

Aplicação da Linguística Computacional nas Ciências Forenses

Adelino Pinheiro Silva

Núcleo de Linguística Computacional (ADA)
Faculdade de Letras (FALE)
Universidade Federal de Minas Gerais (UFMG)

13 de junho de 2023



Sumário

- 1 Sumário
- 2 Introdução
- 3 Linguística Computacional Forense
- 4 Aplicações da Linguística Computacional nas Ciências Forenses
- 5 Encerramento
- 6 Dúvidas

Assuntos

- 1 Sumário
- 2 **Introdução**
- 3 Linguística Computacional Forense
- 4 Aplicações da Linguística Computacional nas Ciências Forenses
- 5 Encerramento
- 6 Dúvidas

Visão Geral das Ciências Forenses

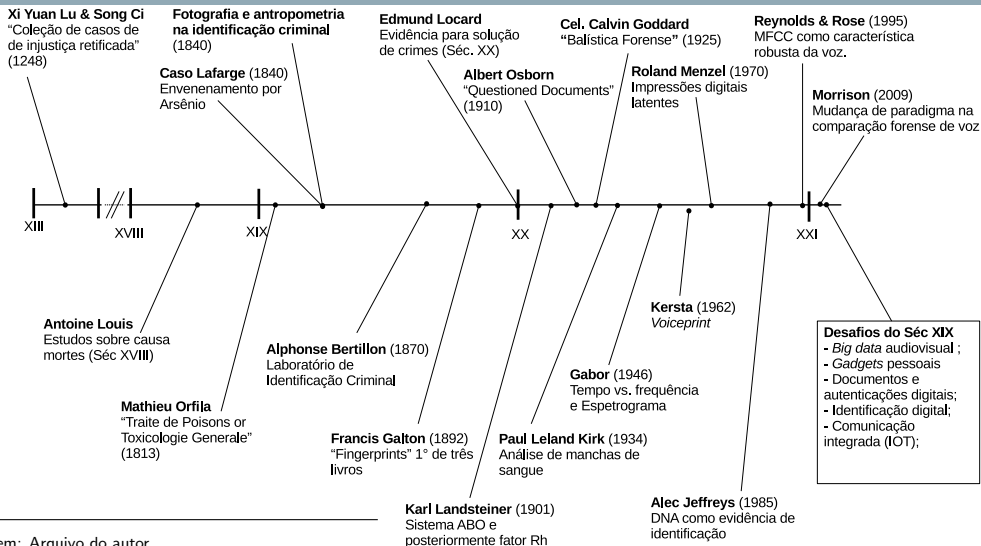


Imagem: Arquivo do autor.

Visão Geral das Ciências Forenses

- Conjunto de conhecimentos e técnicas.
- Aplicadas a assuntos legais.
- Livro Xi Yuan Lu (Registro de injustiça ou lavagem dos erros).
- Escola alemã *Kriminalistik*.
- Escola Francesa Locard.
- Arquimedes, Poincaré,
- Conan Doyle, Maurice Leblanc, Barry Allen.



Imagem: <https://noticias.paginas.ufsc.br/files/2022/09/Foto-Henrique-Almeida-11.jpg>

Assuntos

- 1 Sumário
- 2 Introdução
- 3 Linguística Computacional Forense**
- 4 Aplicações da Linguística Computacional nas Ciências Forenses
- 5 Encerramento
- 6 Dúvidas

Definição

“a parte da ciência linguística que se preocupa com o tratamento computacional da linguagem, e suas aplicações.”

Gabriel de Ávila Othero.

- escritos e manuscritos;
- entrevistas de testemunhas vulneráveis no sistema legal;
- atribuição de autoria, plágio, edição e manipulação;
- fonética forense e identificação do locutor;
- fala para texto e traduções;
- localização sociolinguística de falantes;
- imagem para texto.

Na prática

- corpus;
- dados empíricos;
- estudos constantes;
- NLP;
- programação;
- reconhecimento de padrões e estatística.

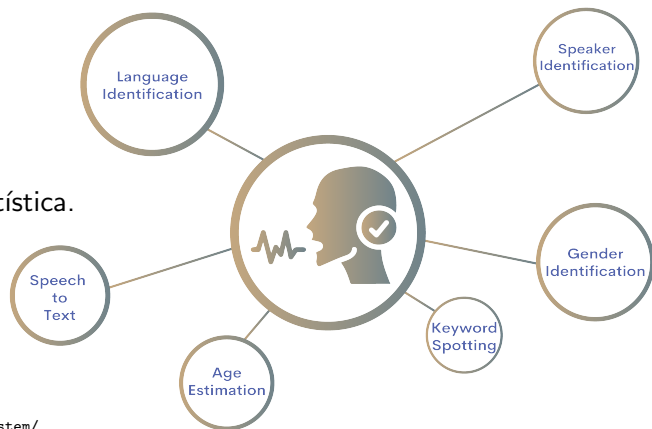


Imagem: <https://www.stratign.com/voice-forensic-system/>

Escritos e manuscritos

- grafotécnica
- atribuição de estilo.

Meu nome é Natália, tenho 11 anos, e estou enviando essa carta
 para te pedir ajuda, Presidente Lula.
 Estou preocupada com meu pai. Ele é metrôviário em Belo Horizonte.
 Um dia desses, perguntei para ele porque estava tão calado e pensa-
 tivo, e ele me disse que estava preocupado e chateado com tudo que
 está acontecendo por causa da privatização do metrô de Belo Horizonte.
 Eu não entendi direito o que ele me disse, mas sei dizer que mu-
 itos não perder o emprego, e que só o Senhor Presidente Lula pode
 salvar o emprego deles. Me lembro das eleições do ano passado, o
 meu pai dizia que todos os metrôviários daqui de Belo Horizonte
 estavam com suas famílias lutando pelo Lula. No dia da vitória
 do Senhor, ele chorou de alegria.
 Hoje, essa alegria se foi e só restou tristeza e decepção. Eu não
 quero mais ver o meu pai assim, e nem os colegas dele de trabalho.
 Lula! Vou pedir ao Papai do Céu para te ajudar e tocar o seu
 coração para salvar os empregos de todos os metrôviários do metrô
 de Belo Horizonte. Amém.
 Belo Horizonte, quatro de junho de dois mil
 e vinte e três.

Imagem: <https://www.otempo.com.br/cidades/em-carta-filha-de-metroviario-de-bh-faz-apelo-a-lula-salve-os-empregos-1.2885744>

Testemunhas vulneráveis

- espaço de apoio e acolhimento;
- análise do discursos;
- cuidado com eventos sociais sensíveis.



Imagem: <https://www.capistranoadvogados.com.br/images/noticias/grd/119334b48feeb38af04a349145591eb9.jpg>

Plágio, edição e manipulação

- reconhecimento de padrões;
- música;
- atribuição de estilo.

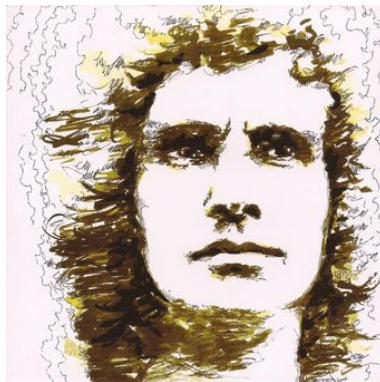


Imagem: <https://genius.com/Roberto-carlos-traumas-lyrics>

Localização sociolinguística de falantes



Imagem: <https://localingual.com/> e <https://www.thefools.com.br/blog/post/dialetos-do-portugues-brasileiro>

Perspectivas

Paradigma indutivo:

- acúmulo de convergências e divergências;
- baseada no material observado; e
- difícil inferir sobre a possibilidade de erro.

Paradigma hipotético-dedutivo:

- observação são ponderadas pela semelhança vs. tipicidade;
- necessita de uma base de dados para estimação de parâmetros; e
- possível estimar erros

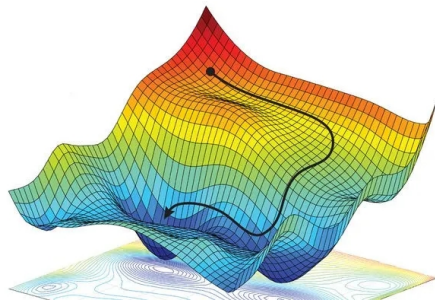


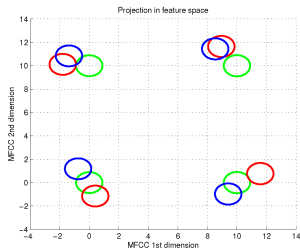
Imagem: https://miro.medium.com/v2/resize:fit:720/format:webp/1*IxMLWG1xsZ50b91M5VJXSA.jpeg

Assuntos

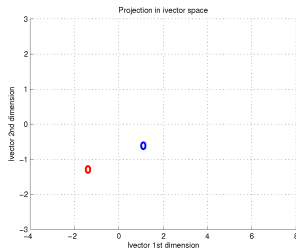
- 1 Sumário
- 2 Introdução
- 3 Linguística Computacional Forense
- 4 Aplicações da Linguística Computacional nas Ciências Forenses**
- 5 Encerramento
- 6 Dúvidas

Comparação de Locutor

- métodos linguísticos;
- medidas acústicas;



(a) Data domain



(b) i-vector domain

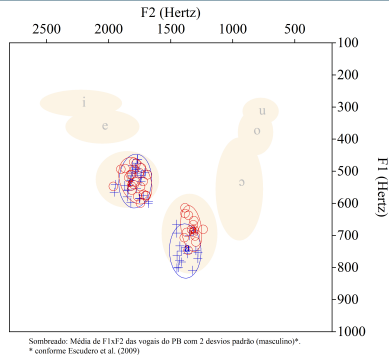


Imagem: <https://doi.org/10.3390/app9183697> e arquivo pessoal

Interpretação de falas em imagem

- estimar o som da fala a partir de pistas articulatórias
- taxas de 51% a 81%.



Imagem: Arquivo pessoal.

Conversão de áudio em texto

Multitask training data (680k hours)

English transcription

- 🗣️ "Ask not what your country can do for ..."
- 📝 Ask not what your country can do for ...

Any-to-English speech translation

- 🗣️ "El rápido zorro marrón salta sobre ..."
- 📝 The quick brown fox jumps over ...

Non-English transcription

- 🗣️ "언덕 위에 올라 내려다보면 너무나 넓고 넓은 ..."
- 📝 언덕 위에 올라 내려다보면 너무나 넓고 넓은 ...

No speech

- 🎧 (background music playing)
- 📝 🚫

Sequence-to-sequence learning

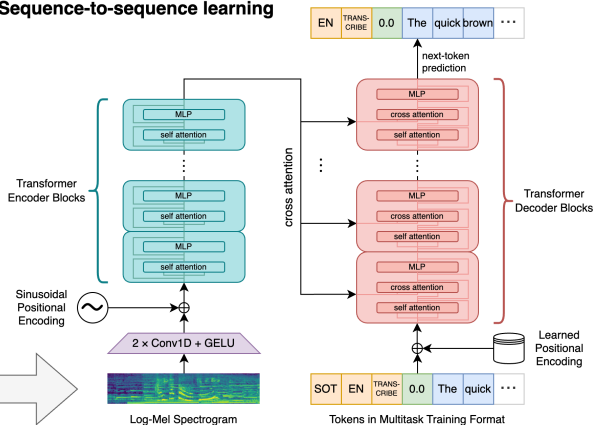


Imagem: <https://github.com/openai/whisper>.

Assuntos

- 1 Sumário
- 2 Introdução
- 3 Linguística Computacional Forense
- 4 Aplicações da Linguística Computacional nas Ciências Forenses
- 5 Encerramento**
- 6 Dúvidas

Agradecimentos



Fim!

Contatos:

- e-mail: adelinocpp@gmail.com, adelinocpp@yahoo.com ou adelino@ufmg.br;
- Whatsapp (31) 98801-3605;
- Setor de Perícias em Áudio e Vídeo - Instituto de Criminalística. Av. Augusto de Lima, 1833, Barro Preto, Belo Horizonte-MG. Tel.: (31) 3330-1887;

Sobre este material

Esta obra está licenciada sob a licença *Creative Commons* CC BY-NC-SA 4.0

Favor fazer referência a este trabalho como:

Silva, A. P. (2023), *Aplicação da Linguística Computacional nas Ciências Forenses*. Online:
<https://github.com/adelinocpp/estatistica-para-linguistica>

```
@Misc{Silva2022,
  title={Aplicação da Linguística Computacional nas Ciências Forenses},
  author={Adelino Pinheiro Silva},
  howPublished={\url{https://github.com/adelinocpp/ada-lccf}},
  year={2022},
  note={Version 1.0; Creative Commons BY-NC-SA 4.0.},
}
```



Assuntos

- 1 Sumário
- 2 Introdução
- 3 Linguística Computacional Forense
- 4 Aplicações da Linguística Computacional nas Ciências Forenses
- 5 Encerramento
- 6 Dúvidas**

Dúvidas



Imagem: <https://www.hevcon.com.br/duvidas-frequentes-relacionado-ao-novo-bem-e-as-alteracoes-das-mps/>.