

RELATO DE PESQUISA

# COMENTCORPUS: DESCRIÇÃO E ANÁLISE DE IRONIA EM UM CORPUS DE OPINIÃO PARA O PORTUGUÊS DO BRASIL

Gabriela WICK-PEDRO  

Universidade Federal de São Carlos (UFSCar)

Oto Araújo VALE  

Universidade Federal de São Carlos (UFSCar)



## OPEN ACCESS

### EDITADO POR

– Miguel Oliveira, Jr. (UFAL)  
– René Almeida (UFS)

### REVISADO POR

– Elisa Duarte Teixeira (UnB)  
– Bruno de Melo Rocha (UFMG)

### SOBRE OS AUTORES

– Gabriela Wick-Pedro  
Conceptualização, Metodologia e  
Escrita – Rascunho Original.  
– Oto Araújo Vale  
Conceptualização e Supervisão.

### DATAS

– Recebido: 05/02/2020  
– Aceito: 24/09/2020  
– Publicado: 23/11/2020

### COMO CITAR

WICK-PEDRO, Gabriela; VALE, Oto  
Araújo. (2020). *Comentcorpus:*  
descrição e análise de ironia em um  
*corpus* de opinião para o Português do  
Brasil. *Cadernos de Linguística*,  
v. 1, n. 2, p. 01-15.

## RESUMO

O aumento de dados na Web tem despertado grande interesse na Linguística e nas Ciências da Computação e, dentre as principais tarefas, destaca-se a extração de opinião em grandes volumes de material textual encontrados na rede. A Mineração de Opinião, ou Análise de Sentimentos, busca processar automaticamente opiniões, emoções, sentimentos e subjetividades presentes em textos (LIU, 2012). No entanto, observa-se que muitas das opiniões encontradas podem carregar um sentido irônico, transformando o significado da sentença. Esta pesquisa tem como objetivo investigar expressões irônicas em mídias sociais, focando em descrições de dispositivos linguísticos que servem como pistas de ironia em textos opinativos no português do Brasil. Para a compreensão do funcionamento de tal mecanismo figurado, partiremos da busca em um *corpus* composto por comentários de notícias do portal da *Folha de S. Paulo*. Conjuntamente, foi desenvolvido um esquema de anotação de *corpus*, apoiado em teorias pragmáticas e cognitivas, referente à intenção das opiniões: i) irônicas – quando há oposição de sentido; ii) outro tipo de ironia – quando há na sentença mecanismos linguísticos figurados, mas não há uma oposição do sentido na sentença e iii) não irônica – quando não há mecanismos que alteram a literalidade da sentença. Como resultado, obteve-se uma lista de pistas linguísticas que podem ser

usadas para caracterizar expressões de ironia, colaborando, assim, para o desenvolvimento da área do PLN e da Mineração de Opinião.

#### ABSTRACT

The increase in data on the Web has aroused great interest in Linguistics and Computer Science and, among the main tasks, the extraction of opinion in large volumes of textual material found on the network stands out. Opinion Mining, or Sentiment Analysis, seeks to automatically process opinions, emotions, feelings and subjectivities present in texts (LIU, 2012). However, it is observed that many of the opinions found can express ironic, transforming the meaning of the sentence. This research aims to investigate ironic expressions in social media focusing on descriptions of linguistic devices that serve as clues of irony in opinionated texts in Brazilian Portuguese. In order to understand how this figurative mechanism works, we will start by searching a corpus composed of news comments from the Folha de S. Paulo portal. Together, a corpus annotation scheme was developed, supported by pragmatic and cognitive theories, referring to the intention of the opinions: i) ironic - when there is opposition of meaning; ii) another type of irony - when there are figurative linguistic mechanisms in the sentence, however there is no opposition from the meaning in the sentence and iii) non-ironic - when there are no mechanisms that alter the literalness of the sentence. As a result, a list of linguistic clues was obtained that can be used to characterize expressions of irony, thus collaborating for the development of the area of NLP and Opinion Mining.

#### PALAVRAS-CHAVE

Ironia; Mineração de Opinião; *Corpus*; Anotação de *Corpus*; Opinião.

#### KEYWORDS

Irony; Opinion Mining; Corpus; Corpus Annotation; Opinion.

## INTRODUÇÃO

Nos últimos anos, devido ao crescimento da *Web 2.0*, é possível observar um aumento na variedade de material textual disponível *on-line*. Parte desses textos são opiniões geradas por usuários da Internet em fóruns, mídias sociais e outros. Dada a abundância de dados publicados na rede, a partir de ferramentas linguístico-computacionais, torna-se viável (e necessária) a automatização da extração, síntese e análise de conteúdos gerados por usuários. Em vista disso, a área de Mineração de Opinião, ou Análise de Sentimentos, que busca interpretar e analisar automaticamente opiniões, sentimentos, avaliações e emoções (LIU, 2010; PANG e LEE, 2008), tem ganhado grande importância no âmbito do Processamento de Língua Natural (PLN). A análise dessas opiniões pode ser útil para pesquisas de mercado e avaliação de reações de clientes sobre determinados produtos ou pessoas públicas, por exemplo.

De um modo geral, Liu (2010) classifica os conteúdos textuais disponíveis na rede como “fato” e “opinião”. Para o autor, fatos são sentenças objetivas e desprovidas de sentimento, tendo a função de apenas transmitir informações sobre algo. Em contraposição, opiniões são sentenças subjetivas que carregam uma avaliação, sentimento, opinião ou emoção sobre algo ou alguém. Entretanto, uma das primeiras dificuldades encontradas na área da Análise de Sentimentos é exatamente a diferenciação entre o que é um texto informativo/factual e o que é subjetivo/opinião, uma vez que a linguagem não se apresenta tão exata e binária como propõe a consideração acima. Consequentemente, os conceitos de subjetividade e emoções estão intimamente ligados (LIU, 2010). Enquanto o propósito de uma sentença objetiva é apresentar uma informação factual, a sentença subjetiva pode se configurar de várias formas, tais como: opiniões, desejos, crenças, suspeitas, emoção, pensamentos ou especulações (WIEBE, 2000; RILOFF *et al.*, 2006).

Uma das formas de o falante expressar sua opinião é por meio da ironia. No entanto, compreender a ironia não é uma tarefa tão simples, embora os humanos a realizem com uma certa facilidade, dependendo de seus conhecimentos extralinguísticos. Apesar de estar presente cotidianamente na comunicação entre falantes, é comum existir mal-entendidos, caso haja falhas comunicativas daquilo que se quer fazer entender. Pragmaticamente, um enunciado irônico é entendido como uma forma de comunicar o oposto do significado literal das palavras, ou quando ocorre uma aparente violação dos princípios pragmáticos (GRICE, 1975; SEARLE, 1969). Entretanto, de uma perspectiva psicolinguística (KREUZ e GLUCKSBERG, 1989; SPERBER e WILSON, 1981), é possível considerar como a informação é processada cognitivamente e como é representada por inferências. Dessa forma, o atual trabalho apoia-se em dois princípios básicos: o lógico, a partir dos pressupostos lógicos das máximas conversacionais de Grice (1975) e o cognitivo, com base na afirmação de Sperber e Wilson (1981) de que a comunicação humana é resultante

de um processo interativo entre falantes e ouvintes e que, no processamento de informações, é possível uma alteração de seus espaços cognitivos. Segundo Grice (1975), o efeito irônico pode estar na violação da Máxima de Qualidade<sup>1</sup> e está associado à expressão de atitude, sentimento e avaliação dos falantes. Entretanto, do ponto de vista cognitivo, Sperber e Wilson (1981), baseando-se na Teoria da Relevância, acreditam que o principal elemento característico da ironia são os usos particulares de desempenho interpretativo, isto é, quando o falante se utiliza de seu enunciado para atribuir implicitamente o pensamento representativo em seu enunciado para outra pessoa que não seja ele próprio.

Desse modo, estudar a ironia pode transpassar as dificuldades de interpretá-la. É necessário, primeiramente, definir seus limites conceituais. Porém, por ser um fenômeno multifacetado, objeto de estudo de diversas abordagens das mais variadas áreas de conhecimento, essa tarefa pode ser vista como um tanto quanto hermética (GRICE, 1975; SPERBER e Wilson, 1981; KREUZ e GLUCKSBERG, 1989, ATTARDO, 2000; GIBBS e COLSTON, 2007). No entanto, não é o objetivo aqui abordar essas questões teóricas, uma vez que o interesse deste trabalho é fazer uma descrição de sentenças irônicas e não irônicas, e dentre estas, as que geram a ironia por meio de mecanismos que não sejam a oposição. Para tanto, a definição não especializada do dicionário Houaiss (2009), em que ironia é definida como uma “figura por meio da qual se diz o contrário do que se quer dar a entender; uso de palavra ou frase de sentido diverso ou oposto ao que deveria ser empregado para definir ou denominar algo” é suficiente.

No entanto, há enunciados que podem carregar um sentido figurado, mas não se encaixam na definição, aqui utilizada, de ironia. Segundo Attardo (2000), sarcasmo é um tipo de ironia abertamente agressiva, com pistas mais claras. Singh (2012) afirma que a sátira é uma figura de linguagem que enfatiza as fraquezas das pessoas. O autor ainda define humor como a capacidade de expressão de uma sensação de algo divertido e inteligente. Apoia-se, principalmente, no reconhecimento de incongruências ou peculiaridades presentes em uma situação.

Grandes conjuntos de textos opinativos têm sido classificados automaticamente para identificar o sentimento (positivo, negativo ou neutro) de opiniões neles presentes, caso em

1 O que se conhece na retórica por sentido figurado, a Pragmática irá tratar como *implicatura*. Para Grice (1975), existem dois tipos de implicatura: a *convencional* e a *conversacional*. A primeira refere-se ao significado convencional de um enunciado e a última está ligada ao significado não usual de um enunciado e é determinada pela diferença do que é dito pelo falante e o que ele realmente quer dizer. Além disso, o autor postula que, mesmo que involuntariamente, os interlocutores respeitam quatro *Princípios de Cooperação* para que a informação seja a mais homogênea possível. São elas: i) *Máxima de Qualidade*: tente que a sua contribuição conversacional seja a mais verdadeira possível, para tal, não afirme o que crê ser falso e não afirme aquilo de que não tem provas suficientes para confirmar a sua veracidade; ii) *Máxima de Quantidade*: faça com que a sua contribuição conversacional seja tão informativa quanto necessário; iii) *Máxima de Relevância*: seja relevante ao objetivo da conversa, estabelecendo relações de pertinência entre os enunciados e iv) *Máxima de Modo*: seja claro e evite ambiguidades, obscuridades; seja breve e ordenado.

que a análise pode gerar um falso resultado. Uma opinião como a expressa em (1), a seguir, por exemplo, pode ser interpretada erroneamente, em seu sentido literal. Ou seja, o sentido real da sentença carrega um sentimento negativo; porém, ferramentas de análise automática podem classificá-la como positiva por causa dos itens lexicais positivos que a compõem (“Como é bom...”). Isso ocorre pelo fato de a ironia ter como efeito uma expressão avaliativa cuja polaridade é alterada entre a avaliação literal e a pretendida, resultando em uma incongruência entre a significação literal e seu contexto (CARVALHO et al, 2009). Em outras palavras, quando se tem a intenção de produzir uma sentença irônica, o autor expressa um sentimento positivo, ainda que a opinião implícita seja negativa, ou vice-versa.

(1) Como é bom ver os políticos brasileiros gastarem o dinheiro público com propina.

No exemplo (1), a partir de um conhecimento de mundo e de uma contextualização política dos acontecimentos no país, a compreensão da ironia na sentença é possível, tendo em vista que políticos gastarem dinheiro público com propina não é algo realmente bom. Outro fator observável para inferir o sentido figurado é a presença de elementos lexicais que se opõem na mesma sentença, como “bom” *versus* “gastar dinheiro público” e “propina”.

Desse modo, o objetivo principal deste trabalho é analisar e descrever mecanismos subjacentes à expressão de ironia em textos avaliativos no português do Brasil. Além disso, busca identificar as principais características linguísticas utilizadas para expressar ironia em textos de opinião no domínio político. A hipótese é de que a presença de elementos lexicais contrastantes em um dado contexto pode ser um indicador da presença de ironia.

## 1. OS DESAFIOS DO PLN NA DESCRIÇÃO E DETECÇÃO AUTOMÁTICA DA IRONIA

A máquina atinge boa performance de acertos na análise e descrição automática de uma língua natural em seus níveis mais concretos, como a morfológica e a sintaxe, apesar de apresentarem determinadas complexidades. Contudo, ao partir para níveis mais abstratos, como o semântico, o pragmático e o discursivo, depara-se com grandes obstáculos, a serem resolvidos no âmbito do Processamento Natural da Linguagem. Dessa forma, o processamento computacional de linguagem não-literal (e também de opiniões, sentimentos e emoções) exige uma análise detalhada de informações linguísticas subjacentes ao texto. Além disso, pela linguagem figurada estar intrinsecamente ligada ao contexto e ao conhecimento extralinguístico, que por sua vez estão em constante modificação, há a

dificuldade de se criarem padrões aplicáveis a algoritmos de Aprendizado de Máquina (EMMZ e LUZ, 2011), por exemplo.

Em casos de sintetização de textos opinativos, a presença de ironia pode ser responsável por erros na classificação automática da polaridade de uma opinião (CARVALHO et al., 2009). Isso não significa que é impossível identificá-la de forma automática. Mas, primeiramente, é necessário criar padrões mais genéricos, capazes de identificar possíveis enunciados irônicos em textos. Ou seja, reconhecer aspectos linguísticos suscetíveis à análise computacional de linguagem figurada e, por meio de tratamentos individuais de cada caso, partir para uma solução mais generalizável.

Outro ponto importante a se destacar nos estudos envolvendo análises automáticas de ironia é a escassez de pesquisas voltadas para o português do Brasil (PB). É comum encontrar uma abundância de trabalhos voltados para o processamento do inglês, o que exige uma adaptação dos resultados e dos dados já existentes para o PB.

Para o inglês, algumas pesquisas são baseadas em léxicos para a identificação de pistas linguísticas para o reconhecimento de sentenças irônicas. Gonzáles-Ibáñez *et al.* (2011) sugerem que características lexicais por si só não são suficientes para identificar sarcasmo e que características pragmáticas e contextuais merecem mais atenção. Hogenboom et al. (2013) fazem uma análise de como os *emoticons* podem expressar sentimentos. Para isso, os autores criaram manualmente um léxico de sentimentos composto por *emoticons*, com vistas a melhorar o método de classificação de sentimento baseado em léxico.

Ainda para o inglês, Veale e Hao (2010) investigaram a relação entre ironia e conhecimento estereotipado de um domínio, mostrando como esse tipo de percepção ajuda no reconhecimento e entendimento de enunciados irônicos. Reyes et al. (2013) analisam a ironia em termos de um modelo multidimensional de elementos textuais, com a identificação de um conjunto de características discriminativas para gerar automaticamente textos irônicos a partir de textos não-irônicos.

Com relação aos trabalhos voltados para o PB, destaca-se o de Carvalho et al. (2009), que elaboraram um conjunto de padrões linguísticos para identificar ironia no português europeu, como *emoticons*, expressões onomatopaicas, pontuação e aspas. Algumas dessas pistas são específicas para o português (padrões morfológicos, como os diminutivos), enquanto outras parecem ser independentes da linguagem e presentes em todas as mídias sociais, por exemplo, *Facebook* ou *Twitter*. Vanin et al. (2013) apresentam um trabalho inicial para o português do Brasil sobre alguns padrões de detecção de ironia em *tweets*. Em seu artigo, os autores identificaram padrões com formas diminutivas, que podem tanto expressar sentimentos positivos, como afeto, ternura e intimidade, mas também conotações sarcásticas ou irônicas, quando a intenção é desvalorizar ou insultar uma determinada entidade.

Nota-se que a maioria das pesquisas aqui relatadas são voltadas para redes sociais. Essa preferência ocorre pela facilidade na coleta dos dados por *crawlers* (programas computacionais que “varrem” sites de modo metódico e automático). Além disso, plataformas como o *Twitter* permitem a extração de *tweets* em tempo real por meio de APIs<sup>2</sup>. Entretanto, *tweets* possuem um número limitado de caracteres, razão pela qual este artigo optou por trabalhar com dados advindos de comentários de notícias de jornal *on-line*, em que não há restrição ao tamanho do conteúdo textual.

## 2. CONSTRUÇÃO E ANOTAÇÃO DO *CORPUS* OPINATIVO

A ascensão das redes sociais alavancou a utilização de plataformas como Facebook, Twitter e YouTube, além de outros conteúdos gerados pelo usuário (em inglês, *user-generated content*). Essas mídias sociais permitem que milhares de pessoas se expressem sobre os mais diferentes assuntos, e suas opiniões podem ser acessadas rapidamente por qualquer outra pessoa ou por ferramentas computacionais.

O material textual encontrado nas mídias sociais consiste em textos mais reduzidos e escritos em uma linguagem mais informal, muitas vezes contendo desvios ortográficos e gramaticais da norma culta, que podem ser acidentais ou mesmo deliberados. Ao mesmo tempo, os textos são mais espontâneos. Para Bisognin (2009), a linguagem utilizada em plataformas digitais (ou em todas as mídias sociais) possui um modo único de o usuário expressar suas emoções, opiniões ou impressões, possuindo cada vez mais características da oralidade. Procurou-se, nesta pesquisa, trabalhar com textos em que o usuário pudesse expressar sua opinião de forma direta. Isso justifica a opção por comentários de um jornal de grande circulação, que tem uma versão *on-line* e que dá espaço para seus leitores manifestarem suas opiniões, como é o caso da Folha de S. Paulo.

O *ComentCorpus* é um *corpus* composto por comentários de notícias no PB. Para sua construção foram compilados comentários de 90 notícias referentes ao período de *impeachment* da ex-presidente Dilma Rousseff, presentes no caderno *Poder* do jornal *Folha de S. Paulo*, no período de janeiro a junho de 2016. Conteúdo gerado por usuários pode ser considerado subjetivo, e por se tratar de comentários referentes ao domínio político, a hipótese é de que o uso de recursos como a ironia seja inevitável. O *corpus* é composto por 6.185 comentários, totalizando aproximadamente 14 mil sentenças e 207 mil palavras.

2 API é uma sigla para *Application Programming Interface* – é um conjunto de instruções, rotinas e padrões de programação usados para que se possa acessar um aplicativo na Web.

Antes do processo de anotação, foi realizada, manualmente, uma identificação de cada comentário por meio da criação de um cabeçalho, em formato XML, incluindo dados como o número do comentário, do autor e da data em que foi postado. Desse modo, cada comentário foi delimitado com as etiquetas `<coment id:"xxx">` indicando o início do comentário e `</coment>` indicando o fim de cada comentário do usuário. As etiquetas `<author>` e `</author>` marcam o autor de cada comentário e as etiquetas `<date>` e `</date>` indicam a data. Para a anotação das etiquetas criadas na identificação dos comentários, dos autores, das datas e das sentenças foram criadas macros (conjunto de operações codificadas) no software Notepad++, facilitando a anotação das etiquetas.

Todas as sentenças também foram identificadas manualmente<sup>3</sup>, em uma primeira etapa de anotação, sendo delimitadas com a etiqueta `<sentence>`, indicando o início e `</sentence>`, indicando o fim. Considera-se sentença, toda sequência de palavras seguidas por um ponto final, exclamação, interrogação ou reticências.

```
<coment id:"01610">
  <author>Éder</author>
  <date>05/03/2016</date>
  <sentence>Sempre que existe a possibilidade de punir políticos importantes, mesmo com provas ca-
  bais e inequívocas, as elites reagem com esse discurso sobre crise institucional</sentence>
  <sentence>Seremos eternos reféns das quadrilhas que tomam o poder a séculos neste país?</sen-
  tence>
  <sentence>Estão apenas usando o medo para se perpetuar no poder.</sentence>
</coment>
```

**Quadro 1.** Exemplo de anotação utilizada para delimitar o cabeçalho e as sentenças anotadas.

No entanto, observou-se que em alguns casos a pontuação indica hesitação, pausa ou reprodução da oralidade na escrita dentro de uma mesma sentença. Nesse caso, como mostra o exemplo (2), apesar da pontuação, a sequência foi anotada como uma única sentença.

(2) Afinados?... duas pessoas que envergonham a nação!

Inicialmente, todas as sentenças identificadas na etapa anterior foram anotadas manu-  
almente de acordo com a percepção do anotador quanto à intenção e classificadas como:

- **Opinião não irônica:** sentenças que não contêm mecanismos linguísticos que alterem o seu referencial, por exemplo:

<sup>3</sup> Todo o processo de anotação dos comentários e das sentenças foi manual, pois acreditava-se naquele momento que seria necessário uma leitura atenta do material textual que seria analisado posteriormente.



(3) Se o povo votar em candidatos processados ou que tenham tido os nomes envolvidos com as empreiteiras, então vai ser um problema daqueles que votaram.

- **Opinião irônica:** sentenças opinativas em que há **oposição** de sentido entre a intenção que é pretendida e o que é realmente dito.

(4) A Sra. Presidente é tão honesta que vai pro céu...Tadinha...

- **Opinião com outro tipo de ironia:** sentenças opinativas em que **não** há oposição de sentido literal e significado pretendido, embora o texto carregue algum dispositivo figurado que altera sua literalidade. Casos de outro tipo de ironia incluem o sarcasmo, a sátira e o humor.

(5) Já já começará aquela conversinha de ele é "pelsseguido pulíticu" e blá blá blá.

A análise foi realizada apenas no nível da sentença, não recorrendo às outras sentenças do comentário. A escolha justifica-se pelo propósito de isolar o contexto da sentença, tendo em vista que esse seria o processo realizado pela máquina.

É importante ressaltar que uma anotação deve seguir determinados processos para não existirem inconsistências entre os anotadores, sendo necessário etapas-padrão, como treinamento dos anotadores, rodadas de anotação e intersecção entre uma rodada e outra para esclarecimentos de dúvidas entre os anotadores (HOVY e LAVID, 2010). No entanto, devido ao tempo restrito e à falta de anotadores disponíveis, foi necessária uma nova etapa de anotação em uma pequena porção do *corpus* (aproximadamente 2 mil sentenças), para validar o modelo de anotação proposto anteriormente. Desta vez, além das categorias anteriores, foi incluída também a categoria "não soube responder".

Com base nas 198 sentenças provenientes de opiniões anotadas como irônicas, no conjunto das quase 2 mil sentenças e de acordo com o conceito de ironia adotado, constatou-se a presença de dois tipos de oposição na estrutura das opiniões: implícita e explícita. Ou seja, há sentenças em que a ironia ocorre em uma camada mais superficial, sendo explicitada a partir de itens lexicais que se opõem. Em outros casos, é preciso acessar uma camada mais profunda do significado, sendo necessário, nesse momento, recorrer ao contexto e ao conhecimento de mundo do leitor.

Visto que o foco desta pesquisa é modelar a ironia em textos opinativos, uma nova anotação foi realizada, classificando as opiniões irônicas por oposição em "oposição implícita" ou "oposição explícita". Como é possível observar em (6), a oposição explícita é quando a ironia é compreendida devido à oposição de pistas presentes no próprio enunciado. Já em

(7), a oposição implícita é quando a ironia é compreendida através de pistas presentes e por um contexto pragmático adicional ao enunciado.

(6) Viva a corrupção, o país de babacas onde vivemos.

(7) Agora são trilionários, ganharam na Mega-Sena...

### 3. APRESENTAÇÃO E ANÁLISE DOS RESULTADOS DOS DADOS

Das 198 sentenças irônicas analisadas, 68% são opiniões irônicas com oposição implícita. Já as opiniões irônicas com oposição explícita totalizam 32% das anotadas.

A partir dos dados extraídos com base nas anotações realizadas, procurou-se estabelecer uma análise linguística do funcionamento da ironia em um *corpus* opinativo em um domínio político. É necessário enfatizar o domínio, pois o que pode ser compreendido como ironia em um, em outro pode ser entendido em seu sentido literal, o que demonstra mais um obstáculo na caracterização de ironia em textos com vistas à aplicação computacional.

Os métodos baseados em Aprendizado de Máquina consistem em oferecer grandes quantidades de exemplos. Cada um deles pode ser visto como um vetor de características (*features*) – *input*, seguido por um rótulo que o classifica – *output* correlacionados por um algoritmo classificador determinando o rótulo para entradas nunca vistas, ou seja, textos a serem classificados. Alguns mecanismos linguísticos foram observados nas opiniões irônicas anotadas neste trabalho, podendo ser utilizados futuramente como características em algoritmos de Aprendizado de Máquina. Todavia, estas aplicações não serão abordadas neste trabalho.

Conforme foi mencionado anteriormente, no processo de anotação, observou-se dois tipos de oposição em opiniões irônicas, as implícitas e as explícitas. Para uma modelagem linguística, as oposições explícitas são muito mais fáceis de serem formalizadas, uma vez que elas se encontram presentes no corpo da sentença. Já para a compreensão automática da oposição implícita, é necessário um banco de dados, como outro *corpus* ou acesso à *Web*, ou outro método para que a máquina entenda informações extralinguísticas.

Em (8) é possível inferir o sentido irônico por meio do contraste entre “são muito nobres” versus “usou dinheiro de corrupção”. Nesse caso, o autor da opinião faz a oposição de uma sequência positiva e uma sequência negativa para construir a ironia. Em (9), não é possível compreender o sentido figurado ou literal da sentença sem o contexto da qual foi extraída.

(8) São muito nobres, afinal, a chapa usou dinheiro de corrupção.

(9) Nossa, como ela é legal!

Há ainda casos em que não há necessariamente um contraste de polaridade entre os elementos da sentença, nem oposição de proposições. Em (10), nota-se que a ironia se apresenta na contraposição de elementos que, habitualmente, não pertencem a mesma esfera semântica.

(10) Dilma é muito boa para governar o Brasil, tanto que faliu uma lojinha de 1,99!

No exemplo acima, além da oposição entre as proposições “muito boa para governar o Brasil” e “faliu uma lojinha de 1,99”, que são positiva e negativa, respectivamente, há também um contraste entre o domínio político e o domínio do mercado e/ou comércio.

Nas opiniões de oposição explícita ou implícita, são necessários dispositivos linguísticos para que exista essa oposição de que uma opinião irônica necessita. Estes dispositivos são itens lexicais que podem coocorrer no mesmo enunciado.

(11) Aumentar impostos é fácil, né?

No exemplo (11), se fosse uma opinião afirmativa, ela poderia ser compreendida como uma oposição implícita ou até mesmo ser entendida em seu sentido literal. Ao transformá-la em uma interrogativa, com o uso do item “né?”, o autor recorre a um dispositivo linguístico retórico, quando uma questão é feita sem a necessidade de obter uma resposta. Neste trabalho, acredita-se que, de uma forma genérica, toda questão retórica é um mecanismo considerado como uma opinião irônica, porque não existe um elemento linguístico evidente que indique a ironia a não ser a própria questão.

É possível observar, com base nos dados desta pesquisa, que as questões retóricas se relacionam mais às ironias implícitas (12,9%), em comparação com as ironias explícitas (4,2%). Esse resultado pode ser explicado pelo fato de que quando o autor de uma opinião utiliza uma questão retórica, ele busca ecoar contextos que não estão necessariamente no texto, como é possível observar no exemplo (12).

(12) Devemos ser grato ao P.T e seus eleitores, não é verdade?

O uso dos diminutivos parece ser um recurso mais evidente em uma opinião irônica. Geralmente, é utilizado para expressar afeto, carinho, intimidade ou outros sentimentos

positivos. No entanto, dependendo da sua aplicação na sentença, percebeu-se que foi usado para transmitir um sentido irônico ou sarcástico.

(13) Estão preocupados com o coitadinho do Cunha.

Em (13) sabe-se que Cunha é um político brasileiro envolvido em casos de corrupção. Logo, o uso do termo “coitadinho” não pode ser compreendido como algo positivo, ativando, conseqüentemente, o sentido irônico.

Outro mecanismo utilizado foram as aspas, comumente empregadas em ambientes virtuais para indicar que aquilo que está sendo dito é uma ironia. Alguns trabalhos para o português (FREITAS et al., 2014; CARVALHO et al., 2009) também salientam o uso das aspas para gerar efeito de ironia.

(14) O PMDB é um partido tão “pequeno” que não consegue ser medíocre!

As aspas são frequentemente usadas para enfatizar um conteúdo, principalmente a polaridade em relação ao resto da sentença. No exemplo (14), o termo “pequeno” não pode ser usado para qualificar o partido PMDB, por exemplo. Isso acontece por ser atualmente o maior partido brasileiro.

Para enfatizar uma opinião, é comum o uso de adjetivos. Apesar disso, muitas vezes os adjetivos podem fazer uma oposição a outro adjetivo (ou a todo um enunciado) de polaridade contrastiva. Alguns adjetivos com alto grau de polaridade, quando usados na mesma sentença, podem ressaltar o sentido irônico de uma opinião, como ocorre em (15).

(15) Nossos políticos corruptos aceitam de bom grado, pois vivemos um momento extraordinário nesse país!!!

No decorrer da análise foi possível observar, ainda, a existência de alguns itens lexicais essenciais para estabelecer a oposição. Esses itens, na maioria das vezes, possuem uma polaridade bem marcada e são restritos a um determinado domínio, mas ativam a ironia do enunciado, sendo classificados como **disparadores**. Palavras como “corrupção”, “impeachment”, “imposto” quando em contraste com outras palavras ou outro contexto que não seja o político, por exemplo, geram uma sentença irônica.

No exemplo (16), os termos “corrupção” e “campanha” se unem e se opõem com “não é grave”, dando o sentido irônico para a opinião.

(16) Dinheiro de corrupção para financiar campanha de Dilma não é grave.

Em (17), a ironia ocorre na contraposição entre “muito nobres” e “usou dinheiro de corrupção”. Como no exemplo anterior, o termo “corrupção” é considerado um disparador, pois, se houvesse a substituição por outro elemento lexical, como “trabalho”, não seria possível a compreensão irônica do enunciado.

(17) São muito nobres, afinal, a chapa usou dinheiro de corrupção.

O uso da hipérbole ou superlativos pode também ser uma característica para indicar a ironia por meio de exageros, como em (18). Ademais, como pode ser visto em (19), algumas interjeições, principalmente positivas, também podem ser um indicador de ironia.

(18) Belíssimo modelo de ética e moral dos vermelhos chupins.

(19) Viva a decência.

Por fim, (20) exemplifica a representação do riso, muito utilizada em conteúdos gerados por usuários e que pode evidenciar a ironia de uma opinião.

(20) Muito boa essa Data Folha... KKKKK

Todos os recursos linguísticos aqui descritos são evidências encontradas no *CommentCorpus*, mas que só podem ser analisadas como possíveis indicadores de ironia de acordo com a intuição do leitor, além da coocorrência com outros itens lexicais na estrutura da opinião, pois até o momento não foram testados em Aprendizado de Máquina ou outros métodos computacionais.

## 4. CONCLUSÕES

O presente trabalho procurou estudar e descrever a ironia em um corpus de comentários com o propósito de auxiliar na construção de recursos linguísticos-computacionais para a Mineração de Opinião. Para tanto, foi criado um esquema de anotação para opiniões irônicas de conteúdo gerado por usuários no domínio político que busca esclarecer como é o processo de identificação desses casos e se existem dispositivos linguísticos capazes de facilitar sua detecção. Por meio da análise feita na atual pesquisa, também foi possível identificar características de expressão da ironia no domínio escolhido, como o uso de itens

lexicais em um grupo limitado dentro de um contexto – os chamados “disparadores” – e adjetivos com polaridade positiva e negativa, por exemplo.

É pertinente salientar que um dos principais mecanismos encontrados na identificação de ironia são fatores extralinguísticos e sociais intrínsecos a declarações irônicas, que tornam a tarefa de detecção automática desses mecanismos figurados em instâncias bastante complexa, principalmente pela falta dos contextos e informações extralinguísticas. Ainda assim, foi possível observar que o significado de um enunciado irônico pode ser implicado por essas informações pragmáticas, o que suscita a necessidade de realização de referência a eventos e enunciados anteriores ao que se está analisando.

Tendo em vista o desenvolvimento recente das redes sociais e da importância que ganham em nossas vidas, cremos que o desenvolvimento desses recursos continuará sendo uma tarefa importante no campo do PLN e da Mineração de Opinião. Um próximo passo para a continuidade desta pesquisa seria o desenvolvimento de um léxico de sentimentos voltado à mineração de opinião (e de textos) de mídias sociais, integrando interjeições, adjetivos e advérbios altamente polarizados mais frequentes em português do Brasil, e de outros disparadores de ironia para o domínio político.

## REFERÊNCIAS

ATTARDO, Salvatore. Irony markers and functions: Towards a goal-oriented theory of irony and its processing. *Rask*, v. 12, n. 1, p. 3-20, 2000.

BARBIERI, Francesco; RONZANO, Francesco; SAGGION, Horacio. Pathways for irony detection in tweets. *Proceedings of the 29<sup>th</sup> Annual ACM Symposium on Applied Computing*, New York, NY, USA, p. 628-633, 2014.

BISOGNIN, Tadeu R. *Sem Medo do Internetês*. 1. ed. v. 1. Porto Alegre: AGE, p. 189, 2009.

CARVALHO, Paula; SARMENTO, Luís; SILVA, Mario J.; OLIVEIRA, Eugénio. Clues for detecting irony in user-generated contents: oh...!! it's "so easy" ;-). *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*, New York, NY, USA, p. 53-56, 2009.

CARVALHO, Paula; SARMENTO, Luís; TEIXEIRA, Jorge; SILVA, Mario J. Liars and Saviors in a Sentiment Annotated Corpus of Comments to Political debates. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: shortpapers*, Portland, Oregon, USA, v. 2, p. 564-568, 2011.

EMMS, Martin; LUZ, Saturnino. *Machine Learning for Natural Language Processing*. European Summer School of Logic, Language and Information, course reader, Dublin, p. 90. 2007. Disponível em: <http://ronaldo.cs.tcd.ie/essli07/mlfornlp.pdf>.

FREITAS, Claudia. Sobre a construção de um léxico da afetividade para o processamento computacional do português. *Revista Brasileira de Linguística Aplicada*, v. 13, n. 4, p.1031-1059, 2013.

FREITAS, Claudia; MOTTA, Eduardo; MILIDIÚ, Rui L.; CÉZAR, Juliana. Vampiro que brilha... rá! Desafios na anotação de opinião em um corpus de resenhas de livros. *Caderno de Resumos de XI Encontro de Linguística de Corpus (ELC)*, v. 11, p. 1-13, 2012.

GIBBS JR, Raymond W.; COLSTON, Herbert L. *Irony in Language and Thought: A Cognitive Science Reader*. 1. ed. Nova Iorque: Routledge, p. 618, 2007.

GONZÁLEZ-IBÁÑEZ, Roberto; MURESAN, Smaranda; WACHOLDER, Nina. Identifying Sarcasm in Twitter: A Closer Look. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Portland, Oregon, USA, p. 581-586, 2011.

GRICE, Paul. Logic and conversation. In: COLE, Peter; MORGAN, J. L. (Eds.) *Syntax and semantics*, Vol. 3: Speech acts. New York, Academic Press, p. 225-242, 1975.

HAO, Yanfen; VEALE, Tony. Support Structures for Linguistic Creativity: A Computational Analysis of Creative Irony in Similes. *Proceedings of the 31st Annual Meeting of the Cognitive Science Society 2009 (CogSci 2009)*, Amsterdam, Netherlands., p. 1376-1381, 2009.

HOGENBOOM, Alexander; BAL, Daniella; FRASINCAR, Flavius; BAL, Malissa; DE JONG, Franciska; KAYMAK, Uzay. Exploiting emoticons in sentiment analysis. *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, New York, NY, USA, p. 703-710, 2013.

HOUAISS. *Dicionário eletrônico Houaiss da língua portuguesa*. Rio de Janeiro: Objetiva, 2009.

HOVY, Eduard; LAVID, Julia. Towards a 'Science' of Corpus Annotation: A New Methodological Challenge for Corpus Linguistics. *International Journal of Translation*, v. 22, n. 1, p. 13-36, 2010.

KREUZ, Roger J.; GLUCKSBERG, Sam. How to be sarcastic: The echoic reminder theory of verbal irony. *Journal of experimental psychology: General*, v. 118, n. 4, p. 374, 1989.

LIU, Bing. Sentiment Analysis and Subjectivity. In: INDURKHYA, N.; DAMERAU, F. J. (Eds.). *Handbook of Natural Language Processing*. [s.l.] Chapman and Hall/CRC, p. 627-666, 2010.

LIU, Bing. *Sentiment Analysis and Opinion Mining*. 1. ed. Toronto: Morgan & Claypool Publishers, p. 182, 2012.

PANG, Bo; LEE, Lillian. Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, v. 2, n. 1-2, p. 1-135, 2008.

REYES, Antonio; ROSSO, Paolo; VEALE, Tony. A multidimensional approach for detecting irony in twitter. *Language resources and evaluation*, v. 47, n. 1, p. 239-268, 2013.

RILOFF, Ellen; PATWARDHAN, Siddharth; WIEBE, Janyce. Feature subsumption for opinion analysis. *Proceedings of the 2006 conference on empirical methods in natural language processing*. Association for Computational Linguistics, 2006.

SEARLE, John. *Speech acts: an essay in the philosophy of language*. Cambridge, Cambridge University Press, 1969.

SINGH, Raj Kishnor. Humour, irony and satire in literature. *International Journal of English and Literature*, v. 3, n. 4, p. 63-72, 2012.

SPERBER, Dan; WILSON, Deirdre. Irony and the use-mention distinction. *Philosophy*, v. 3, p. 143-184, 1981.

VANIN, Aline, FREITAS, Larissa A.; VIEIRA, Renata; BOCHERNITSAN, Marco. Some Clues on Irony Detection in Tweets. *Proceedings of the 22nd International Conference on World Wide Web*. Association for Computing Machinery, p. 635-636, 2013.

VEALE, Tony; HAO, Yanfen. Detecting Ironic Intent in Creative Comparisons. *Proceedings of the 2010 conference on ECAI 2010: 19th European Conference on Artificial Intelligence*, Amsterdam, Netherlands, p. 765-770, 2010.

WIEBE, Janyce. Learning Subjective Adjectives from Corpora. *Proceedings of the Seventeenth National Conference on Artificial Intelligence*, Austin, Texas, USA, 2000.