

# UNIDADE DE APRENDIZAGEM:

## Aplicações e desafios do processamento de linguagem natural

### Apresentação

O processamento de linguagem natural (PLN) é um ramo da inteligência artificial que ajuda os computadores a entender, interpretar e manipular a linguagem humana. Muitas organizações usam técnicas de PLN para otimizar o suporte ao cliente, melhorar a eficiência da análise de texto, encontrando facilmente as informações necessárias e aprimorando

o monitoramento de mídia social. Os bancos, por exemplo, podem implementar algoritmos de PNL para otimizar o suporte ao cliente.

Já uma grande marca de produtos de consumo pode combinar o processamento de linguagem natural e a análise semântica para melhorar suas estratégias de monitoramento de mídia social para verificar, por exemplo, os produtos mais procurados. Nesse contexto, é possível perceber que essa área resolve um conjunto enorme de problemas e que só tende a crescer.

Nesta Unidade de Aprendizagem, você identificará as principais características que tornam o processamento de linguagem natural adequado para resolver um problema, compreenderá os desafios a serem tratados pelas técnicas de PLN e também poderá verificar se o processamento de linguagem natural é aplicado de forma adequada em problemas reais.

Bons estudos.

**Ao final desta Unidade de Aprendizagem, você deve apresentar os seguintes aprendizados:**

- Identificar as principais características que tornam o processamento de linguagem natural (PLN) adequado para resolver um problema.
- Listar os desafios a serem tratados pelas técnicas de PLN.
- Avaliar se o PLN é aplicado de forma adequada em problemas reais.

# Desafio

Existem várias aplicações do dia a dia que podem utilizar conceitos de PLN. Assistentes virtuais e *chatbots* são exemplos de aplicações de *software* que ajudam a resolver os problemas das pessoas de maneira rápida e eficiente. E, dependendo do problema, pode ser resolvido em poucos minutos por meio do atendimento.

Imagine que você acabou de ser contratado como analista de testes de *software* em uma empresa.



Você foi alocado em um projeto de desenvolvimento de um *chatbot*. Seu gerente pediu para você estudar um pouco sobre *chatbots* para melhor conhecer aplicativos do tipo, e depois voltar para conversar com ele, respondendo algumas questões sobre a interação com esse tipo de aplicação. Você está livre para escolher o *chatbot* ou assistente virtual que deseja conversar.

Imagine que você fez os seguintes testes:

- Conversou sobre assuntos do domínio da aplicação. Por exemplo: se for um assistente virtual de um banco, digitou conteúdo relacionado ao domínio de bancos, como extrato da conta, saldo bancário, transferências, etc.
- Escreveu mensagens aleatórias e texto sem sentido no *chat*, como a seguinte mensagem: "o-ljd-kbmfhysddrew".



De acordo com os dados fornecidos, você deve relatar ao seu chefe:

a) Após conversar com *chatbots* ou assistentes virtuais, como você acha que essas aplicações conseguem responder de forma satisfatória? Qual o processo envolvido? Justifique sua resposta.

b) Por que ao digitar mensagens aleatórias ou texto sem sentido as aplicações não conseguiram responder de forma satisfatória? Justifique sua resposta.

# Infográfico

---

A área de processamento de linguagem natural é uma das mais promissoras atualmente. Há muitas pesquisas em torno desse tema, pois essa área de conhecimento ajuda as empresas a melhorarem a qualidade de seus produtos e serviços. Existem diversos domínios diferentes que utilizam a PLN para a resolução de problemas.

A automação pode ajudar a transformar rapidamente seus negócios. Quando você aprimora a navegação de um *site*, facilita o uso de produtos com o suporte de *chatbots* ou desenvolve serviços analisando comentários, sua empresa tende a crescer. Nesse contexto, é extremamente importante que você entenda as diversas áreas nas quais a PLN consegue ajudar na resolução de problemas.

No Infográfico a seguir, você vai ver e entender as possíveis aplicações de PLN e os seus diferentes domínios.

# APLICAÇÕES DO PROCESSAMENTO DE LINGUAGEM NATURAL E SEUS DIFERENTES DOMÍNIOS

Os conceitos do processamento de linguagem natural (PLN) podem ser aplicados em diversas áreas e domínios. E quando aplicar o processamento de linguagem natural para resolver um problema?

A seguir, algumas aplicações e características que tornam o uso do PLN adequado para resolver um problema.

## NECESSIDADE DE ENTENDER O SER HUMANO



Se você, em algum momento, estiver trabalhando em alguma aplicação que precise entender o que uma pessoa for falar ou escrever, utilizar o processamento de linguagem natural pode ser adequado para essa situação.

**Exemplos:** ar-condicionado, assistentes virtuais, aparelhos em geral que precisam entender comandos de voz.

## NECESSIDADE DE RESPOSTA POR PARTE DAS MÁQUINAS



Da mesma maneira que as máquinas podem nos entender, elas também podem nos fornecer respostas. Ou seja, a partir de algum comando de voz, é possível dar entrada a aplicação informando que ela precisa fazer alguma tarefa.

**Exemplos:** chatbot, motores de busca, assistentes virtuais.

## NECESSIDADE DE PROCESSAR GRANDE QUANTIDADE DE DADOS PARA FAZER PREVISÃO



Algumas aplicações necessitam processar uma grande quantidade de informação para fazer alguma previsão ou trazer uma síntese dos dados ou fornecer alguma sugestão.

**Exemplos:** aplicações de monitoramento de redes sociais, nas quais são analisados posts e comentários; preenchimento automático de formulários.

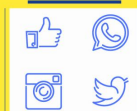
Cada uma dessas aplicações pertencem a áreas e a domínios diferentes, mas com um objetivo em comum: **facilitar a vida das pessoas**.

A seguir, são apresentados os diferentes domínios:

### NEGÓCIOS



### MÍDIAS SOCIAIS



### DISPOSITIVOS INTELIGENTES



É possível perceber que a PLN é importante em contextos de áreas diferentes. Por exemplo: na área de **negócios**, podem-se destacar *chatbots* e assistentes virtuais para melhorar o atendimento em uma empresa.

Em **mídias sociais**, é importante ter uma síntese de dados das redes sociais, pois ajuda a identificar a opinião das pessoas sobre determinado assunto. **Dispositivos inteligentes** são uma tendência atual e para o futuro, como o ar-condicionado que funciona por meio de comando de voz.



Aponte a câmera para o código e acesse o link do conteúdo ou clique no código para acessar.

# Conteúdo do Livro

---

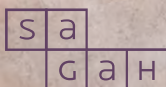
O processamento de linguagem natural é uma subárea da inteligência artificial, a qual está preocupada com o entendimento das máquinas a respeito da linguagem natural conversada diariamente pelos seres humanos. Essa é uma das áreas que mais cresce atualmente, pois envolve um conjunto de aplicações cotidianas. Lembre-se da última vez que você utilizou um serviço de tradução de idiomas, um buscador de *sites*, um *site* de *e-commerce* ou um serviço de *streaming* de filmes e séries. Todas essas aplicações utilizam a PLN de alguma maneira.

No capítulo Aplicações e desafios do processamento de linguagem natural, da obra *Processamento de linguagem natural*, você vai identificar as principais características que tornam o processamento de linguagem natural adequado para resolver um problema, compreender os desafios a serem tratados pelas técnicas de PLN e vai avaliar se o processamento de linguagem natural é aplicado de forma adequada em problemas reais.

Boa leitura.

# PROCESSAMENTO DE LINGUAGEM NATURAL

Júlio Serafim Martins



SOLUÇÕES  
EDUCACIONAIS  
INTEGRADAS



# Aplicações e desafios do processamento de linguagem natural

## Objetivos de aprendizagem

Ao final deste texto, você deve apresentar os seguintes aprendizados:

- Identificar as principais características que tornam o processamento de linguagem natural (PLN) adequado para resolver um problema.
- Listar os desafios a serem tratados pelas técnicas de PLN.
- Avaliar se o PLN é aplicado de forma adequada em problemas reais.

## Introdução

Cohen e Demner-Fushman (2014) afirmam que o processamento de linguagem natural (PLN) representa a habilidade da máquina para processar o que foi dito, estruturar a informação recebida, determinar a resposta necessária e responder de modo que o ser humano consiga entender, ou seja, são várias aplicações do PLN com o objetivo de melhorar a vida do ser humano de maneira geral. Essa área tem tido uma enorme evolução nos últimos anos, cujos velhos e novos desafios as pesquisas têm buscado resolver.

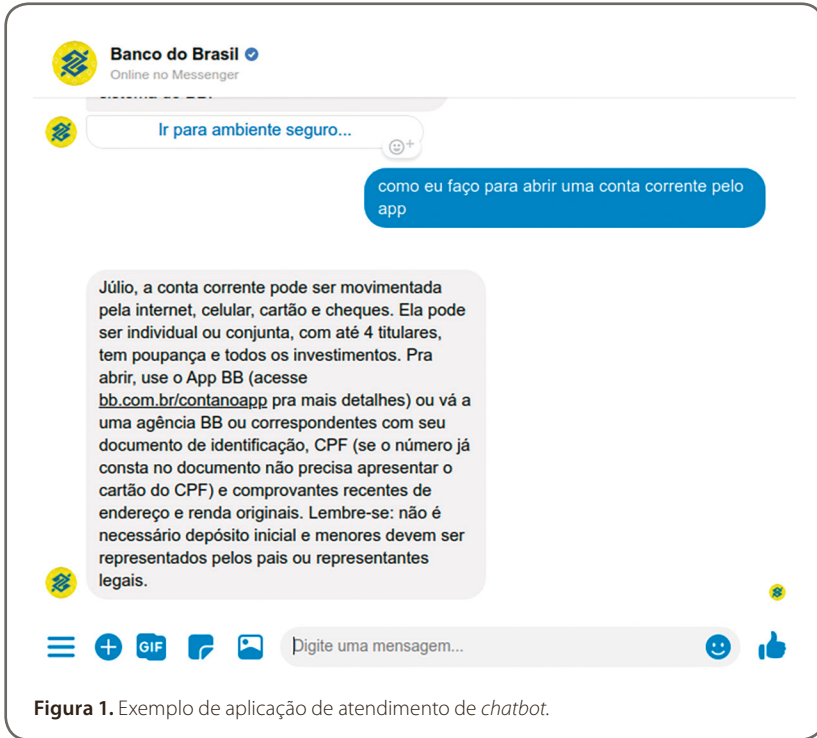
Neste capítulo, você verá as características que tornam o PLN adequado para resoluções de problemas, entenderá os desafios a serem tratados pelas técnicas de PLN e conhecerá a avaliação da adequada aplicação do PLN em problemas reais.



## 1 Características que tornam o processamento de linguagem natural adequado para resolução de problemas

Segundo Cohen e Demner-Fushman (2014), o processamento de linguagem natural (PLN) constitui o estudo de programas de computador que tomam a linguagem natural ou humana como entrada. As aplicações no contexto do PLN podem abordar tarefas que variam de um processamento de baixo nível, como atribuir partes do discurso a palavras, a tarefas de alto nível, como responder a perguntas.

PLN representa a tecnologia utilizada para que computadores entendam a linguagem do ser humano, contudo não é uma tarefa fácil ensinar as máquinas a entenderem como nós nos comunicamos. Então, um dos objetivos atuais dessa área é ler, decifrar, entender e dar sentido à linguagem humana de maneira geral. Tudo que nós expressamos, pela voz ou pela escrita, carrega consigo um conjunto enorme de informações; o assunto escolhido, o tom e a seleção das palavras constituem fatores que podem ser interpretados para definir o sentido. Para ilustrar essa situação, pense na última vez que você utilizou uma aplicação de *chatbot*, que, conforme Brandtzaeg e Følstad (2017), são *software* capazes de conduzir uma conversa por texto projetados para desenvolver uma conversa da maneira mais semelhante possível a de um ser humano. Por exemplo: imagine que Júlio está com problemas no banco e quer entrar em contato com este pela internet, atendimento inicial que é feito, em geral, por *chatbot*. Então, Júlio digita a frase “Eu não reconheço uma cobrança na minha fatura do cartão!”; o *chatbot* “entenderá” que João quer falar sobre o seu cartão e, assim, fornece um conjunto de opções relacionadas ao cartão. Contudo, nenhuma delas ajuda a resolver o problema de Júlio, que, então, começa a escrever muito texto com palavras agressivas e exclamações e o manda para o *chatbot*. Nesse momento, o PLN pode ajudar o *chatbot* a “entender” que Júlio está impaciente pela maneira como escreve o texto, redirecionando-o ou sugerindo a ele a opção de atendimento humano. Essa é uma situação que mostra o quanto essa área se preocupa em analisar todo tipo de informação que traga algum valor para o entendimento da máquina, que, no caso, era o atendente *chatbot*, conforme o exemplo da Figura 1.



**Figura 1.** Exemplo de aplicação de atendimento de *chatbot*.

A Figura 1 mostra o atendimento *on-line* do *chatbot* do Banco do Brasil, em que, a partir de determinado texto, consegue-se fornecer uma resposta para tentar resolver o problema. Por exemplo, foi digitada a frase “como faço para abrir uma conta corrente pelo app”, a partir da qual o *chatbot* conseguiu “entender” a solicitação e enviou a resposta mais precisa possível para solucionar o problema.

Dados gerados de conversas, declarações e texto bruto são exemplos de dados não estruturados (GHAREHCHOPOGH; KHALIFEHLOU, 2011), que representam a maioria de dados disponíveis atualmente no mundo, difíceis de lidar e manipular, um quadro que tem mudado graças ao avanço de disciplinas como aprendizado de máquina, um método de análise de dados que explora a construção de algoritmos que aprendem com os erros e, a partir disso, podem fazer previsões sobre os dados. Atualmente, além de tentar interpretar um texto pela análise de suas palavras-chave, é possível “entender” o significado delas, o que possibilita até mesmo o entendimento de figuras de linguagem, como a ironia.

Assim, podemos entender que o PLN é uma área que enfatiza a interação entre ciência de dados, máquinas e linguagem humana, a qual tem crescido em razão das melhorias no acesso aos dados e do aumento do poder computacional, possibilitando a obtenção de resultados expressivos por profissionais de diversas áreas, como da saúde, mídia, financeira, mercado de vendas, etc.

Diante disso, existem algumas características, como as listadas a seguir, que tornam o PLN adequado para resolver um conjunto de problemas.

- Necessidade para entender o que um ser humano deseja comunicar, pela **fala** ou **escrita**.
- A máquina deve “responder” da forma mais precisa possível levando em consideração o que foi “entendido” na interação com o ser humano.
- Processamento e análise de uma grande quantidade de dados para fazer previsões ou até mesmo fornecer sugestões de acordo com a situação e o domínio.

Essas características são “pistas” da possibilidade de utilizar o PLN para resolver os mais diversos problemas, facilitando a vida das pessoas na era digital. Porém, você verá que existem alguns desafios na área a ser explorados e resolvidos.

## 2 Desafios nas técnicas de processamento de linguagem natural

Como visto até agora, o PLN é uma área fantástica, em razão de suas várias aplicações efetivas para ajudar no nosso dia a dia. A PLN funciona de maneira parecida com o ser humano, já que, quando falamos com outra pessoa, a maioria das frases em um contexto é entendida por ambos, quando se interpreta facilmente a conversa. No entanto, em alguns momentos um dos participantes da conversa pode “falhar” ao explicar uma ideia e, consequentemente, seu interlocutor “falhar” em entender o contexto da conversa por vários motivos. Similarmente, máquinas podem falhar para compreender o contexto de algumas frases ou textos.

Para humanos, o aprendizado acontece nos primeiros passos da infância e de maneira consistente, em que as crianças aprendem a interagir com dados não estruturados e os processam em informação. Depois de reunirem as informações, os seres humanos passam a analisá-las na tentativa de entender suas implicações a partir de uma situação ou um problema. Por exemplo,

no início da vida, as pessoas não fazem ideia de que o fogo pode machucar, mas, à medida que crescem e têm mais informações sobre o fogo, podem notar o perigo associado a ele, ou seja, depois de entender as implicações, a informação pode ser usada para resolver um conjunto de problemas ou situações da vida.

Humanos interagem por múltiplos cenários para que, consciente ou inconscientemente, consigam simular se uma solução terá sucesso ou falhará. Depois de lidar com dados não estruturados, informações e conhecimentos, os humanos conseguem enfim adquirir “sabedoria”. Por exemplo, durante a evolução da humanidade, as pessoas perceberam que, para chegar a determinado destino de carro, devem colocar uma quantidade de gasolina suficiente para isso, uma sabedoria adquirida com o decorrer do tempo a partir da falha anterior de outras pessoas. Em outras palavras, se o tanque do carro estiver com a quantidade suficiente de gasolina, terá mais chance de sucesso ao realizar a atividade “chegar ao destino”; caso contrário, essa atividade provavelmente falhará.

Máquinas podem aprender por um método similar, traduzindo textos não estruturados em termos com significado, o que possibilita identificar a conexão entre os termos e, finalmente, **compreender** o contexto. A área de PLN teve muitos avanços nas últimas décadas, porém ainda existem desafios a superar, como descrito a seguir.

## Problemas de segmentação

Entender palavra por palavra até compreender o contexto geral em determinados casos pode não ser uma tarefa complexa, mas existem situações em que se torna muito difícil decompor e “quebrar” as frases para que a máquina consiga entender. Assim, uma aplicação pode ser inteligente suficiente para separar parágrafos em unidades de frase — pensando no texto “Mario Kart é um dos jogos mais antigos do console Super Nintendo. Esse jogo consiste em simular corridas de kart com personagens infantis”, como é feita a sua segmentação? Ele é quebrado em sentenças da seguinte maneira:

1. “Mario Kart é um dos jogos mais antigos do console Super Nintendo”
2. “Esse jogo consiste em simular corridas de kart com personagens infantis”

No entanto, dados altamente complexos nem sempre estão disponíveis de modo “fácil” para a correta decomposição das frases, como no caso de dados na forma de tabelas, gráficos, abreviações e notações.

## Análise de contexto

Trata-se de uma das tarefas mais importantes e desafiadoras no contexto do PLN. Por exemplo, considere as seguintes frases:

1. “Amanhã vamos colher as maçãs no pomar”
2. “Amanhã vamos comer as maçãs do pomar com colher”

Os contextos dessas frases são muito diferentes, e existem inúmeros métodos para ajudar a treinar a máquina a entendê-los. Alguns dos mais populares usam gráficos de conhecimento personalizados, em que as possibilidades ocorreriam com base em cálculos estatísticos. Ao analisar um novo documento, a máquina consultaria o gráfico para determinar o contexto.

Com a análise de contexto, outro grande desafio consiste em construir o gráfico de conhecimento da **linguagem específica de um domínio**. Na prática, gráficos não podem ser universalmente utilizados. Nos exemplos dados, a palavra “colher” no contexto da primeira frase sugere que significa pegar as maçãs no pomar, enquanto, na segunda, corresponde ao objeto utilizado para comer, ou seja, duas frases com contextos totalmente diferentes em domínios distintos podem “confundir” a máquina se ela se basear apenas em gráficos de conhecimento. Por isso, é importante aprimorar os métodos empregando uma abordagem estatística para inferir o contexto e o domínio corretos.

## Extração do significado semântico

Outro desafio do PLN, já que a análise linguística de termos em um vocabulário pode não ser suficiente para a máquina aplicar corretamente o conhecimento adquirido. Para ter sucesso no aprendizado, a máquina deve entender a semântica de cada termo dentro de um contexto de documentos. Por exemplo, considere as duas frases:

1. “O gato come o inseto”
2. “O inseto come a comida do gato”

As duas frases têm o contexto de comida de gato e de inseto e também apresentam algumas palavras idênticas, contudo a informação resultada delas é totalmente diferente em razão de sua semântica distinta, ou seja, elas têm significados diferentes apesar de apresentarem praticamente as mesmas palavras! Assim, é possível imaginar a dificuldade de uma máquina em entender

esse tipo de situação, o que leva à necessidade de dispor de uma combinação entre metodologias linguísticas e semânticas que possibilitem que a máquina consiga entender verdadeiramente o significado dos textos.

Choueika (2014) afirma que a **ambiguidade** é, sem dúvida, um dos maiores desafios do PLN. Uma palavra é considerada ambígua se tem dois ou mais significados diferentes que pertencem a campos semânticos distintos. Por exemplo, em “Ela sentou na cadeira e quebrou o braço”, há ambiguidade porque não se pode ter certeza de seu sentido correto: tanto o braço da cadeira quanto o da pessoa podem ter quebrado.

Segundo Gomaa e Fahmy (2013), a similaridade textual pode ser aplicada em diversas situações, como classificação de texto, clusterização, sumarização de texto, etc., contexto no qual no PLN os **erros ortográficos** são um grande desafio. Por exemplo, imagine uma busca no Google digitando “orrída de cavalos”; muito provavelmente, a máquina sugerirá a correção da frase para “corrida de cavalos”, já que o termo “orrída” não existe, embora seja muito “parecido” com “corrida”, tornando-se possível assumir que “corrida” seria a palavra correta.

## Neologismos

Outro desafio no PLN, o neologismo corresponde à formação de novos termos sobre um novo conceito. Por exemplo, a frase “Eu fui para o lugar combinado, mas ela me deu um bolo” (não apareceu no local combinado): imagine fazer a máquina entender expressões como esta ou outras que vão surgindo? Por isso, trata-se de uma tarefa muito complexa “ensinar” neologismos para as máquinas.

## Figuras de linguagem e o uso de abreviações e acrônimos

Pensando, por exemplo, na **ironia**, uma das figuras de linguagem mais conhecidas: como uma máquina conseguiria identificá-la? Para essa tarefa, é preciso também **analisar o contexto**, outro desafio na PLN. Humanos podem identificar ironia pela expressão facial, o que, para as máquinas, representa uma atividade complexa, o que torna as figuras de linguagem um desafio importante na área de PLN.

E o uso de abreviações, cada vez mais comum com a evolução da internet, com expressões como “vc” (você), “td bem” (tudo bem), “blz” (beleza), também dificulta o processo de “leitura” e “entendimento” da máquina, além do emprego de acrônimos, como “SQN” (só que não).

Segundo Cruz, Rocha e Cardoso (2020), a *coreference resolution*, uma tarefa da PLN que identifica todas as expressões que apontam para uma mesma entidade do mundo real, também constitui um dos grandes desafios enfrentados no entendimento das máquinas da linguagem humana. Por exemplo, imagine que uma pessoa disse a seguinte frase: “Eu votei no Zezinho porque ele tem uma visão de governo parecida com a minha”. Como fazer para identificar entidades do mundo real nesse caso? As palavras “eu” e “minha” se referem à pessoa que está dizendo a frase, ou seja, uma entidade do mundo real. Já as palavras “Zezinho” e “ele” se correspondem ao candidato Zezinho, outra entidade do mundo real. Este é mais um desafio na PLN porque nem sempre será fácil identificar as entidades nas frases.

## Uso de negações e especulação

Segundo Zou, Zhou e Zhu (2014), especulação e negação são comuns no PLN, utilizados em muitas aplicações de biomedicina. A negação se refere a uma categoria gramatical que compreende a negação de uma proposição e especulação é a categoria gramatical que denota certo grau de confiança e perspectiva. Por exemplo, veja as seguintes frases:

1. “O médico alertou que fumar **pode** prejudicar os pulmões”
2. “Ela **não** gosta de jogar tênis, mas gosta de praticar natação”

A frase 1 “especula”, por meio da palavra “pode”, que fumar pode prejudicar os pulmões. E, na frase 2, a palavra “não” nega que a pessoa gosta de jogar tênis. Esse tipo de ferramenta ajuda a identificar relatórios médicos para saber, por exemplo, se não existe ou se determinado tratamento é especulado para alguma pessoa. O grande desafio é conseguir identificar de maneira precisa a negação e a especulação nos textos, tarefa em que ainda residem muitos erros (QIAN *et al.*, 2016).

É importante dizer que a PLN ainda representa um desafio na **língua portuguesa**, em virtude dos recursos limitados para trabalhar com essa língua e de a maioria das ferramentas e estudos utilizar a **língua inglesa** como base. Porém, ferramentas como a WordNet.Br, um banco de dados léxicos que ajuda a trabalhar com PLN na língua portuguesa, têm surgido para mudar esse panorama.

A maioria dos desafios apresentados resultam da complexidade dos dados, da definição do contexto e do entendimento do significado semântico de palavras em frases. Apesar de ter evoluído bastante nos últimos anos, o PLN ainda é uma área com um escopo vasto de oportunidades para engenheiros, pesquisadores e profissionais da indústria para lidar com muitos desafios abertos na implementação de aplicações que utilizam a PLN. Com o devido foco na abordagem dos desafios, as empresas podem construir e melhorar dicionários e trabalhar em fatores capazes de diminuir a lacuna entre o vocabulário utilizado nos mais diferentes domínios.

### **3 Avaliação da adequada aplicação do processamento de linguagem natural em problemas reais**

O objetivo do PLN consiste em especificar uma linguagem de compreensão e promover uma teoria que esteja em um nível de detalhamento em que uma pessoa consegue escrever um programa de computador capaz de entender e produzir linguagem natural.

Boa parte do escopo do PLN envolve mensurar o nível de similaridade entre dois fragmentos de texto. Imagine que um usuário pronuncie “Blasil”, mas que, na verdade, queria falar outra palavra parecida com a palavra dita. Como a palavra “Brasil” tem apenas uma única letra diferente em relação a palavra emitida, isso a torna a uma potencial candidata para a correção (JURAFSKY; MARTIN, 2019).

A pesquisa que tem sido feita sobre PLN abrange várias áreas, como a da saúde, de entretenimento e, principalmente, de negócios. Para isso, a máquina interpreta elementos importantes das frases na linguagem humana para retornar respostas adequadas. Em outras palavras, o PLN representa uma abordagem automática para lidar com a linguagem humana, podendo ajudar em **muitas tarefas e na resolução de problemas básicos diários**. Por exemplo, uma de suas aplicações mais comuns diariamente reside no buscador Google para fazer pesquisas (Figura 2).



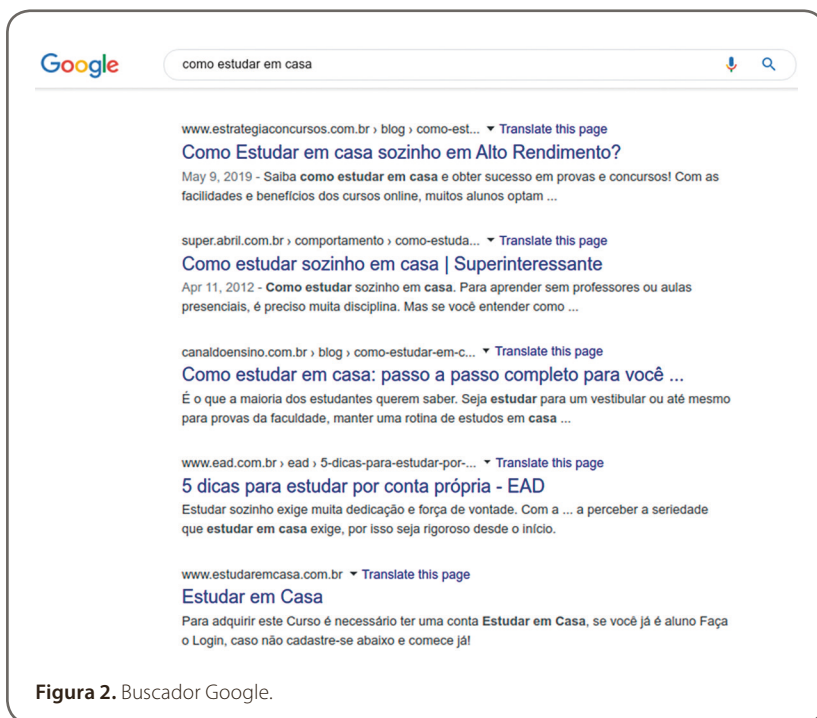
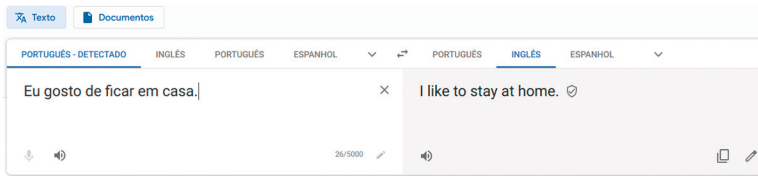


Figura 2. Buscador Google.

Quando você digita um texto na caixa de pesquisa e clica no botão de pesquisa, o PLN pode ser útil para “entender” o que você está querendo pesquisar e, a partir desse entendimento, apresentar resultados satisfatórios para a pesquisa. Por exemplo, foi digitado o seguinte texto para pesquisa: “como estudar em casa”; ao clicar no botão de pesquisar, o buscador Google retornou os principais resultados para a busca. Isso ilustra uma das principais áreas que a PLN busca resolver: a *information retrieval* (“recuperação de informação”), recuperando a informação a partir de um texto de pesquisa.

Você já pode ter precisado traduzir um trecho escrito em outro idioma para o seu idioma nativo, ou vice-versa, como do inglês para a língua portuguesa. Atualmente, não é mais necessário utilizar dicionários de inglês manualmente para conseguir realizar essa atividade. A Figura 3 mostra o Google Tradutor, um exemplo de aplicação para esse mesmo fim.



**Figura 3.** Aplicação do Google Tradutor.

Mais uma vez, como o PLN pode ajudar nessa situação? Por exemplo, imagine que você quer traduzir uma frase em inglês para o português. Nesse caso, a aplicação Google Tradutor deve “compreender” o significado das palavras no contexto da frase e, a partir disso, sugerir a tradução mais próxima da frase na língua portuguesa. Na Figura 3, ao digitar a frase “Eu gosto de ficar em casa”, automaticamente se detectou o idioma português; a partir disso, você pode selecionar o idioma para o qual deseja traduzir a frase, como o idioma inglês no exemplo. A *machine translation* (“máquina de tradução”) é um exemplo que necessita da PLN para funcionar corretamente.

Você se lembra da última vez que fez alguma compra *on-line*? Quando você clica em um produto, percebe que o *site* começa a sugerir produtos que “talvez você também se interesse”? A Figura 4 ilustra esse tipo de situação.

#### Cientes que compraram este item também compraram



**Figura 4.** Sistema de recomendação da Amazon.

Fonte: Amazon (2020, documento *on-line*).

No *site* da Amazon, mostrado no exemplo da Figura 4, ao buscar um livro e clicar nele na parte de baixo da página, é possível ver sugestões de compra de outros livros. Esta é uma das áreas na qual o PLN é utilizado: a de *recommendation systems* (“sistemas de recomendação”), que sugerem conteúdo a partir de **pesquisas anteriores** do usuário (na Figura 4, são sugeridos vários livros a partir de um livro pesquisado). E como funciona esse processo? A PLN leva em consideração a categoria do livro, faz a “leitura” do título e da descrição do livro e, depois, apresenta sugestões de livros similares. Nesse sentido, podemos pensar em *sites* de notícias, plataformas de vídeo (p. ex., YouTube), serviços de *streaming* (p. ex., Netflix), que geralmente sugerem novas notícias, vídeos, filmes ou séries a partir de informação consumida anteriormente.

Outra aplicação bem conhecida que utiliza PLN é o Gmail da Google, que sugere respostas inteligentes a partir de um *e-mail* recebido (Figura 5).

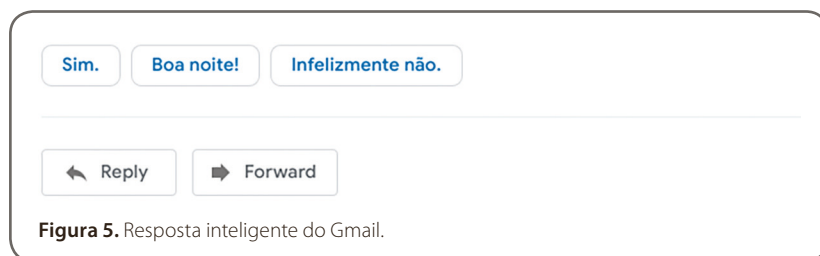


Figura 5. Resposta inteligente do Gmail.

É possível perceber que essa aplicação “sugere” respostas automáticas a partir das palavras-chave recebidas no *e-mail*.

Além dos exemplos apresentados, existem outros que utilizam a PLN de alguma forma, como os descritos a seguir.

- Aplicações que precisam fazer o **reconhecimento de voz** para realizar determinadas atividades, como a Siri, dos dispositivos da Apple, e a Cortana, da Microsoft.
- Uma empresa pode verificar se um produto é bem visto pelos consumidores pela **análise de sentimento** em comentários dos clientes no *site* da empresa e em redes sociais, como Facebook e Twitter, em que, a partir do que é dito, se pode “entender” se o sentimento é positivo ou negativo sobre aquele produto.

- **Sites de perguntas e respostas** também utilizam a PLN para verificar se determinada pergunta já foi respondida ou mostrar perguntas similares, a partir do entendimento, da compreensão e da análise de um conjunto de dados para “responder” ou “sugerir” informações para o usuário.
- **Extração de informação**, uma importante aplicação no campo do PLN e da linguística, refere-se à extração automática de informação estruturada, como entidades, relacionamentos entre entidades e atributos descrevendo as entidades. Por exemplo, dada uma empresa, é possível saber suas localizações; dada uma localização, pode-se saber quais empresas operam nessa localização.
- **Sumarização de texto** é uma técnica do PLN que permite “encurtar” grandes textos. A ideia consiste em criar um resumo com os principais pontos abordados no texto. As aplicações de *chatbots* utilizam sumarização de texto para entender os pontos-chave de uma conversa com um ser humano.
- **Classificação de texto** é o processo de atribuir *tags* ou categorias ao texto de acordo com o seu conteúdo. Trata-se de uma das principais tarefas no PLN, pois recebe como entrada dados não estruturados, como *e-mails*, mídia social, páginas *web*, etc. Extrair informação relevante de dados não estruturados representa uma das principais necessidades das empresas atualmente, já que ajuda a criar estratégias de mercado e no processo de tomada de decisões.
- **Modelagem de tópicos** é a tarefa de extrair os principais tópicos (representados como um conjunto de palavras) em uma coleção de documentos. É importante para sistemas de recomendação pelo fato de recomendar conteúdos de tópicos similares ao que está sendo consumido.
- **Optical character recognition (OCR)** é a principal tecnologia para reconhecimento automático de texto. Com o OCR, pode-se converter documentos impressos, manuscritos e digitalizados em um formato legível pela máquina (p. ex., reconhecer e realizar a leitura automática de cartões e passaportes).
- **Named entity recognition** é uma técnica popular usada na extração de informação para identificar, segmentar, classificar ou categorizar entidades, que, em um texto, são objetos do mundo real, como pessoas, lugares, empresas, etc.
- **Natural language generation** é um processo que transforma dados estruturados em linguagem natural, ou seja, a linguagem do ser humano. Pode ser utilizado para produzir um grande conteúdo para as empresas (p. ex., para um *site*) com o objetivo de automatizar relatórios personalizados.

Como você pôde observar, o PLN auxilia em um conjunto de aplicações, que provavelmente utilizamos no nosso dia a dia sem perceber, o que mostra se tratar de uma área com um potencial enorme, pelas evoluções tecnológicas atuais e futuras, já que cada vez mais as máquinas farão parte da sociedade com o objetivo de ajudar os seres humanos.



## Referências

AMAZON. [Site]. 2020. Disponível em: <https://www.amazon.com.br/>. Acesso em: 13 abr. 2020.

BRANDTZAEG, P. B.; FØLSTAD, A. Why people use chatbots. *In: INTERNATIONAL CONFERENCE ON INTERNET SCIENCE*, 4., 2017, Thessaloniki. *Proceedings* [...]. Thessaloniki: [s. n.], 2017.

CHOUKEA, Y. Natural language processing challenges, achievements and problems: a sober personal perspective. *In: DERSHOWITZ, N.; NISSAN, E. (ed.). Language, culture, computation: computational linguistics and linguistics*. Berlin: Springer, 2014. p. 1–13.

COHEN, K. B.; DEMNER-FUSHMAN, D. *Biomedical natural language processing*. Amsterdam: John Benjamins, 2014. (Natural Language Processing, v. 11).

CRUZ, A. F.; ROCHA, G.; CARDOSO, H. L. Coreference resolution: toward end-to-end and cross-lingual systems. *Information*, [s. l.], v. 11, n. 2, p. 1–24, 2020.

GHAREHCHOPOGH, F. S.; KHALIFEHLOU, Z. A. Analysis and evaluation of unstructured data: text mining versus natural language processing. *In: INTERNATIONAL CONFERENCE ON APPLICATION OF INFORMATION AND COMMUNICATION TECHNOLOGIES*, 5., 2011. *Proceedings* [...]. [S. l.]: IEEE, 2011.

GOMAA, W. H.; FAHMY, A. A. A survey of text similarity approaches. *International Journal of Computer Applications*, [s. l.], v. 68, n. 13, p. 13–18, 2013.

JURAFSKY, D.; MARTIN, J. H. *Speech and language processing*. 3rd ed. [S. l.: s. n.], 2019.

QIAN, Z. *et al.* Speculation and negation scope detection via convolutional neural networks. *In: CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING*, 2016. *Proceeding* [...]. [S. l.: s. n.], 2016.

ZOU, B.; ZHOU, G.; ZHU, Q. Negation and speculation target identification. *In: ZONG, C. (ed.). Natural language processing and Chinese computing*. Berlin: Springer, 2014. p. 34–45

**Fique atento**

Os *links* para *sites da web* fornecidos neste capítulo foram todos testados, e seu funcionamento foi comprovado no momento da publicação do material. No entanto, a rede é extremamente dinâmica; suas páginas estão constantemente mudando de local e conteúdo. Assim, os editores declaram não ter qualquer responsabilidade sobre qualidade, precisão ou integralidade das informações referidas em tais *links*.

Encerra aqui o trecho do livro disponibilizado para esta Unidade de Aprendizagem. Na Biblioteca Virtual da Instituição, você encontra a obra na íntegra.

Conteúdo:



SOLUÇÕES  
EDUCACIONAIS  
INTEGRADAS



# Dica do Professor

---

A pesquisa em processamento de linguagem natural gira em torno da pesquisa corporativa. Isso envolve permitir que os usuários consultem conjuntos de dados na forma de uma pergunta que eles possam fazer para outra pessoa. A máquina interpreta os elementos importantes nas frases de linguagem natural para que possa retornar com uma resposta. Com a evolução das tecnologias inteligentes, essa área pode ser considerada umas das mais promissoras no futuro. E nesse contexto, existem diversas aplicações que utilizam os conceitos de processamento de linguagem natural.

Nesta Dica do Professor, você irá conhecer o processo básico de entendimento das máquinas da linguagem dos seres humanos.



Aponte a câmera para o código e acesse o link do conteúdo ou clique no código para acessar.

# Exercícios

- 1) Existem algumas características que fazem o processamento de linguagem natural adequado para resolver um conjunto de problemas de aplicações.

Marque a alternativa que contém essas características.

- A) Necessidade de entender o ser humano por meio de escrita e fala, responder ao ser humano da forma mais precisa possível e fornecer sugestões e previsões com base no processamento de um grande conjunto de dados.
- B) Necessidade de entender o ser humano por meio de escrita, fala e sinais, responder ao ser humano com ironia e fornecer sugestões e previsões com base no processamento de um grande conjunto de dados.
- C) Necessidade de entender o ser humano por meio de sinais, responder ao ser humano da forma mais precisa possível e fornecer sugestões e previsões de forma aleatória.
- D) Necessidade de entender o ser humano por meio de escrita e fala, responder ao ser humano com ironia e fornecer sugestões e previsões com base no processamento de um grande conjunto de dados.
- E) Necessidade de entender o ser humano por meio de escrita e fala, responder ao ser humano da forma mais precisa possível e fornecer sugestões e previsões de forma aleatória.

- 2) Existem diversas áreas com aplicações que podem utilizar conceitos de processamento de linguagem natural.

Marque a opção em que as duas aplicações podem utilizar o processamento de linguagem natural.

- A) Google Tradutor e algoritmo de ordenação.
- B) Gmail e site com informações estáticas.
- C) Google Tradutor e Gmail.
- D) Rádio convencional e reconhecimento de voz.
- E) Gmail e lousa de giz.

**3) Existem vários desafios no contexto de processamento de linguagem natural. Um deles é decompor frases, processo que consiste em:**

- A) entender o significado semântico de palavras em uma frase.**
- B) ordenar frases com base em seu significado.**
- C) definir o contexto de uma frase.**
- D) quebrar frases de forma que as máquinas possam entender.**
- E) processar imagens.**

**4) O robô Ed é uma aplicação criada em 2004, com o objetivo de falar sobre o uso racional de energia, petróleo, gás natural, meio ambiente, entre outros assuntos. Ao enviar a seguinte mensagem "ola td bem cm vc?", ele responde da seguinte maneira "Comigo tudo em paz. Quais as novidades?".**

**Assinale a alternativa que contém um dos desafios do PLN com que o robô Ed conseguiu lidar de forma satisfatória nesse caso.**

- A) Figuras de linguagem.**
- B) *Coreference resolution*.**
- C) Ambiguidade.**
- D) Algoritmo de ordenação.**
- E) Uso de abreviações e de acrônimos.**

**5) Empresas utilizam o processamento de linguagem natural (PLN) para aumentar o faturamento e melhorar seus produtos e serviços. No contexto do PLN, uma técnica consiste em analisar comentários de clientes em redes sociais para saber o grau de satisfação de determinado produto. A técnica descrita é conhecida como:**

- A) máquinas de tradução.**
- B) recuperação de informação.**

- C) aplicação de *chatbot*.
- D) reconhecimento de fala.
- E) análise de sentimento.

# Na prática

O processamento de linguagem natural veio para facilitar a vida das pessoas em suas atividades diárias, como na tradução de frases de um idioma para outro de forma rápida, sem precisar perder tempo buscando informações em dicionários físicos. O objetivo básico da PLN é fazer com que as máquinas consigam entender os seres humanos. Quando essa comunicação acontece, muitas possibilidades podem ser exploradas, como aplicações, aparelhos domésticos e lâmpadas que podem utilizar o conceito de processamento de linguagem natural para deixar as atividades mais fáceis.

Neste Na Prática, confira um estudo de caso no qual é possível perceber o poder da utilização da PLN em diversas partes para a melhoria de uma empresa.

Conteúdo interativo disponível na plataforma de ensino!

# Saiba mais

---

Para ampliar o seu conhecimento a respeito desse assunto, veja abaixo as sugestões do professor:

## Robô Ed

O robô Ed, criado em 2004, com a tecnologia InBot pelo CONPET (Petrobras), é um robô com inteligência artificial capaz de conversar com milhares de usuários simultaneamente, como um atendente real, e falar sobre assuntos ligados ao uso racional de energia. Neste *link*, você poderá interagir com o robô Ed.



Aponte a câmera para o código e acesse o link do conteúdo ou clique no código para acessar.

## Google Assistente

Google Assistente é uma plataforma de assistência virtual que dispõe de uma grande lista de utilidades, como a criação de lembretes, informações sobre o trânsito, previsão do tempo, entre muitas outras. Clique aqui para conhecer mais sobre essa plataforma.



Aponte a câmera para o código e acesse o link do conteúdo ou clique no código para acessar.

## Processamento de linguagem natural (PLN)

PLN é um conjunto de teorias e de técnicas computacionais para analisar e representar naturalmente textos, em um ou mais níveis da análise linguística, com o propósito de realizar o processamento da linguagem humana para um conjunto de tarefas e aplicações. Aqui, você encontrará uma aula sobre processamento de linguagem natural.



Aponte a câmera para o código e acesse o link do conteúdo ou clique no código para acessar.

## **Técnicas de processamento de linguagem natural aplicadas às ciências sociais**

O objetivo desse trabalho é criar uma ferramenta baseada em processamento de linguagem natural que ajude pesquisadores a navegar uma base de dados. Aqui você encontrará uma base muito importante, que pode trazer uma nova luz sobre vários momentos-chave do início do século XXI a respeito desse assunto.



Aponte a câmera para o código e acesse o link do conteúdo ou clique no código para acessar.

## **Introdução ao processamento de linguagem natural usando Python**

O objetivo principal deste artigo é apresentar uma introdução ao processamento de linguagem natural (PLN) usando a ferramenta Natural Language Toolkit (NLTK) do Python. Aqui, você encontrará, a partir da página 3, a abordagem clássica do processamento de linguagem natural.



Aponte a câmera para o código e acesse o link do conteúdo ou clique no código para acessar.