

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

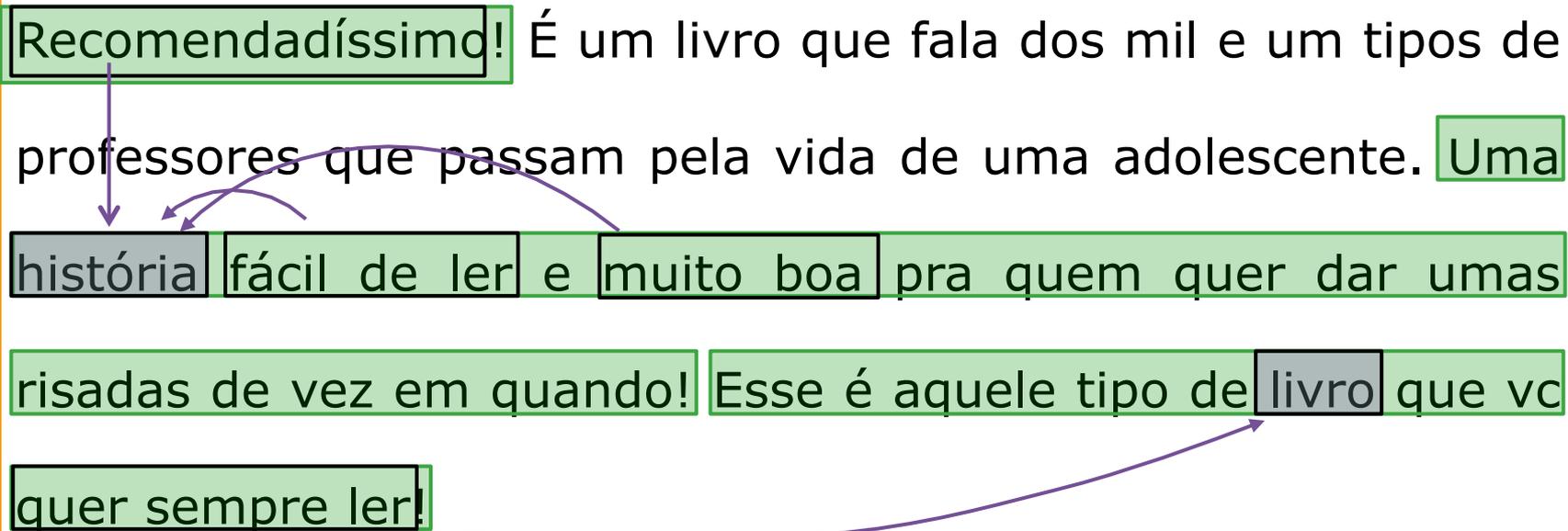
Autor	Título	Resenhas	Frases	Palavras
Stephenie Meyer	Crepúsculo	409	3.266	62.268
Thalita Rebouças	Fala sério, amiga!	161	910	16.864
	Fala sério, amor!			
	Fala sério, mãe!			
	Fala sério, pai!			
	Fala sério, professor!			
Sidney Sheldon	O Outro lado da meia noite	230	1.569	31.712
	O Reverso da Medalha			
	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
TOTAL		1600	12.470	259.978

- 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
 - 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*

Opinião indireta:

Quando comecei a ler não consegui mais parar.

Parei na metade do livro.

Não tem como descrever tudo o que o livro aborda.

Processo de anotação

- go_1984029.cbxt
- go_1984030.cbxt
- go_1984031.cbxt
- go_1984032.cbxt
- go_1984033.cbxt
- go_1984034.cbxt
- go_1984035.cbxt
- go_1984036.cbxt
- go_1984037.cbxt
- go_1984038.cbxt
- go_1984039.cbxt
- go_1984040.cbxt
- go_1984041.cbxt
- ja_capitaesdaareia001.cbxt
- ja_capitaesdaareia002.cbxt
- ja_capitaesdaareia003.cbxt
- ja_capitaesdaareia004.cbxt
- ja_capitaesdaareia005.cbxt
- ja_capitaesdaareia006.cbxt
- ja_capitaesdaareia007.cbxt
- ja_capitaesdaareia007.cbxt.rtx
- ja_capitaesdaareia008.cbxt
- ja_capitaesdaareia009.cbxt
- ja_capitaesdaareia010.cbxt
- ja_capitaesdaareia011.cbxt
- ja_capitaesdaareia012.cbxt
- ja_capitaesdaareia013.cbxt
- ja_capitaesdaareia014.cbxt
- ja_capitaesdaareia015.cbxt
- ja_capitaesdaareia016.cbxt

ja_capitaesdaareia007.cbxt

BODY[30] Por: Junior Nascimento Correndo o risco de parecer nostálgico, afirmo que essa é a minha resenha ma importante, por vários motivos. Capitães da Areia foi o primeiro livro que li inte fiz, sem nem saber o que era isso e de uma forma extremamente forçada (contraria aluno fã de literatura) em o auge de os meus 15 anos. Esta resenha é importante ju vou poder entregar uma resenha de verdade sobre uma obra que me marcou. Depois q 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 proporci importante para mim. Apesar de suas quase trezentas páginas, que demoraria nori acabei demorando mais de 1 semana para reler a obra e isso se deve a um único mo para "devorá-lo" e sim para apreciá-lo e foi exatamente isso que eu fiz durante es unico livro que cheguei a ler um unico capitulo mais de 3 vezes. Como disse em o i extremamente nostálgico. Jorge Amado em os apresenta a os Capitães de a Areia extremamente organizados e amigos. Durante a leitura é impossivel não se colocar afinal, apesar de os pesares são crianças comuns que foram forçadas a amadure barreira familiar, sem limites, sem carinho, sem um lar. Isso fica claro em determ eles esperam é um colo, alguém que cuide de eles, ou simplesmente poder passar u como as demais crianças. Em contrapartida o livro é extremamente critico quanto a políticas e uma midia manipuladora de a verdade (isso fica claro logo em os prim isso que torna o livro atemporal, todas as mazelas de a sociedade ainda hoje é vis esses meninos de a realidade atual e que hoje temos um agravante, as drogas il apresentados por Jorge Amado sejam "limpos", porém não temos esse contato direto d crianças de rua são em grande maioria marionetes de traficantes, os Capitães da A 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+	+	
"	"	"		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. [+]
 - *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. [-]
 - **Bom**, mas **meio melancólico**.

Dificuldades e opções da anotação (3)

- Anáforas

- Frequentemente eu me **decepçiono** 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		
B A	87.2	72.3	83.1	98.1	99.9
Média A B	85	71.1	78.5		
A C	77.7	74.4	78.6		
C A	86.1	76.1	84.5	98.4	99.6
Média A C	81.9	75.3	81.5		
B C	78.5	72.9	84.6		
C B	78.4	69.6	75.5	98.4	100
Média B C	78.5	71.3	79.8		

Resultados da concordância

	ReLi*
Conteúdo	resenhas de livros
idioma	português
Resenhas	1.600
Palavras	260.000
Frases	12.470
Frases [+]	2.883
Frases [-]	596
Frases mistas	212
Opin. [+]	4.210
Opin. [-]	1.024

- 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
<u>noossa</u>	ADJ	I-NP	0	0	+	0	0	0
,	,	B-NP	0	0	+	0	0	0
personagens	N	B-NP	OBJ01	0	+	0	0	0
<u>incríveis</u>	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.**
- **Tenso**. É o único livro que me deixou **nervoso** e **aprensivo** enquanto lia.

“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 última página louca pra terminar .[+]

Variação na polaridade

- 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 em poucas semanas em
http://learn.inf.puc-rio.br/index.php/Datasets_and_Corpora

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

- Opinião desfavorável (emPB) → português para estrangeiros

- Com uma narrativa **fácil** e **contagante** , você é levado a viver a paixão de Bella' . É uma leitura **leve** , **gostosa** . Mas **perde pontos** por a **perfeição** de Edward , tudo bem.
- Enredo mirabolante, trama envolvente mas chega a haver exagero .
- Polaridade do mas
 - Gostei de o livro , leitura simple , mas esperava mais emoção e ação a história é muito parada .

Variação na polaridade

definição mencionada deixa claro que está vinculada, ao termo opinar, a noção de julgamento de valor, a qual, por sua vez, já estava presente na palavra latina *opīno*□, *ās*, *āre*, cujo sentido era 'julgar, pensar, crer' (Torrinha: s.d., p. 588).

Vale destacar que os trabalhos de Pomerantz (1984) e Mulkey (1985) abordam casos do inglês, enquanto Oliveira (1996), Bastos e Pereira (1998) e Almeida (2003) apresentam dados do português, buscando todos identificar estratégias empregadas na elaboração da avaliação ou opinião. É também digno de nota o fato de que esses trabalhos, por se dedicarem a questões de concordância e discordância, estão voltados para a análise das segundas avaliações ou opiniões.

endo em vista o exposto nesta sessão intitulada "Estudos sobre a expressão da opinião", foi possível verificar que a elaboração da opinião em português do Brasil é, ainda hoje, assunto de um número reduzido de pesquisas, sendo estas, na maior parte das vezes, direcionadas para o estudo das segundas opiniões.