

# Palavras Cruzadas e Quebra de Códigos

## A Teoria da Informação na Linguagem e Criptografia

Luís Felipe Ramos Ferreira<sup>1</sup>, Gabriel Fialho<sup>2</sup>, Diego Pereira<sup>3</sup>

Universidade Federal de Minas Gerais (UFMG)

23 de julho de 2024

# Sumário

## 1 Introdução

- Zodíaco

## 2 Palavras Cruzadas

- Tipos de palavras cruzadas
- Modelo “palavro-chinês”
- Estudo de caso: viabilidade de palavras cruzadas no “palavro-chinês”
- Hipótese sobre palavras cruzadas em chinês

## 3 Quebra de Códigos

- História e Contexto
- Métodos de Quebra de Código

## 4 Conclusão

# Sumário

## 1 Introdução

- Zodíaco

## 2 Palavras Cruzadas

- Tipos de palavras cruzadas
- Modelo “palavro-chinês”
- Estudo de caso: viabilidade de palavras cruzadas no “palavro-chinês”
- Hipótese sobre palavras cruzadas em chinês

## 3 Quebra de Códigos

- História e Contexto
- Métodos de Quebra de Código

## 4 Conclusão

# Introdução

- Teoria da informação: quantificação, armazenamento e comunicação de informações.
- Aplicações: palavras cruzadas e quebra de códigos.
- Importância da entropia e redundância na linguagem.
- Nas palavras cruzadas, entropia e redundância afetam criação e resolução.
- A redundância e entropia da linguagem podem ajudar a decifrar mensagens criptografadas.

# Zodíaco

- Em agosto de 1969, o assassino zodíaco enviou mensagens para jornais, detalhando seus crimes.
- Junto às mensagens, enviou criptogramas e exigiu que esses criptogramas fossem publicados nos jornais.

4 San Francisco Chronicle ☆☆ Sat., Aug. 2, 1969

## Coded Clue in Murders

A man who claimed he shot and killed two Vallejo teenagers last December and a young woman on July 4 threatened yesterday to kill 12 more this weekend.

The menacing message came in unsigned letters mailed to the editors of The Chronicle, the Vallejo Times-Herald and the San Francisco Examiner.

"Here is a copy of a cipher," the letter said in part. "In this cipher is my identity."

"If you do not print this cipher by the afternoon of Friday, I will go on a kill rampage."

PEOPLE

"I will chase ferulsi around all week end killing lone people in the night, then I will again to kill again until I end up with dozen people over the weekend."

The cover letter listed what the writer called "some facts which only I and the police know such as the brand of ammunition used and the locations in which the bodies were found."

Vallejo police said most of the material was actually common public knowledge, but officers took the letter seriously because it was the second letter learned last month that the killer of Darlene Ferrin, 22, who was slain July 4, was a man with a bizarre craving for attention.

Half an hour after the girl was killed, the police received another anonymous phone call from a man who said he had her and a young com-

N K Q S C E / A D B O Z F A P B V  
9 3 X P W D F □ A C + □ Δ A Δ B  
□ O T O R U □ + □ D A Y □ □ S □ W  
V Z E G Y K E □ □ T Y A □ □ □ L □ □  
H □ F B X □ □ □ X A D C □ □ □ L □ □  
□ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □  
Z □ J T □ □ □ □ □ □ □ □ □ □ □ □ □ □ □  
V E X A W I □ □ □ □ □ □ □ □ □ □ □ □ □ □

This code may conceal Vallejo killer's identity

pation while they were parked in Blue Rock Springs Park.

VICTIMS

The other victims the man claimed were Thomas Faray, 17, and Bettie Jensen, 16, who were shot while walking home from a baseball game at the same place.

Faray was a man who said he had her and a young com-

the letter was written by the murderer, but it could have been written by a child or a senseless Vallejo fool. He requested the writer to send a second letter "with more facts to prove it."

The three newspapers turned over their letters to Vallejo police, and the ciphered message in turn was passed along. No one stepped over in the hope that he could decode it.

# Zodíaco

- O casal de professores Donald e Bettye Harden logo solucionaram o primeiro criptograma.
- Foi simples resolver o primeiro criptograma, pois ele continha muitos caracteres e o zodíaco usou uma simples cifra de substituição.
- A redundância da linguagem foi um fator crucial na decifração do criptograma
  - **Frequência de Letras:** "E" e "T" são muito comuns no inglês;
  - **Palavras Comuns e Padrões de Frase:** "THE", "AND", "IS"
  - **Contexto e Significado:** Uma carta escrita por um assassino provavelmente contém "KILL"

# Zodíaco

Figura: Nerdologia (YouTube)

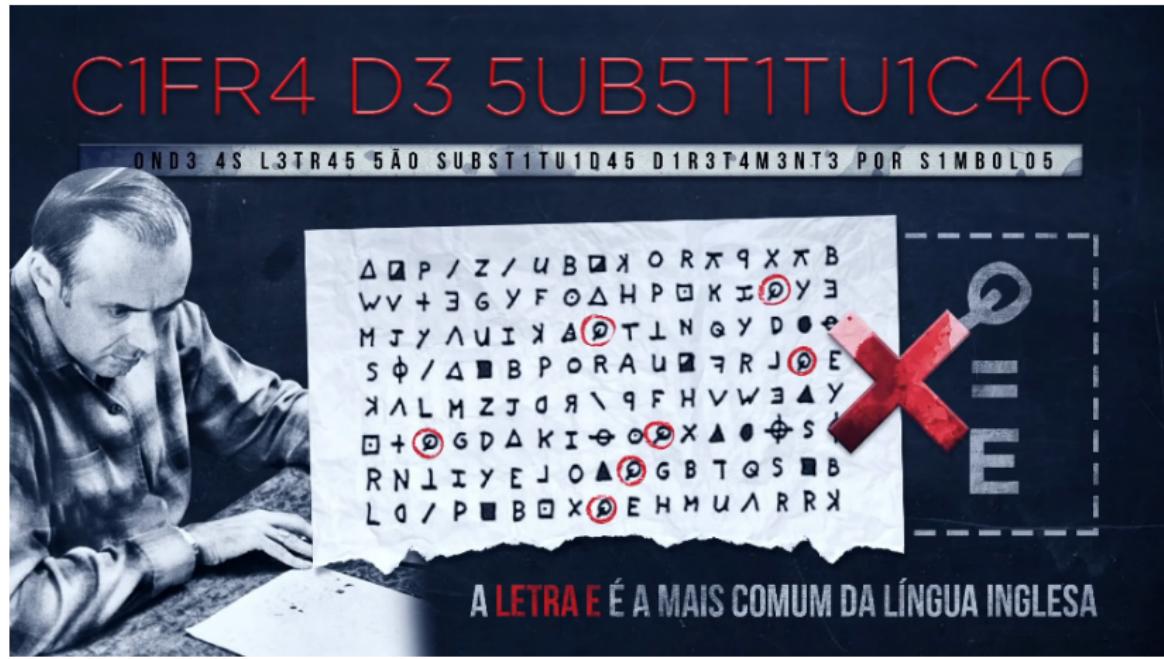
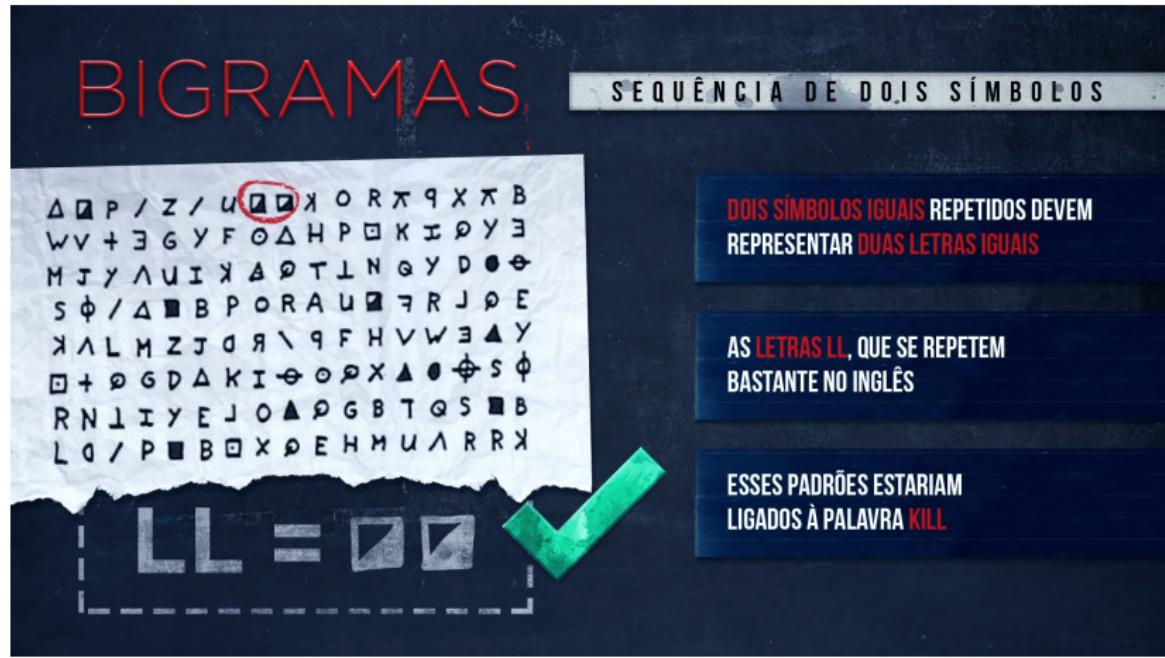
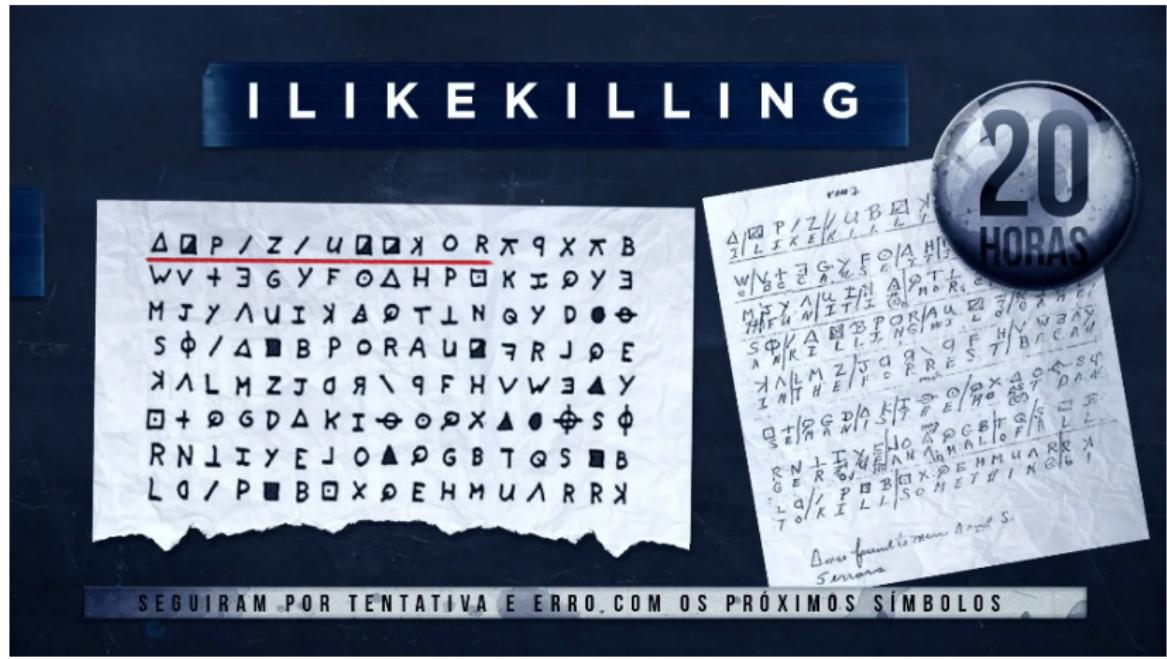


Figura: Nerdologia (YouTube)



# Zodíaco

Figura: Nerdologia (YouTube)



Solução do Z480:

"I LIKE KILLING PEOPLE BECAUSE IT IS SO MUCH FUN IT IS MORE FUN THAN KILLING WILD GAME IN THE FORREST BECAUSE MAN IS THE MOST DANGEROUS ANIMAL OF ALL TO KILL SOMETHING GIVES ME THE MOST THRILLING EXPERENCE IT IS EVEN BETTER THAN GETTING YOUR ROCKS OFF WITH A GIRL THE BEST PART OF IT IS THAT WHEN I DIE I WILL BE REBORN IN PARADICE AND ALL THAT I HAVE KILLED WILL BECOME MY SLAVES I WILL NOT GIVE YOU MY NAME BECAUSE YOU WILL TRY TO SLOW DOWN OR STOP MY COLLECTING OF SLAVES FOR MY AFTERLIFE EBEORIETEMETHHPITI"

# Sumário

## 1 Introdução

- Zodíaco

## 2 Palavras Cruzadas

- Tipos de palavras cruzadas
- Modelo “palavro-chinês”
- Estudo de caso: viabilidade de palavras cruzadas no “palavro-chinês”
- Hipótese sobre palavras cruzadas em chinês

## 3 Quebra de Códigos

- História e Contexto
- Métodos de Quebra de Código

## 4 Conclusão

# Palavras cruzadas

A existência de palavras cruzadas depende das características da linguagem

- Quanto maior a entropia do idioma (relacionado ao número de palavras), maiores as possibilidades de palavras cruzadas
- Quanto maior a redundância num idioma, mais difícil criar palavras cruzadas (em uma língua sem redundância qualquer combinação de letras seria uma palavra válida)

# Tipos de palavras cruzadas



Figura: Palavras cruzadas dos tipos A (americano) e B (britânico).

# Modelo “palavro-chinês”



Utilizamos um idioma simplificado chamado “palavro-chinês”, que consiste em  $W = 4000$  palavras com  $L = 4$  ideogramas cada, com entropia  $H_W \equiv \frac{\log_2 W}{L+1} = 2,4$ . As palavras que compõem este idioma são originadas de uma seleção de caracteres aleatória feita por um canal fonte considerando 2000 ideogramas.

## Estudo de caso: viabilidade de palavras cruzadas no “palavro-chinês”

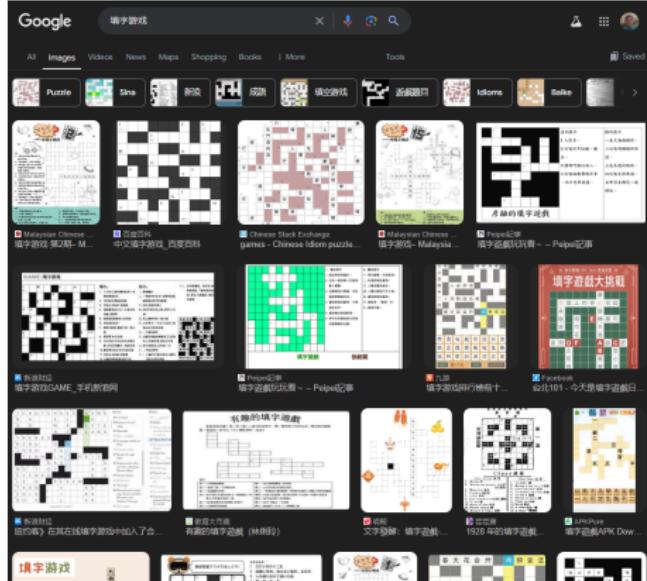
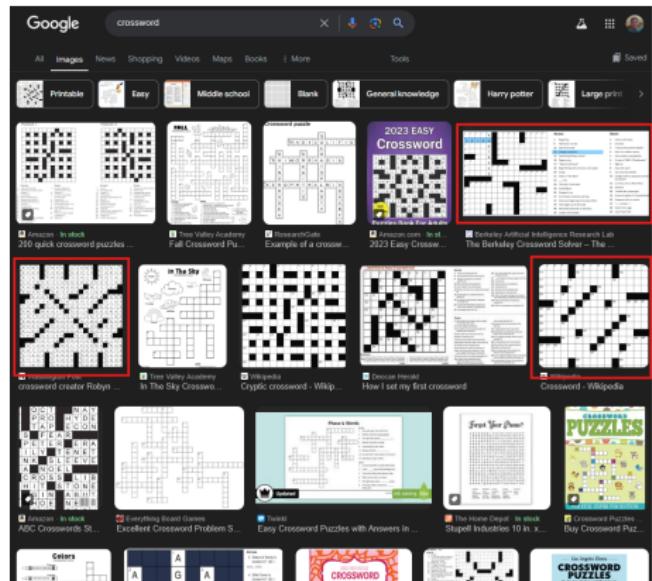
- Para uma instância de palavras cruzadas com  $S$  quadrados, seja  $w = f_w S$  o número de palavras e  $I = f_1 S$  o número de espaços ocupados com caracteres.
- De quantas formas é possível preencher aleatoriamente o quadro:  $|T| = 2^{IH_0}$
- Probabilidade de que uma dessas palavras seja válida é  $\beta = \frac{W}{2^{LH_0}}$
- Chance de que todas as palavras estejam preenchidas de forma válida:  $\beta^w$
- O número total de palavras cruzadas é  
$$\log \beta^w |T| = w(L+1)H_W + H_0(I-Lw) = S[(f_1 - f_w L)H_0 + f_w(L+1)H_W]$$

## Estudo de caso: viabilidade de palavras cruzadas no “palavro-chinês”

- O número total de palavras cruzadas é uma função crescente de  $S$  quando  $(f_1 - f_w L)H_0 + f_w(L+1)H_W > 0$
- Para haver palavras cruzadas do tipo B:  $H_W > \frac{1}{4} \frac{L}{L+1} H_0$
- Para haver palavras cruzadas do tipo A:  $H_W > \frac{1}{2} \frac{L}{L+1} H_0$
- Se assumirmos uma distribuição com probabilidades iguais para cada ideograma, teríamos uma entropia de  $\log_2 2000 = 10,97$ , para simplificar consideramos  $H_0 = 10$
- É possível criar palavras cruzadas do tipo B:  $2,4 > 2$
- Não é possível criar palavras cruzadas do tipo A:  $2,4 < 4$

# Hipótese sobre palavras cruzadas em chinês

É fácil achar exemplos de palavras cruzadas do tipo A em inglês, mas não conseguimos encontrar exemplos desse tipo em chinês



# Sumário

## 1 Introdução

- Zodíaco

## 2 Palavras Cruzadas

- Tipos de palavras cruzadas
- Modelo “palavro-chinês”
- Estudo de caso: viabilidade de palavras cruzadas no “palavro-chinês”
- Hipótese sobre palavras cruzadas em chinês

## 3 Quebra de Códigos

- História e Contexto
- Métodos de Quebra de Código

## 4 Conclusão

# Quebra de Códigos

- Aplicação da Teoria da Informação na criptografia.
- Exemplo: máquina Enigma na Segunda Guerra Mundial.

- Máquina Enigma: vastas configurações iniciais. Aproximadamente  $8 \times 10^{12}$ .
- Mensagens cifradas pelo exército alemão.
- Decifragem dependia de falhas no processo de cifragem e redundância do idioma.

# Métodos de Quebra de Código

- Gigantesco número de mensagens diariamente → duas máquinas poderiam estar na mesma configuração ao codificar uma mensagem.
- Hipótese nula( $H_0$ ): estados das máquinas não possuem correlação.
- Hipótese alternativa( $H_1$ ): máquinas codificaram suas mensagens a partir do mesmo estado.

# Métodos de Quebra de Código

- Hipótese nula ( $\mathcal{H}_0$ )

- $x = x_1 x_2 x_3 \dots = c_1(u_1) c_2(u_2) c_3(u_3) \dots$
- $y = y_1 y_2 y_3 \dots = c'_1(v_1) c'_2(v_2) c'_3(v_3) \dots$

- Hipótese alternativa ( $\mathcal{H}_1$ )

- $x = x_1 x_2 x_3 \dots = c_1(u_1) c_2(u_2) c_3(u_3) \dots$
- $y = y_1 y_2 y_3 \dots = c_1(v_1) c_2(v_2) c_3(v_3) \dots$

- Queremos saber o valor de  $\log \frac{P(x,y|\mathcal{H}_1)}{P(x,y|\mathcal{H}_0)}$

# Métodos de Quebra de Código

- Dados  $x$  e  $y$  de comprimento  $T$ , vindos de um alfabeto com  $A$  caracteres, temos que:

$$\log \frac{P(x, y | \mathcal{H}_1)}{P(x, y | \mathcal{H}_0)} = M \log m A + N \log \frac{(1 - m)A}{A - 1}$$

- $m$  é a probabilidade de dois caracteres codificados serem idênticos.

# Exemplo

u LITTLE-JACK-HORNER-SAT-IN-THE-CORNER-EATING-A-CHRISTMAS-PIE--HE-PUT-IN-H  
v RIDE-A-COCK-HORSE-TO-BANBURY-CROSS-TO-SEE-A-FINE-LADY-UPON-A-WHITE-HORSE

matches: .\*.....\*...\*\*\*\*\*.\*.....\*.....\*.....\*

# Sumário

## 1 Introdução

- Zodíaco

## 2 Palavras Cruzadas

- Tipos de palavras cruzadas
- Modelo “palavro-chinês”
- Estudo de caso: viabilidade de palavras cruzadas no “palavro-chinês”
- Hipótese sobre palavras cruzadas em chinês

## 3 Quebra de Códigos

- História e Contexto
- Métodos de Quebra de Código

## 4 Conclusão

# Conclusão

- Teoria da informação: essencial para comunicação e criptografia.
- Entropia e redundância influenciam a criação e resolução de palavras cruzadas.
- Quebra de códigos na Segunda Guerra Mundial alterou o curso da história.

# Referências

- Alvim, M. S., et al. (2020). *The Science of Quantitative Information Flow*.
- Cover, T. M., Thomas, J. A. (2006). *Elements of Information Theory*.
- Hinsley, F. H., Stripp, A. (2001). *Codebreakers: the inside story of Bletchley Park*.
- MacKay, D. J. C. (2005). *Information Theory, Inference and Learning Algorithms*.
- Singh, S. (1999). *The code book*.