



Processamento de Linguagem Natural

Introdução ao Processamento de Linguagens Naturais

Marlo Souza¹

¹Universidade Federal da Bahia - Brasil

22 de julho de 2024



Summary

Introdução

PLN, Linguagem e
Linguística

PLN em Português

Introdução

PLN, Linguagem e Linguística

PLN em Português



O que é PLN?

Introdução

3

PLN, Linguagem e
Linguística

PLN em Português

Qual o escopo?

Qual os objetivos?

Com o que se relaciona?



Aplicações

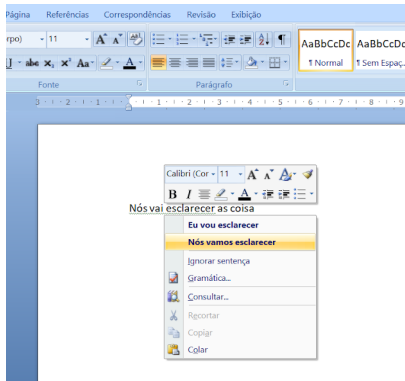
Introdução

4

PLN, Linguagem e Linguística

PLN em Português

25





Aplicações

Introdução

5

PLN, Linguagem e
Linguística

PLN em Português

The screenshot displays the Google Translate web interface. At the top, the Google logo is on the left, and a search bar contains the word 'translate'. Below the search bar are navigation links: 'Todas', 'Notícias', 'Imagens', 'Shopping', 'Vídeos', 'Mais', and 'Ferramentas'. A line of text indicates 'Aproximadamente 5.550.000.000 resultados (0,32 segundos)'. The main area shows two language selection dropdowns: 'Português' on the left and 'Inglês' on the right, separated by a double-headed arrow. Below these, the input text 'Eu adoro a aula de PLN' is shown with a close button (X). A suggestion box below it says 'Você quis dizer: Eu adoro a aula de PNL'. To the right, the translated text 'I love NLP class' is displayed in a light gray box. At the bottom of the interface, there are icons for voice input and output, and a footer with the links 'Abrir no Google Tradutor' and 'Feedback'.

25



Aplicações

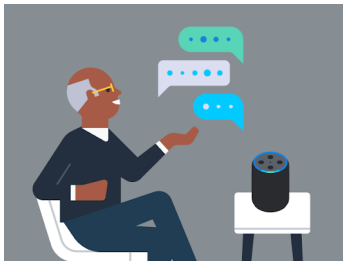
Introdução

6

PLN, Linguagem e
Linguística

PLN em Português

25





Linguagem Natural e Linguagem Artificial

Introdução

7

PLN, Linguagem e
Linguística

PLN em Português

A Computação emprega diversas linguagens artificiais para construir sistemas como:

Linguagens de programação: Python, C, R, Javascript

Linguagens de especificação: UML, BPMN, etc.

Linguagens de serialização: JSON, XML, HTML

25



Linguagem Natural e Linguagem Artificial

Introdução

7

PLN, Linguagem e
Linguística

PLN em Português

A Computação emprega diversas linguagens artificiais para construir sistemas como:

Linguagens de programação: Python, C, R, Javascript

Linguagens de especificação: UML, BPMN, etc.

Linguagens de serialização: JSON, XML, HTML

O que as diferencia de uma linguagem natural?

25



PLN, IA, Computação e Linguística

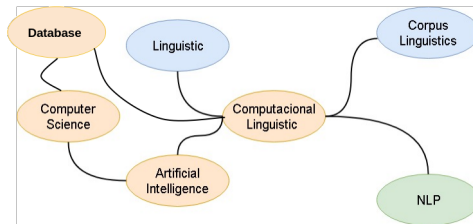
Introdução

8

PLN, Linguagem e
Linguística

PLN em Português

Entre a Computação e a Linguística...



25



Processamento de Linguagens Naturais (1/2)

Introdução

9

PLN, Linguagem e
Linguística

PLN em Português

“[...] a área de conhecimento que explora as relações entre linguística e informática, tornando possível a construção de sistemas com capacidade de reconhecer e produzir informação apresentada em linguagem natural”.

VIEIRA, Renata; LIMA, Vera LS. Lingüística computacional: princípios e aplicações. In: Anais do XXI Congresso da SBC. I Jornada de Atualização em Inteligência Artificial. sn, 2001. p. 47-86.



Processamento de Linguagens Naturais (1/2)

Introdução

10

PLN, Linguagem e
Linguística

PLN em Português

“O Processamento de Linguagem Natural (PLN) é um campo de pesquisa que tem como objetivo investigar e propor métodos e sistemas de processamento computacional da linguagem humana.”

CASELI, Helena de M.; NUNES, Maria das Graças V.; PAGANO, Adriana. O que é PLN?. In: Processamento de Linguagem Natural. BPLN, 2024.

25



Processamento de Linguagens Naturais

Introdução

11

PLN, Linguagem e
Linguística

PLN em Português

Divide-se normalmente em duas grandes áreas de acordo com os problemas estudados em cada uma delas:

25



Processamento de Linguagens Naturais

Introdução

11

PLN, Linguagem e
Linguística

PLN em Português

Divide-se normalmente em duas grandes áreas de acordo com os problemas estudados em cada uma delas:

Interpretação de Linguagem Natural

Geração de Linguagem Natural

25



Interpretação de Linguagem Natural (NLU)

Introdução

12

PLN, Linguagem e
Linguística

PLN em Português

NLU é o estudo de métodos que permitem processar dados de natureza linguística e acessar o conteúdo codificado através da linguagem, focam portanto na análise linguística e discursiva e na interpretação dos elementos linguísticos.

25



Interpretação de Linguagem Natural (NLU)

Introdução

12

PLN, Linguagem e
Linguística

PLN em Português

NLU é o estudo de métodos que permitem processar dados de natureza linguística e acessar o conteúdo codificado através da linguagem, focam portanto na análise linguística e discursiva e na interpretação dos elementos linguísticos. Estão associados a essa área, problemas como Extração de Informação, Detecção de Discurso Opinativo e Avaliativo, Classificação de Textos, etc.



Figura: Interpretação de Linguagem Natural

25



Geração de Linguagem Natural (NLG)

Introdução

13

PLN, Linguagem e
Linguística

PLN em Português

NLG é o estudo de técnicas que permitem organizar e apresentar informações através de linguagem natural, ou seja, objetivo gerar conteúdo linguístico.

25



Geração de Linguagem Natural (NLG)

Introdução

13

PLN, Linguagem e
Linguística

PLN em Português

NLG é o estudo de técnicas que permitem organizar e apresentar informações através de linguagem natural, ou seja, objetivo gerar conteúdo linguístico. Estão associados a essa área, problemas como Sumarização de Documentos, Síntese de Fala.



Figura: Geração de Linguagem Natural

25



Introdução

PLN, Linguagem e 14
Linguística

PLN em Português

PLN, Linguagem e Linguística



A linguagem e a análise linguística

Introdução

PLN, Linguagem e 15
Linguística

PLN em Português

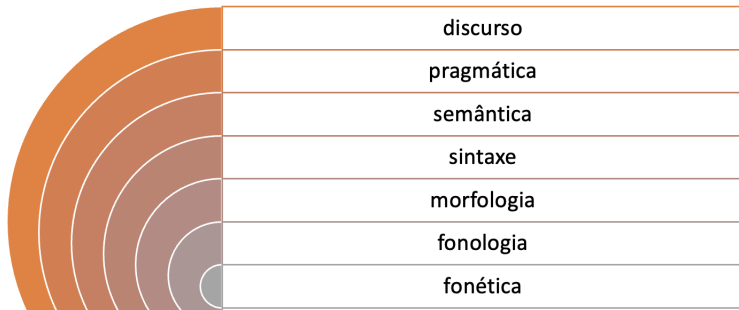


Figura: Os níveis de Análise Linguística

Fonte: (Caseli et al., 2024)



Linguística de Corpus e PLN

Introdução

PLN, Linguagem e 16
Linguística

PLN em Português

A Linguística de Corpus preocupa-se basicamente com o trabalho a partir de corpora eletrônicos que contenham amostras de linguagem natural. Essas amostras podem ser de diferentes fontes.

..há corpora de linguagem falada, corpora de linguagem escrita literária, corpora com textos de jornal, corpora compostos exclusivamente por falas de crianças em estágio de desenvolvimento lingüístico, etc. . ." (OTHERO, 2006)



Recursos, Ferramentas e Aplicações

Introdução

PLN, Linguagem e 17
Linguística

PLN em Português

Quando nos referimos à construção de sistemas de software para o processamento de linguagem natural, é importante separarmos três noções:

Ferramenta: auxiliam na construção de uma aplicação

Método: técnica computacional empregada para resolver um problema na construção de uma aplicação

Aplicação: fornecem um resultado ao usuário tendo uma entrada ou saída em linguagem natural

Ferramentas: são usadas como base para construção de aplicações



Recursos, Ferramentas e Aplicações

Introdução

PLN, Linguagem e 18
Linguística

PLN em Português

Conceito	Caracterização	Exemplos
Aplicações	processam uma entrada (input) em linguagem natural e a transformam produzindo um determinado resultado	tradutor automático corretor ortográfico ou gramatical
Recursos	são fontes de informação linguística para sistemas	léxicos corpora anotados
Método	forma através da a qual a aplicação foi projetada computacionalmente	Classificação Modelos de Linguagem
Ferramentas	auxiliam na construção de uma aplicação ou até de outras ferramentas	segmentadores textuais bibliotecas (NLTK, Pytorch)



Recursos, ferramentas e Aplicações

Introdução

PLN, Linguagem e 19
Linguística

PLN em Português

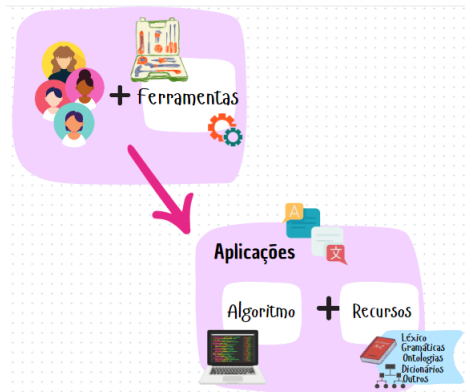


Figura: Recursos, Ferramentas e Aplicações

(Caseli et al, 2024)



Introdução

PLN, Linguagem e
Linguística

PLN em Português 20

PLN para a Língua Portuguesa



Por que PLN para Português

Introdução

PLN, Linguagem e
Linguística

PLN em Português 21

As línguas variam em dispositivos utilizados para construir sentido.

Em nível de morfologia

Em nível de estrutura



Por que PLN para Português

Introdução

PLN, Linguagem e
Linguística

PLN em Português 21

As línguas variam em dispositivos utilizados para construir sentido.

Em nível de morfologia

Em nível de estrutura

Métodos que funcionam bem para uma língua, podem não variar para outra.



Por que PLN para Português

Introdução

PLN, Linguagem e
Linguística

PLN em Português 21

As línguas variam em dispositivos utilizados para construir sentido.

Em nível de morfologia

Em nível de estrutura

Métodos que funcionam bem para uma língua, podem não variar para outra.

falta de recursos disponíveis

a solução explora uma particularidade da língua/amostra empregada

o fenômeno se manifesta de forma diferente em português



Recursos para PLN em português

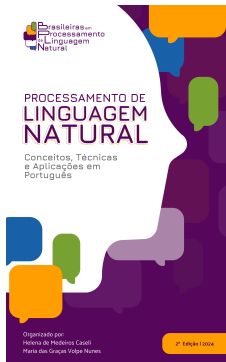
Introdução

PLN, Linguagem e
Linguística

PLN em Português 22

Livro: Caseli, H et al. (Org). Processamento de Linguagem Natural. 2a. ed.
BPLN, 2024

<https://brasileiraspln.com/livro-pln>





Recursos para PLN em português

Introdução

PLN, Linguagem e
Linguística

PLN em Português 23

Lingueca: conjunto de corpora, recursos linguísticos computacionais, livros e relatórios técnicos - <https://www.lingueca.pt/>

HuggingFace-Hub: modelos e datasets disponíveis em diversas línguas - <https://huggingface.co/>

NILC: recursos, literatura e ferramentas - <https://sites.google.com/view/nilc-usp/>



Fontes para pesquisar

Introdução

PLN, Linguagem e
Linguística

PLN em Português 24

25

EMNLP, ACL, NAACL, COLING

Natural Language Engineering, Computational Linguistics

STIL, PROPOR

Linguagmática (em português ou espanhol)



Contatos

Introdução

PLN, Linguagem e
Linguística

PLN em Português 25



Marlo Souza

Email: msouza1@ufba.com

Sala: 213 PAF II