

# **Vampiro que brilha... rá!**

## **Desafios na anotação de opinião em um corpus de resenha de livros**

Cláudia Freitas (PUC-Rio/Dep. Letras)

Eduardo Motta (PUC-Rio/Dep. Informática)

Juliana Cesar (PUC-Rio/Dep. Letras)

Ruy Luiz Milidiú (PUC-Rio/Dep. Informática)



# Corpus de **Resenhas** de **Livros**

publicadas na internet em um site colaborativo

## **Linguística**

- Texto não-jornalístico
- Linguagem não padronizada
- Linguagem avaliativa

## **Computacional**

- Identificação automática de opinião em textos

## **Motivação**

- Léxicos
  - Gerais
  - Específicos
- Regras
- Aprendizagem automática

CORPUS ANOTADO

**Abordagens para identificação automática da opinião**

## ReLi

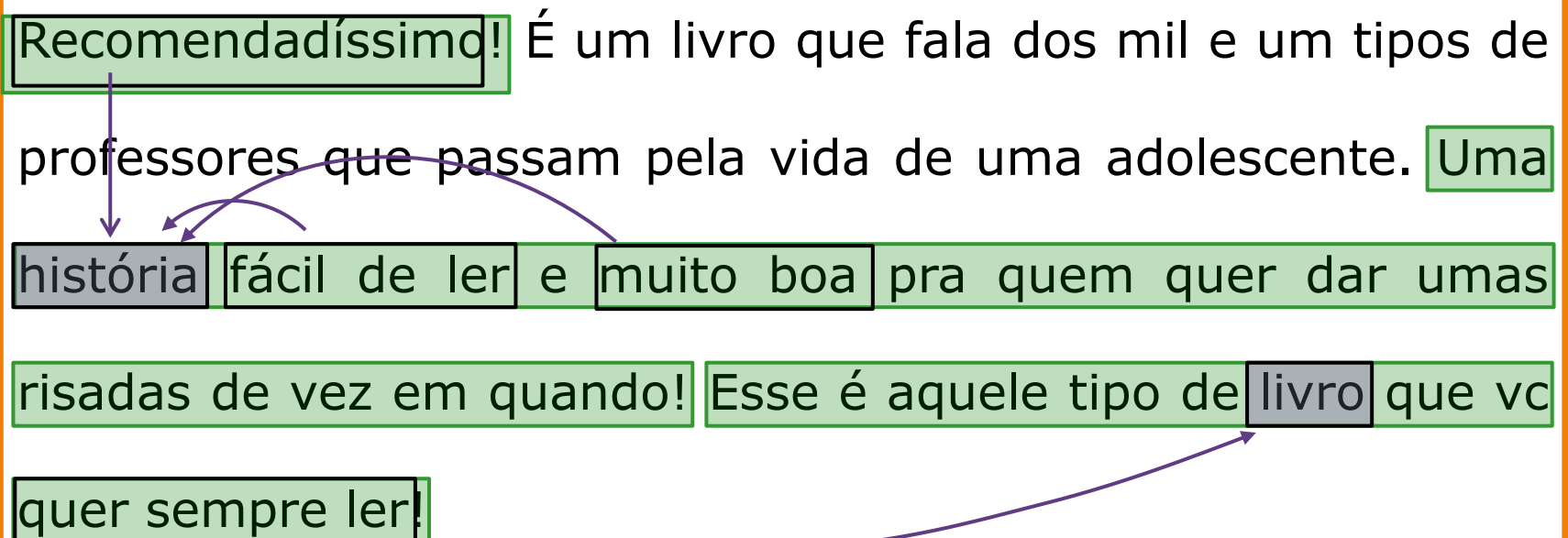
Autor	Título	Resenhas	Frases	Palavras
Stephenie Meyer	Crepúsculo	409	3.266	62.268
Thalita Rebouças	Fala sério, amiga!	161	910	16.864
	Fala sério, amor!			
	Fala sério, mãe!			
	Fala sério, pai!			
	Fala sério, professor!			
Sidney Sheldon	O Outro lado da meia noite	230	1.569	31.712
	O Reverso da Medalha			
	Se houver Amanhã			
Jorge Amado	Capitães da Areia	187	1.348	32.117
George Orwell	1984	202	2.228	51.320
José Saramago	Ensaio sobre a Cegueira	271	1.991	42.152
J.D. Salinger	O Apanhador nos Campos de Centeio	140	1158	23725
<b>TOTAL</b>		<b>1600</b>	<b>12.470</b>	<b>259.978</b>

- Objeto da opinião
  - Livro
  - Partes do livro: personagens, capítulos, narrativa, início, *parte romântica*...
- Opinião propriamente
  - Sintagmas: palavras ou expressões
- Polaridade
  - Positiva ou Negativa
    - Da frase
    - Do sintagma: o menor núcleo possível
  - A mesma frase pode conter várias opiniões, com polaridades diferentes.
  - A frase pode ter apenas uma polaridade.
  - SEMPRE considerada em contexto

**Anotação**

Eu ODIEI esse livro, simplesmente porque eu nunca mais vou ler um livro tão empolgante!

Recomendadíssimo! É um livro que fala dos mil e um tipos de professores que passam pela vida de uma adolescente. Uma história fácil de ler e muito boa pra quem quer dar umas risadas de vez em quando! Esse é aquele tipo de livro que vc quer sempre ler!



Gostei da história em si, ela me prendia, mas toda essa descrição me cansava.



- **Anotadores**
  - 3 -2- 1
  - Familiarização
    - Com a tarefa
    - Com a anotação
- **Criação de um manual**
  - Definição da tarefa
  - Regras
  - Exemplos
  - Casos difíceis
  - Dúvidas
- Discussão de dúvidas
- **Revisão do manual**
- Revisão da anotação
  - Busca por inconsistências
    - polaridade sem objeto
    - objeto sem polaridade
  - Revisão geral do conteúdo\*

# Processo de anotação



FileEditNavigateSearchProjectRunWindowHelp

Quick Access

JavaCo

Navigator

go\_1984029.cbt

go\_1984030.cbt

go\_1984031.cbt

go\_1984032.cbt

go\_1984033.cbt

go\_1984034.cbt

go\_1984035.cbt

go\_1984036.cbt

go\_1984037.cbt

go\_1984038.cbt

go\_1984039.cbt

go\_1984040.cbt

go\_1984041.cbt

ja\_capitaesdaareia001.cbt

ja\_capitaesdaareia002.cbt

ja\_capitaesdaareia003.cbt

ja\_capitaesdaareia004.cbt

ja\_capitaesdaareia005.cbt

ja\_capitaesdaareia006.cbt

ja\_capitaesdaareia007.cbt

ja\_capitaesdaareia007.cbt.rtg

ja\_capitaesdaareia008.cbt

ja\_capitaesdaareia009.cbt

ja\_capitaesdaareia010.cbt

ja\_capitaesdaareia011.cbt

ja\_capitaesdaareia012.cbt

ja\_capitaesdaareia013.cbt

ja\_capitaesdaareia014.cbt

ja\_capitaesdaareia015.cbt

ja\_capitaesdaareia016.cbt

ja\_capitaesdaareia007.cbt

\_BODY[30]\_ Por: Junior Nascimento Correndo o risco de parecer nostálgico, afirmo que essa é a minha resenha mais importante, por vários motivos. Capitães da Areia foi o primeiro livro que li inteiramente, sem nem saber o que era isso e de uma forma extremamente forçada (contrariando o aluno fã de literatura) em o auge de os meus 15 anos. Esta resenha é importante porque vou poder entregar uma resenha de verdade sobre uma obra que me marcou. Depois de sendo dominado por o vício de a leitura e devo isso a o mestre Jorge Amado. Mas gostaria de deixar registrada a minha felicidade por a Cia das Letras me proporcionar importante para mim. Apesar de suas quase trezentas páginas, que demoraria normalmente acabar demorando mais de 1 semana para reler a obra e isso se deve a um único motivo para "devorá-lo" e sim para apreciá-lo e foi exatamente isso que eu fiz durante esse único livro que cheguei a ler um único capítulo mais de 3 vezes. Como disse em o início extremamente nostálgico. Jorge Amado em os apresenta a os Capitães de a Areia extremamente organizados e amigos. Durante a leitura é impossível não se colocar afinal, apesar de os pesares são crianças comuns que foram forçadas a amadurecer barreiras familiares, sem limites, sem carinho, sem um lar. Isso fica claro em determinados eles esperam é um colo, alguém que cuide de eles, ou simplesmente poder passar como as demais crianças. Em contrapartida o livro é extremamente crítico quanto a políticas e uma mídia manipuladora de a verdade (isso fica claro logo em os primeiros isso que torna o livro atemporal, todas as mazelas de a sociedade ainda hoje é visíveis esses meninos de a realidade atual e que hoje temos um agravante, as drogas ilicítas apresentados por Jorge Amado sejam "limpos", porém não temos esse contato direto com as crianças de rua são em grande maioria marionetes de traficantes, os Capitães da Areia malandragem. Sabe aquele tipo de livro que em os faz colocar o pé em o chão e o que me sinto a o ler Capitães de a Areia. Esta resenha possui mais um motivo para

Segment Features (Column format)

word	pos	ck	object	opinion	polarity	help
livro	N	I-NP	OBJ00		+	
não	ADV	I-NP		op00+	+	
foi	VAUX	B-VP		op00+	+	
feito	PCP	I-VP		op00+	+	
para	PREP	B-PP		op00+	+	
"	"	I-NP		op00+	+	

Segment Annotations

Segment Relations

Add op05- annotation

Add HELP annotation

Add OBJ03 annotation

Add OBJ04 annotation

Add op05+ annotation

Add OBJ01 annotation

Add OBJ02 annotation

Add op03+ annotation

Add OBJ05 annotation

Add op04+ annotation

Add + annotation

Add op04- annotation

Add op03- annotation

Add OBJ00 annotation

Add - annotation

Add op00+ annotation

Add op02- annotation

Add op01+ annotation

Add op00- annotation

Add op02+ annotation

Add op01- annotation

Add ARQX annotation

Remove + annotation

Export artifact as XML golden standard...

Export artifact as Duke...

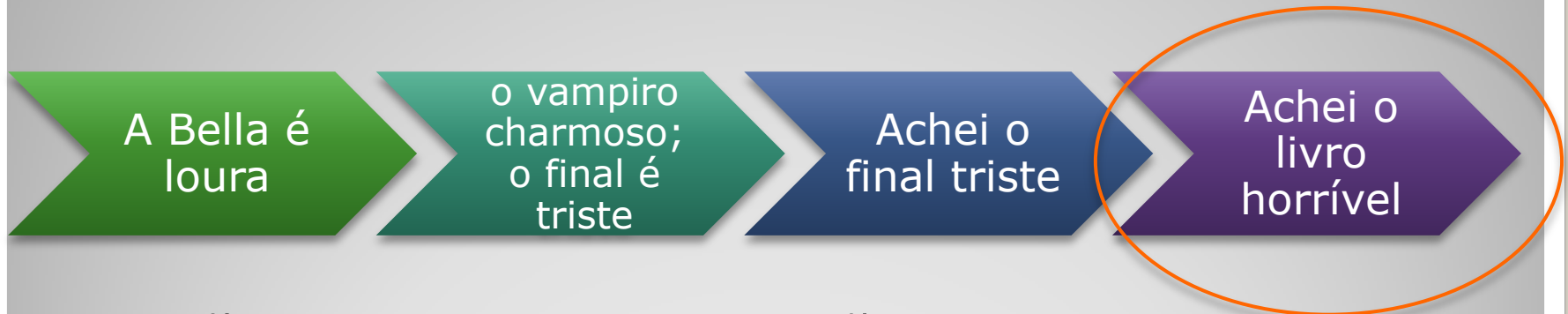
Close artifact and move it to target folder



- Opinião vs Fato/Descrição

Gostou  
do livro?

- Só consideramos opinião com polaridade




- Não consideramos *opiniões neutras*

**Dificuldades  
e opções da anotação (1)**

- (...) o imaginário encontro com um vampiro *todo charmoso* não é algo que me desperte a vontade de acompanhar a história.
- Eis o início desta trama sobrenatural. Um vampiro charmoso, cavaleiro, um digno lorde, que se apaixona por uma “pobre” mortal.
- Uma das leituras mais *complexas* e *difíceis* que eu tive.
- Continuo achando a Bella Swan uma **Emo idiota** e **oferecida**.
- Assim começa "O Apanhador em o Campo de Centeio", um **verdadeiro** *clássico* de a literatura mundial.

## Dificuldades e opções da anotação

- Quais são as “partes do livro” ?
  - O diálogo com o professor Antolini **melhor parte**
  - A cena do campeonato de xadrez **minha preferida**
  - **Gostei do** trecho em que o personagem Holden revela o fato que serve de ...
- Comparações
  - Livro vs filme
  - Livro vs saga

não  
anotamos

## Dificuldades e opções da anotação (2)

- Determinação da polaridade
  - É claro que o livro ***não é perfeito***, mas é muito bom.[+]
  - É um livro **bom**, mas **não é excelente**. [+]
  - É ***apenas um bom*** livro.
  - **Tenso**. É o único livro que me deixou ***nervoso*** e ***apreensivo*** enquanto lia.
  - **Bom**, mas ***meio melancólico***.

**Dificuldades e opções  
da anotação (3)**

- Anáforas

- Frequentemente eu me **decepçãoo** com best-sellers. Talvez por conta das expectativas criadas sobre obras já muito comentadas.

**E isso** aconteceu novamente, com **Crepúsculo**.

- Perguntas

- Se o livro tem méritos? Eu diria que sim.
- O que falta para uma boa história? Nada!

**Dificuldades e opções  
da anotação (4)**

- 2 conjuntos de dados
  - 390 resenhas; 48.000 palavras (anotadores A e B)
    - (subconjunto de) 107 resenhas; 1200 palavras (anotadores A, B e C)
- Concordância:
  - Frases selecionadas
  - Polaridade das frases selecionadas
  - Objetos selecionados
  - Opiniões selecionadas
  - Polaridade das opiniões selecionadas

## Concordância entre anotadores

- **Concordância:**
  - Objetos selecionados
  - Opiniões selecionadas

→ Identificação da mesma opinião (ou objeto), mas divergência quanto aos limites da unidade.

Wiebe et al.[2005] : MPQA Opinion Corpus  
equivalência entre “parte final” | “final” |  
“parte”

- “agr”: que proporção de A também foi marcada por B.

**Concordância entre  
anotadores**



Anotadores	IDENTIFICAÇÃO			ATRIBUIÇÃO DE POLARIDADE	
	Frase	Objeto	Opinião	Frase	Sintagma
A B	82.9	69.9	73.9	98.1	99.9
B A	87.2	72.3	83.1		
<b>Média A B</b>	<b>85</b>	<b>71.1</b>	<b>78.5</b>		
A C	77.7	74.4	78.6	98.4	99.6
C A	86.1	76.1	84.5		
<b>Média A C</b>	<b>81.9</b>	<b>75.3</b>	<b>81.5</b>		
B C	78.5	72.9	84.6	98.4	100
C B	78.4	69.6	75.5		
<b>Média B C</b>	<b>78.5</b>	<b>71.3</b>	<b>79.8</b>		

**Resultados da  
concordância**

	<b>ReLi*</b>	<b>MPQA*</b> (Wiebe et al. 2005)	<i>Pang &amp; Lee</i> (2004)	<b>Zagibalov et al.</b> (2010)	<b>SentiCorpus-PT</b> (Carvalho et al. 2011)
Conteúdo	resenhas de livros	artigos de jornal	resenhas de filmes	resenhas de livros	posts de leitores de jornal
idioma	português	inglês	inglês	ing. e russo	português
Resenhas	<b>1.600</b>	535 (artigos)	2.000 (1000 +; 1000 -)	1.500 (cada corpus)	2.795 posts
Palavras	<b>260.000</b>	---	---	7.500	---
Frases	<b>12.470</b>	10.000	10.000 (5.000 <i>subj</i> ; 5.000 <i>obj.</i> )	750 resenhas +/ 750 resenhas - (cada corpus)	8.000
Frases [+]	<b>2.883</b>	---			3.888 frases com opinião
Frases [- ]	<b>596</b>	---			
Frases mistas	<b>212</b>	---			
Opin. [+]	<b>4.210</b>	---	---	---	
Opin. [-]	<b>1.024</b>	---	---	---	

\* Anotação manual

- txt
- POS (96.7% precisão)
- Chunk (NP: 88%)

#45

## MASK:False - REVS:1 - DOC:bc4583177 - S

_BODY[22]_	0	0	0	0	0	0	0	0
noossa	ADJ	I-NP	0	0	+	0	0	0
,	,	B-NP	0	0	+	0	0	0
personagens	N	B-NP	OBJ01	0	+	0	0	0
incríveis	ADJ	I-NP	0	op01+	+	0	0	0
!	!	0	0	0	+	0	0	0

ReLi

# ReLi: breve exploração

XI ELC - São Carlos, Setembro  
2012

- Personagens
- Capítulos
  - Cada capítulo; capítulos finais; últimos capítulos; começo; início; fim; etc
- História
- Cenário
- Conteúdo
- Descrições
- Desenvolvimento
- Diálogos / Falas
- Enredo
- Escrita
- Estilo
- Frases
- Leitura
- Linguagem
- Narrativa
- Ritmo
- Tema
- Texto
- Título
- Trama
- Vocabulário
- Trechos/partes
  - "outros"

## **Taxonomia do livro (ficção)**

- **Lamentável** tê-lo lido somente agora.
- O livro é **sufocante**.
- Passei dias me recuperando do **soco no queixo** que tomei ao terminar o livro. Fui à nocaute, sem direito a (re)contagem. **Chorei. Fiquei meio deprê.**
- *Historia **água com açúcar** e **bem comportada*** sobre uma princesa de um reino fictício os dias atuais - bem, a gente tem de ler o que nossos filhos leem.

**“Falsas pistas” lexicais**

- Vampiro que brilha... rá! Daqui a pouco ele vive de tofu!
- Depois de atravessar as 380 páginas dessa *meleca*, tive que tomar café pra virar macho.
- Se não houver coisa melhor.
- ... a tradução deveria ser CRAPúsculo .
- Lobisomens depilados?
- Stephenie Meyer acabou com o mito do vampiro, depois desse livro só o que Conde Vlad deseja é ter uma estaca de madeira no peito.
- Tudo o que vi foi blábláblá os vampiros eram lindos blábláblá e eram alvos como o leite blábláblá e Edward era a coisa mais linda que Bella já vira e blábláblá eram realmente brancos blábláblá ...

**Muita criatividade!**



- Verbo SOFRER
  - **Sofri** para *terminar* o livro. [-]
  - Nunca **sofri** tanto para *ler* um livro. [-]
  - **Sofri** com a *protagonista* a cada nova página. [+]
  - Eu **sofria** a cada vez que tinha que adiar a leitura. [+]
  - simplesmente **sofri** até a ultima página louca pra terminar .[+]

## Opinião e emoções

- Criação de regras (?)
  - “Como expressamos a opinião em português?”
- Argumentação e adversativas
- O léxico afetivo do português
- Exploração dos atributos associados a
  - pessoas
  - Objetos

## Caminhos futuros

- Subestimação da tarefa
  - Difícil para as pessoas.. E para as máquinas?
- “erro é inerente ao processo de anotação manual”
  - Disponível na semana que vem em

## Considerações Finais

- Bo Pang & Lillian Lee. *A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts*, Proceedings of ACL 2004.
- Carvalho, Paula, Luís Sarmento, Jorge Teixeira, Mário J. Silva: Liars and Saviors in a Sentiment Annotated Corpus of Comments to Political Debates. ACL (Short Papers) 2011: 564-568
- Fernandes, E., Milidiú, R., Santos, C., 2009, "Portuguese Language Processing Service", In: 18th International World Wide Web Conference. Madrid, 2009
- Liu, Bing, Hu, Minqing e Cheng, Junsheng. (2005) "Opinion Observer: Analyzing and Comparing Opinions on the Web". Proceedings of the 14th international World Wide Web conference (WWW-2005). Chiba, Japan.
- Wiebe, Janyce, Wilson, Theresa e Cardie, Claire (2005). Annotating expressions of opinions and emotions in language. Language Resources and Evaluation, volume 39, issue 2-3, pp. 165-210.
- Zagibalov, Taras, Belyatskaya, Katerina e Carroll, John. (2010) "Comparable English-Russian book review corpora for sentiment analysis". In Proceedings of the 1st Workshop on Computational Approaches to Subjectivity and Sentiment Analysis.

## Referências