

UMA INTRODUÇÃO AO MUNDO DA LINGUÍSTICA COMPUTACIONAL E NLP

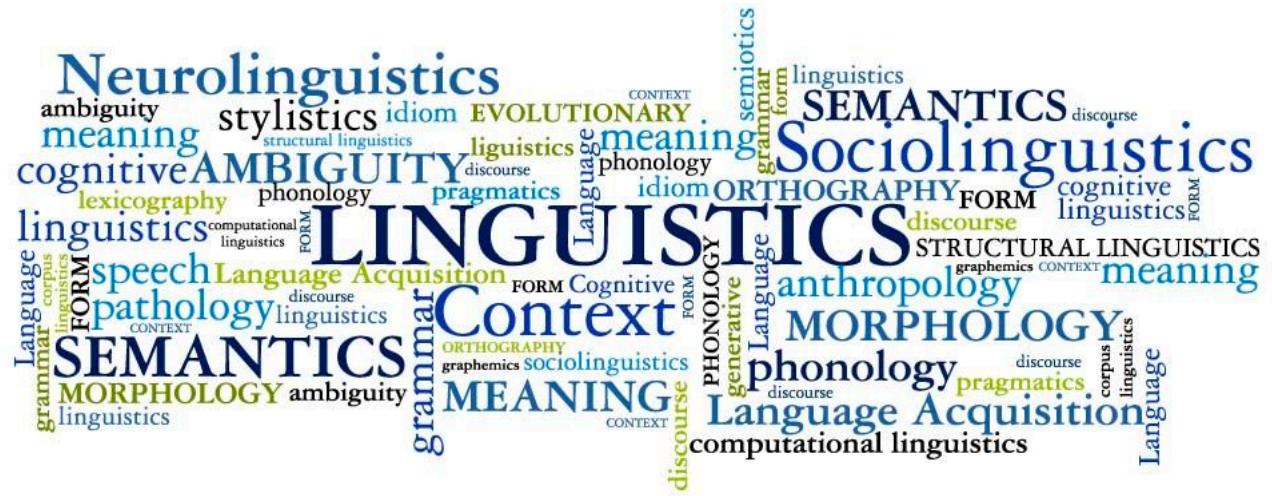
Beatriz Albiero



O QUE É
LINGUÍSTICA?

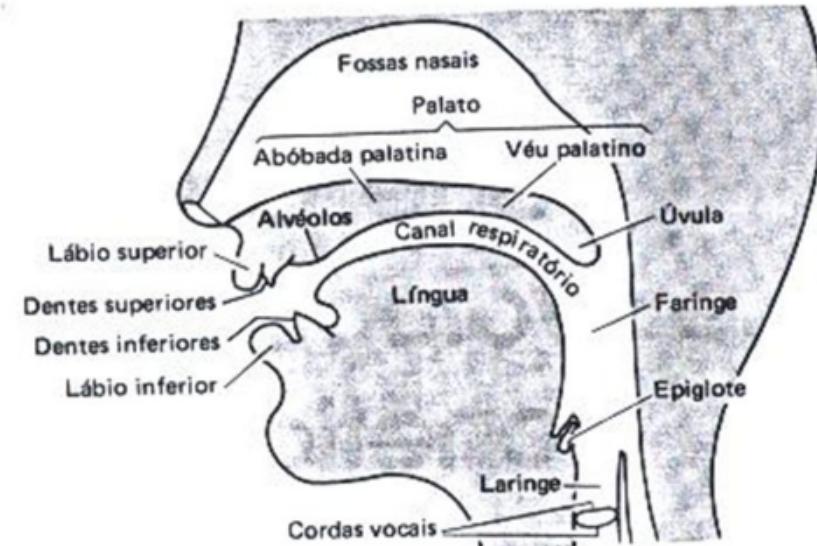
Fonética
Fonologia
Morfologia
Sintaxe
Semântica
Pragmática

Fonética
Fonología
Morfología
Sintaxis
Semántica
Pragmática





Fonética
Fonologia
Morfologia
Sintaxe
Semântica
Pragmática





Fonética
Fonologia
Morfologia
Sintaxe
Semântica
Pragmática



(Yupic)

Quais são os sons importantes dessa
língua?



Fonética
Fonologia
Morfologia
Sintaxe
Semântica
Pragmática

Pato e Gato





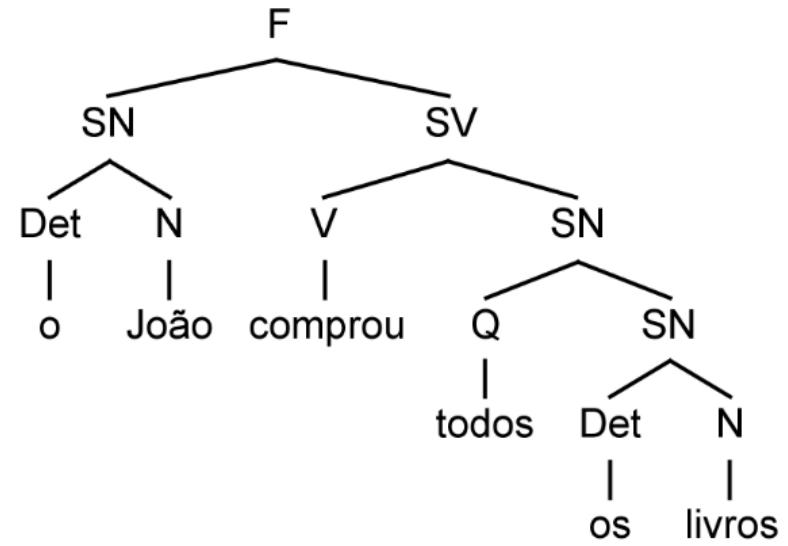
Fonética
Fonologia
Morfologia
Sintaxe
Semântica
Pragmática

ńbá	“eu chego”
àbá	“você chega”
òbá	“ele/ela chega”
èbá	“nós chegamos”
ámùbá	“vocês chegam”
bèbá	“eles/elas chegam”

Língua Baulê (Níger-Congo)



Fonética
Fonologia
Morfologia
Sintaxe
Semântica
Pragmática





Fonética
Fonología
Morfología
Sintaxe
Semântica
Pragmática

“green colorless ideas sleep furiously”
Noam Chomsky



Fonética
Fonologia
Morfologia
Sintaxe
Semântica
Pragmática

“A Magali é doida mas eu gosto dela.”

Relações entre palavras:
Belo = Bonito



Fonética
Fonologia
Morfologia
Sintaxe
Semântica
Pragmática

Frases podem **descrever** ações



Fonética
Fonologia
Morfologia
Sintaxe
Semântica
Pragmática

Frases podem **ser** ações



Fonética
Fonologia
Morfologia
Sintaxe
Semântica
Pragmática

“Eu prometo que não acontecerá novamente.”

“Eu te perdoo”

“Eu vos declaro marido e mulher”

“Eu ordeno que você saia daqui”

Atos de Fala

E LINGÜÍSTICA
COMPUTACIONAL?

How is computational linguistics different from natural language processing?

This question previously had details. They are now in a comment.

 Answer

 Follow · 157

 Request

 1



...

8 Answers



Jason Eisner, computer science professor at Johns Hopkins



Updated Aug 20, 2016 · Upvoted by Moukthika Yerramilli, [MA Computational Linguistics & Linguistics, English and Foreign Languages University \(2018\)](#) and Elynn Lee, [Master's Computer Science, The University of Texas at Austin \(2015\)](#)

<https://www.quora.com/How-is-computational-linguistics-different-from-natural-language-processing>

A linguística computacional desenvolve métodos computacionais para responder a questões científicas da linguística.

Exemplos:

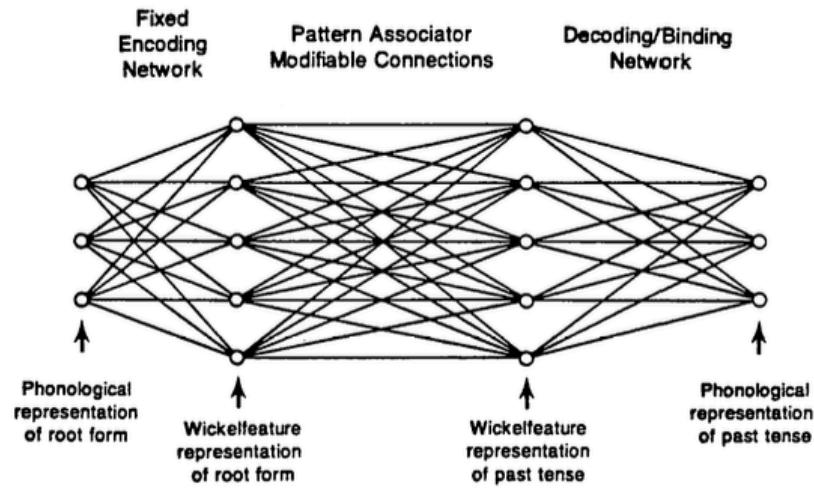


FIGURE 1. The basic structure of the model.

On Learning the Past Tenses of English Verbs
D. E. RUMELHART and J. L. McCLELLAND

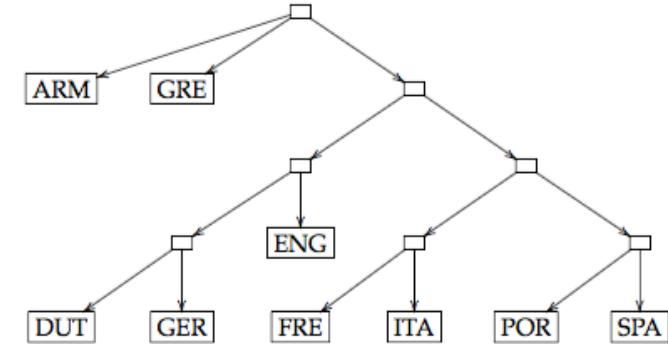


Figure 7.7: 95% consensus tree obtained from NCSMC applied to the languages Armenian (ARM), Dutch (DUT), English (ENG), French (FRE), German (GER), Greek (GRE), Italian (ITA), Portuguese (POR) and Spanish (SPA) in [Wiktionary \[2015\]](#).

A Statistical contribution to Historical Linguistics
Rafael Bassi Stern

Natural language processing is the art of solving engineering problems that need to analyze (or generate) natural language text. Here, the metric of success is not whether you designed a better scientific theory or proved that languages X and Y were historically related. Rather, the metric is whether you got good solutions on the engineering problem.

Quais problemas?

Problema #1

Entender a opinião dos usuários do seu produto.

Campanha de marketing voltada para necessidades dos seus clientes.

Problema #1

Word Embeddings



- Comentário Positivo
- Comentário Negativo
- Comentário Neutro

Problema #1

Word Embeddings

Rome = [1, 0, 0, 0, 0, 0, ..., 0]
Paris = [0, 1, 0, 0, 0, 0, ..., 0]
Italy = [0, 0, 1, 0, 0, 0, ..., 0]
France = [0, 0, 0, 1, 0, 0, ..., 0]

One-hot



Problema: Essa representação
não preserva as relações entre
as palavras.

$$notebook^T laptop = 0$$

Problema #1

Word Embeddings

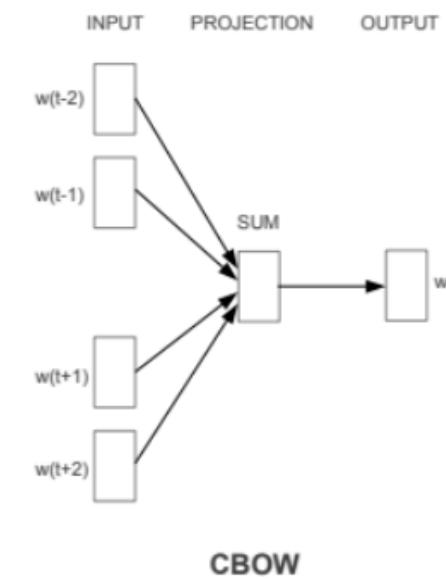
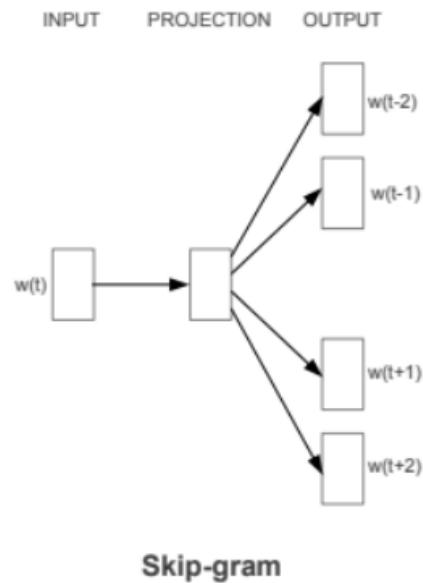
"You shall know a word by the company it keeps"
J. R. Firth

Intuição:

*"I love him, he is such a lovely, **pulchritudinous**, nice man."*

Problema #1

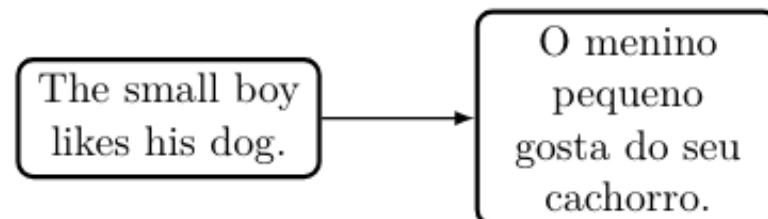
Word Embeddings



Efficient Estimation of Word Representations in Vector Space
Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean
Google Inc., Mountain View, CA

Problema #2

Tradução Automática:



Problema #2

Modelos de Linguagem

N- Gramas

Essa é uma frase de exemplo.

Essa é uma outra frase de exemplo.

Essa também é uma frase de exemplo.

N = 1

	Essa	é	também	uma	frase	de	exemplo	outra
Essa	0	2	1	0	0	0	0	0
é	0	0	0	3	0	0	0	0
também	0	1	0	0	0	0	0	0
Uma	0	0	0	0	2	0	0	1
frase	0	0	0	0	0	3	0	0
de	0	0	0	0	0	0	3	0
exemplo	0	0	0	0	0	0	0	0
outra	0	0	0	0	1	0	0	0

Problema #2

Modelos de Linguagem

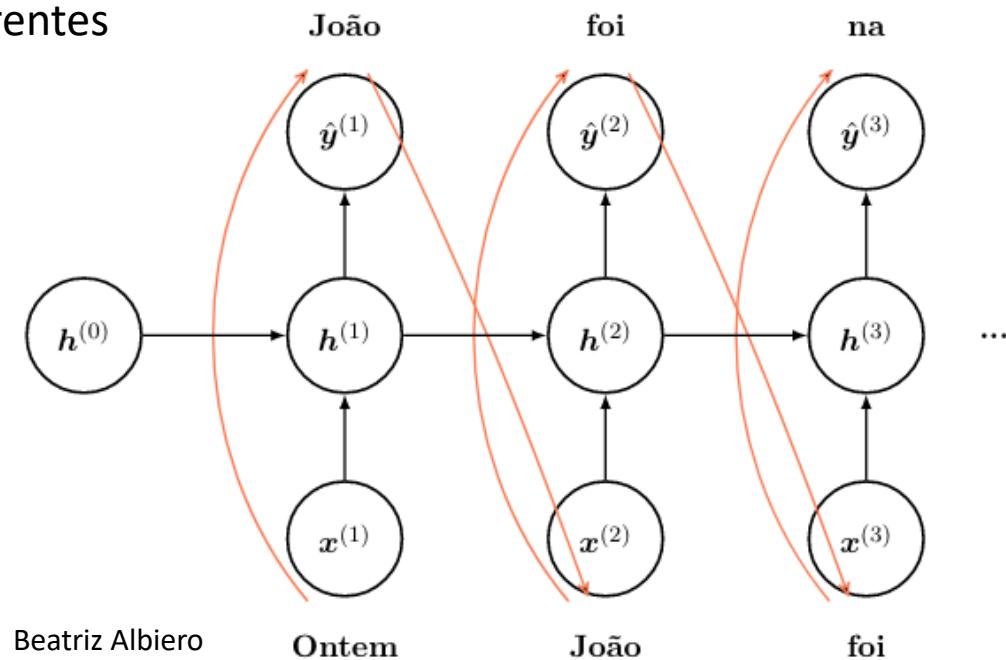
N- Gramas

Essa é uma frase de exemplo.

Problema #2

Modelos de Linguagem

Redes Neurais Recorrentes

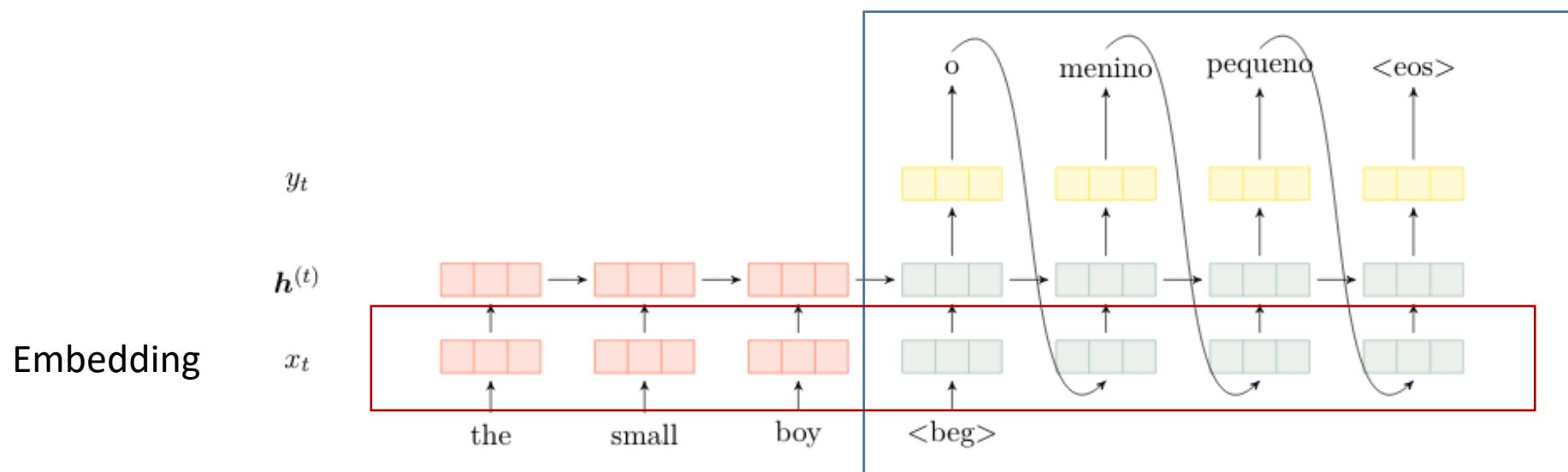


Problema #2

Modelos de Linguagem

Tradução Automática:

Modelo de Linguagem com Rede Neural Recorrente



Problema #2

Modelos de Linguagem

Completion

i want to have dinner tonight with my friends and I've got a pretty good head on my shoulders now. We just hit a few places and all I want is to come back to my parents and my brother. Now that I know that my Mom wants to stay the hell away from me, I feel guilty as hell to be a part of your family. And I don't think he even likes me! I'm just mad that I've wasted away this week without him, you know what I mean?" I can hear him groaning again and again and I wonder what I'm going to do about this now, so I decide to ignore his comment and tell him that I love him for saying this and we'll see where the rest of the evening goes. I tell him it's just that, I really feel bad for hurting his feelings and it might work. When we get back out the back door, Jack is already walking towards his room in a panicked state. "Wh-what were you doing?" I ask him as I start to turn my room key in the lock, hoping that he will at least remember to look for me after he makes his way into my room. He shakes his head and smiles sadly, trying to hide his expression as we both begin to walk towards the door. "I-I'm sorry Mr. Fluffy. I just don't feel comfortable around this

<https://talktotransformer.com/>

Built by [Adam King \(@AdamDanielKing\)](#) as an easier way to play with OpenAI's new machine learning model. In February, OpenAI unveiled a [language model called GPT-2](#) that generates coherent paragraphs of text one word at a time.

Problema #3

Problema #3

Automação de Legendas



animals



theatre



cooking



news



autos



sports

A woman giving a speech on a news channel.

Cars racing on a road surrounded by lots of people.

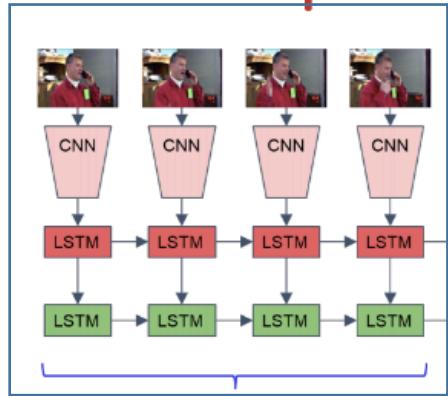
A basketball team is playing in front of speculators.

Problema #3

Automação de Legendas

Video Captioning

Image Detection



Sequence to Sequence Video to Text
[Venugopalan et. al. ICCV'15]

Slides: Kate Saenko
University of Boston

Modelo de Linguagem



[AAAI13] [ICCV13] [COLING14]
[NAACL15] [ICCV15] [CVPR15]

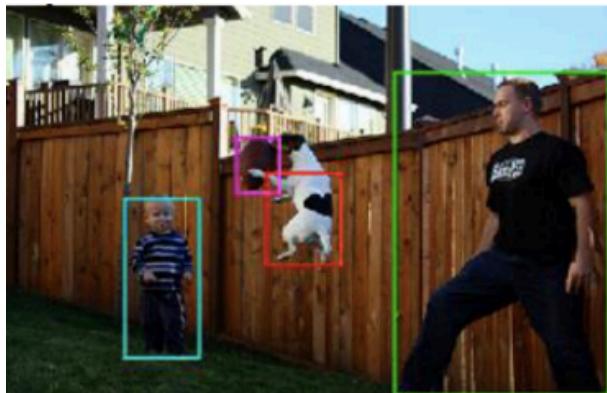


Problema #3

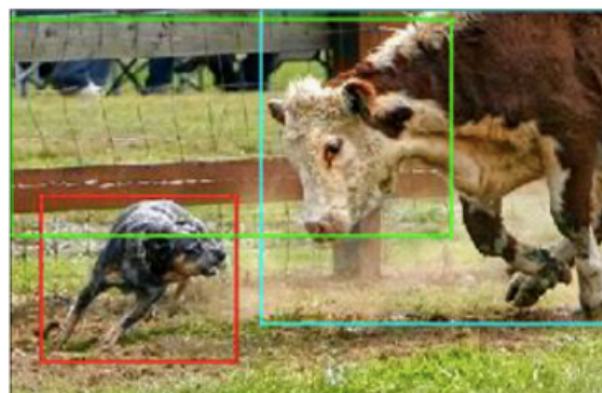
Automação de Legendas

Phrase detection

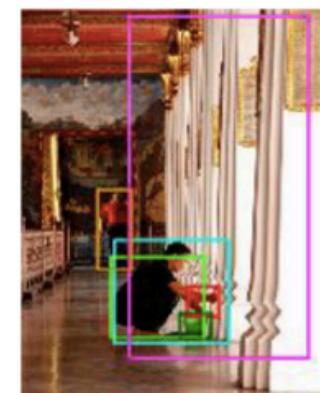
with CCA Initialization



dog, man, child, football



dog, fence area, large brown
white cow



orange gloves, woman,
black clothes, columns,
someone, green bucket

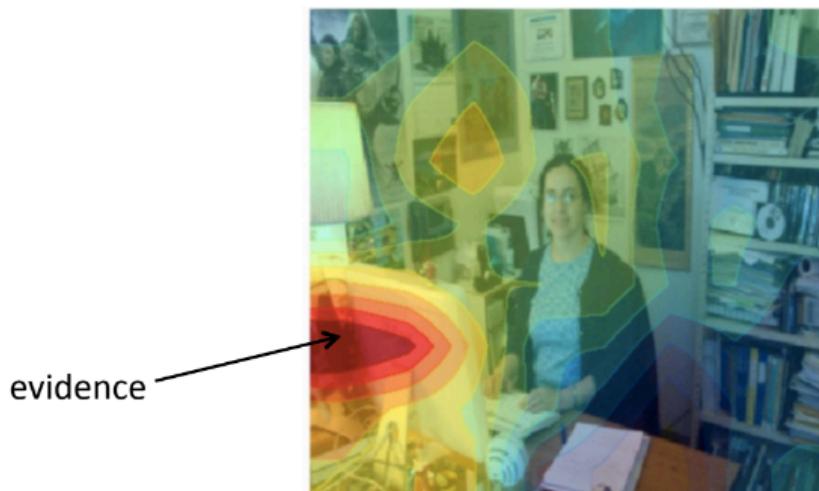
Slides: Kate Saenko
University of Boston

Problema #3

Automação de Legendas

Why do we need explainable models?

- Captioning models may make errors due to **bias**:



Model: A **man** sitting at a desk with a laptop computer.

Wrong

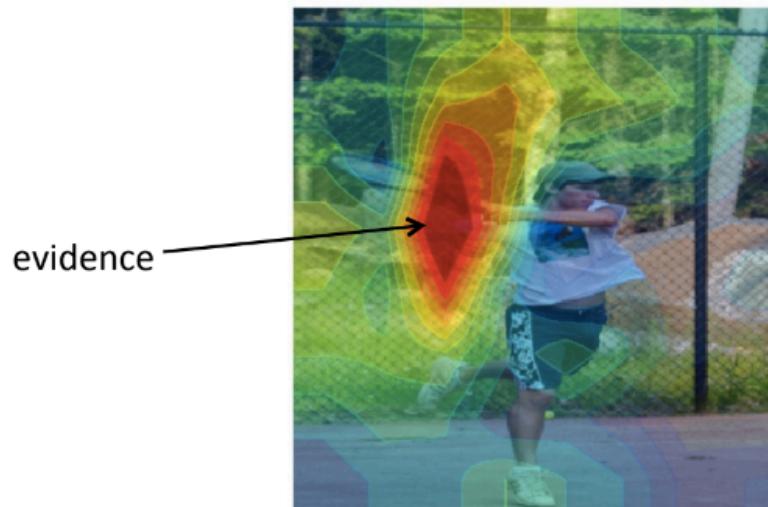
Slides: Kate Saenko
University of Boston

Problema #3

Automação de Legendas

Why do we need explainable models?

- Captioning models may base correct prediction on wrong evidence:



Model: *A man holding a tennis racquet on a tennis court.*

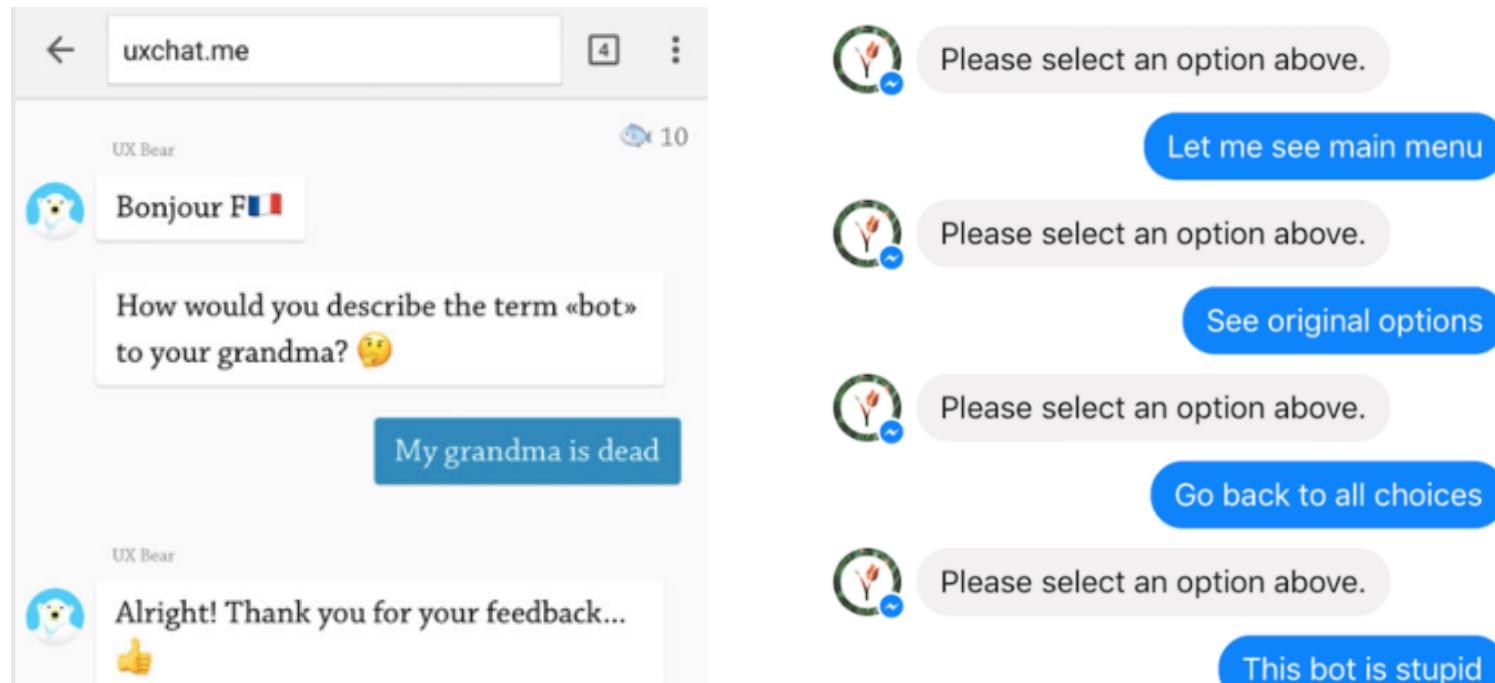
Right for the Wrong
Reasons

Slides: Kate Saenko
University of Boston

Problema #4

Problema #4

Automação de Diálogos



<https://slides.com/beatrizalbiero/o-que-sao-chatbots>

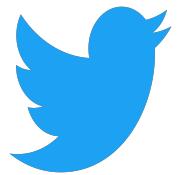
Problema #4

Automação de Diálogos

Desafios:

- Memória
- Identificação de Entidades
- Identificação de sentimentos
- Personalidade
- Entender as sutilezas da língua (Pragmática)

Vamos estudar NLP?



@b_albiero

Medium

<https://medium.com/@beatrizalbiero>



<https://www.linkedin.com/in/beatriz-albiero/>



Referências

Introdução a Linguística I e II – José Luis Fiorin

The Language Instinct – Steven Pinker

A Statistical contribution to Historical Linguistics - Rafael Bassi Stern

On Learning the Past Tenses of English Verbs - D. E. RUMELHART and J. L. McCLELLAND

E

fficient Estimation of Word Representations in Vector Space

Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean

Google Inc., Mountain View, CA

<https://taktotransformer.com/>

<https://www.quora.com/How-is-computational-linguistics-different-from-natural-language-processing>

Why Chatbots Fail <https://chatbot.fail/>

Kate Saenko Slides:

<https://drive.google.com/file/d/1G6DXv5JHtrvpuJtxbDLUJuKBrm6vJT8a/view>