, 2016, p.53.

³⁷ Então, alka, por meio do fator TEMPO, termina por ser ôk. No fundo, onde está a LIGAÇÃO entre alka e ôk? Ao entrar nesse caminho, que é inflexivelmente necessário de entrar, veremos logo que é preciso se perguntar onde está ^{ela mesma} a LIGAÇÃO entre alka e alka e, nesse momento, compreendemos que não há ponto das coisas precisamente outra coisa que as ligações ou os gêneros de relações que nós estabeleceremos, em parte alguma, ^{como fato primordial}, uma coisa que seja alka^{nem coisa alguma}: mas existe, antes, um gênero de relações que nós estabeleceremos, por exemplo a relação entre alka e ôk[...].

53

Esse trecho, com bastante marcas de rasura, evidencia uma tese saussuriana em formulação sobre a importância do aspecto do tempo para as transformações linguísticas. Curiosamente, ao fazer isso, Saussure já antecipa, indiretamente, o que ele considera como o próprio trabalho **do** linguista que, à época, como relembraria Mejía (1997) em análise desse manuscrito, parecia se confundir com o que seria o trabalho **de** linguista.

O principal propósito de Saussure neste texto é definir o objeto linguístico, mas ao fazer isso ele se encontra em uma posição bastante difícil. Isto é, como um linguista ele questiona o objeto linguístico, mas esta reflexão de linguista o conduzirá a colocar em questão o trabalho do linguista propriamente dito. Saussure desempenha aqui o papel do observador de um objeto (I), ou do epistemologista que descreve o conhecimento de um objeto por um sujeito (II)? É sobre esta corda bamba que o equilíbrio das novas ciências humanas é estruturado.

(MEJÍA, 1997, p. 98, traduzido por Marques, 2016, p.54).

Ainda no excerto 6, Saussure se questiona sobre onde exatamente estaria a ligação entre alka e ôk. Em outras palavras, o que lhe possibilitaria estabelecer uma **relação** entre esses dois termos? Marques (2016, p.54), assim, assume

que essa questão nos conduz a outra, a saber, qual seria a **ligação** entre *alka* e *alka*. Nas palavras da autora, com a qual concordamos: “Isso leva Saussure a concluir que não existe nada intrinsecamente fixo que seja *alka*, mas sim diferentes **gêneros de relações** que criamos, os quais nos permitem associar, por exemplo, *alka* a *ôk*” (Marques, 2016, p.54).

Essas transformações entre *alka* e *ôk*, por exemplo, já nos possibilita uma articulação prévia com a semântica vetorial ao abordar a questão central de como as palavras e seus significados estão relacionados. A semântica vetorial, como abordagem computacional, assume que as palavras não têm significados fixos intrínsecos, mas que seus significados emergem de relações entre elas em um espaço vetorial. No caso específico do excerto, Saussure argumenta que a ligação entre *alka* e *ôk* não é intrínseca, mas criada a partir de relações estabelecidas pelo sistema. Na semântica vetorial, algo semelhante ocorre: as palavras são representadas como vetores em um espaço multidimensional, onde a proximidade ou a direção desses vetores refletem relações semânticas ou contextuais baseadas em padrões de uso. Assim, o significado de uma palavra não é fixo, mas dependente de sua posição relativa em relação a outras palavras.

Dito isso, essa abordagem de significado relacional na semântica vetorial parece dialogar diretamente com a conclusão de Saussure, a saber: o significado de *alka*, ou sua associação com *ôk*, é compreendido em função das relações entre diferentes estados de língua, e não por um conceito essencial ou fixo. Portanto, o espaço vetorial funciona como uma formalização dessas “diferentes relações” mencionadas por Saussure, permitindo modelar como associações entre palavras surgem, se organizam e podem ser utilizadas como base para a compreensão dos sentidos na(s) língua(s).

Seguindo em suas análises com os manuscritos buscando a delimitação da noção de relação em Saussure, Marques (2016) recupera, agora, outro manuscrito do conjunto *Notes de linguistique*, a saber, o *Notes pour un livre sur la linguistique générale* 19f, o qual apresenta indícios mais explícitos de quais são as relações linguísticas que interessam à própria ciência da linguagem.

A autora, então, explica, baseando-se em Godel (1969[1957]), que o referido manuscrito é composto por anotações de dois cadernos, um de cor verde, identificado com o número 11, e outro de cor azul, identificado com o

número 12¹⁹. Marques, então, infere, a partir desses documentos, que é possível observar como a noção de relação desempenha um papel fundamental especialmente na formulação de importantes noções de Saussure, notadamente a sincronia e a diacronia.

Nas primeiras folhas do caderno verde, isto é, o Manuscrito 11, Saussure afirma que há, em nosso espírito, uma tendência em considerarmos sempre os acontecimentos, isto é, uma sequência histórica ou uma sucessão de coisas no tempo, e nos desinteressarmos pelos estados no desenvolvimento de qualquer objeto. Segundo ele, embora essa tendência funcione em outras ciências, na Linguística, ela causa desordem (Marques, 2016, p.60).

No caderno 12, todavia, a *noção de relação* começa a adquirir uma conotação mais formal, o que, inclusive, auxilia Saussure a solucionar um problema epistemológico, isto é, a crítica estabelecida no excerto 12 do *Notes pour un livre sur la linguistique générale* (Ms. Fr. 3951/11.f. 4), que incide sobre a não delimitação que a Linguística faz de seu objeto, tratando de dois objetos distintos - um *acontecimento* e um *estado* - como se fossem um único: “Ao caracterizar a diferença entre as duas ordens pelas quais um objeto pode ser tomado, Saussure recorre à noção de relação.” (Marques, 2016, p.62).

Figura 7. *Notes pour un livre sur la linguistique générale* 19f (Ms. Fr. 3951/11.f. 4) , excerto 12, transcrito por Marques, 2016, p.63.

¹⁹ Marques (2016, p.59) afirma que esses cadernos podem ter sido catalogados em conjunto devido à similaridade de seus conteúdos, o que justifica o mesmo nome atribuído a ambos.

Le premier objet qui peut frapper

Soit donc une ~~la forme~~ prise au hasard εγνω. Il n'y a rien à dire de cette forme tant qu'elle n'est mise en rapport avec rien.

Avec quoi peut elle être mise en rapport

qu'on ne l'oppose à rien, que l'on ne désigne pas [x]^{le} second terme avec lequel il y aurait à examiner son rapport. Sans doute cette vérité préliminaire est déjà ce qui échappe à la linguistique tradition de l'école, en même temps qu'elle qui-[]

Avec qui peut-elle être mise en rapport? [x] par ex. avec egnōt. Certainement, et quelle que soit le nature de ^{ce premier} rapport, avec egnōt que existait à une autre époque.

Mais certainement aussi avec egnōn qui lequel règne à la même époque.⁴⁷

Figura 8. Notes pour un livre sur la linguistique générale 19f (Ms. Fr. 3951/11.f. 4), excerto 12, transcrito e traduzido por Marques, 2016, p.63.

47 O primeiro objeto que pode chamar atenção: Seja, então, uma forma tomada ao acaso εγνω. Não há nada a dizer dessa forma enquanto ela não é colocada em relação com nada. Com o que ela pode ser colocada em relação ela não for oposta a nada, enquanto não se designar o segundo termo, com o qual ele terá sua relação examinada. Sem dúvida essa verdade preliminar escapa a escola tradicional linguística, ao mesmo tempo que ela [] Com o que ela pode ser posta em relação? [x] por exemplo com egnōt. Certamente, e seja qual for a natureza ^{dessa primeira} relação, com egnōt, que existia numa outra época. Mas certamente também com egnōn, que ^{o qual} reina na mesma época.

Como se nota, a presença da rasura no trecho que concerne à *relação*, no excerto 12, faz com que creiamos que, primeiramente, a noção de relação, à primeira vista, ainda não possuía uma delimitação muito clara para Saussure. Porém, ao leremos esse trecho em conjunto com as outras folhas deste conjunto de manuscritos, poderíamos dizer, na verdade, que a relação, a essa altura, já parece admitir um estatuto de *princípio* em meio aos fatos linguísticos.

No início do excerto 12, o mestre genebrino dá indícios de que a construção do “poder de significar” (Marques, 2016, p.62) reside no jogo de relações entre diferentes estados de língua. Ou seja, a identidade de uma forma ao acaso, como o exemplo grego que Saussure traz na primeira linha, nada significa (“Não há nada a dizer”) se este não estiver em oposição a outro signo. O mesmo se dá com “egnot”, uma variação de “egnon”, que representa, cada forma, um estado de língua diferente, de modo que resta ao linguista, então, examinar as possíveis relações que constituem e permeiam esses estados. Além disso, esse mesmo excerto parece antecipar as discussões sobre sincronia e diacronia trazidas no CLG, e mais do que isso, ele demonstra o quanto esses

conceitos necessariamente estão intimamente conectados para o funcionamento das línguas e para o entendimento de suas transformações.

Ora, se consideramos que os estados de língua podem ser uma predição do que, posteriormente, viria a ser os fatos sincrônicos (cf. Saussure, 2012 [1916]), precisamos reconhecer, então, que esses recortes de língua estão unidos por uma linha bastante firme: o tempo; ou melhor, a diacronia. Isso mostra que, embora Saussure busque privilegiar o estudo sincrônico como a tarefa do linguista, ao mesmo tempo, ele não deixa de assinalar a indissociabilidade dos fatos diacrônicos²⁰, uma vez que a sincronia contempla, na verdade, as transformações diacrônicas.

Nesse sentido, até aqui, temos demonstrado algumas evidências documentais, por meio de manuscritos autógrafos, sobre o quanto a noção de *relação* inquietou Saussure ao longo da formulação de suas teorias. Dito isso, essa noção, refletida, primeiramente, na delimitação da diacronia e da sincronia, após a publicação do CLG, começa a possuir uma maior clareza. Por isso, julgamos fundamental continuarmos a examinar como a *relação* aparece no Curso de Linguística Geral, mantendo-nos em consonância a Marques (2016) para nossas análises.

Para nós, a importância de evidenciar o que se entende por *relação* no CLG admite uma conotação especial neste trabalho. Isso porque sabemos que é o Curso a obra de Saussure que mais circulou, com a qual expoentes dos estruturalismos estadunidense e inglês - principais influências linguísticas no Processamento de Linguagem Natural (PLN) - tiveram contato.

Nesse ínterim, quando se analisa o Curso de Linguística Geral, pode-se encontrar indícios de que a noção de *relação*, na verdade, não se limita à diacronia e à sincronia, mas fundamenta muitos outros pressupostos, como o conceito de língua como sistema, bem como as noções de valor e de significação - teorias essas que parecem, desde já, primordiais para o desenvolvimento de

²⁰ Aqui está, aliás, a diferença entre a sincronia de Hermann Paul e a de Saussure. É verdade que Paul, em meados do século XIX, já havia preconizado o funcionamento das línguas por meio de recortes sincrônicos. Contudo, a noção de relação que Saussure estabelece entre a sincronia e a necessidade da diacronia para a sua existência demarca uma diferença teórica peculiar. A sincronia de Paul, é então, uma antecipação do que seriam os estados de língua saussurianos. A sincronia de Saussure, por sua vez, proponho, é o produto visível ao falante da soma de diferentes estados diacrônicos - sendo esta mais uma inovação saussuriana, diga-se de passagem.

novas técnicas de PLN (cf. Giamarusti, 2024).

Em linhas gerais, os modelos de linguagem artificial, conhecidos como *Word Embeddings* (WE), técnica de PLN que também faz parte da arquitetura do modelo *Generative Pre-Trained Model* (GPT), operam com base em análises estatísticas para determinar quão provável é uma palavra em uma sentença dada (cf. Mikolov et al, 2017). O raciocínio por trás dessa probabilidade, assim, parece advir da Hipótese Distribucional (Harris, 1955), em que os significados das palavras estão relacionados, na verdade, aos outros signos que estão em co-ocorrência no *corpus*. Dito isso, o cerne desses modelos reside na maneira como as palavras se associam umas às outras: quanto mais similar o significado, maior a probabilidade de co-ocorrência.

Em outros termos, os referidos modelos de linguagem parecem considerar uma perspectiva teórica que comprehende o funcionamento da linguagem a partir de *relações* do sistema linguístico - entre analogias e diferenças linguísticas - para que, enfim, possam inferir significados, compreender e produzir textos. Para nós, essa interconexão entre as “palavras” para gerar significação toca diretamente naquilo que consideramos ser o fundamento da(s) língua(s), o qual já havia sido esboçado por Saussure no final do século XIX e ao longo dos cursos no século XX: **o princípio relacional**; quer dizer, a importância da relação como mecanismo motriz de todo o funcionamento da língua. É esse princípio, assim, que parece sustentar o intercâmbio de valores e de significados entre diferentes estados de língua. É ele, também, quem oferece aporte à Semântica Distribucional (cf. Firth, 1968 [1955]), como veremos, bastante aplicada ao PLN atualmente²¹ como estratégia de modelagem de significados lexicais.

No Curso de Linguística Geral, o princípio relacional - ao qual Marques (2016, p. 81) prefere chamar de *natureza relacional* - já havia sido percebido por alguns saussurianistas, como Normand (2009) e Ducrot (1968). Não é surpreendente, porém, que essa noção tenha sido notada atrelada aos capítulos dedicados ao valor linguístico no CLG, já que a noção de valor é um dos pressupostos que mais sintetiza a importância da relação para a teoria saussuriana.

²¹ Conferir a Parte 2 para mais detalhes sobre as articulações entre Saussure e o PLN.

Em contrapartida, para além do *valor*, há diferentes momentos em que a natureza relacional da teoria saussuriana pode ser reconhecida no CLG. Marques (2016), assim, relembra-nos algumas vezes em que o termo *relação* aparece na edição brasileira de 2012.

Tabela 2. Recorte de algumas ocorrências do termo “relação”, e seus sinônimos, no Curso de Linguística Geral (2012), adaptado de Marques, 2016.

A ciência dos sons não adquire valor enquanto dois ou mais elementos não se achem implicados numa relação de dependência interna ; pois existe um limite para as variações de um conforme as variações do outro; somente o fato de que haja dois elementos engendra uma relação e uma regra, o que é muito diferente da simples verificação [...]	CLG, 2012, p.88
A etimologia é, pois, antes de tudo, a explicação das palavras pela pesquisa de suas relações com outras. Explicar quer dizer: reduzir a termos conhecidos, e em Linguística explicar uma palavra é reduzi-la a outras palavras, porquanto não existem relações necessárias entre o som e o sentido. [...] A etimologia não se contenta em explicar palavras isoladas; faz a história de famílias de palavras, assim, como a faz dos elementos formativos, prefixos, sufixos etc.	CLG, 2012, p.250
Cumpre, sobretudo, notar que o termo emprestado não é considerado mais como tal desde que seja estudado no seio do sistema; ele existe somente por sua relação e oposição com as palavras que lhe estão associadas, da mesma forma que qualquer outro signo autóctone.	CLG, 2012, p.55
O eixo das simultaneidades, concernentes às relações entre coisas existentes, de onde toda intervenção do tempo se exclui e o eixo das sucessões, sobre o qual não se pode considerar mais que uma coisa por vez, mas onde estão situadas todas as coisas do primeiro eixo com suas respectivas transformações.	CLG, 2012, p.88
A Linguística sincrônica se ocupará das relações lógicas e psicológicas que unem os termos coexistentes e que formam sistemas, tais como são percebidos pela consciência coletiva. A Linguística diacrônica estudará, ao contrário, as relações que unem termos sucessivos percebidos por uma mesma consciência coletiva e que substituem uns aos outros sem formar sistema entre si.	CLG, 2012, p.142
Os termos implicados no signo linguístico são psíquicos e estão unidos , em nosso cérebro, por um vínculo de associação. Insistamos nesse ponto. O signo linguístico une não uma coisa e	CLG, 2012, p.106

<p>uma palavra, mas um conceito e uma imagem acústica. Esta não é o som material, coisa puramente física, mas a impressão (empreinte) psíquica desse som, a representação que dele nos dá o testemunho de nossos sentidos; tal imagem é sensorial e, se chegamos a chamá-la “material”, é somente nesse sentido, e por oposição ao outro termo da associação</p>	
<p>Esses dois elementos [o conceito e a imagem acústica] estão intimamente unidos e um reclama o outro.</p>	CLG, 2012, p.107

Embora na Tabela 2 não estejam contempladas todas as ocorrências, a partir dela, já podemos ter uma dimensão da presença do princípio relacional em diferentes pressupostos, desde a delimitação do signo e da conceituação da língua como um sistema, até os eixos de funcionamento de *la langue*, como as relações sintagmáticas e associativas. Isso nos mostra que, afinal, no saussurianismo, **tudo está relacionado e tudo se relaciona**. Em outras palavras, todas as teorias, de alguma forma, acabam complementando-se entre si e entre o todo. Contudo, sabemos que ainda pode restar o questionamento sobre como a noção de *relação* saussuriana pode estar articulada com a nossa proposta de significação nas línguas.

Para nós, essa articulação é, em primeiro lugar, indireta. A noção de relação perpassa, notadamente, a constituição dos signos linguísticos - como veremos a seguir. Os signos linguísticos, por sua vez, formados a partir da relação arbitrária entre um significante e um significado, geram os valores linguísticos. Os valores linguísticos, assim, são os responsáveis por engendrar as possíveis relações (positivas e negativas) entre os signos. Essas relações, portanto, marcadas por ligações de semelhança e dessemelhança, garantem, *a priori*, um espaço determinado para cada signo na língua - isto é, o seu valor. A posição de cada signo, desse modo, revela alguns limites entre as relações dos signos - e, quando mapeamos todas as possíveis ligações entre um signo e a sua totalidade sistêmica, temos, por conseguinte, o que entendemos por significação em matéria de línguas naturais. Isso porque, para nós, um signo isolado, que podemos chamar de X, não é suficiente para dar sentido às línguas. O mecanismo de construção da significação, assim, só é possível quando se analisa todas as possíveis relações de X com seus similares e seus diferentes.

Nesse sentido, a seguir, iremos nos ater a como o princípio relacional

está vinculado aos signos linguísticos. Feita essa discussão, seguiremos para as nossas próximas seções, dedicadas àquilo que os signos formam, a saber: o sistema, os valores e, enfim, a significação.

3. O signo linguístico e suas relações internas e externas

Iniciamos esta discussão sobre o signo linguístico trazendo luz para o fato de que, na verdade, este conceito não nasce com Saussure, como talvez alguns imaginem, tão pouco com o Curso de Linguística Geral. Flores (2017) relembra que a ideia de signo já fazia parte, bem antes do linguista suíço, da tradição filosófica geral (*ibidem*, p. 1006).

Para nós, esse interesse da filosofia para o funcionamento da linguagem²² remonta ao “problema da representação” (Pinheiro, 2003, p.31), o qual já havia sido, primeiramente, discutido por Platão e, depois, por Aristóteles, no tocante às reflexões sobre a *mímesis* e nos diálogos do Crátilo.

Independentemente da origem dessa discussão, é fato que a teoria dos signos é basilar para o saussurianismo, uma vez que ela fundamenta uma das concepções²³ de língua trazidas por Saussure no CLG.

Apegando-se a essa segunda definição, poder-se-ia dizer que não é a linguagem que é natural ao homem, mas a faculdade de constituir uma língua, vale dizer: um sistema de signos distintos correspondentes a ideias distintas (Saussure, 2012, p.42).

Mas, afinal, como Saussure delimita o que é o signo linguístico? Assim como várias outras questões que temos feito ao longo deste trabalho, esta também não admite respostas imediatas. É perigoso sintetizar, em poucas linhas, conceitos saussurianos em formato de verbetes. Todos aqueles que assumiram essa missão falharam ao longo da história das ideias linguísticas, tendo em vista a multiplicidade de fontes disponíveis e as diferentes possibilidades de leituras da obra do mestre genebrino.

²² Importante notar que essa reflexão da tradição filosófica parece recair mais para a linguagem, em seu sentido ampla, do que para a língua (*la langue*). Contudo, no cerne dos estudos gregos e medievais sobre esse tema (cf. Pinheiro, 2003), a distinção que se faz entre língua e linguagem ainda não havia sido feita. Essa diferenciação é consolidada apenas no final do século XIX por Ferdinand de Saussure.

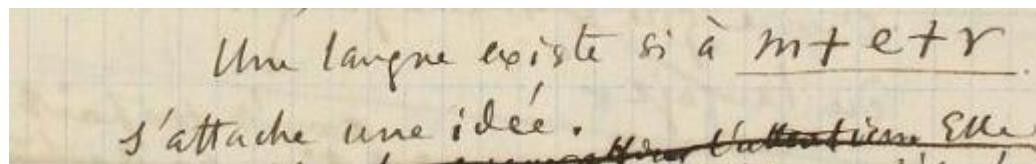
²³ O uso do plural “concepções” é proposital aqui. Alguns autores, como De-vitto (2018), indicam, pelo menos, 2 concepções de língua presentes no CLG para além da definição de língua como sistema.

Nesse sentido, a definição de signo com a qual trabalhamos refere-se aos binômios clássicos conceito-significado e imagem acústica-significante, bastante comuns no Curso de Linguística Geral e presentes também em alguns manuscritos. Vale dizer, porém, que a terminologia para “significante” e “significado” parece ter sido transformada ao longo do tempo até chegar no que se entende por esses pressupostos no CLG, isto é, o conceito como sinônimo de *ideia*, assim como o significante como a *imagem acústica*.

Nesse sentido, o que se entende por “conceito” também parece ter sofrido interpretações diversificadas. As Notas de Constantin para o curso III (*Troisième Cours de Linguistique Générale: d'après les cahiers d'Emile Constantin*), publicado a primeira vez em Francês por Komatsu em 1993, revela uma referência ao significado como “parte invisível do signo” (p.285) e também como “face espiritual” (p.291), trazendo a discussão para um plano bastante filosófico.

Ainda que uma definição sobre os constituintes do signo seja de difícil precisão, é fato que a noção de relação, por meio da união de uma imagem acústica a uma ideia, garante um dos principais conceitos de língua na teoria saussuriana. No conhecido manuscrito *Dupla Essência da Linguagem* (EDL), arquivado na biblioteca pública de Genebra sob a inscrição Arch. de Saussure 372, 1891, há um pequeno excerto que parece aduzir a essa concepção de língua pautada na relação entre os constituintes do signo. Na folha 14 do EDL, em especial, destacamos esse trecho para análise.

Figura 9. Reprodução EDL Arch.372/1891 14f, disponível em formato digital.



Uma língua existe se a m+e+r lhe corresponder uma ideia.²⁴

Nessa folha (14), Saussure, um pouco antes desse trecho, escrevia sobre a natureza fônica de *mer* (mar), a qual ele usa de exemplo para refletir sobre o domínio acústico (*le domaine de l'acoustique*) da língua. Inferimos, diante disso,

²⁴No original: “Un langue existe si à m+e+r l'attache une idée.”

que a união entre um significante e uma ideia não só garante uma identidade para o signo, como também demarca um conceito de língua que advém da correspondência entre essas “duas faces de uma mesma moeda” (cf. Saussure, 2012 [1916]). Mais do que isso, é apenas essa junção, essa relação, que permite ao falante reconhecer, nessa identidade, um valor e, por consequência, uma significação. Em outras palavras, nessa concepção saussuriana ainda em formulação, a construção de significados, interpretamos, só é possível quando se entende que o signo linguístico é produto de relações internas, entre seus constituintes, e externas, entre um signo e os outros elementos do sistema.

Ao analisar o Curso de Linguística Geral, encontramos algumas possíveis definições de signo linguístico que corroboram essa interpretação, principalmente no “Capítulo III - Objeto da Linguística” e, exaustivamente, nos capítulos I e II da Primeira Parte, “Natureza do signo linguístico” e “Imutabilidade e mutabilidade do signo” respectivamente.

De forma geral, podemos dizer 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*, p.107). Carvalho (1976), também partilha dessa concepção, como vemos: “Em outras palavras, para Saussure, conceito é sinônimo de significado” (*ibidem*, p.61); “Melhor dizendo, a *imagem acústica* é o significante” (*ibidem*, p.61).

Tais autores, portanto, aduzem a uma interpretação do signo linguístico como formado a partir de uma dualidade, o que também é confirmado por Saussure no CLG. Vale dizer, porém, que afirmar a dualidade da natureza do signo não permite isolar seus constituintes, já que o significante reclama o significado e vice-versa.

Nessa perspectiva, para além da própria definição de significante e significado, o princípio relacional também é bastante evidente quando analisamos a natureza do signo linguístico, especialmente no que concerne à arbitrariedade. Henriques (2012) explica, contudo, que, mais uma vez, essa discussão não nasce com a linguística saussuriana: “O princípio da arbitrariedade envolve discussões sobre sua natureza desde a Antiguidade Clássica” (*ibidem*, 2012).

Sendo assim, propomos que o ponto-chave para entendermos a

arbitrariedade é tomá-la como fruto da coletividade. Em outras palavras, o signo linguístico é arbitrário por natureza. Isso implica a reflexão de que os signos, por serem de natureza coletiva, também retratam um importante aspecto social da(s) língua(s). Um nome, então, na teoria saussuriana, corresponde a determinado conceito não por relação direta entre forma e substância, entre *significante* e *significado*; mas porque foi aceito, pelo sistema e pela massa de falantes. Tal dinâmica dos signos permite, inclusive, uma particularidade de análise entre as línguas. Por exemplo, “exquisito”, em Espanhol, possui uma dimensão semântica próxima à “saboroso” e “delicioso” em Português. Já na Língua Portuguesa, “esquisito” faz referência ao aposto, a algo “estranho”, “de mau gosto”.

Desse modo, essa comparação evidencia que a essência da significação não está contida nas unidades gramaticais, isto é, em seus morfemas - como parecem projetar alguns estruturalistas como Câmara Jr (2011 [1968]) e Bloomfield (1961 [1933]). Isso porque, ainda que ambas as palavras admitam os mesmos radicais ex(s)quis- elas se nutrem de significados totalmente diferentes- e a explicação para esse fenômeno reside justamente no jogo de diferenças possibilitado pela arbitrariedade. “Exquisito”, em Espanhol, tem uma conotação positiva porque, no sistema da língua espanhola, a união entre esse significante e esse significado, alude a algo positivo. O mesmo para “esquisito” em Português, que, em nosso sistema, tem uma conotação pejorativa.

Em contrapartida, não podemos deixar de mencionar que, em alguns casos, o próprio Saussure fala sobre uma relativa motivação em alguns signos. É o caso de palavras como “Dezesseis” (dez + seis), que, à primeira vista, parecem motivadas. Ora, “16” é a junção de “10 + 6”, então, é factível dizer que este signo é motivado; mas não é bem este o entendimento. De forma bem clara, questionamos: porque o signo “10” significa “dez” e não “um (1)? Por que “6” implica meia dúzia e não duas dúzias completas? Simplesmente porque o “significante é imotivado” (Saussure, 2012, p.109) e, portanto, a sua relação com o significado é, mais uma vez, arbitrária.

Cumpre, assim, distinguirmos, brevemente, o que entendemos por significado e o que entendemos por *significação*. Em primeiro lugar, ainda que esses termos sejam parecidos e, alhures, utilizados por Saussure como sinônimos, como nota Coelho (2015), notamos uma diferença conceitual entre eles. Isto é, o significado é, em suma, a contraparte do significante. É nele que

reside o aspecto conceitual do signo. É ele quem traz substância para as imagens acústicas que, sozinhas, não significam nada, sendo, portanto, apenas “massas amorfas” (Saussure, 2012 [1916]).

A significação, por outro lado, neste trabalho, tratamos como o fruto das relações externas, como anteriormente mencionamos, entre um signo e todos os outros do sistema linguístico. Por isso, para nós, a significação, ainda que beba da fonte dos significados, está além das relações internas que constituem os signos. Ela está inserida, também, nas relações externas do signo: ou seja, nas possíveis ligações entre todos os valores semelhantes e dessemelhantes de cada elemento do sistema linguístico. Sendo assim, é essencial que, para que continuemos nossa investigação sobre o que se entende por significação nas línguas naturais - e o que disso tem sido emulado por IAs - devemos passar pelo terreno do sistema linguístico e dos valores que emergem dele, afinal, é apenas através da noção de sistema que enxergamos ser possível encontrar os valores de um signo e, posteriormente, a significação na língua.

4. Delimitando “sistema” em Saussure

Como dissemos, para que cheguemos a uma melhor clareza sobre o que entendemos por significação, foi necessário, primeiro, delimitar o princípio relacional. Essa delimitação, por consequência, nos guiou pelo caminho que leva à constituição dos signos. Porém, como também já discutimos, os signos linguísticos, isoladamente, não possuem nenhum efeito de sentido. Esses sentidos, por sua vez, só são possibilitados quando se analisa a totalidade de todos os signos de uma língua. E, quando nos referimos a essa totalidade, indubitavelmente, referimo-nos ao fato de que as línguas naturais são um conjunto de signos refletidos num sistema de partes interconectadas.

Nessa perspectiva, refletir sobre a noção de *sistema* baseando-nos em Saussure, principalmente a partir do CLG, não é uma tarefa fácil. É verdade que, na obra, muitas são as passagens nas quais aparece o termo “sistema”, mas, em nenhuma delas, é claro o que se entende por essa noção, o que torna a missão de descrever o que se entende por *sistema* uma atividade complexa, confusa e, em certa medida, incompleta - haja vista a impossibilidade de se resgatar a integralidade desse pressuposto saussuriano que admite várias

interpretações.

Nesse sentido, para propor uma interpretação para a noção de sistema, Flores (2023) elenca algumas vezes em que essa palavra aparece no Curso de Linguística Geral, valendo-se da edição brasileira de 2012 como fonte saussuriana:

Tabela 3. Ocorrências do termo “sistema” no Curso de Linguística Geral (Saussure, 2012 [1916]), adaptado de Flores (2023, p.101-102).

“Uma língua, vale dizer: um sistema de signos distintos correspondentes a ideias distintas.”	CLG, p.18
“A língua é um sistema de signos que exprimem ideias.”	CLG, p.24
“A tarefa do linguista é definir o que faz da língua um sistema especial no conjunto dos fatos semiológicos.”	CLG, p.24
“Nossa definição da língua supõe que eliminemos dela tudo o que lhe seja estranha ao organismo, ao seu sistema.”	CLG, p.29
“A língua é um sistema que conhece apenas a sua ordem própria.”	CLG, p.31
“Uma língua constitui um sistema”; “a língua, sistema de signos arbitrários.”	CLG, p.87
“A língua é um sistema do qual todas as partes podem e devem ser consideradas em sua solidariedade sincrônica.”	CLG, p.102

“Sistema da língua”	CLG, p.127, 154
“Sistema de equivalência”	CLG, p.102
“Sistema de valores”	CLG, p. 95, 96, 104, 139
“Sistema de relações”	CLG, p.100
Sistema de elementos	CLG, p.138

Como se nota, há uma diversidade de lugares em que o termo “sistema” aparece no CLG. Em alguns momentos, ele parece se comportar como um substantivo, como em “sistema de valores” e “sistema de relações”; mas, em outros momentos, a terminologia se comporta como um predicativo, uma qualidade/característica da própria língua, como no caso de “A língua é um sistema de signos”.

Vale dizer também que essa falta de delimitação clara sobre a noção de sistema não se limita apenas ao Curso de Linguística Geral, mas se estende a trabalhos anteriores de Saussure, como observado por Coelho (2015), que esclarece:

No Mémoire, por exemplo, a falta de uma definição direta de noções que são centrais para o trabalho, como a de sistema e a de valor, pode incidir como uma marca de movimento da teorização saussuriana. Isso porque essa falta não compromete o entendimento da forma como essas noções contribuem para a tese do linguista, mas, apesar disso, deixa em aberto um caminho para a procura de caracterizações e definições dessas noções, procura esta que funciona, a nosso ver, como um combustível para a trajetória de elaboração das reflexões de Saussure. Essa falta, contudo, não é uma característica exclusiva do Mémoire (Coelho, 2015, p.14).

Nessa perspectiva, Coelho (2015), em seu trabalho “A noção de sistema na fundação da Linguística Moderna”, também aventura-se nessa empreitada

epistemológica que busca uma melhor delimitação sobre o que é sistema, partindo, também, da teoria saussuriana. Inicialmente, a autora destaca que a “noção de sistema já era considerada essencial desde os primeiros estudos linguísticos no Ocidente” (ibidem, p.9). Assim, ao focar na análise da noção de sistema presente no CLG, Coelho (2015) infere:

A noção de sistema é um elemento central para que o mecanismo de funcionamento da língua proceda conforme pensado por Saussure. Por outro lado, as reflexões do linguista que fundamentam a distinção entre os aspectos sincrônicos e diacrônicos da língua, apresentados no CLG, indicam que a noção de sistema parece não se diferir daquela de sincronia. Isto é, **um sistema de língua nada mais é que um estado de língua** (p.37-38, grifo nosso).

Para avançar na proposta de que o sistema saussuriano, na verdade, está relacionado aos estados de língua, é necessário esclarecermos algumas questões. Em primeiro lugar, essa interpretação de Coelho, em grande medida, constrói bastante intertextualidade com os manuscritos analisados anteriormente, especialmente o do conjunto *Notes de Linguistique*. Isso porque os estados de língua, como o mestre genebrino exemplificou com “egnot” e “egnon” (cf. Ms. Fr. 3951/11.f. 4), na verdade, são entidades diferentes entre si que, ao serem postas em relação com o todo, demonstram ser apenas partes, um recorte, de um sistema linguístico. E, se entendemos que sistema é, então, um conjunto de elementos, ou de entidades, que estão interconectados, poderíamos dizer que essas entidades, esses signos, são, também, os próprios estados de língua. Ou seja, falar de estados de língua é, ao mesmo tempo, discorrer sobre a própria composição do sistema linguístico.

Essa interpretação, todavia, não é uníssona em níveis de estudos linguísticos. Para Silveira (2009), por exemplo, a noção de sistema está intimamente relacionada à ideia de “sistema de valores” (Saussure, 2012 [1916], p.92). Ou seja, para compreendermos o que entendemos por *sistema*, antes, precisa-se reconhecer que a língua é formada por uma rede de valores que incidem, posteriormente, sobre a construção da significação. Portanto, na próxima seção, discutiremos um pouco mais sobre o sistema de valores saussurianos e a relação deles com a construção dos sentidos nas línguas naturais.

6. A Teoria do Valor: entre semelhanças e dessemelhanças

As relações entre valor, significado e significação nem sempre são fáceis de serem pontuadas. É fato, porém, que, com uma leitura mais acurada sobre o CLG, por exemplo, já podemos notar que, pelo menos, esses termos são carregados de marcas teóricas e, portanto, não devem ser confundidos. Temos visto, nesse sentido, até aqui, que o significado é o constituinte conceitual do signo e, por lógica, a contraparte do significante. A significação, por outro lado, pode ser considerada como fruto das relações externas do signo, a saber, as ligações entre um signo e todos os outros elementos do sistema. Porém, essa ligação entre os signos não é vazia; ela é constituída, necessariamente, de valores linguísticos, especialmente, valores *semelhantes* e *dessemelhantes*.

Para nós, há uma importância acentuada da Teoria do Valor para essa investigação, haja vista que o argumento que sustenta essa pesquisa é que a hipótese distributiva, amplamente utilizada em WE e no modelo GPT, parece ter forte inspiração na proposta dos valores linguísticos de Saussure. Em primeiro lugar, o mestre genebrino é citado, rotineiramente, nos textos clássicos que preconizam a hipótese distributiva- como veremos na parte 2 deste trabalho. E, não obstante, é a teoria do valor também quem parece nos fornecer indícios de que, ainda que as inteligências artificiais não possam, ainda, atribuir sentido aos fatos linguísticos tal como humanos o fazem, elas podem, por sua vez, reconhecer uma noção de *significação* baseada nos valores das *palavras*; e, com base nos valores semelhantes e dessemelhantes, determinar a probabilidade de ocorrência de signos em uma sentença. Como já dissemos: quanto maior a similaridade, ou seja, o seu valor, maior a chance de ocorrência simultânea.

É nessa perspectiva que vamos até a "Teoria do Valor" de Saussure, ela que foi apresentada principalmente no último dos três cursos que o mestre ministrou no início do século XX, na Universidade de Genebra. Embora essa teoria já tenha sido largamente investigada por outros pesquisadores da fortuna saussuriana (cf. De Mauro, 1967; Normand, 2009; Silveira, 2009; Coelho, 2013; 2015; Coelho e Lima, 2014), argumentamos ainda a existência de possibilidades interpretativas. Trata-se de um princípio central para o entendimento da língua como sistema, o qual só pôde ser desenvolvido por Saussure após ele ter definido diversos outros conceitos e princípios linguísticos abordados em seus

cursos, como a arbitrariedade do signo, a linearidade do significante, a distinção de significado e significante como partes do signo linguístico, além das clássicas distinções entre língua, fala e linguagem (cf. Coelho, 2013).

No artigo "A Teoria do Valor no Curso de Linguística Geral", Silveira (2009) explora a concepção de Ferdinand de Saussure sobre a teoria do valor, a partir de uma análise do Curso de Linguística Geral. A autora parte de uma reflexão sobre qual seria a natureza da língua, destacando que as noções de organização, estrutura e sistema foram recorrentes na busca por essa natureza, as quais possuem direta relação com a constituição e o funcionamento do sistema de valores.

Um dos pontos centrais do artigo de Silveira é a ideia de que, para Saussure, o valor linguístico não é determinado apenas por signos isolados, mas por um sistema de relações diferenciais entre eles. No CLG, há vários momentos que sustentam essa afirmação, de modo que destacamos: "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). Nessa passagem, contida no capítulo IV da segunda parte, dedicado à Teoria do Valor, é possível inferir que o sentido gerado por uma palavra só pode ser compreendido quando este estiver em relação aos outros do sistema linguístico. Isso, mais uma vez, fundamenta aquilo que já havíamos proposto, a saber, o princípio relacional como mecanismo de funcionamento de todos os pressupostos de Saussure - e o valor não lhe escapa.

No Curso de Linguística Geral, são recorrentes às vezes em que a noção de valor aparece. Contudo, é, de fato, no capítulo IV da segunda parte em que este pressuposto é desenvolvido e também desdobrado em outros conceitos, como o valor considerado em seu aspecto conceitual e em seu aspecto material. É nesse capítulo, também, que Saussure começa a indicar que há distinções entre *valor* e *significação*.

Nesse sentido, no que se refere ao valor em seu aspecto conceitual, Saussure admite que "o **valor** constitui, sem dúvida, um elemento da **significação**, e é difícil saber como esta se distingue dele, apesar de estar sob sua dependência" (Saussure, 2012 [1916], p.161, grifo nosso). Com base nesse trecho, Coelho (2013) argumenta que "existe uma relação de dependência entre esses dois princípios que não é apresentada de forma clara na edição

[CLG]" (ibidem, p.2).

Ora, se a distinção entre valor e significação não é clara no CLG, por onde poderíamos, então, começar a delimitá-la? Para isso, Coelho (2013), em trabalho intitulado “Significação em Saussure: os três cursos de linguística geral”, percorre a trajetória do termo “significação” em alguns cadernos de alunos dos três cursos de Genebra, com especial atenção às anotações de Riedlinger, editadas e publicadas por Komatsu e Wolf em 1996, quando da análise do primeiro e do segundo cursos; bem como as anotações de E. Constantin para fundamentar suas reflexões sobre a análise do terceiro curso, publicado por Komatsu e Harris em 1993.

Da análise de Coelho (2013), nota-se que, no primeiro curso (1907), a “significação” aparece ainda de maneira incipiente, com foco em aspectos relacionados à possível substituição de termos como “ideia” e “conceito”, enquanto que, no segundo curso (1908-1909), há uma profundidade nessa aproximação e uma distinção mais clara entre “valor” e “significação”. No terceiro curso (1910-1911), por sua vez, essa distinção torna-se mais sólida, mas ainda assim não isenta de certas ambiguidades.

A análise dos cursos começa com a reflexão de que a “significação”, tal como apresentada no Primeiro Curso de Linguística Geral (PCLG), ora é sinônimo de termos como “conceito”, “ideia” ou “significado”, ora é diferenciada desses termos. Essa interpretação surge, assim, a partir de evidências documentais contidas nas anotações de Riedlinger, em que, ao abordar o que seria um “fato de significação”, Saussure utiliza o termo “conceito”, o que nos leva a inferir que, neste primeiro momento, a “significação” parece estar muito próxima da contraparte do signo, o “significado”, e, por consequência, à ideia de “conceito”.

<Deve-se notar> um fato de significação que é apenas outro que aquele mencionado por -ιστο-: se um conceito composto é dado em um signo determinado, a tendência mecânica da língua de o tornar mais simples, indecomponível, e a tendência de tomar o caminho mais direto, a simplificação da ideia: de dois ou três dados acaba-se percebendo nada mais que essa <que é> entendida. (Saussure apud Riedlinger, 1996, p. 92, traduzido por Coelho, 2013, p.6)²⁵

²⁵ No original: “Un fait de signification qui n'est autre que celui mentionné pour –ιστο –: la tendance mécanique de la langue, si un concept composé lui est donné dans un signe déterminé, de le rendre simple, indécomposable, la tendance de prendre le chemin de traverse, la

Nessa perspectiva, a autora explora um pouco mais sobre essa aproximação entre “significação” e “conceito” ao analisar as anotações de Riedlinger referentes ao Segundo Curso de Linguística Geral (SCLG), em que se encontra o seguinte trecho:

O momento da gênese não é em si mesmo conhecível, nem mesmo visível. O contrato primitivo se confunde com o que <se> passa todos os dias na língua, <com as condições permanentes da língua:> se você aumenta um signo na língua, você diminui proporcionalmente a significação dos outros. <Reciprocamente: se, por impossível que seja, tivessem sido escolhidos em princípio apenas dois signos, todas as significações repartir-se-iam entre estes dois signos. Um designaria a metade de um objeto e o outro a outra metade (Saussure apud Riedlinger, 1997, p. 12, traduzido por Coelho, 2013, p.8)¹⁷

De acordo com esse excerto, inferimos que a significação de um signo específico está proporcionalmente ligada ao número de signos existentes na língua; ou, em outras palavras, à totalidade do sistema linguístico. Saussure afirma que, se a língua fosse composta por apenas dois signos, eles teriam a função de representar todas as significações. Com isso em mente, é relevante citar um trecho do CLG em que Saussure aborda, de forma similar, a limitação mútua de significado entre palavras aparentemente “sinônimas”.

No interior de uma mesma língua, todas as palavras que exprimem ideias vizinhas se limitam reciprocamente: sinônimos como recear, temer, ter medo só têm valor próprio pela oposição; se recear não existisse, todo seu conteúdo iria para os seus concorrentes. [...] Assim, o valor de qualquer termo que seja está determinado por aquilo que o rodeia; nem sequer da palavra que significa “sol” se pode fixar imediatamente o valor sem levar em conta o que lhe existe em redor; (Saussure, [1916] 2012, p. 163).²⁶

Com relação à diferença entre *valor* e *significação*, é neste segundo curso que uma distinção mais formal é consolidada: “O valor é dado por outros dados; pela situação recíproca das partes na língua e assim por diante” (Saussure apud Riedlinger, 1997, p. 29, traduzido por Coelho, 2013, p.8).²⁷ Essa distinção,

simplification de l'idée: de deux ou trois données on finit par ne plus apercevoir que celle entendue.”

²⁶ No original: “Le moment de la genèse n'est lui-même pas saisissable, on ne le voit pas. Le contrat primitif se confond avec ce qui passe tous les jours dans la langue, si vous augmentez d'un signe la langue vous diminuez d'autant la signification des autres”

²⁷ No original: “Il y a un phénomène déjà par le fait que cette différence est une des choses qui contribuent à la signification.”

assim, é mantida no Terceiro Curso de Linguística Geral (TCLG), publicado por Komatsu e Harris em 1993, em que, em notas de E. Constantin, lê-se:

Quando se fala em valor, sentimos que ele se torna <aqui> sinônimo de sentido (significação) e isso indica um outro terreno de confusão (<aqui a confusão> será mais duradoura nas coisas propriamente ditas). O valor é bem um elemento do sentido, mas é importante não tomar o sentido de outra forma que como um valor. Essa é, talvez, uma das operações mais delicadas de se fazer em linguística, que é ver como o sentido depende e, contudo, resta distinto do valor (Saussure apud Constantin, 1993, p. 134, traduzido por Coelho, 2013, p.10).

Nesse ínterim, a inferência a que podemos chegar a partir da leitura de alguns trechos desses cadernos é que há uma evolução coerente no pensamento de Saussure sobre a "significação", que, em vários momentos, parece estar próxima da ideia de "conceito", embora não se confunda com o que se entende por "valor".

. No CLG, por sua vez, a confluência entre esses termos parece mais conflituosa, haja vista que, no subcapítulo dedicado ao aspecto conceitual do valor linguístico, o mestre genebrino, mais uma vez, retoma discussões sobre o signo e seus constituintes, com ênfase para a contraparte da imagem acústica, isto é, o significado, assemelhando-o ao que seria a significação na língua. É justo nesse momento que é apresentado, no CLG, os indícios das diferenças entre *valor* e *significação/significado*, com os quais concordamos.

Em outras palavras, o significado apresenta-se, assim, como a contraparte do significante, estabelecendo, necessariamente, uma relação entre os constituintes internos do signo. Por outro lado, esse mesmo signo possui níveis de relação que extrapolam a sua interioridade, admitindo, por consequência, ligações com a totalidade de todos outros signos - como dissemos, relações externas. Sendo assim, quem sustenta essa conexão entre diferentes signos é a dinâmica de valores. Isto é, o significado está contido na constituição do signo, e as relações externas dele, por sua vez, são quem constitui o valor, o qual, por lógica, também apresenta o significado como uma de suas partes - sendo este o lado conceitual. Na figura a seguir, apresentamos um esquema apresentado no CLG o qual parece ilustrar este entendimento:

Figura 10. Esquema de valores entre os signos.



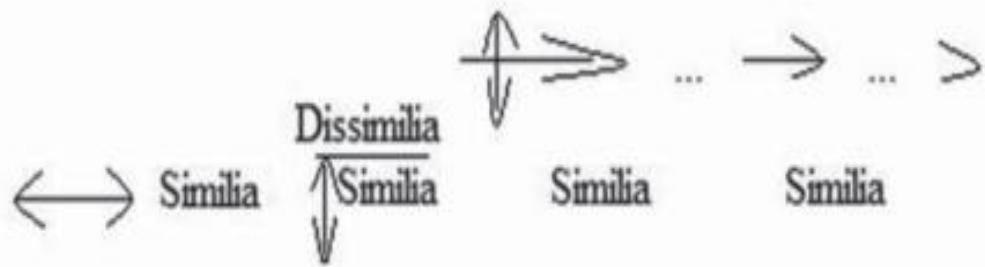
Fonte: Saussure, 2012, p.161.

Na figura, nota-se que há uma linha que “corta” os signos. Este conjunto de traços, assim, trata-se, para nós, do efeito na língua do princípio relacional, isto é, a noção de relação perpassando a ligação entre o significante e o significado, bem como compondo as relações externas entre eles. As flechas, por outro lado, parecem indicar os valores que são construídos entre os signos, os quais podem se organizar por semelhança ou dessemelhança. Nesse sentido, a organização dos valores por meio dessa via de relações aparece em alguns momentos na produção saussuriana.

Na segunda parte do conjunto de manuscritos *Notes pour le cour III* (1910-1911), como mostra Coelho (2015), lemos:

Valor é, na verdade, eminentemente sinônimo a cada instante de termo situado em um sistema de termos similares, do mesmo modo que é, na verdade também na verdade, eminentemente sinônimo a cada instante de coisa trocável, certo um X objeto X é o que faz de tempo X. Não há nele um ponto [] Tomando a coisa trocável de fato X os termos adjacentes ao val de outro os termos co-sistêmáticos, que não oferecem nenhum parentesco. É próprio do valor colocar em relação essas duas coisas. Ele as coloca em relação de uma maneira que é até tal que se pode desafiar, que se pode dizer desesperadora perigosa para o espírito pela impossibilidade de investigar se essas duas faces do valor diferem por elas, visto em que, a única coisa indiscutível certa evidente é que o valor se encontra nesses dois eixos, é determinado segundo esses dois eixos concorrentes.²⁸

²⁸ Valeur est éminemment tout à fait synonyme à chaque instant de terme situé dans un système de termes similaires, de même qu'il est tout à fait aussi tout à fait éminemment synonyme à chaque instant de chose échangeable, contre un X objet dissimilaire ce qui fait de temps X Il n'y a point de [...] Prenant la chose échangeable en fait X les termes adjacents à la val de l'autre les termes co-systématiques, cela n'offre aucune parenté. C'est le propre de la valeur de mettre en rapport ces deux choses. Elle les met en rapport d'une manière qui va jusqu'à telle qu'on peut défi er, qu'on peut dire désespérant dangereuse pour l'esprit par la impossibilité de scruter si ces deux faces de la valeur diffèrent pour elles, vu en quoi, la seule chose indiscutable certaine évidente est que la valeur va dans ces deux axes, est déterminé selon ces deux axes concurrents.



(Saussure, 1910-1911, f. 27, traduzido por Coelho, 2015, p.142)

No Curso de Linguística Geral, por sua vez, também encontramos a determinação do valor linguístico via relações de semelhança (“similia”) e dessemelhança (“dissimilia”):

[...] 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)

Sendo assim, Saussure retoma, em suas explicações no Curso, que tanto os signos semelhantes quanto os dessemelhantes são necessários para a existência de um valor. O mestre, ainda, utiliza de algumas metáforas para explicar esse entendimento, como o valor de uma moeda de 5 francos (cf. Saussure, 2012, p.162). Para determinar o valor deste signo, por exemplo, é necessário, em primeiro lugar, trocá-la por uma *coisa* diferente, como “pão”; e, em seguida, compará-la com algo de valor semelhante no sistema, como uma moeda de 1 franco.

Em meio a essa explicação, além disso, Saussure preconiza que os valores nos sistemas não estão fixados, haja vista que é necessária uma dinâmica de relações com a totalidade semelhante e dessemelhante dos signos para determiná-los, de modo a afirmar, como já dissemos na apresentação, que “Seu conteúdo [valor] só é verdadeiramente determinado pelo concurso do que existe fora dela [da palavra]” (*ibidem*).

Em continuidade à discussão trazida sobre o valor linguístico no capítulo IV da segunda parte do CLG, cumpre também trazermos algumas reflexões

quanto ao valor em seu aspecto material. Durante este subcapítulo, Saussure dedica-se a explicar que, embora a materialidade do valor pareça estar contida no significante, conferindo-lhe uma delimitação que ultrapassa a matéria fônica, a imagem acústica, por si, não representa nada, a menos que esta esteja em relação com as outras imagens do sistema e comportando-se como a contraparte do significado.

Isso é ainda mais verdadeiro no que diz respeito ao significante linguístico; em sua essência, este não é de modo algum fônico; é incorpóreo, constituído não por sua substância material, mas unicamente pelas diferenças que separam sua imagem acústica de todas as outras (Saussure, 2012 [1916], p.166).

Além do mais, o aspecto material do valor parece, para Saussure, também aproximar-se do próprio sistema de escrita que, igualmente como signos, são arbitrários; ou seja, “o valor das letras é puramente negativo e diferencial; assim, a mesma pessoa pode escrever *t* com variantes [...] (ibidem)”.

A partir dessas discussões, o mestre genebrino começa, então, um percurso teórico, que podemos acompanhar no CLG, em busca de justificar a ideia de diferenças nos sistemas linguísticos. O subcapítulo “O signo linguístico considerado na sua totalidade” inicia-se com uma máxima bastante conhecida nos estudos saussurianos, a de que as relações entre os signos são, por essência, puramente fruto de diferenças: “tudo o que precede equivale a dizer que na língua só existem diferenças.” (Saussure, 2012 [1916], p.167).

As diferenças na língua, notadamente, as relações negativas e positivas, só são possíveis de serem identificadas ao se analisar o todo do sistema. Sobre isso, Saussure indica que o sistema linguístico é uma série de diferenças de significantes combinadas com uma série de significados. Quando essas diferenciações são confrontadas, por exemplo, ao se analisar um signo com seus semelhantes e dessemelhantes, tem-se, enfim, um sistema de valores.

Nessa perspectiva, Saussure chega a afirmar que, ainda que o significado e o significante, quando isolados, apresentem características diferenciais e negativas, a combinação entre esses constituintes engendra um “fato positivo” (ibidem, p.168), o qual temos interpretado como o *valor similia* (Saussure, 1910-1911, f. 27), isto é, a relação com signos semelhantes. Esse fato positivo, por conseguinte, contrai relações com o *valor dissimilia* (ibidem), ou seja, dessemelhante e, portanto, negativo por sua natureza opositiva. Dito isso,

embora Saussure não despreze a existência de valores positivos, o mestre preconiza que, *na língua*, só existem diferenças. Isso pode, à primeira vista, levar o leitor a uma certa confusão entre a noção de “diferença” e a de “dessemelhança”. Dessemelhança, então, seriam as relações com signos que ocupam espaços distantes na língua, como bem mostra o funcionamento de embeddings de palavras e a própria metáfora do jogo de xadrez, contida no CLG (cf. Giamarusti, 2024). A diferença, contudo, assumimos como o efeito dos valores positivos e negativos; ou, em outras palavras, como a função do valor dentro do sistema linguístico, que é delimitar o signo por relações de *símile* e *dissimile*. Sendo assim, quando Saussure enfatiza que, na língua, só há diferenças, seria o mesmo que afirmar que, na língua, só há valores.

Essa perspectiva apresentada anteriormente, então, parece evidenciar o fato de que a semântica vetorial pode ser compreendida como uma formalização das ideias saussurianas sobre o funcionamento da língua. No espaço vetorial, as palavras não possuem significado intrínseco, mas são delimitadas por suas relações diferenciais com outras palavras, tal como Saussure argumenta que o valor dos signos linguísticos emerge apenas de suas diferenças no sistema.

Assim, a ideia de que “na língua só há diferenças” parece encontrar correspondência direta no princípio segundo o qual os vetores das palavras só adquirem significado ao serem posicionados em relação aos outros vetores no espaço multidimensional. Além disso, os valores positivos e negativos descritos por Saussure, traduzidos pelas noções de *símile* e *dissímile*, refletem-se nos agrupamentos e distâncias que configuram a estrutura do espaço vetorial: palavras semanticamente próximas compartilham regiões similares no espaço, enquanto palavras distantes ilustram relações de oposição ou dessemelhança. Essa organização sistêmica, que delimita o signo por relações de proximidade e oposição, é reforçada com a analogia da metáfora do jogo de xadrez, onde o valor de cada peça é inteiramente dependente de sua relação com as demais e das regras que organizam o jogo. Portanto, o que propomos é que o espaço vetorial pode ser interpretado como uma representação matemática do sistema de diferenças que define a língua, confirmado que, assim como na linguística saussuriana, na semântica vetorial só existem valores determinados pelas relações diferenciais entre os elementos.

Ainda sobre o valor, Silveira (2009), aponta que, por meio da teoria dos

valores linguísticos, encontra-se um distanciamento da máxima neogramática de que a língua representa o pensamento. Para a autora, há indícios de que *la langue* saussuriana não representa diretamente a mente, mas sim organiza o pensamento em unidades significativas. Ela resgata, no Curso, uma passagem que se refere ao capítulo IV da segunda parte, em que, para Saussure, a língua é "um intermediário entre o pensamento e o som" (Saussure, 2012 [1916], p.131), onde o pensamento, naturalmente caótico, é organizado pela língua.

Para nós, essa organização inerente ao próprio sistema linguístico se dá, em primeiro lugar, por meio dos valores estabelecidos entre os signos. Isso porque a língua, enquanto conjunto de signos, não possui entidades com substância, sendo apenas formas, de modo que todos os seus componentes adquirem valor a partir de suas diferenças e semelhanças com outros elementos do sistema. Antes desse jogo de relações, o pensamento é apenas uma massa amorfa, a qual precisa da língua para se organizar em entidades significativas: "Psicologicamente, abstração feita de sua expressão por meio das palavras, nosso pensamento não passa de uma massa amorfa e indistinta" (Saussure, 2012 [1916], p.158).

Esse argumento de Saussure, assim, parece demarcar uma interpretação diferente da visão tradicional de língua advinda desde o século XVIII (cf. Ducrot, 1968), a saber, a de que a língua seria uma imitação do pensamento. O mestre genebrino, em vez disso, parece sugerir que *la langue* organiza e intermedia a construção do pensamento através da sua organização em um sistema que é capaz de engendrar valores diferenciais, como vemos:

O papel característico da língua diante do pensamento não é criar um meio fônico material para a expressão das ideias, mas servir de intermediário entre o pensamento e o som, em condições tais que uma união conduza necessariamente a delimitações recíprocas de unidades. (*ibidem*, p.159).

Em suma, ao tratarmos a língua como um sistema de diferenças que intermedia o pensamento e o som, não podemos nos esquecer de que essas diferenças só se manifestam a partir das relações estabelecidas dentro do sistema. Por isso, Saussure insiste que, embora o valor e a significação estejam profundamente conectados, é essencial que o linguista não os confunda. O valor, sendo uma função do sistema de relações, depende da posição dos signos em relação à totalidade sistêmica, enquanto que o significado é a contraparte do

significante, e a significação, o produto final gerado a partir dos valores estabelecidos pelo sistema.

Essa diferenciação teórica, assim, aponta para a complexidade e a dificuldade de delimitar a noção de sistema na obra saussuriana, bem como a noção de valor e de significação. As encruzilhadas epistemológicas presentes no CLG, portanto, revelam que, embora a noção de sistema seja central para a teoria, ela talvez permaneça inacabada e, portanto, aberta a diferentes interpretações - e o mesmo ocorre com os valores linguísticos. Essa complexidade advinda das lacunas epistemológicas, talvez da própria edição, entretanto, é o que confere à teoria saussuriana uma capacidade de renovação e adaptação a outros campos do conhecimento, como o Processamento de Linguagem Natural (PLN).

7. Considerações parciais

Nesta primeira parte, fizemos um breve panorama teórico sobre uma teoria da significação da língua que reside na ideia de que a língua é um sistema de relações e valores. Para isso, partimos do pressuposto de que é o princípio relacional de Saussure - a noção de *relação* (Marques, 2016) - o principal mecanismo pelo qual se chega ao sentido na língua. Isso porque sustentamos a ideia de que a significação na língua se constitui numa cascata de fenômenos linguísticos, a saber: partimos da noção de relação, que constitui, em primeiro nível, os signos linguísticos. Os signos linguísticos, em segundo nível, são organizados por meio de relações internas, entre o significante e o significado, e relações externas, entre os próprios signos, engendrando valores positivos (semelhantes) e negativos (dessemelhantes). Em terceiro nível, os signos constituem o sistema, que é um conjunto de elementos intrinsecamente relacionados, em que cada signo ocupa um espaço determinado na língua por meio de seus valores. Os valores, por sua vez, num possível quarto nível, tornam perceptíveis ao falante o jogo de diferenças na língua; o que, por consequência, gera a “significação”.

Nesse sentido, por meio de evidências documentais e comparação de autores, buscamos traçar este mecanismo da significação, analisando cada um destes níveis com base no CLG, em alguns manuscritos autógrafos e em alguns

cadernos de alunos de Saussure.

Tendo essas discussões em mente, a segunda parte desta pesquisa, então, discutirá como esse entendimento da significação em Saussure pode servir de aporte teórico para interpretarmos, linguisticamente, o modo como é feita a modelagem semântica de itens lexicais em alguns modelos de linguagem que são baseados na Semântica Distribucional, como a técnica de embeddings de palavras conhecida por Word2Vec (Mikolov et al, 2013).

Na seção seguinte, então, propomos um breve panorama teórico acerca dos fundamentos do distribucionalismo linguístico de Firth e Harris e de que forma tais pressupostos parecem estar intimamente relacionados à noção saussuriana de relação e o que dela deriva: os signos, o sistema e os valores. Em seguida, iremos demonstrar, sucintamente, o que é a técnica Word2Vec e como ela pode ser utilizada. Nesta mesma seção, iremos analisar alguns frameworks já trabalhados por Gastaldi (2021)²⁹, quem também argumenta que o Word2Vec remonta, em grande medida, a princípios e a pressupostos da linguística saussuriana.

Na sequência, queremos demonstrar a aplicação da Semântica Distributiva num curioso projeto desenvolvido pela Universidade de Genebra, a qual está trabalhando, juntamente com o Centro de Estudos Informáticos, em uma ferramenta para transcrição e tradução de manuscritos, começando, desde já, com manuscritos saussurianos. Vale dizer, nesse sentido, que Giussep Cosenza já demonstrou, em artigo intitulado, *Les projets de “Digital HUmanities” relatifs à l’œuvre de Ferdinand de Saussure*, publicado no Cahier Ferdinand de Saussure nº70, em 2017, que há vários projetos de humanidades digitais atualmente em andamento relacionados com o desenvolvimento de softwares para trabalhar com manuscritos históricos, tendo como base de início os manuscritos autógrafos de F. de Saussure. Nas próximas seções, nos dedicaremos um pouco mais a fundo sobre estes projetos.

Esperamos, nessa perspectiva, que a Parte II deste trabalho seja um convite ao leitor para refletirmos sobre o percurso que vai do signo ao algoritmo, da noção de valores linguísticos à semântica distribucional; e, por fim, da semântica distribucional à modelagem semântica de incorporações de palavras.

²⁹ No original: “It is impossible not to recognize in Saussure’s diagram the mechanism at play in embedding spaces as we characterized them earlier [...].”

PARTE 2

DO SIGNO AOS VETORES DE PALAVRAS

É impossível não reconhecer no diagrama de Saussure o mecanismo presente nos espaços de embeddings.
(Gastaldi, 2021, p.199)

1. A Semântica Distribucional sob a égide do saussurianismo

A teoria saussuriana sobre a relação na língua e suas implicações para a significação, com ênfase nas relações externas do signo, ganha uma nova dimensão no contexto da semântica distribucional e, consequentemente, na Linguística Computacional. Enquanto que, em Saussure (2012), como vimos na Parte I, é possível propor uma teoria da significação que reside na posição relativa do signo dentro do sistema, gerando valores positivos e negativos a partir das relações entre os signos, as abordagens distribucionais levam essa ideia que pressupõe uma espacialidade na língua adiante, transformando relações de valores semelhantes e dessemelhantes em representações numéricas em um espaço vetorial - isto é, dimensionável - que, enfim, pode ser lido pelo computador para desempenhar tarefas de PLN.

Nesse sentido, na semântica distribucional (SD), a ideia central é que o significado das palavras é amplamente determinado por seu contexto e pelas formas como elas aparecem distribuídas em determinado *corpus* de análise. Neste trabalho, propomos que a referida abordagem semântica possui, basicamente, três fortes inspirações teóricas, a saber:

- (i) a linguística saussuriana, com ênfase para a Teoria do Valor;
- (ii) a linguística hariana, com ênfase para a noção de distribucionalismo contido na obra *Distributional Structures* (1954);
- (iii) a linguística firthiana, com ênfase para a ideia de que o significado das palavras depende da rede de relações que ela mantém (Firth, 1957).

Nesse sentido, Saussure (2012 [1916]) já havia dado indícios de que a significação, na língua, não está contida no significado da palavra, mas sim, como vimos, no concurso do que existe fora dela, a saber, as relações entre os signos, que, na Parte I, nomeamos como “relações externas”. Essas relações, portanto, criam uma rede dinâmica de valores positivos, baseados em semelhanças, e valores negativos, baseados em dessemelhanças. Esse princípio, consequentemente, parece antecipar algumas ideias fundamentais exploradas mais tarde no campo da semântica distribucional, especialmente aquelas trazidas no texto seminal de Zelling S. Harris (Harris, 1955).

No *Distributional Structure* (ibidem), Harris demonstra, no início da década de 50, nos EUA, os princípios fundamentais que deveriam ser considerados para

a formação de uma linguística distribucional, isto é, que vê o funcionamento da língua em função das relações e da distribuição de um signo em meio aos outros constituintes do sistema. Quanto ao que se entende por “distribuição”, Harris explica que:

A **DISTRIBUIÇÃO** de um elemento é o conjunto de todos os ambientes em que ele ocorre, ou seja, a **soma de todas as posições (ou ocorrências) diferentes de um elemento em relação à ocorrência de outros elementos**. Duas enunciação ou características serão consideradas linguisticamente, descritivamente ou distribucionalmente equivalentes se forem idênticas em termos de seus elementos linguísticos e das relações distribucionais entre esses elementos.³⁰ (Harris, 1954, p.16, tradução nossa, grifo nosso).

Como se nota no trecho acima destacado, já é possível perceber uma intertextualidade com a noção de sistema de valores desenvolvida por Saussure, notadamente aquela presente no CLG; isto é, a ideia de que os constituintes do sistema, os signos, são delimitados exclusivamente via o plano das relações com os outros signos.

Embora, na citação acima, a menção a Saussure possa estabelecer, a princípio, uma menção indireta à linguística saussuriana, em outro fragmento do *Distributional structure*, o próprio Harris faz menção nominal a Saussure em uma nota de rodapé³¹ ³², creditando ao mestre genebrino e aos linguistas de Praga, a base teórica relativa à delimitação do objeto de análise da linguística distribucional. Na mesma página dessa nota, encontra-se mais uma intertextualidade com Saussure:

Os elementos [os fonemas e os morfemas] são, portanto,

³⁰ No original: “The DISTRIBUTION of an element is the total of all environments which it occurs, i.e. the sum of ;ill the (different) positions (or occurroncos) of an element rehitive to the occurrence of other elements. Two utterances or features will be said to be ling:uistically, descriptively, or distributionally equivalent if they are identical as to their lin-guistic elements and the distributional relations among these elements.”

³¹ A nota de rodapé diz: “A declaração mais explícita sobre o caráter relativo e padronizado dos elementos fonológicos foi apresentada por Edward Sapir em *Sound Patterns in Language*, publicado em *Language*, vol. 1, pp. 37–51 (1925), também incluído em *Selected Writings of Edward Sapir*, pp. 33–45. Além disso, ver o tratamento dos elementos fonológicos em Ferdinand de Saussure, *Cours de linguistique générale*, e Nikolai Trubetzkoy, *Grundzüge der Phonologie* (*Travaux du Cercle Linguistique de Prague*, vol. 7, 1939)” (Harris, 1954, p.7,tradução nossa).

³² Nota de rodapé no original: “The most explicit statement of the relative and patterned character of the phonologic elements is given by Edward Sapir in *Sound Patterns in Language*, L\nguage. 1.37-51 (1925); now also in *Selected Writings of Edward Sapir* 33-45. See also the treatment of phonologic elements in Ferdinand de Saussure, *Cours de linguistique generale*; Nikolai Trubetzkoy, *Grundzüge der Phonologie* (*Travaux du Cercle Linguistique de Prague* 7, 1939)” (Harris, 1954, p.7).

determinados em relação uns aos outros e com base nas relações distribucionais entre eles. É de fundamental importância que esses elementos sejam definidos em relação aos outros elementos e às interrelações entre todos eles.”³³ (Harris, 1954, p.7-8, tradução nossa, grifo nosso).

Um pouco antes da citação acima, Harris está discutindo sobre o que torna um elemento linguístico relevante para a análise distribucional. Nesse ponto, o autor parte da fonologia e da morfologia do estruturalismo praguense, e da noção de fonologia contida no CLG, para inferir que: “x e y são incluídos no mesmo elemento A, se a distribuição de x em relação aos outros elementos B, C, etc., for, de algum modo, equivalente à distribuição de y.”³⁴

Em níveis matemáticos, a proposição de Harris poderia ser interpretada via um sistema de funções. Ou seja: seja $D(x)$ a função que representa a distribuição de um elemento x em relação aos outros elementos B, C, \dots, B, C , então:

$$A = \{x, y \mid D(x) \approx D(y)\}$$

Isso indica que x e y pertencem ao mesmo elemento A se, e somente se, $D(x)$ for aproximadamente igual a $D(y)$, onde $D(x)$ é definida como:

$$D(x) = \{(x, b), (x, c), \dots \mid b \in B, c \in C, \dots\}$$

Aqui, $D(x)$ descreve o conjunto das relações de x com os outros elementos B, C, \dots . Assim, a equivalência $D(x) \approx D(y)$ reflete a similaridade nas distribuições relativas de x e y.

Como se nota, a teoria distribucional de Harris pode ser vista como uma ponte entre a Linguística e a Matemática, sendo essa articulação bastante cara à semântica computacional, oferecendo bases consistentes para o desenvolvimento da semântica distribucional no Processamento de Linguagem Natural. Em outras palavras, a maneira como Harris transformou conceitos linguísticos tradicionais em operações formais baseadas em distribuição e padrões, utilizando ferramentas matemáticas para descrever relações entre entidades linguísticas, constitui forte inspiração para os cientistas da

³³ No original: “The elements are thus determined relatively to each other, and on the basis of the distributional relations among them. It is a matter of prime importance that these elements be defined relatively to the other elements and to the interrelations among all of them”.

³⁴ No original: “x and y are included in the same element A if the distribution of x relative to the other elements B, C, etc., is in some sense the same as the distribution of y.” (Harris, 1954, p.7).

computação, nos anos subsequentes, proporem modelos de linguagem baseados em uma noção espacial de língua, notadamente, os modelos vetoriais de linguagem.

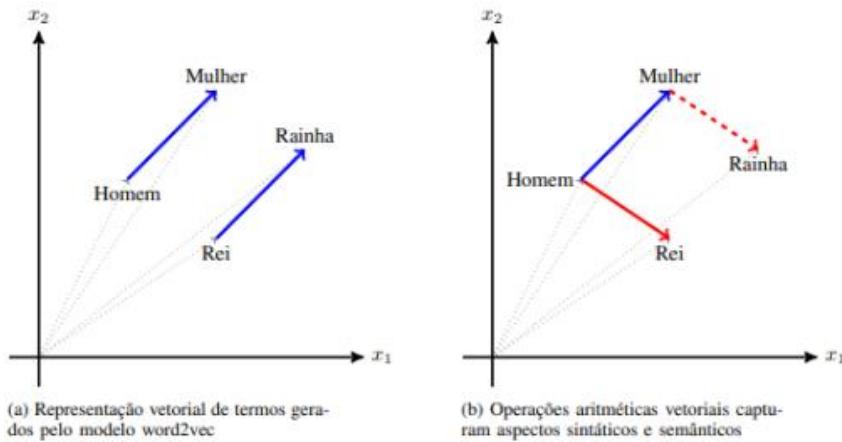
No entanto, é intrigante que, apesar de Harris ter tido acesso ao *Curso de Linguística Geral*, visível em alusões diretas e indiretas a algumas noções saussurianas, como o *sistema* e o *valor*, as menções ao mestre genebrino em suas obras são escassas. Para além das menções já referidas anteriormente, o nome de Saussure aparece outras duas vezes no *Distributional Structure* - em ambos os casos, trata-se de notas de rodapé que relacionam Saussure ao desenvolvimento de uma linguística descritiva, como se lê: “O estudo independente da estrutura descritiva foi esclarecido em grande parte pelo ‘Cours de linguistique générale’ de Ferdinand de Saussure.” (Harris, 1954, p.149, tradução nossa)³⁵.

Ainda que, então, seja aparentemente fácil estabelecer referências de Harris a Saussure, não podemos negar que o linguista norte-americano avançou na abordagem saussuriana para a significação. Enquanto o mestre genebrino parece partir da noção de relação e do que dela deriva para pensar em relações positivas e negativas que engendram valores e, por consequência, a significação na língua; Harris, por sua vez, sem desprezar o caráter sistêmico da língua, propõe que a significação é determinada, quantitativamente, pelas ocorrências dos signos em um corpus e pela natureza das relações com os signos que estão em co-ocorrência no mesmo sintagma.

Na Linguística Distribucional, por exemplo, os signos “Rei” e “Rainha” compartilham traços semânticos comuns, como serem monarcas e possuírem, em geral, as mesmas funções sintáticas na oração. Essas semelhanças permitiriam substituições em frases como “Vida longa ao rei” por “Vida longa à rainha”, sem comprometer a complexidade sintática e as relações semânticas da oração. No contexto do Processamento de Linguagem Natural (PLN), essa proximidade semântica entre dois termos se reflete em sequências numéricas próximas dentro de um espaço matemático, isto é, vetorial, como demonstramos na Figura 11 a seguir:

³⁵ No original: “The independent study of descriptive structure was clarified largely by Ferdinand de Saussure's Cours de linguistique general”.

Figura 11. Exemplo de representação vetorial criada pelo word2vec.



Fonte: Aguiar e Prati, 2015, p.4.

A Figura 11 apresenta uma representação vetorial de palavras geradas pelo modelo de linguagem Word2Vec, o qual, como veremos nas seções subsequentes, é fortemente inspirado por esta teoria da significação de Harris que estamos analisando; teoria essa conhecida como “Hipótese Distributiva” (Harris, 1954), a saber, a ideia de que palavras utilizadas em contextos semelhantes admitem, por consequência, traços semânticos comuns. No painel (a), observa-se que palavras como “Homem”, “Mulher”, “Rei” e “Rainha” são representadas como vetores. Nesse sentido, a proximidade entre esses vetores reflete similaridades semânticas entre os termos, enquanto a direção dos vetores pode indicar relações específicas. Por exemplo, as palavras “Homem” e “Mulher” possuem uma relação próxima, assim como “Rei” e “Rainha”, evidenciada por vetores paralelos e proporcionalmente distantes.

No painel (b), por sua vez, a imagem ilustra como operações aritméticas vetoriais que permitem capturar tanto aspectos semânticos quanto sintáticos das palavras. Um exemplo clássico é a subtração entre vetores, como “Rei - Homem” e “Rainha - Mulher”. Essas operações resultam em vetores equivalentes, indicando que o modelo é capaz de identificar a relação entre gênero e realeza de forma sistemática. Isso revela, portanto, uma das maiores inovações dos modelos baseados em word embeddings: a capacidade de representar as relações externas do signos, ou seja, seus valores, por meio de operações matemáticas simples no espaço vetorial. Ou seja, “mulher” é o mesmo que: mulher = rainha – homem, o que nos faz retomar a Saussure, no entendimento

de que *mulher* é, em níveis computacionais, o resultado da diferença dos signos que a circunscreve.

Como vimos, o valor de um signo linguístico é definido por suas relações de semelhança e dessemelhança com os outros signos do sistema, e não exclusivamente por seu significado, isto é, a contraparte do significante que, juntos, formam as relações internas do signo. De forma análoga, os modelos de word embeddings, como ilustrado na imagem, capturam essas relações diferenciais em um espaço vetorial, em que a posição de cada palavra, portanto os seus valores, é definida em relação às outras palavras do sistema.

No primeiro painel, as relações entre os vetores que representam palavras como "Homem", "Mulher", "Rei" e "Rainha" refletem a estrutura relacional do sistema. Assim como na teoria saussuriana, em que o valor de uma palavra depende de suas relações de *simile* e *dissimile* (Saussure, 1910-1911, f.27) com os outros signos, a proximidade e a direção dos vetores expressam relações semânticas e sintáticas baseadas na co-ocorrência das palavras em um corpus textual - que, a grosso modo, é sempre um recorte de determinado sistema linguístico, isto é, uma amostragem quantitativa dos signos e dos seus valores em níveis pragmáticos.

Ademais, o segundo painel reforça a conexão com a teoria do valor ao demonstrar que as operações aritméticas entre vetores, como a subtração, podem reforçar relações também gramaticais entre os signos, como o gênero ou a posição social. Esse processo, assumimos, é bastante semelhante à forma como Saussure postula que o valor linguístico emerge de uma rede de relações que perpassam e constituem o sistema da língua. Aqui, "Rei" e "Rainha" compartilham um valor similar em relação a "Homem" e "Mulher", uma dinâmica que pode ser visualizada nos vetores paralelos e proporcionais, alinhando-se à noção saussuriana de que o sentido é derivado do concurso do que existe fora da palavra, ou seja, as relações externas do signo: os valores que constituem o sistema.

Portanto, temos discutido que a representação vetorial de palavras, conhecida como *word embeddings*, pode ser entendida como uma formalização computacional da TdV de Saussure, em que as palavras são identificadas como elementos interdependentes em um sistema, ou seja, como signos linguísticos interconectados, reafirmando a relevância das relações diferenciais e

semelhantes na construção da significação na língua natural e, ao mesmo tempo, contribuindo para a modelagem semântica de modelos de linguagem baseados na Semântica Distribucional.

Nessa perspectiva, voltando-nos às principais influências teórico-lingüísticas para a consolidação da abordagem distribucionista nos estudos da linguagem, é impossível não refletirmos sobre as contribuições de John R. Firth para a SD. Concomitantemente a Harris, Firth desenvolve, no apogeu da Escola Linguística de Londres, entre 1930 e 1950, um aprofundamento do distribucionalismo, propondo um método de análise que serviu de base, posteriormente, para o desenvolvimento do que se convencionou chamar de Funcionalismo Inglês, na metade da década de 50.

Diferentemente de Harris, Firth fez recorrentes menções a Saussure ao longo de seu percurso teórico (Firth, 1934; 1949; 1950), o que já evidencia, do ponto de vista documental e histórico, que ambos os autores compartilhavam raciocínios comuns em certos pontos, notadamente a noção de que a língua é um sistema. Porém, enquanto Saussure aprofundou-se nessa noção buscando meios para descrever os elementos que constituem o sistema, bem como as leis internas que regem o seu funcionamento, Firth se inclinou a pesquisas que buscavam descrever a *função* de um elemento dentro de um sistema de relações. Essa noção de *função*, por consequência, permitiu-lhe propor uma teoria da significação que reside na dinâmica de conexão entre os signos, muito parecida com a Hipótese Distributiva de Harris (1954), propondo uma máxima bastante conhecida no PLN atual, a saber: “Reconhecerás uma palavra pela companhia que ela mantém” (Firth, 1955, p.108).

No trabalho *Linguistics and the functional point of view*, Firth estabelece uma conexão significativa entre sua abordagem e a teoria saussuriana. Inspirado pela máxima de Saussure de que "o ponto de vista cria o objeto" (cf. Saussure, 2012 [1916]), o autor desenvolve uma perspectiva que combina elementos do estruturalismo praguense e da sua perspectiva funcional para compreender a língua, o que nos possibilita afirmar que, em primeiro lugar, Firth não só teve contato com o CLG, como também se dedicou a estudá-lo e, em alguma medida, avançar na proposição saussuriana de sistema em direção a uma linguística funcionalista.

Qualquer sistema de conhecimento é uma forma de pensamento

sobre os fatos. As estrelas não fazem a astronomia. As estrelas podem ser o objeto, os fatos, mas é o astrônomo quem faz a astronomia. Alguns de nossos astrônomos estão, inclusive, criando seus próprios universos, compondo, por assim dizer, sua própria música das esferas. E isso não é tão absurdo quanto parece à primeira vista. Sem conhecimento prévio, o que um homem vê que realmente importa? Como observa Saussure: “Bem longe de o objeto preceder o ponto de vista, poder-se-ia dizer que é o ponto de vista que cria o objeto”³⁶ (Firth, 1934, p.18).

Por outro lado, as menções a Saussure feitas por Firth, nem sempre, estão no plano da concordância. É preciso mencionar, inclusive, como já demonstramos em outra ocasião (cf. Giamarusti, 2024), que a tese funcionalista de Firth para a língua pressupõe, também, uma certa ruptura com preceitos elementares de Saussure. Em artigo publicado em 1949, intitulado *The Semantic of Linguistic Science*, o autor londrino esclarece esses pontos de divergência com Saussure:

Apontaria que as noções, algo análogo, previamente associadas aos termos franceses *langage*, *langue* e *parole*, na obra de Saussure, sob a influência da sociologia durkheimiana, não fornecem uma base filosófica satisfatória para as técnicas de análise linguística que proponho. Elas não estão em consonância com a filosofia ou a teoria geral da linguagem aqui apresentadas. Além disso, ao enfatizar a natureza sistêmica da linguagem, não proponho um sistema *a priori* de categorias gerais para descrever os fatos de todas as línguas. Em vez disso, diferentes sistemas podem ser encontrados na atividade da fala e, ao serem enunciados, devem dar conta adequadamente dessa atividade. A ciência não deve impor sistemas às línguas, mas buscar sistemas na atividade da fala e, ao encontrá-los, descrever os fatos em uma linguagem apropriada³⁷ (Firth, 1949, p. 400, tradução nossa).

³⁶ No original: “Any system of knowledge is a form of thought about facts. The stars do not make astronomy. The stars may be the subject matter, the facts, but it is the astronomer who makes astronomy. Some of our astronomers are making their own universes as well, composing as it were their own music of the spheres. And this is not so absurd as it seems at first sight. Without previous knowledge what man sees anything that matters ? As de Saussure remarks "Bien loin que l'objet précède le point de vue, on dirait que c'est le point de vue qui crée l'objet".

³⁷ No original: “I would point out that the somewhat analogous notions previously coupled with the French words *langage*, *langue*, and *parole* in the work of de Saussure under the influence of Durckheimian sociology, are not a satisfactory philosophical basis for the techniques of linguistic analysis I have in mind. They do not accord with the philosophy or general theory of language here put forward. Moreover, in emphasizing the systemic nature of language, I do not propose an *a priori* .system of general categories by means of which the facts of all languages may be stated. Various systems are to be found in speech activity and when stated must adequately account for such activity. Science should not impose,systems on languages, it should lo& for systems in speech activity,’ and, having found them, state the facts in a suitable language”.

No excerto apresentado, é perceptível que Firth não rejeita completamente a noção de sistema em Linguística, como vemos em “não proponho um sistema *a priori* de categoriais gerais...”. Ora, se o autor não propõe um sistema de categorias gerais, ele propõe qual tipo de sistema? Isso indica que, quanto a noção de sistema do londrino pareça diferente da do mestre genebrino, ambos admitem a necessidade de uma delimitação de sistema para as suas teorias linguísticas. Não obstante, no excerto acima, podemos acompanhar uma crítica que Firth faz a Saussure relacionada à aplicação de sua noção de sistema como um suposto modelo universal e pré-determinado para a análise linguística das línguas - o que, argumenta Firrh (1949), seria impossível dada à complexidade das línguas particulares.

Além disso, o mestre londrino também faz críticas às distinções *língua*, *fala* e *linguagem*, partindo da ideia de que Saussure se limita a uma noção de sistema que se estenderia, somente, à língua em seu aspecto geral. Em outras palavras, Firth (1949, p.400) propõe que a tarefa do linguista não deve ser propor leis ou categorias gerais às línguas, mas identificar os sistemas particulares que emergem da própria atividade da fala, isto é, do próprio uso.

Em outras palavras, Firth, para propor suas teorias, parece partir da linguística da fala e do uso, enquanto Saussure toma como ponto de partida o sistema linguístico da língua e dos princípios que atravessam o seu funcionamento. Além disso, o mestre londrino, em linhas gerais, entende que a fala, no sentido de *la parole*, forma, por sua própria natureza, espécies de “microssistemas”, distribuídos e unicamente percebidos no uso dos falantes. Embora essa visão firthiana de que a fala engendra sistemas próprios possa parecer contrastar com a perspectiva de Saussure, a saber, a noção de que a língua é um aparente único sistema de leis universais que “conhece a sua própria ordem” (Saussure, 2012 [1916], p.31), deve-se evidenciar, também, que a própria linguística saussuriana não exclui a possibilidade de se pensar nos sistemas que se formam a partir do uso da *língua*, tal como Firth propõe.

Coelho (2019), nesse sentido, demonstrou, em sua tese de doutorado, intitulada *Ferdinand de Saussure: entre a língua e as línguas*, a partir do CLG e de alguns manuscritos autógrafos, que Saussure admite leis gerais para a língua, isto é, *la langue*, e, ao mesmo tempo, também para as línguas, os *idiomas*; sendo estes considerados também sistemas, já que a noção de *línguas* se entrelaça

com a ideia de que a *língua* é um sistema.

A análise dos documentos utilizados como fontes deste trabalho – tanto o CLG como os conjuntos de manuscritos de Saussure – nos mostrou que, embora o conceito saussuriano de língua seja recorrentemente tomado como um objeto estritamente formal, é possível depreender a incidência da noção de línguas particulares em sua delimitação. Essa incidência, que pode ser vislumbrada mesmo partindo unicamente do CLG, indica a existência de um entrelaçamento entre a língua, enquanto um objeto formal, e as línguas particulares – ou idiomas – enquanto objetos empíricos (Coelho, 2019, p.122)

Nessa perspectiva, Firth (1949) não despreza a ideia de que a língua constitui um sistema de regras gerais, mas tece críticas a Saussure contra uma suposta aplicação rígida ou universalizante de leis universais para o funcionamento das línguas particulares; interpretação esta do CLG que se apresenta como uma leitura, talvez mais literal, da noção de que a língua seria um único sistema de regras próprias, o que, como sugere Coelho (2019), há controvérsias.

Sendo assim, Firth adota uma abordagem na qual ele sugere que, em sua linguística distribucional, o objeto deve ser os sistemas linguísticos materializados, e também quantificados, na prática do uso, pois, só assim, para o autor, seria possível identificar as reais funções das relações entre os signos (cf. Firth, 1949) e mapear as relações próximas e as relações mais díspares.

Nessa perspectiva, o desenvolvimento das ideias distribucionais de Harris e Firth alcançou o Processamento de Linguagem Natural em meados da década de 70, desdobrando-se numa teoria semântica e computacional que vê, nas relações entre os signos, o mecanismo-chave para que uma IA, por meio de técnicas de aprendizado automático, possa emular significados e gerar, por fim, identificar relações entre termos ou até gerar textos coerentes - tal como o modelo GPT, do ChatGPT, hoje o faz, por meio de técnicas de redes neurais e transformadores.

Com relação à pertinência da SD para o PLN, Seno et. al explicam que:

A semântica distribucional tem sido atualmente a principal abordagem de representação do significado lexical adotada nas mais diversas tarefas do processamento de linguagem natural. Nessa abordagem, os itens lexicais (palavras) são representados por meio de vetores de **valores** reais, conhecidos por vetores semânticos, que codificam o significado das palavras a partir de sua distribuição em textos (Seno et. al, 2020, p.1, grifo

nosso).

Nesse sentido, o artigo "A Vector Space Model for Automatic Indexing", de Salton, Wong e Yang (1975) é um marco no desenvolvimento da semântica vetorial, principalmente aplicada à recuperação automática de informações³⁸. O estudo introduz o conceito de **modelo de espaço vetorial** como uma abordagem para representação e recuperação de documentos em sistemas de informação. Dito isso, a ideia central é que documentos e palavras podem ser representados como vetores³⁹ em um espaço multidimensional, permitindo cálculos de similaridade baseados na proximidade desses vetores:

Dado os vetores de índice para dois documentos, é possível calcular um coeficiente de similaridade entre eles, $s(D_i, D_j)$, que reflete o grau de similaridade nos termos correspondentes e nos pesos atribuídos a esses termos. Essa medida de similaridade pode ser o produto interno dos dois vetores ou, alternativamente, uma função inversa do ângulo entre os pares de vetores correspondentes; quando a atribuição de termos para dois vetores é idêntica, o ângulo será zero, produzindo uma medida de similaridade máxima (Salton, Wong, Yang, 1975, p.613, tradução nossa)⁴⁰.

Os autores (*ibidem*) propõem um método em que cada documento é representado por uma lista de números, chamada de vetor, que indica a importância de certas palavras ou termos no texto. A relação entre documentos é avaliada com base na semelhança entre essas listas, usando cálculos matemáticos como a comparação de direções dos vetores ou do grau de proximidade entre eles. Esse método organiza documentos em grupos (*clusters*),

³⁸ Recuperação Automática de Informação (RAI) é uma área da ciência da computação que se concentra no desenvolvimento de sistemas capazes de encontrar informações relevantes em grandes conjuntos de dados, geralmente em resposta a consultas específicas feitas por usuários. Esses sistemas são amplamente utilizados em mecanismos de busca, bancos de dados e plataformas de recuperação de documentos.

³⁹ Um vetor é uma entidade matemática que possui magnitude (tamanho) e direção, usada para representar grandezas como força, velocidade ou posição no espaço. Ele é geralmente representado por uma seta em diagramas ou por uma sequência de números (coordenadas) em um sistema cartesiano. Em computação e estatística, vetores são usados para armazenar e manipular conjuntos de valores, como características de dados. No contexto de PLN, vetores representam palavras ou documentos em espaços multidimensionais para análise de relações semânticas. Essencialmente, vetores permitem cálculos e representações precisas em diversas áreas científicas.

⁴⁰ No original: Given the index vectors for two documents, it is possible to compute a similarity coefficient between them, $s(D_i, D_j)$, which reflects the degree of similarity in the corresponding terms and term weights. Such a similarity measure might be the inner product of the two vectors, or alternatively an inverse function of the angle between the corresponding vector pairs; when the term assignment for two vectors is identical, the angle will be zero, producing a maximum similarity measure.

facilitando a recuperação daqueles mais relevantes quando alguém faz uma busca específica.

Além disso, o artigo também aborda outras questões fundamentais para a consolidação de uma chamada “semântica vetorial” - advinda dos princípios da semântica distribucional de Harris e Firth, bem como da noção de espacialidade da língua que a TdV de Saussure supõe. São elas:

- (i) Representação de documentos: cada documento é associado a um vetor D_i , onde cada dimensão corresponde a um termo de índice com um peso atribuído.
- (ii) Similaridade entre documentos: a proximidade entre dois documentos é inversamente proporcional à distância vetorial entre eles, indicando maior relevância.
- (iii) Otimização do espaço: os autores propõem métodos para maximizar a separação entre documentos irrelevantes, enquanto agrupam documentos relacionados em clusters compactos.
- (iv) Peso de termos: apresenta estratégias de ponderação, como a frequência inversa do documento (IDF), para aumentar a precisão da recuperação.

Nesse sentido, o modelo introduzido por Salton, Wong e Yang é fundamental porque fornece uma base matemática, computacional e, ao mesmo tempo, linguística para o entendimento das relações semânticas entre as palavras. Para isso, os autores incorporaram princípios da semântica distribucional ao definir que o significado de um termo é inferido por sua relação com outros termos no espaço vetorial. Essa abordagem, portanto, tornou-se a precursora de técnicas de PLN modernas, como *word embeddings*, com o Word2Vec, que também utiliza representações vetoriais para capturar relações semânticas e, posteriormente, modelar significados.

O Word2Vec, assim, foi desenvolvido por uma equipe do Google, em 2013, liderada por Tomás Mikolov, um jovem cientista da computação extremamente influente no cenário atual de *machine learning* e IA, atualmente associado ao departamento de informática da Czech Technical University (CIIRC CTU).

Na próxima seção, portanto, vamos nos dedicar um pouco mais sobre o modelo Word2Vec, procurando evidenciar de que modo esta técnica de PLN pode se articular com a semântica distribucional e, por consequência, com a noção de *relação* (Marques, 2016) de Saussure e do que dela deriva: os signos,

o sistema, os valores e a significação, respectivamente.

2. Word Embeddings

Antes de discorrermos um pouco mais sobre o Word2Vec, é importante retomarmos os paradigmas do Processamento de Linguagem Natural que o antecederam a fim de entendermos o contexto de produção deste modelo de linguagem e, por fim, sua funcionalidade e aplicação.

Em primeiro lugar, precisamos pontuar que o Word2Vec é o produto de sucessões de avanços em técnicas de vetorização de palavras. Nesse sentido, chamamos este processo de transformar palavras em vetores numéricos de **Word Embeddings** (WE), o qual atua “permitindo que máquinas compreendam e processem texto de maneira eficaz” (Araújo, 2024, p.28). Trata-se, portanto, de uma etapa fundamental no PLN para qualquer atividade que envolva transformar palavras, termos ou documentos em vetores.

Dito isso, pode-se dizer que o Word2Vec:

pertence a uma classe de modelos de linguagem baseados em redes neurais que abordam o problema da relação semântica entre palavras e a maldição da dimensionalidade, explorando o poder das redes neurais para aprender suas representações distribuídas (Abubakar, Umar; Bakale, 2021, p.28)
⁴¹.

Mas, afinal, o que é um modelo de linguagem? Várias são as possíveis respostas para este questionamento, mas ressaltamos o fato de que um modelo é “uma simplificação de um fenômeno complexo, no nosso caso, uma simplificação da língua que possa ser representada por ferramentas computacionais” (Paes; Vianna; Rodrigues, 2023, p.2). O Word2Vec, portanto, busca “simplificar” relações semânticas e sintáticas entre diferentes signos para facilitar análises probabilísticas de uma sequência de palavras, além de facilitar a realização de análises de similaridade entre termos e documentos, busca e indexação semânticas, entre outras atividades do Processamento de Linguagem Natural.

Vale dizer, neste sentido, que o Word2Vec consolidou-se com uma inovação na área de PLN tendo em vista que, por meio de redes neurais, o

⁴¹ No original: *Word2Vec belongs to a class of neural network language models that address the problem of semantic relatedness of words and the curse of dimensionality by exploiting the power of neural networks to learn their distributed representations* (

modelo conseguiu oferecer soluções para questões complexas, como a “maldição da dimensionalidade”, apontada por Abubakar, Umar e Bakale (2021).

Nesse sentido, o paradigma da dimensionalidade reside no fato de que, quando trabalhamos com textos, cada palavra pode ser transformada em um número, ou seja, em um vetor. Contudo, ao tentarmos representar uma palavra com muitas informações, por exemplo, gênero, função sintática, contextos de uso etc, os cálculos se tornam muito lentos e difíceis, haja vista que, quanto mais informações, maior a dimensionalidade do vetor. O que o Word2Vec faz, então, é, em linhas gerais, reduzir uma quantidade de informações consideradas desnecessárias para o processamento do texto, logo, redimensionando o vetor, tornando o modelo de embeddings mais eficiente e, por consequência, mais ágil.

Poderíamos, assim, fazer uma analogia com o dicionário. Se quisermos encontrar uma palavra, é mais fácil num dicionário pequeno do que em um livro gigantesco com milhares de palavras desorganizadas (e aqui, pode-se ler também *desestruturadas*). Sendo assim, o Word2Vec faz algo que parece resolver esta questão da dificuldade em encontrar uma palavra relevante em meio a um compilado gigante de dados, reduzindo as informações para que os cálculos fiquem mais rápidos e precisos.

Um fato muito importante que não pode escapar desta discussão sobre o Word2Vec é a proximidade teórica que ele contrai com a Semântica Distribucional, uma vez que ele trabalha a partir da análise simultânea de co-ocorrências; e a “coocorrência pode ser traduzida para: itens que aparecem próximos uns dos outros ou ainda itens que aparecem em contextos similares” (Paes, Vianna e Rodrigues, 2023, p.2), indicando que a probalidade de ocorrência de uma palavra está intimamente relacionada às palavras que estão a sua volta, tal como já fora preconizado por Saussure (2012 [1916]) na determinação dos valores linguísticos, e por Firth (1957), por meio da Hipótese Distributiva.

Em contrapartida, embora o Word2Vec possa auxiliar o pesquisador das interfaces Linguística-PLN com evidências claras sobre quais são os aspectos teóricos do saber linguístico que estão envoltos em modelos vetoriais, nem sempre são perceptíveis as influências linguístico-teóricas imbricadas no desenvolvimento de um modelo de linguagem. Um reflexo disso são os poucos os trabalhos que refletem sobre a relação entre teorias linguísticas e PLN de

forma geral. Por outro lado, autores importantes na área de Processamento de Linguagem Natural, como Russel e Norvig (2013 [2010]), não ignoram a contribuição da Linguística teórica para o desenvolvimento do PLN. Contudo, os autores partem das correntes cognitivistas, notadamente representadas por Chomsky (1956), chegando a afirmar que o desenvolvimento do campo das ciências cognitivas é concomitante ao desenvolvimento do PLN, devido a possíveis interesses comuns.

Todavia, embora o cognitivismo tenha dominado boa parte das pesquisas em PLN durante muito tempo, tem-se notado, nos últimos anos, um aparente interesse, no Processamento de Linguagem Natural, por princípios teóricos do formalismo linguístico da primeira metade do século XX, dando vazão ao surgimento de modelos de linguagem, como o Word2Vec e o modelo GPT⁴², fortemente inspirados por princípios da Linguística Moderna, como a linguística saussuriana - influência essa que temos buscado demonstrar ao longo deste trabalho; do Círculo Linguístico de Londres, representado pela semântica distribucional de John R. Firth; e do Círculo Linguístico de Nova Iorque, com grande influência da linguística distribucional de Zellig S. Harris.

Como dissemos, o Word2Vec nasce em meio a esse possível retorno à linguística formalista promovido pelo PLN. Nesse sentido, o trabalho seminal que origina este modelo é um artigo de Mikolov e seus pares, intitulado *Efficient Estimation of Word Representations in Vector Space*, em 2013. É este o artigo que apresenta o Word2Vec para a comunidade especializada, fortalecendo o desenvolvimento de modelos de linguagem de representações vetoriais de palavras em espaços multidimensionais⁴³, os quais, como também já vimos, foram previamente previstos por Salton, Wong e Yang (1975).

Por outro lado, o modelo de Mikolov revolucionou o Processamento de

⁴² Em outro trabalho (cf. Giamarusti, 2024b), demonstramos que o modelo GPT, arquitetura básica do ChatGPT, da OpenAi, também é um modelo de linguagem baseado, em grande grau, em embeddings de palavras.

⁴³ Espaços multidimensionais são estruturas matemáticas que permitem representar e analisar dados ou objetos que possuem múltiplas características ou dimensões. Eles são uma generalização do conceito de espaço tridimensional (altura, largura e profundidade) para um número arbitrário de dimensões. No caso dos vetores no Processamento de Linguagem Natural, cada dimensão contém uma “informação” da palavra, ou seja, pode conter informações sobre seu gênero gramatical, as relações de semelhança com outras palavras, as relações de dessemelhança, os contextos de uso pragmáticos etc.

Linguagem Natural (PLN)⁴⁴ e, em grande medida, superou a proposta de Salton, Wong e Yang (*ibidem*), uma vez que o modelo mikoloviano, além de ser baseado em redes neurais⁴⁵, oferece uma maneira mais eficiente de capturar relações semânticas e sintáticas entre as palavras em camadas mais profundas por meio de técnicas de *deep learning*, possibilitando que uma IA possa compreender relações complexas entre os signos, redimensionando os vetores. Em contrapartida, Kenneth Church (2016), professor titular do Institute for Experiential AI (EAI) da Northeastern University, nos EUA, alerta para algumas limitações do modelo, como a dificuldade de trabalhar com palavras polissêmicas e com ambiguidades. Contudo, o professor também considera que, por mais que o Word2Vec não seja o primeiro nem o último modelo de linguagem que trabalha com espaços vetoriais, tão pouco o “melhor”, ele ainda sim é o mais simples tecnicamente e, portanto, ainda bastante aplicado:

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)⁴⁶

Mas, afinal, como o Word2Vec funciona? Em primeiro lugar, precisa-se lembrar que um modelo de linguagem nada mais é que um conjunto de algoritmos; e o Word2Vec, por consequência, também. Por *algoritmo*, entende-se “um conjunto de instruções para realizar uma tarefa, produzindo um resultado final a partir de algum ponto de partida.” (Doneda; Almeida, 2018, p.141).

O fluxo comum de um algoritmo sempre obedece a uma lógica padrão, com diferentes variações em cada uma das etapas a depender da finalidade e da complexidade do projeto. Contudo, em linhas gerais, todo algoritmo precisa de “dados de entrada”, que serão tratados para uma maior extração de

⁴⁴ A ideia de “revolução” aqui não é exagero. Até o presente momento desta escrita, o artigo de Mikolov ao qual nos referimos já foi citado mais de 45 mil vezes em trabalhos acadêmicos entre 2013 e 2024, de acordo com dados do Google Scholar.

⁴⁵ Redes neurais são sistemas computacionais inspirados no funcionamento do cérebro humano. Elas são compostas por “neurônios” artificiais organizados em camadas, que processam informações aprendendo padrões a partir de dados. Esse aprendizado permite que as redes realizem tarefas como reconhecer imagens, traduzir textos ou prever resultados, ajustando-se continuamente para melhorar sua precisão.

⁴⁶ 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).”

informações (ou mineração). Os dados de entrada, então, quando prontos para uso, passarão por uma etapa de “processamento”. Nesta fase, insere-se diferentes códigos que irão atuar no cumprimento da instrução dada. No caso do Word2Vec, usa-se tradicionalmente SKIP-GRAM e CBOW como os principais conjuntos de códigos de processamento, os quais, adiante, iremos retomar. Feito isto, o modelo entrega o resultado final do processamento, o que consideramos como os dados de saída.

4. Funcionamento do Word2Vec e a determinação de palavras com base nos valores semelhantes e dessemelhantes

Agora que já entendemos brevemente o que é o Word2Vec, vamos nos dedicar um pouco mais sobre sua etapa de processamento. Antes, porém, é preciso elucidar o paradigma anterior ao Word2Vec, conhecido como Bag-of-Words (BoW), uma vez que o Word2Vec pode ser entendido como uma evolução do BoW.

Em seções anteriores, discutimos de que modo o artigo de Salton et. al (1975) trouxe grandes avanços no PLN, propondo um modelo de linguagem baseado em vetores. Com alguns conhecimentos mais amadurecidos, agora, podemos dizer que o método introduzido por esses autores é o de “saco” de palavras, isto é, Bag-of-Words. Nesse sentido, o BoW trata-se de algoritmos representação de texto, que convertem palavras em vetores numéricos, ignorando a ordem das palavras e considerando apenas a frequência de ocorrência de termos no texto. É uma técnica fundamental em Processamento de Linguagem Natural (NLP) e foi amplamente utilizada em tarefas como classificação de texto (Sebastiani, 2002; Joachims, 1998), recuperação de informação (Salton et. al, 1975) e análise de sentimentos (Pang & Lee, 2008).

Contudo, embora o BoW seja útil para muitas tarefas, suas limitações já são objeto de discussão há algum tempo. Uma das maiores discussões a esse respeito está na dificuldade de captação semântica entre os termos, já que o vetor é construído a partir do número de ocorrências da palavra, indiferentemente das relações que ela contrai com outros signos no corpus. O modelo Word2Vec (Mikolov et. al, 2013), por sua vez, difere do BoW justamente por conseguir captar as relações semânticas entre as palavras; e para isso,

aplica-se a lógica saussuriana de que o valor de um signo é sempre determinado pelo concurso que existe fora dele, além da hipótese distributiva de Harris e Firth, conforme já discutimos. Além disso, o Word2Vec, justamente por captar as relações semânticas, também é útil para identificar a probabilidade de ocorrência de termos.

Nessa perspectiva, os dois métodos principais na abordagem com o Word2Vec podem ser divididos em: Skip-Gram e CBOW. Ambos os métodos são usados para modelar a relação entre uma palavra e seu contexto. A diferença, contudo, está na direção da previsão: no Skip-Gram, o modelo recebe uma palavra-alvo e tenta prever as palavras ao redor; já no CBOW, o modelo recebe as palavras de contexto para prever a palavra-alvo.

Por exemplo, façamos uma previsão simplificada utilizando o método CBOW. Dado o enunciado “O gato caça ratos”, suponha-se que o verbo “caça” esteja ausente, de modo a termos a seguinte lacuna: “O gato _____ ratos”. Nesse caso, a palavra ausente “caça” constitui a palavra-alvo, dada por w_t . As demais palavras da sentença — “O”, “gato”, “ratos” — formam o contexto da predição, isto é, das palavras possíveis. Em seguida, deve-se determinar um “tamanho de janela” para que haja uma probabilidade mais precisa. No exemplo, vamos tomar como tamanho a janela = 1. Portanto, o modelo vai utilizar como entrada uma palavra do contexto que esteja antes e uma palavra que esteja depois, dado por: w_{t-1}, w_{t+1} , respectivamente “gato”, “rato”.

Imaginemos, então, que o enunciado analisado esteja contido dentro de um grande conjunto de dados. Nesse caso, quanto maior a quantidade de textos antes e depois da palavra-alvo, maior deve ser o tamanho da janela de contexto. Por isso, a função objetiva do CBOW pode ser expressa por $\max \frac{1}{T} \sum \log[P(w_t | w_{t-n}, \dots, w_{t-1}, \dots, w_{t+1}, \dots, w_{t+n})]$, buscando-se maximizar a probabilidade logarítmica da palavra-alvo dada as palavras de seu contexto, até um número n de janelas. Assim, o modelo bem treinado tenderá a atribuir maior probabilidade à palavra “caça”, por reconhecer sua adequação semântica e sintática em relação ao contexto apresentado. Com isso, é possível identificar, com base nos vetores de contexto, a palavra que melhor preenche a lacuna dentro da sentença dada.

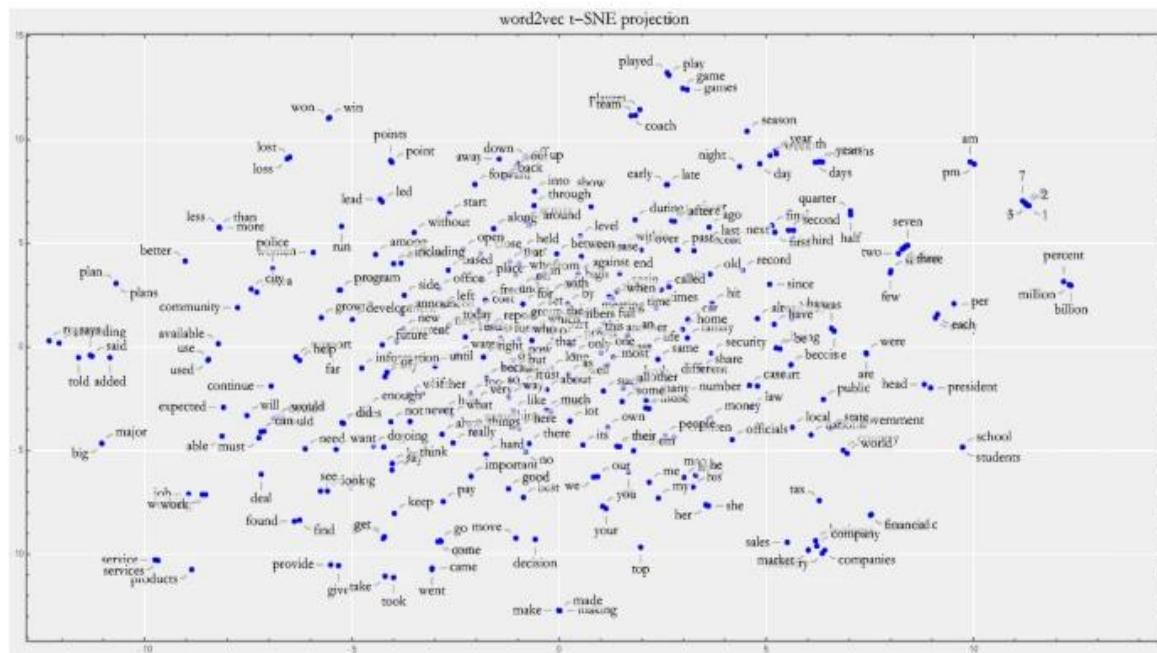
Nesse sentido, o fato mais relevante deste método para essa pesquisa é

a reflexão sobre a impossibilidade de que um signo possa se autodeterminar. Em outras palavras, para que “caça” seja identificado, é necessário que se analise a rede de signos que estão à sua volta, isto é, seus semelhantes, comparando-os com os signos que estão mais distantes dela, ou seja, os seus dessemelhantes. Nesse ponto, ao aplicarmos uma abordagem saussurianista, poderíamos dizer que “caça” preencheria a lacuna “O gato _____ ratos”, porque, em primeiro lugar, os valores de “caça” estão, nesse pequeno corpus, relacionados às palavras de contexto “gato” e “ratos”, contraindo, com esses signos, valores em *similia*.

Todavia, como sabemos, um computador não entende a língua tal como seres humanos, embora o mecanismo de produção de significados seja parecido numa perspectiva saussuriana, a saber, a ideia de que o significado está relacionado com os valores semelhantes e dissemelhantes do signos; isto é, suas relações externas. Para que uma máquina nos compreenda, seria necessário transformar cada uma das palavras do *corpus* (“o gato caça ratos”) em sequências numéricas, que são chamadas de “vetor” ou, em nosso caso, *embeddings de palavras*. Chama-se de “vetor” porque empresta-se o termo da álgebra linear, entendendo os vetores como um matriz coluna que indica direção e movimento num plano dimensional.

Nesse sentido, os dados de saída do Word2Vec poderiam, então, ser representados em um mapa vetorial, indicando as relações semânticas entre as palavras com base na proximidade ou distância entre elas, conforme vemos na figura a seguir:

Figura 12. Representação vetorial de um saco de palavras utilizando Word2Vec.



Fonte: Gastaldi, 2020, p.157.

A imagem em questão é uma visualização do Word2Vec gerada pelo método **t-SNE** (*t-Distributed Stochastic Neighbor Embedding*), um conjunto de algoritmos para mapear a rede de similaridade de vetores/embeddings. O objetivo do *framework* da Figura 12 é representar o nível de semelhança entre essas palavras, utilizando-se da hipótese distributiva para suportar a ideia de que palavras próximas, no espaço vetorial, indicam sentidos ou usos parecidos. Mais uma vez, essa proximidade entre os signos revela, também, os valores simile (cf. Saussure, 1910-1911, f. 27), de semelhança, e *dissimilia* (ibidem), dessemelhança. Quanto mais distante no espaço vetorial, maior o valor *dissimilia*; enquanto que quanto mais perto as palavras, maior o valor de *similia*.

Do ponto de vista de uma interpretação saussurianista, poderíamos dizer, então, este framework representa pode demonstrar computacional alguns aspectos fundamentais da noção de sistema linguístico relacional proposta por Saussure. No sistema linguístico saussuriano, como vimos, o valor de um signo é determinado por suas relações de similaridade e disparidade com os demais signos do sistema, sendo atravessados por um princípio relacional que admite relações internas e externas para o signo. Além disso, tal como se demonstra a

partir da metáfora do jogo de Xadrez (cf. Saussure, 2012 [1916]), cada palavra - cada *signo* - admite uma posição única no sistema, demarcando-lhe valores semelhantes e dessemelhantes. De forma análoga, o Word2Vec, por meio do t-SNE⁴⁷, pode traduzir essas relações diferenciais em proximidades vetoriais, onde palavras com valores semânticos semelhantes se agrupam e aquelas com valores distintos se afastam - numa clara alusão à hipótese distributiva e, em nossa leitura, à Teoria do Valor.

Essa representação computacional é, no mínimo, curiosa, pois permite visualizar e medir relações de valor linguístico de forma quantitativa e gráfica. Enquanto Saussure descreveu o valor como uma relação de oposição e semelhança em um sistema linguístico, aqui vemos um exemplo concreto no qual essas relações se materializam em dimensões reduzidas que capturam, com precisão matemática, as propriedades relacionais - valores - entre os signos, isto é, suas relações externas.

Portanto, essa abordagem parece conectar as ideias de Saussure à semântica distribucional moderna, permitindo que os valores semânticos, derivados de contextos de uso, sejam reinterpretados como valores relacionais ou não relacionais no espaço vetorial. Isso, além de evidenciar o conceito de valor de maneira aplicada, também possibilita ampliar o alcance do saussurianismo para além da linguística teórica.

Nessa perspectiva, para além do Word2Vec, o uso da semântica distribucional e, por consequência, o uso da linguística saussuriana por filiação, tem sido utilizado em grandes projetos de PLN ao redor do mundo atualmente. Conhecidos como projetos de *Digital Humanities*, eles têm chamado a atenção devido ao uso contemporâneo de técnicas de PLN para a construção de ferramentas baseadas em IA para se trabalhar com manuscritos históricos. Apoiando-nos no trabalho de Cosenza (2017), poderíamos afirmar diferentes benefícios dessas ferramentas para os estudos saussurianos contemporâneos, como também algumas limitações. Contudo, neste trabalho, não vamos nos ater a todas as ferramentas de PLN disponíveis para trabalhar com manuscritos de

⁴⁷ O t-SNE é uma técnica de redução de dimensionalidade desenvolvida por Laurens van der Maaten e Geoffrey Hinton em 2008. Ela é particularmente útil para visualizar dados de alta dimensionalidade, como embeddings de palavras, em duas ou três dimensões. O t-SNE preserva as relações de proximidade entre os dados, facilitando a identificação de agrupamentos naturais.

Saussure, mas queremos dar destaque a uma em especial, que utiliza a Semântica Distribucional como parte de seu desenvolvimento e arquitetura.

O artigo de Aljabout Sahar⁴⁸ e Gilles Falquet⁴⁹ (2018), intitulado *A Semantic Model for Historical Manuscripts*, apresenta um modelo semântico para a análise de manuscritos históricos, com foco nos manuscritos de Ferdinand de Saussure. Este modelo é parte de um projeto interdisciplinar que combina tecnologias semânticas e estudos humanísticos para lidar com desafios relacionados à compreensão, organização e publicação de manuscritos. A abordagem busca atender às necessidades específicas de especialistas saussurianos, incluindo categorização temática, enriquecimento semântico e organização temporal dos textos.

O projeto visa criar uma arquitetura semântica que facilite o estudo dos manuscritos de Saussure. Entre os principais objetivos estão:

1. Indexação semântica: associar palavras e conceitos encontrados nos manuscritos a recursos de conhecimento, como ontologias e tesouros.
2. Organização temporal: determinar cronologicamente os manuscritos, inferindo datas e organizando o fluxo de ideias de Saussure.
3. Publicação e visualização: prover uma interface para consulta e anotação de manuscritos, permitindo navegação intuitiva entre textos e conceitos relacionados.
4. Infraestrutura multi-conhecimento: construir uma base de dados com informações interconectadas, incluindo transcrições, terminologias e bibliografias

No caso desta dissertação, a técnica de indexação semântica muito nos interessa. Isso porque a indexação semântica é um processo de organização e anotação de conteúdos, como textos ou documentos, para associar palavras e frases a seus significados e contextos dentro de um domínio específico. Diferentemente da indexação tradicional, que é baseada em palavras-chave simples, a indexação semântica leva em conta o significado das palavras e suas

⁴⁸ Aljabout Sahar é professora na Université de Géneve e pesquisadora sênior na mesma instituição no Institute of Information Service Science. Sua formação e experiência se concentram nas áreas das ciências da computação, machine learning e inteligência artificial.

⁴⁹ Gilles Falquet é professor adjunto do Centro de Informática da Universidade de Genebra. Suas pesquisas se centram, sobretudo, em Processamento de Linguagem Natural e Machine Learning.

relações contextuais. Em vista disso, o principal aporte teórico desta estratégia de PLN é a própria semântica distribucional, abordagem referida, inclusive, pelos autores do estudo:

Assim, na implementação atual, o indexador calcula e armazena pontuações para as associações palavra-conceito, mantendo todos os significados cuja pontuação seja maior que um limite. Essas pontuações podem ser exibidas nas interfaces do usuário e usadas para classificar os resultados ou ajudar especialistas a decidir sobre os significados corretos. A pontuação de uma associação palavra-conceito depende de dois fatores: (I) A similaridade entre o contexto da palavra na transcrição e os "contextos de uso" fornecidos para o conceito na terminologia (veja a Figura 3). **Isso corresponde a avaliar a similaridade em termos de semântica distribucional (uma aproximação muito simplificada disso);** (II) Se o momento de escrita de um manuscrito é conhecido, os conceitos pertencentes a terminologias usadas pelo autor naquele período recebem uma pontuação mais alta⁵⁰ (Sahar; Falquet, 2018, p.9, grifo nosso, tradução nossa).

Nesse sentido, o uso da semântica distribucional neste projeto nos permite duas reflexões consideráveis. Primeiramente, embora seja uma teoria da significação desenvolvida a partir de trabalhos da década de 1950, a semântica distribucional continua a sustentar diversos trabalhos em PLN, especialmente no que diz respeito à modelagem e indexação semânticas. Em segundo lugar, e não menos importante, a aplicação da semântica distribucional em projetos de Digital Humanities relacionados a Saussure revela uma perspectiva quase metalinguística: utilizar fundamentos teóricos saussurianos como base para criar ferramentas destinadas ao próprio estudo do saussurianismo. Essa abordagem não é meramente incidental; como buscamos demonstrar ao longo deste trabalho, a semântica distribucional foi profundamente influenciada pelos ideais saussurianos de relação, sistema, valor e significação. Assim, o uso da SD nesses projetos não apenas reforça a relevância contemporânea de Saussure, mas também destaca o caráter cíclico

⁵⁰ No original: "Therefore, in the current implementation, the indexer computes and stores scores for the word - concept associations, keeping all the meanings whose score is higher than a threshold. These scores can then be shown in the user interfaces and used to sort the results or to help experts decide on the correct meanings. The score of a word - concept association depends on two factors: 1. The similarity between the word's context in the transcription and the "contexts of use" provided for the concept in the terminology (see Fig.3). This corresponds to evaluating the similarity in terms of distributional semantics (in fact a very rough approximation of it) 2. If the writing time of a manuscript is known, the concepts that belong to terminologies used by the author at that time are given a higher score."

e integrador de suas contribuições teóricas, que agora retornam como alicerce para novas tecnologias aplicadas ao estudo de sua obra.

5. Considerações finais

As reflexões desenvolvidas ao longo desta dissertação tiveram como ponto de partida a possível pertinência da linguística saussuriana para o Processamento de Linguagem Natural. Propusemos que a significação, na perspectiva saussuriana, oferece argumentos sólidos para sustentarmos a hipótese de que a semântica saussuriana pode residir na ideia de que a língua é um sistema de relações. Para chegar a essa formulação, organizamos nossa investigação em etapas que permitiram compreender como diferentes elementos teóricos da fortuna saussuriana se articulam para formar uma teoria da significação, cuja base epistemológica é a Teoria do Valor e os pressupostos que a circulam. Inicialmente, exploramos a ideia de *relação*, fundamento central na teoria saussuriana, que opera tanto na constituição interna dos signos quanto nas conexões externas entre eles. Foi a partir dessa noção que elucidamos como a produção de sentidos depende intrinsecamente das relações que os signos estabelecem entre si.

Na sequência, nos concentrarmos no que essas relações produzem: o sistema linguístico. Para Saussure, o sistema não é uma mera organização estática, mas um conjunto dinâmico de interações diferenciais que conferem valor aos signos. Avançamos, então, para a análise dos valores gerados por esse sistema. Os valores, como demonstrado, são resultados do jogo de semelhança (*similia*) e dessemelhança (*dissimilia*), compondo um campo relacional que delimita o significado de cada elemento sempre à margem dos outros signos. Finalmente, analisamos como o jogo de diferenças entre os valores culmina na significação, elemento central da língua, que emerge das relações diferenciais e nunca de forma isolada.

Já na segunda parte desta pesquisa, nos dedicamos a examinar o diálogo entre a linguística saussuriana e o campo do Processamento de Linguagem Natural (PLN) na prática. Buscamos, em primeiro lugar, investigar as possíveis filiações epistemológicas entre a Semântica Distribucional (SD), base teórica de muitos modelos computacionais contemporâneos, como o Word2Vec, e as

ideias de Saussure. Percebemos, então, que o cerne epistemológico da SD, conforme desenvolvido por Harris e Firth, mantém estreita conexão com o princípio relacional saussuriano. Essa conexão é especialmente visível na ideia de que a língua é um sistema de relações diferenciais, no qual os significados não são fixos, mas determinados pelas interações contextuais e pelas diferenças entre os signos.

Além disso, na parte 2, também realizamos algumas análises de embeddings de palavras gerados pelo modelo Word2Vec a fim de avaliar a real relação entre Saussure, a Semântica Distribucional e o PLN. Ao analisar *clusters* gerados pelo Word2Vec, constatamos o caráter relacional do significado em modelos vetoriais de linguagem, uma vez que a proximidade ou distância entre os vetores demonstrou algumas relações de similaridade e oposição semântica entre os signos. Nossos resultados, assim, sugerem que, assim como no sistema proposto por Saussure, os significados em modelos baseados na SD emergem de relações diferenciais entre os signos, reafirmando a possível aplicabilidade da linguística saussuriana em determinados projetos e técnicas de PLN cujo o mapeamento das relações entre as palavras seja relevante.

Adicionalmente, exploramos uma aplicação prática da Semântica Distribucional em um projeto desenvolvido por professores da Universidade de Genebra, que consiste num modelo semântico para trabalhar com manuscritos históricos. Esse exemplo reforçou como os conceitos saussurianos podem subsidiar avanços na tecnologia de processamento linguístico, ampliando sua utilidade para tarefas como análise semântica, busca textual e reconstrução de textos históricos.

Essa pesquisa, ao estabelecer pontes entre a teoria saussuriana e as tecnologias de PLN, pode contribuir para atualizar o legado do mestre de Genebra. Ao mesmo tempo, evidenciamos como suas noções de relação, valor e sistema fornecem bases teóricas robustas para interpretar o funcionamento de modelos computacionais de linguagem baseados na SD. Mais do que um resgate histórico, este trabalho buscou demonstrar a relevância interdisciplinar da linguística saussuriana, tanto para a compreensão dos fenômenos da língua quanto para o avanço de inteligências artificiais.

Por fim, admitimos que o pensamento de Saussure permanece atual e fértil para o entendimento de questões contemporâneas relacionadas à

linguagem e à tecnologia. Sua visão de língua como um sistema relacional e dinâmico não só contribui para uma compreensão mais profunda do funcionamento da natureza relacional da língua humana, como também oferece uma teoria da significação, baseada na Teoria do Valor, que pode orientar novas investigações no campo do PLN, permitindo caminhos cada vez mais transdisciplinares no futuro da semântica computacional e, por consequência, na forma com que máquinas entendem e compreendem a complexidade da significação das línguas humanas.

REFERÊNCIAS

AARSLEFF, H. Bréal, la Sémantique et Saussure. *Histoire, epistémologie, langage*, v. 3, nº 2, p. 115-134, 1981.

<https://doi.org/10.3406/hel.1981.1077>

ABUBAKAR, H. D; UMAR, M; BAKALE, M. A. Sentiment classification: review of text vectorization methods: Bag of Words, Tf-Idf, Word2vec and Doc2vec. *Sule Lamido University Journal of Science & Technology*, v. 4, n. 1&2, p. 27-33, jul. 2022. Disponível em: <https://doi.org/10.56471/slujst.v4i.266>. Acesso em: 17 mar. 2025.

<https://doi.org/10.56471/slujst.v4i.266>

ALJALBOUT, S.; FALQUET, G. A Semantic Model for Historical Manuscripts. 2018. Disponível em: <https://arxiv.org/abs/1802.00295>. Acesso em: 19 dez. 2024.

ALTMAN, C. No ano de 1894, uma carta de Saussure a Meillet. *Estudos Semióticos*, São Paulo, Brasil, v. 19, n. 3, p. 1-27, 2023. DOI: 10.11606/issn.1980-4016.esse.2023.214524. Disponível em: <https://www.revistas.usp.br/esse/article/view/214524>. Acesso em: 9 out. 2024.

<https://doi.org/10.11606/issn.1980-4016.esse.2023.214524>

ALTMAN, C. Saussure e o Memóire. "III Saussure in focus - Simpósio 3" [vídeo]. III Saussure In Focus, Universidade Federal de Uberlândia, 18 a 20 de setembro de 2024. Disponível em: <https://www.youtube.com/watch?v=be2sE8ZXq5M&t=2s>. Acesso em 25/09/2024.

ARAÚJO, D. P. Abordagens de processamento de linguagem natural e aprendizado profundo para classificação de atos administrativos de diário oficial. 2025. 75 f. Dissertação (Mestrado em Ciências da Computação) - Universidade Federal de Uberlândia, Uberlândia, 2025. DOI <http://doi.org/10.14393/ufu.di.2024.726>.

[https://doi.org/10.14393/ufu.di.2024.726](http://doi.org/10.14393/ufu.di.2024.726)

ARRIVÉ, M. Em busca de Ferdinand de Saussure. Trad. Marcos Marcionilo. São Paulo: Parábola Editorial, 2010

BRÉAL, M. Ensaio de semântica. São Paulo: Pontes/Educ, 1992 [1891].

BLOOMFIELD, L. Language. New York: Holt, Rinehart and Winston, 1961 [1933].

CAMARA JR., J. M. O estruturalismo. ALFA: Revista de Linguística, São Paulo, v. 11, 2001 [1968]. Disponível em: <https://periodicos.fclar.unesp.br/alfa/article/view/3298>. Acesso em: 19 maio. 2024.

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 out. 2024.

COELHO, M.P. "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 out. 2024.

COELHO, M.P. A noção de sistema na fundação da linguística moderna. 2015. 129 f. Dissertação (Mestrado em Estudos Linguísticos) - Universidade Federal de Uberlândia, Uberlândia, 2015. DOI: <https://doi.org/10.14393/ufu.di.2015.307>

COELHO, M.P. Ferdinand de Saussure: entre a língua e as línguas. 2019. 141 f. Tese (Doutorado em Estudos Linguísticos) - Universidade Federal de Uberlândia, Uberlândia, 2019. DOI <http://doi.org/10.14393/ufu.te.2019.2512>

COELHO, M. P.; SILVA E LIMA, T. R. Língua, linguagem e fala na "Teoria do Valor" de Ferdinand de Saussure. *Estudos Linguísticos* (São Paulo. 1978), [S. I.], v. 43, n. 01, p. 347-357, 2015. Disponível em: <https://revistas.gel.org.br/estudos-linguisticos/article/view/443>. Acesso em: 9 out. 2024.

CORREIA, T. T. S. O lugar do sentido em Saussure. 2020. 111 f. Dissertação (Mestrado em Linguística e Literatura) - Faculdade de Letras, Programa de Pós-Graduação em Linguística e Literatura, Universidade Federal de Alagoas, Maceió, 2020

COSENZA, G. Les projets de 'Digital Humanities' relatifs à l'œuvre de Ferdinand de Saussure. *Cahiers Ferdinand de Saussure*, n. 70, p. 25-39, 2017. Disponível em: <http://www.jstor.org/stable/45212068>. Acesso em: [data de acesso].

CRUZ, M. A. A filologia saussuriana: debates contemporâneos. Alfa, São José do Rio Preto, São Paulo, v. 53, n. 1, p. 107-126, 2009.

DALL'CORTIVO-LEBLER, C. Do sentido ao valor: relações teóricas entre a Semântica de Michel Bréal e o Estruturalismo de Ferdinand de Saussure. *Revista de Estudos da Linguagem*, Belo Horizonte, v. 25, nº4, p. 1965-1987, 2017.

<https://doi.org/10.17851/2237-2083.25.4.1965-1987>

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

DE MAURO, T. Ancora Saussure e la semantica. Cahiers Ferdinand de Saussure, v. 45, p. 101-109, 1991. Disponível em:
<http://www.jstor.org/stable/27758440>. Acessado em 25 de janeiro de 2025.

DONEDA, Danilo; ALMEIDA, Virgílio A. F. O que é a governança de algoritmos? In: BRUNO, Fernanda (Org.). Tecnopolíticas da vigilância: perspectivas da margem. São Paulo: Boitempo, 2018, p. 141-148.

DUCROT, O. Estruturalismo e Lingüística. São Paulo: Cultrix, 1968.

ENGLER, R. Rôle et place d'une sémantique dans une linguistique saussurienne. Cahiers Ferdinand de Saussure, n. 28, p. 35-52, 1973. Publicado por: Librairie Droz. Disponível em:
<https://www.jstor.org/stable/27758124>. Acessado em 25 de janeiro de 2025.

FREITAS, C. Linguística Computacional. Parábola: São Paulo, 2022.

FIRTH, J. R. Linguistics and the functional point of view. English Studies, v. 16, n. 1-6, p. 18-24, 1934. Disponível em:
<https://doi.org/10.1080/00138383408596618>. Acesso em: 4 nov. 2024.
<https://doi.org/10.1080/00138383408596618>

FIRTH, J. R. Personality and language in society. The Sociological Review, v. 42, n. 1, p. 37-52, 1950. Disponível em: <https://doi.org/10.1111/j.1467-954X.1950.tb02460.x>. Acesso em: 4 nov. 2024.
<https://doi.org/10.1111/j.1467-954X.1950.tb02460.x>

FIRTH, J. R. The semantics of linguistic science. Lingua, v. 1, p. 393-404, 1949. Disponível em: [https://doi.org/10.1016/0024-3841\(49\)90085-6](https://doi.org/10.1016/0024-3841(49)90085-6). Acesso em: 4 nov. 2024.
[https://doi.org/10.1016/0024-3841\(49\)90085-6](https://doi.org/10.1016/0024-3841(49)90085-6)

FIRTH, J. R. A synopsis of linguistic theory 1930-1955. Studies in Linguistic Analysis, Special Volume/Blackwell, 1957.

FLORES, V. do N. A Linguística Geral de Ferdinand de Saussure. São Paulo, SP: Editora Contexto, 2023.

GASTALDI, J. L. Why Can Computers Understand Natural Language? Philosophy & Technology, v. 34, p. 149-214, 2020. Disponível em:
<https://doi.org/10.1007/s13347-020-00393-9>. Acesso em: 4 nov. 2024.
<https://doi.org/10.1007/s13347-020-00393-9>

GIAMARUSTI, L. A noção saussuriana de sistema linguístico em modelos vetoriais de Linguagem. IN: VALOZ, J. T. da Silva (Org.). Saussure: termos, conceitos e noções. São Carlos: Pedro & João Editores, 2024a.

GIAMARUSTI, L. 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, 2024b. DOI: 10.5335/rdes.v20i3.16461. Disponível em:
<https://seer.upf.br/index.php/rd/article/view/16461>. Acesso em: 20 dez. 2024.
<https://doi.org/10.5335/rdes.v20i3.16461>

HARRIS, Z. S. Distributional Structure. WORD, v. 10, n. 2-3, p. 146-162, 1954. Disponível em: <https://doi.org/10.1080/00437956.1954.11659520>. Acesso em: 4 nov. 2024.
<https://doi.org/10.1080/00437956.1954.11659520>

JOACHIMS, T. Text Categorization with Support Vector Machines. In: European Conference on Machine Learning (ECML), 1998. DOI: 10.1007/BFb0026683.
<https://doi.org/10.1007/BFb0026683>

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

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
<https://doi.org/10.14393/ufu.di.2016.180>

MEILLET, A. Nécrologie de Michel Bréal. Annuaire 1916-17 de l'E.P.H.E., p. 38-42.
<https://doi.org/10.3406/ephe.1916.9296>

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 2024.

MONARD, M. C.; BARANAUSKAS, J. A. Conceitos de aprendizado de máquina. In: REZENDE, S. O. (org.). Sistemas Inteligentes - Fundamentos e Aplicações. Barueri: Editora Manole, 2003. p. 89-114.

NORMAND, C. Benveniste: linguistique saussurienne et signification. In: MONTAUT, Annie; NORMAND, Claudine (dir.). Lectures d'Émile Benveniste. Linx, n. 26, 1992, p. 49-75. DOI: <https://doi.org/10.3406/linx.1992.1237>. Acesso em: 19 dez. 2024.
<https://doi.org/10.3406/linx.1992.1237>

NORMAND, C. Saussure. Trad. Ana de Alencar e Marcelo Diniz. São Paulo: Estação Liberdade, 2009[2000].

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.

<https://doi.org/10.1561/1500000011>

PAUL, H. Princípios fundamentais da história da língua. Trad. Maria Luisa Schemann. Lisboa: Fundação Calouste Gulbenkian, 1966 [1880].

PINHEIRO, P. Linguagem e conhecimento em Platão: estudo sobre a correção dos nomes no Crátilo. In: Lumina, Juiz de Fora - Facom/UFJF, v. 6, n. 1/2, p. 31-56, jan./dez. 2003. ISSN 1516-0785.

PUECH, C. Palestra de Abertura: receber, herdar e transmitir [vídeo]. III Saussure In Focus, Universidade Federal de Uberlândia, 18 a 20 de setembro de 2024. Disponível em:
<https://www.youtube.com/watch?v=8I2J5swfh9s&t=518s>. Acesso em 25/09/2024.

QUIJANO, C. M. Le cours d'une vie: portrait diachronique de Ferdinand de Saussure. Nantes: Éditions Cécile Dafaut, 2008.

REES, Tobias. Non-Human Words: On GPT-3 as a Philosophical Laboratory. Dédalo, v. 151, n. 2, p. 168-182, 2022. Disponível em:
<https://www.jstor.org/stable/48662034>. Acesso em: 8 out. 2024.
https://doi.org/10.1162/daed_a_01908

SANTAELLA, L. A inteligência artificial é inteligente?. São Paulo: Almedina, 2023.

SALTON, G.; WONG, A.; YANG, C. A vector space model for automatic indexing. Communications of the ACM, v. 18, p. 613-620, 1975.
<https://doi.org/10.1145/361219.361220>

SAUSSURE, F. Cours de Linguistique Générale - Édition critique préparé par Tullio de Mauro. Paris: Payot, 1967.

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

SAUSSURE, F. Deuxième Cours de Linguistique Générale (1908-1909):

d'après les cahiers d'Albert Riedlinger et Charles Patois / Saussure's second course of lectures on general linguistics (1910-1911): from the notebooks of Albert Riedlinger and Charles Patois. French text edited by e English text edited by Eisuke Komatsu George Wolf. Pergamon Press, 1997.

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

SAUSSURE, F. Notes pour un livre sur la linguistique générale. In: Papiers Ferdinand de Saussure, 3951: Notes de Linguistique Générale. Bibliothèque de Genève, 1893-1984.

SAUSSURE, F. Première Cours de Linguistique Générale (1907): d'après les cahiers d'Albert Riedlinger / Saussure's first course of lectures on general linguistics (1907): from the notebooks of Albert Riedlinger. French text edited by Eisuke Komatsu e English text edited by George Wolf. Pergamon Press, 1996.

SAUSSURE, F. Troisième Cours de Linguistique Générale (1910-1911): d'après les cahiers d'Emile Constantin / Saussure's third course of lectures on general linguistics (1910-1911): from the notebooks of Emile Constantin. French text edited by Eisuke Komatsu e English text edited by Roy Harris. Pergamon Press, 1993.

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.

<https://doi.org/10.1145/505282.505283>

SORTICA, M. M. Saussure frente a seus contemporâneos: uma análise das questões relativas ao sentido nos primeiros capítulos do manuscrito De l'Essence Double du Langage. Revista Investigações, v. 34, n. 2, 2021. DOI: 10.51359/2175-294x.2021.251212. Disponível em: <https://periodicos.ufpe.br/revistas/index.php/INV/article/view/251212>. Acesso em: 17 mar. 2025.

<https://doi.org/10.51359/2175-294x.2021.251212>

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.

SILVEIRA, E. A aventura de Saussure. Campinas, SP: Editora Abralin, 2022. Disponível em: <https://editora.abralin.org/wp-content/uploads/2022/12/A-Aventura-de-Saussure.pdf>. Acessado em 28 de janeiro de 2025.

<https://doi.org/10.25189/9788568990285>

SILVEIRA, E. As marcas do movimento de Saussure na fundação da Linguística Moderna. Campinas, SP: Mercado de Letras, 2007.

SILVEIRA, E. A teoria do valor no Curso de Linguística Geral. In: Revista Letras & Letras, Uberlândia, Edufu, v. 25, n. 1, p. 39-54, 2009.

SOFIA, E. A noção de 'sistema' no Mémoire sur le système primitif des voyelles dans les langues indo-européennes de F. de Saussure (1879). Trad. Lygia Testa-Torelli. In: SOFIA, E. Le problème de la définition des entités linguistiques chez Ferdinand de Saussure [tese de doutorado]. Nanterre: Univ. Paris-Ouest, 2009. Disponível em: <https://cedoch.fflch.usp.br/cadernos>. Acessado em 25 de janeiro de 2025.

SOFIA, E. Le problème de la définition des entités linguistiques chez Ferdinand de Saussure. Cahiers Ferdinand de Saussure, n. 62, p. 203-214, 2009. Disponível em: <http://www.jstor.org/stable/41427849>. Acessado em 25 de janeiro de 2025

RUSSEL, S., NORVIG, P. Inteligência Artificial. Trad. MACEDO, R. C. S. Artificial Intelligence: A Modern Approach. 3^a edição. Elsevier Editora Ltda: Rio de Janeiro, 2013 [2010].

TURING, A. M. Computing Machinery and Intelligence. Mind, New Series, vol. 59, no. 236, 1950, pp. 433-460. Disponível em:
<https://courses.cs.umbc.edu/471/papers/turing.pdf>. Acessado em 28 de janeiro de 2025.
<https://doi.org/10.1093/mind/LIX.236.433>