

VAMPIRO QUE BRILHA... RÁ! DESAFIOS NA ANOTAÇÃO DE OPINIÃO EM UM CORPUS DE RESENHAS DE LIVROS

Cláudia Freitas (PUC-Rio), Eduardo Motta (PUC-Rio), Ruy Luiz Milidiú (PUC-Rio)
Juliana César (PUC-Rio)

1. Introdução

Dentre as tarefas do processamento automático de linguagem natural (PLN), a extração de informação em textos tem, relativamente, uma longa tradição. De maneira geral e simplificada, a extração de informação busca, explorando grandes volumes de texto, responder a perguntas como *quem; o quê; quando; onde; para quem*. No entanto, textos não veiculam apenas informação factual. Opiniões, impressões e avaliações também constituem boa parte do material textual, e são um tipo de informação de grande valia. Nesse contexto, a Linguística Computacional/PLN tem se debruçado recentemente na identificação de aspectos subjetivos da linguagem, destacando-se a identificação automática de opinião, ou de sentimento, em textos.

Do lado linguístico, especificamente de uma linguística que busca descrever a língua a partir de ocorrências em corpus, também se tem verificado o interesse na construção de corpora de textos outros que não o jornalístico.

O presente trabalho descreve a construção do ReLi – um corpus de **Resenhas de Livros** manualmente anotado quanto à expressão de opinião. Especificamente, relatamos as opções linguísticas tomadas durante o processo de anotação, apresentamos os resultados de testes de concordância e exploramos brevemente o corpus criado.

A criação do ReLi foi duplamente motivada: do lado dos estudos da linguagem, um corpus composto por textos produzidos na internet, não jornalísticos, se justifica pelo crescente interesse na descrição da linguagem utilizada na rede, não padronizada, além de contribuir para estudos com base em corpus que tratem da linguagem avaliativa. Do lado do PLN, a criação de um corpus anotado quanto à expressão de opinião interessa principalmente à Mineração de Opinião e à Análise de Sentimento, que lidam com a identificação de opiniões, avaliações e atitudes com relação a entidades como pessoas, produtos e organizações, expressas em texto.

Na área de Mineração de Opinião, a maioria dos trabalhos parte de léxicos de emoções/sentimentos, que podem ser gerais, como SentiWordNet (Esuli & Sebastiani 2006), ou específicos para determinados domínios ou tarefas (Riloff & Wiebe 2003; Poirier et al. 2011). Para o português, destacamos o Senti-Corpus (Carvalho et al. 2011) e Senti-Lex (Silva et al. 2012), corpus e léxico construído a partir da anotação de comentários (posts) em matérias sobre eleições, no domínio da política e em uma escrita também típica da internet.

No entanto, são poucos os trabalhos que consideram corpora anotados, devido principalmente ao seu alto custo de elaboração. O MPQA Opinion Corpus (Wiebe et al. 2005) é o corpus que mais se assemelha ao apresentado neste trabalho. Trata-se de um corpus de 10 mil frases de artigos de jornal, ricamente anotado no nível da palavra e do sintagma, com informação relativa a estados mentais e emocionais, visando à distinção entre informação subjetiva de informação factual. Especificamente com relação à identificação de opinião em avaliações de produtos, destacamos Zagibalov et al. (2010), que criaram corpora comparáveis em inglês e russo a partir de resenhas de livros publicadas na Amazon.com. Os corpora foram manualmente anotados quanto à polaridade geral do documento, isto é, cada resenha foi considerada positiva ou negativa, e cada corpus contém 1500 resenhas anotadas.

Liu et al. (2005) classificam as avaliações de produtos em 3 tipos: nas do tipo (1) é pedido ao avaliador que escreva os prós e contras separadamente; no tipo (2), além dos prós e contras, é pedido ao avaliador que escreva uma avaliação detalhada; o tipo (3) tem formato livre, e não há separação formal entre prós e contras.

As resenhas que compõem o ReLi se enquadram no tipo (3), e foram extraídas do sítio Skoob.com, uma rede social de livros e leitores, na qual os leitores/colaboradores participam de forma ativa e entusiasmada comentando os livros que leram.

Por estarmos lidando com resenhas, esperávamos encontrar textos apenas com opinião, ou com uma delimitação relativamente clara entre opinião e descrição/informação, mas não foi o que aconteceu. Há resenhas que contêm apenas resumos do livro. Há resenhas em que há apenas uma ou duas frases com opinião, e todo o resto é um resumo. Há resenhas em que a (escassa) opinião ocorre ao lado de divagações pessoais e opiniões

variadas que não têm como alvo o livro resenhado. Resenhas livres são imprevisíveis, e os autores de resenhas podem ser muito mais criativos do que gostaríamos.

As opiniões anotadas no corpus referem-se exclusivamente às opiniões sobre o livro resenhado. Assim, opiniões sobre outros livros, ou sobre filmes, por exemplo, não foram consideradas. Também não consideramos opiniões que envolviam comparações, mesmo que um dos objetos comparados fosse o livro resenhado. Sabemos que essa foi uma decisão arbitrária, mas nada impede que no futuro ampliemos a anotação para considerar a opinião em estruturas de comparação.

2. O corpus

O corpus é composto por 1600 resenhas de 13 livros (7 autores), totalizando cerca de 260 mil palavras e 12 mil frases. Para cada livro coletamos cerca de 200 resenhas e, quando esse número não pôde ser atingido, completamos com outras obras do mesmo autor até chegarmos a um número próximo a 200. A tabela 1 apresenta a composição do ReLi, com a distribuição dos livros e autores.

A escolha pelos livros e autores foi motivada pela quantidade de resenhas disponíveis por livro. No entanto, alguns autores/livros, como Machado de Assis, foram descartados devido ao grande número de resenhas idênticas. A variação na escolha dos livros reflete-se também na variação da linguagem utilizada: há desde registros altamente informais e com grande uso de gírias, abreviações, neologismos e emoticons, até resenhas mais formais, com um vocabulário mais rebuscado.

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

Tabela 1: Composição do ReLi

2. 1. Opções da anotação

Um primeiro ponto fundamental diz respeito ao alvo da opinião. No ReLi anotamos, exclusivamente, opiniões relativas ao livro ou parte dele, como *personagens*, *capítulos*, *linguagem*, etc. Opiniões relativas a outros objetos não foram consideradas.

A opinião nas resenhas foi anotada no nível da frase e do sintagma, e acontece em 3 etapas: (i) identificação e anotação da polaridade das frases que contêm opinião; (ii) identificação do alvo da opinião; (iii) identificação e anotação da polaridade do trecho (palavras ou expressões) que contém opinião. Consideramos apenas opiniões positivas ou negativas, e não consideramos opiniões neutras.

Quando a frase não mencionava explicitamente o objeto da opinião (por exemplo, “*Adorei!*”), consideramos, por padrão, que a opinião será sempre sobre o livro.

Com a anotação no nível do sintagma, conseguimos lidar com a presença de diferentes opiniões, de diferentes polaridades, em uma mesma frase. No entanto, no nível da frase, só consideramos uma polaridade. No exemplo (1), temos dois alvos de opinião: *história* e *descrição*. A cada um desses alvos estão associadas diferentes opiniões e polaridades. “*Gostei*” e “*prendia*” são positivas, e relacionam-se à *história*; “*cansava*” é negativa, e refere-se à *descrição*. No entanto, apesar da presença de opiniões com polaridades distintas, interpretamos a frase como, de maneira geral, favorável ao livro. Por isso, no nível da frase, a polaridade atribuída é positiva, ainda que, no nível do sintagma, exista uma opinião negativa.

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

Por fim, a atribuição da polaridade é sempre feita em contexto.

2.2. Dificuldades da anotação

A grande dificuldade deste tipo de tarefa está na (tentativa de) separação entre o conteúdo apresentado como factual e o que é considerado opinião. Diferentemente de Wiebe et al. (2005), que anotam a opinião em notícias, esperávamos que, em resenhas,

gênero claramente opinativo, a separação entre conteúdo factual/descritivo e conteúdo opinativo não fosse um grande dificultador. No entanto, não raro era muito difícil decidir se estávamos diante de uma descrição do personagem, de alguma constatação sobre o livro ou de uma opinião sobre ele.

2. O vampiro é o cara. *Altruísta, elegante, charmoso*, não perde a compostura nunca...
3. (...) o imaginário encontro com um *vampiro todo charmoso* não é algo que me desperte a vontade de acompanhar a história.
4. Eis o início desta trama sobrenatural. Um *vampiro charmoso*, cavaleiro, um digno lorde, que se apaixona por uma “pobre” mortal.
5. Continuo achando a Bella Swan uma Emo *idiota e oferecida*.
6. É um livro *estranho*, tudo se passa tão rápido.

Em (2), não temos como saber se há opinião ou descrição do personagem, embora a primeira frase indique a existência de um sentimento positivo. Em (3), temos a impressão de que ser charmoso é uma característica do personagem, e portanto não consideramos opinião. Em (4), estamos claramente diante de uma descrição do vampiro. Já em (5) não há dúvidas de que “*idiota*” e “*oferecida*” são opiniões sobre a personagem, e em (6), por outro lado, temos claramente uma opinião, mas sem polaridade.

Frequentemente é muito difícil distinguir entre declarações mais factuais (“*o final é triste*”) de informações mais subjetivas (“*achei o final triste*”). Por isso, durante a anotação, assumimos uma posição conservadora e quando não conseguimos saber se o trecho relata uma característica do livro ou uma opinião, optamos por não anotar. Assim, mesmo que “*charmoso*” seja frequentemente considerado um elemento de linguagem expressiva, é apenas no seu uso que poderemos confirmar essa característica (que, no nosso caso, não se confirma). Segundo nossa posição conservadora, só consideramos opinião – e, portanto, anotamos – frases que, de alguma maneira, respondiam à pergunta hipotética “*gostou do livro/dessa parte do livro*”? Assim, por exemplo, se a resposta dada à pergunta é “*achei triste/é triste/a leitura é complexa*”, não temos como saber se a pessoa gostou. Já se a resposta é “*achei horrível/é horrível*”, temos claramente uma opinião negativa.

Outra decisão que precisamos tomar se refere aos limites do que seria considerado parte do objeto e, portanto, alvo da opinião. Se, por um lado, é previsível a ocorrência de partes generalizáveis no domínio dos livros, como *capítulos, personagens, linguagem*, por outro, também foi frequente a menção a partes específicas de cada livro, como “*parte romântica*”, em (7):

7. “Não contente em derrapar na *parte romântica*, Stephenie Meyer também resolve deturpar o clássico e marcante vampiro...”.

Embora nossa primeira intenção tenha sido desconsiderar tais objetos justamente por sua especificidade com relação ao livro, logo percebemos que estávamos impondo um limite talvez artificial à tarefa de identificação de opinião em livros, isto é, estávamos tomando uma decisão arbitrária tendo em vista somente a facilidade na anotação. Assim, optamos por considerar a opinião sobre o livro ou partes dele, por mais específicas que fossem.

Com relação à segmentação, nem sempre é simples (ou possível) identificar o trecho exato da opinião. Nesses casos, a preferência é por anotar todo o trecho (ou frase), como ilustram os exemplos (8 - 10):

8. Mas em um contexto todo, o livro *conseguiu suprir minhas expectativas*.
9. *impossível abandonar o livro pela metade*
10. *a tradução deveria ser CRAPúsculo*.

Ainda que pouco frequentes, a presença de opinião em perguntas foi outra dificuldade encontrada no processo de anotação. Nesses casos, que decidimos não anotar, a informação relativa à opinião/polaridade saía do âmbito da frase, como ilustram os exemplos 11-13:

11. Se o livro tem méritos? Eu diria que sim.
12. O que falta para uma boa história? Nada!
13. Ler os outros?!? Nem se eu tivesse a eternidade toda pela frente...

2.2 A determinação da polaridade

Como mencionado, a polaridade só poderia ser atribuída em contexto. Contudo, mesmo com as pistas contextuais, houve casos em que identificar se estávamos diante de uma opinião positiva ou negativa não era trivial, como ilustram os exemplos 14-16: dependia

de conhecimentos prévios e, principalmente, de uma boa capacidade de leitura dos anotadores:

14. É claro que o livro não é perfeito, mas é muito bom.[+]

15. É um livro bom, mas não é excelente. [+]

16. Bom, mas meio melancólico. [+]

2.3 O processo de anotação

Anotadores – e a afinação entre eles - são um aspecto fundamental do processo de anotação, principalmente quando esta é dependente de interpretação. O corpus foi anotado, inicialmente, por 3 anotadores – A, B e C – , embora tenha terminado com apenas 1. Todos os anotadores eram alunos de graduação do curso de Letras, sem experiência com a tarefa de anotação, mas interessados pelo tema das resenhas (livros). Todos passaram por um processo de treino até que estivessem familiarizados com a tarefa, com as instruções e com a ferramenta de anotação, sendo enfatizada a importância das interpretações serem sempre feitas no contexto particular de cada resenha. Durante o processo de treino e também durante todo o processo de anotação, os anotadores foram encorajados a perguntar e discutir suas opções e, à medida que surgiam casos não previstos, as soluções eram discutidas e incorporadas ao manual. O manual elaborado para a tarefa, portanto, passou por uma revisão depois dos primeiros textos anotados, com refinamento das explicações, incorporação de dúvidas, exemplos e casos difíceis (Freitas & Cesar, 2012). A ferramenta de anotação - adaptada uma ferramenta já desenvolvida pelo grupo – dispunha também de uma opção *help*, onde os anotadores poderiam selecionar frases ou trechos que consideraram difíceis.

Terminada a anotação, o corpus passou por um processo de revisão, em que buscamos por inconsistências. Uma revisão geral do conteúdo também foi feita por um dos autores deste artigo.

3. Estudo da concordância entre anotadores

Depois de cerca de 400 resenhas anotadas, realizamos um estudo da concordância entre os anotadores. O estudo considerou 2 conjuntos de dados: (i) a anotação de 2 anotadores em 390 resenhas (48 mil palavras); (ii) a anotação dos 3 anotadores em um subconjunto de 107 resenhas (1200 palavras).

Os seguintes pontos foram considerados na avaliação da concordância: concordância quanto às frases selecionadas; concordância quanto à polaridade das frases selecionadas; concordância quanto aos objetos selecionados; concordância quanto às opiniões selecionadas; concordância quanto à polaridade das opiniões selecionadas.

Embora as instruções de anotação dessem alguma informação quanto à segmentação, era de se esperar alguma variação quanto à extensão das unidades selecionadas. Um dos desafios, portanto, esteve em definir a concordância nos casos em que os anotadores identificaram a mesma opinião ou alvo da opinião, mas divergiram quanto aos limites da unidade.

De fato, trata-se de uma tarefa cuja avaliação é mais complexa que o julgamento de atribuição de polaridade a sentenças, e nos apoiamos no processo de avaliação de Wiebe et al. (2005) em dois pontos fundamentais: (i) consideramos expressões como *parte final* e *final* expressões equivalentes, e (ii) utilizamos a métrica *agr*, que tem como objetivo avaliar se os anotadores identificaram o mesmo conjunto de objetos e de opiniões, isto é, que proporção do que o anotador A anotou também foi anotada por B. A tabela 2 apresenta os resultados do estudo da concordância entre os anotadores A, B e C.

Separamos a avaliação em dois grupos: a concordância quanto à identificação (se os anotadores identificaram os mesmos sintagmas ou frases, tanto do alvo da opinião quanto da opinião) e quanto à atribuição de polaridade (havendo concordância quanto ao trecho selecionado, se há concordância quanto à polaridade atribuída).

Como era de se esperar, os resultados da atribuição de polaridade são bem mais altos que os da identificação. Ou seja, uma vez que há concordância quanto ao trecho identificado, os anotadores concordaram em quase todos os casos se estavam diante de uma opinião positiva ou negativa. A análise qualitativa dos resultados revelou que os raros casos de discordância na atribuição de polaridade na frase devem-se às adversativas/concessivas, em que a mesma frase contém opiniões com polaridades distintas, e a discordância está atribuição da polaridade da frase. Quanto a divergências no sintagma, trata-se de uma única discordância, relativa à polaridade do adjetivo “mirabolante”.

Pelos resultados da concordância na identificação, vemos que o anotador A tende a considerar mais elementos na anotação. No entanto, dada a complexidade da tarefa, os resultados gerais da concordância são bastante altos. Embora o tipo de anotação não seja exatamente o mesmo, os resultados da concordância entre os anotadores de Wiebe et al. (2005) quanto à identificação do sintagma de opinião são de 72%, o que corrobora a ideia de unidade entre os anotadores do ReLi.

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		
Média A B	85	71.1	78.5		
A C	77.7	74.4	78.6	98.4	99.6
C A	86.1	76.1	84.5		
Média A C	81.9	75.3	81.5		
B C	78.5	72.9	84.6	98.4	100
C B	78.4	69.6	75.5		
Média B C	78.5	71.3	79.8		

Tabela 2: Resultados da concordância entre os anotadores

4. Breve exploração do corpus

A tabela 3 apresenta a distribuição da polaridade nas opiniões do ReLi. Das cerca de 12 mil frases do corpus, quase 24% contém opinião. Dessas, cerca de 80% são positivas, e apenas 5% das frases com opinião contém opiniões contrastivas, isto é, aspectos positivos e negativos do livro na mesma frase.

Dados do ReLi	Quantidade
n. de frases	12.514
n. de frases [+]	2.433
n. de frases [-]	544
n. de frases mistas	169
n. de opiniões [+]	4.268
n. de opiniões [-]	1.023

Tabela 3: distribuição das polaridades no ReLi

Em uma breve exploração do ponto de vista qualitativo, ressaltamos a importância da anotação em contexto para a determinação da polaridade. Os exemplos (17) a (21) ilustram o que chamamos de “falsas pistas” lexicais – especificamente, temos palavras

ou expressões que, descontextualizadamente, poderiam ser consideradas negativas, mas que foram usadas de forma positiva:

17. *Lamentável* tê-lo lido somente agora.
18. O livro é *sufocante*.
19. Fui à nocaute, sem direito a (re)contagem. *Chorei*. Fiquei meio *deprê*.
20. *Tenso*. É o único livro que me deixou *nervoso* e *aprensivo* enquanto lia.
21. *Dolorido, pavoroso, nojento, repugnante e nauseante*. Por todos esses adjetivos que o livro nos causa, ele consegue ser bom.

Em outros casos, no mesmo item lexical, constatamos uma variação na polaridade:

22. Nunca *sofri* tanto para ler um livro. [-]
23. Eu *sofria* a cada vez que tinha que adiar a leitura. [+]

Algumas palavras ou expressões, geralmente consideradas neutras, comparecem com uma polaridade bastante definida, reforçando a pertinência de uma abordagem baseada em corpus. O verbo *engolir* e o adjetivo *incrível*, por exemplo, são usados apenas de maneira negativa e positiva, respectivamente.

Um aspecto com o qual não contávamos foi a presença considerável de textos confusos, mal organizados, mal pontuados, que dificultavam a interpretação. Surpreendeu-nos igualmente o uso de palavras com um sentido diferente do licenciado, o que nos leva a refletir sobre a noção de erro e, de um ponto de vista prático, traz um problema para a anotação¹.

24. “*mamão com açúcar*” quando se quer dizer “*água com açúcar*”
25. “*estilo simplista*” ao invés de “*estilo simples*”
26. “*seu texto é pegajoso até o final.*”, quando se queria dizer que era um texto que prendia o leitor.

Como são casos pontuais, esperamos que fiquem diluídos no corpus, mas servem como um ponto a ser considerado quando lidamos com texto de escrita livre.

O uso de neologismos (“eu realmente não gosto dela, acho ela muito *tonga*) e de gírias e expressões típicas da internet (“*rá!*”; “*Nah neh noh*”), além de contribuir para aspectos

¹ Nesses casos, a anotação sempre corresponde à intenção do autor da resenha. Assim, por exemplo, “*pegajoso*” foi considerado positivo.

da descrição do português, evidencia a necessidade de sistemas anotadores de POS robustos.

Por fim, a análise – ainda que superficial – do ReLi revelou dados interessantes sobre a criatividade na maneira de expressar opinião, sobretudo no que se refere à ironia. A seguir, uma pequena amostra do que encontramos, esperando que sirva ao leitor como um convite à exploração do ReLi.

27. Vampiro que brilha ... rá ! Daqui a pouco ele vive de tofu!
28. A vida de Bella, passa a ser o pior pesadelo que ela já teve: viver em um lugar onde o sol aparece menos que foto do roberto carlos de sunga.
29. 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.
30. 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 ...

5. Considerações Finais

Apresentamos o ReLi, um corpus de resenhas de livros manualmente anotado, no nível da frase e do sintagma, quanto à expressão de opinião. Motivado principalmente pela tarefa de identificação de opinião em textos, o fato de o corpus ser constituído por resenhas de um site também contribuiu para o estudo dos gêneros textuais na web, somando-se aos ainda escassos corpora do português que contêm textos não jornalísticos escritos na internet.

A ausência de material sobre a linguagem avaliativa em português² se, por um lado, aumentou o nosso desafio, por outro, nos obrigou a uma documentação farta em exemplos. O alto índice de concordância entre os anotadores, para uma tarefa assumidamente complexa, é um bom indicativo da clareza das diretrizes.

Por fim, salientamos que a granularidade da anotação pode ser abstraída para os casos em que não se precisa de tanta informação semântica, como o treino de sistemas capazes de detectar a polaridade de frases, por exemplo, dentre outras aplicações de PLN. Por outro lado, de um ponto de vista da descrição do português, a granularidade

² Em nossa busca sobre a expressão de opinião, nos deparamos com estudos sobre a expressão de opinião desfavorável, em uma perspectiva de ensino de português para estrangeiros (Almeida 2007), e da análise de *segundas avaliações* ou opiniões (Oliveira (1996), em um enfoque e interesses diferentes dos abordados aqui.

da anotação pode ser de grande valia. A polaridade no nível do sintagma nos permite capturar informações que, de outra maneira, seriam de difícil exame, como as frases com polaridade mista, bem como de palavras ou expressões que tendem a carregar valores positivos ou negativos. Deste modo, com o presente trabalho, buscamos conciliar a criação de recursos linguísticos para a utilização tanto do PLN quanto da Linguística.

Referências bibliográficas

ALMEIDA , Patricia Maria Campos de (2007). A elaboração da opinião desfavorável em português do Brasil e sua inserção nos estudos de português como segunda língua para estrangeiros. Tese de doutorado em Letras. Rio de Janeiro: Pontifícia Universidade Católica do Rio de Janeiro.

CARVALHO, Paula, SARMENTO, Luís, TEIXEIRA, Jorge e SILVA, Mário J. (2011). “Liars and Saviors in a Sentiment Annotated Corpus of Comments to Political Debates”. **ACL (Short Papers)**, pp. 564-568.

ESULI, Andrea e SEBASTIANI, Fabrizio (2006). “Sentwordnet: A Publicly Available Lexical Resource for Opinion Mining”, in: Proceedings of the 5th Conference on Language Resources and Evaluation, pp. 417—422.

FREITAS, Cláudia e CESAR, Juliana (2012). “Diretivas para a anotação da opinião em resenha de livros”, julho **Versão 2. 3**.

LIU, Bing, MINQUING, Hu, e JUNSHENG, Cheng (2005). "Opinion Observer: Analyzing and Comparing Opinions on the Web", in: Proceedings of the 14th international World Wide Web conference. Chiba, Japão.

OLIVEIRA, M. C. L. de (1996). “A organização de preferência em cartas de pedido de empresas estatais brasileiras.” D.E.L.T.A, vol. 12, nº 2, pp. 265-280.

POIRIER, D., BOTHOREL, C., GUIMIER, E. e BOULLÉ, M. (2011). “Automating Opinion Analysis in Film Reviews: the Case of Statistic versus Linguistic Approach”, in: AHMAD, K. (ed), Affective Computing and Sentiment Analysis: Metaphor, Ontology, Affect and Terminology.

RILOFF, E. e WIEBE, J. (2003). "Learning Extraction Patterns for Subjective Expressions", in: Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing.

SILVA, Mário J., CARVALHO, Paula e SARMENTO, Luís. (2012). "Building a Sentiment Lexicon for Social Judgement Mining", in: Lecture Notes in Computer Science (LNCS), International Conference on Computational Processing of the Portuguese Language (PROPOR), Springer, pp. 218-228.

WIEBE, Janyce, WILSON, Theresa e CARDIE, Claire (2005). "Annotating expressions of opinions and emotions in language", in: Language Resources and Evaluation, vol. 39, issue 2-3, pp. 165-210.

ZAGIBALOV, T., K. BELYATSKAYA, K. e CARROL, J. (2010). "Comparable English-Russian book review corpora for sentiment analysis", in: Proceedings of the 1st Workshop on Computational Approaches to Subjectivity and Sentiment Analysis. Lisboa, Portugal, pp. 67-72.