



Programa de
Pós-Graduação em
Linguística

COMENTCORPUS: IDENTIFICAÇÃO E PISTAS LINGUÍSTICAS PARA
DETECÇÃO DE IRONIA NO PORTUGUÊS DO BRASIL

GABRIELA WICK PEDRO

SÃO CARLOS
2018



Universidade Federal de São Carlos

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE EDUCAÇÃO E CIÊNCIAS HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM LINGUÍSTICA

COMENTCORPUS: IDENTIFICAÇÃO E PISTAS LINGUÍSTICAS PARA DETECÇÃO DE
IRONIA NO PORTUGUÊS DO BRASIL

GABRIELA WICK PEDRO

Bolsista: Coordenação de Aperfeiçoamento de Pessoal (CAPES)

Dissertação apresentada ao Programa de Pós-Graduação em Linguística da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do Título de Mestre em Linguística.

Orientador: Prof. Dr. Oto Araújo Vale

São Carlos - São Paulo - Brasil
2018

Pedro, Gabriela Wick

ComentCorpus: Identificação e pistas linguísticas para detecção de ironia no português do Brasil / Gabriela Wick Pedro. -- 2018.
95 f. : 30 cm.

Dissertação (mestrado)-Universidade Federal de São Carlos, campus São Carlos, São Carlos
Orientador: Oto Araújo Vale
Banca examinadora: Cláudia Dias de Barros, Helena de Medeiros Caseli
Bibliografia

1. Ironia. 2. Anotação de corpus. 3. Processamento de Língua Natural. I. Orientador. II. Universidade Federal de São Carlos. III. Título.

Ficha catalográfica elaborada pelo Programa de Geração Automática da Secretaria Geral de Informática (SIn).

DADOS FORNECIDOS PELO(A) AUTOR(A)

Bibliotecário(a) Responsável: Ronildo Santos Prado – CRB/8 7325



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Educação e Ciências Humanas
Programa de Pós-Graduação em Linguística

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado da candidata Gabriela Wick Pedro, realizada em 16/03/2018:

Prof. Dr. Oto Araujo Vale
UFSCar

Profa. Dra. Cláudia Dias de Barros
IFSP - Sertãozinho

Profa. Dra. Helena de Medeiros Caseli
UFSCar

Agradecimentos

Agradeço, primeiramente, à Coordenação de Aperfeiçoamento de Pessoal - CAPES, pelo apoio financeiro concedido durante todo o período do Mestrado.

Ao Programa de Pós-Graduação em Linguística da UFSCar (PPGL/UFSCar), pelo suporte institucional, ao Núcleo Interinstitucional de Linguística Computacional (NILC/USP) pelas ajudas profissionais e novas perspectivas linguístico-computacionais e aos colegas do grupo LeGOS (Léxico-Gramática, Opinião e Sentimentos) pela ajuda e discussões no processo desta pesquisa e aos colegas pelas conversas e experiências trocadas. Em especial, aos anotadores Jackson, Júlia, Roana, Oto, Ryan e Tayná por dedicarem o tempo de cada um à tarefa de anotação.

Ao Professor Doutor Oto Araújo Vale, pela orientação, paciência, apoio em toda minha trajetória acadêmica e, principalmente, por toda confiança depositada durante todos esses longos anos.

Às Professoras Doutoras Helena Medeiros Caseli e Cláudia Dias de Barros pelo disponibilidade e pelo aceite em fazer parte da banca examinadora desta dissertação. Destino também meus agradecimentos ao Professor Doutor Thiago Alexandre Salgueiro Pardo e à Doutora Amanda Pontes Rassi pelas considerações e durante minha qualificação.

Aos meus pais e à minha família por serem meu porto seguro, pela paciência e pelo colo quando necessário e pelos momentos de alegria divididos. À amiga Sarita Mazzini Bruschi, pela confiança compartilhada, permitindo dar novos passos e abrir novas janelas em minha vida e ao amigo Matheus Lucas Hebling por quase uma vida de amizade, pelas risadas e pelo ombro amigo.

À minha mãe, pelo amor incondicional e por estar ao meu lado, sempre.

Resumo

Opiniões na *Web* têm crescido progressivamente e assim, vem despertando o interesse em áreas de estudo da Linguística e Computação, por exemplo. Nesse contexto, surge a Análise de Sentimentos, ou Mineração de Opinião, que tem como objetivo analisar computacionalmente opiniões, emoções, sentimentos e subjetividades presentes em textos (Liu, 2012), entretanto, certas sentenças subjetivas podem carregar ironia, transformando o sentido de uma sentença. Esta dissertação de mestrado tem como propósito investigar expressões de ironia em mídias sociais, com foco na descrição de dispositivos linguísticos como pistas de ironia em textos opinativos no português do Brasil. Para compreender o funcionamento deste mecanismo figurado, partiremos da busca um *corpus* construído por comentários de notícias do portal da *Folha de S. Paulo*. Juntamente, apoiado em teorias pragmáticas e cognitivas, desenvolvemos um esquema de anotação de *corpus* para opiniões e suas intenções: irônicas, outros tipos de ironia ou não irônica. Como resultado, obtivemos uma lista de subcategorias que caracterizam expressões de ironia que permite colaborar com o desenvolvimento da área de PLN e Análise de Sentimentos e, além disso, aperfeiçoar ferramentas de identificação automática de opinião através das descrições e dos recursos linguísticos aqui elaborados.

Palavras-chave: Análise de Sentimentos, *Corpus*, Ironia, Opinião, PLN

Abstract

Opinions on the Web have been increasing progressively and, thus, has aroused interest in areas of study of Linguistics and Computation, for example. In this context comes the Sentiment Analysis, or Opinion Mining, which aims to analyze computationally opinions, emotions, feelings and subjectivities present in texts (Liu, 2012), however, certain subjective sentences can carry irony, transforming the meaning of a sentence. This dissertation aims to investigate expressions of irony in social media, focusing on the description of linguistic devices as clues of irony in opinion texts in Brazilian Portuguese. To understand the functioning of this figurative mechanism, we will start from the search a *corpus* constructed by news commentaries from the *Folha de S. Paulo* portal. In addition, based on pragmatic and cognitive theories, we developed a *corpus* annotation scheme for opinions and their intentions: ironic, other types of irony or non-ironic. As a result, we have obtained a list of subcategories that characterize expressions of irony that allow to collaborate with the development of the NLP area and Sentiment Analysis and, in addition, to improve tools of automatic identification of opinion through the descriptions and the linguistic resources elaborated here.

Keywords: Sentiment Analysis, *Corpus*, Irony, Opinion, NLP

Lista de Figuras

Figura 1. Interface da caixa de comentários do jornal <i>Folha de S. Paulo</i>	42
Figura 2. Exemplo de comentário do jornal <i>Folha de S. Paulo</i>	42
Figura 3. Exemplo de cabeçalho para cada comentário	45
Figura 4. Visualização do <i>ComentCorpus</i>	45
Figura 5. Exemplo de anotação de sentença	46
Figura 6. Exemplo de cabeçalho para cada comentário	47
Figura 7. Exemplo de anotação de opinião irônica.....	48
Figura 8. Exemplo de anotação com outro tipo de ironia.....	49
Figura 9. Busca pela etiqueta <opinião:“ironica”> no Notepad++	50
Figura 10. Anotação de oposição explícita.....	54
Figura 11. Anotação de oposição implícita	54
Figura 12. Anotação de entidades nomeadas (NE).....	55
Figura 13. Anotação de questão retórica	55
Figura 14. Anotação de pontuação	55
Figura 15. Anotação dos diminutivos.....	56
Figura 16. Anotação das aspas	56
Figura 17. Anotação de adjetivos	57
Figura 18. Anotação das disparadores	57
Figura 19. Anotação de hipérbole.....	58
Figura 20. Anotação de interjeição.....	58
Figura 21. Anotação de intensificador/modificador	58
Figura 22. Anotação de repetições.....	59
Figura 23. Anotação de risos	59
Figura 24. Distribuição das anotações pelos três anotadores na ETAPA 3.....	61

Figura 25. Distribuição das anotações pelos cinco anotadores na ETAPA 3.....	61
Figura 26. Porcentagem de oposição implícita e explícita	66

Lista de Quadros

Quadro 1. Principais aplicações em Análise de Sentimentos	29
Quadro 2. Pistas e padrões utilizados por Carvalho et al. (2009)	36
Quadro 3. Padrões utilizados por Vanin et al. (2013)	37
Quadro 4. . Oposições irônicas – implícitas e explícitas	53
Quadro 5. Relação de sentenças irônicas marcadas por todos os anotadores	65
Quadro 6. Adjetivos com polaridade bem marcada extraídos de sentenças irônicas ..	70

Lista de Tabelas

Tabela 1. Resultados para os padrões usados em experimentos por Carvalho et al. (2009).....	37
Tabela 2. Características do <i>ComentCorpus</i>	43
Tabela 3. Descrição dos meses que compõem o <i>corpus</i>	43
Tabela 4. Dados dos anotadores na ETAPA 2.....	49
Tabela 5. Anotação ETAPA 3: Distribuição de anotação – Três anotadores	51
Tabela 6. Anotação ETAPA 3: Distribuição de anotação – Cinco Anotadores.....	52
Tabela 7. Categorias e anotadores para cálculo do coeficiente kappa	67
Tabela 8. Resultados para Pc da anotação	69
Tabela 9. Resultado do Pi da anotação Categorias e anotadores para cálculo do coeficiente kappa	62
Tabela 10. Resultado para P e Pe da anotação Resultados para Pc da anotação	63
Tabela 11. Pontuação kappa para a tarefa de anotação da ETAPA 3	64
Tabela 12. Distribuição das subcategorias nas sentenças analisadas.....	67
Tabela 13. Distribuição das subcategorias entre as oposições implícitas e explícitas.....	69

Sumário

Capítulo 1	11
Introdução	11
1.1. Objetivos e hipóteses	14
1.2. Estrutura da dissertação	14
Capítulo 2	16
Teorias linguísticas da ironia	16
2.1. Contextualização da ironia	16
2.2. Ironia através da perspectiva pragmática de Grice	17
2.3. Perspectiva ecoica de Sperber e Wilson	21
2.4. Ironia <i>versus</i> sarcasmo.....	24
Capítulo 3	26
Análise de Sentimentos e Recursos Linguísticos	26
3.1. Análise de Sentimentos	26
3.1.1. Principais conceitos da Análise de Sentimentos.....	27
3.1.2. Aplicações em Análise de Sentimentos	28
3.1.3. Desafios	29
3.2. Anotação de <i>Corpus</i>	29
3.2.1. Definições	31
3.2.2. Formatos de anotação	33
3.3. Trabalhos Relacionados.....	34
Capítulo 4	38
Mídias sociais e construção do <i>corpus</i>	38
4.1. Mídias sociais como <i>corpus</i> para Análise de Sentimentos	39
4.2. Construção do <i>corpus</i>	41

Capítulo 5	44
Esquema de anotação para expressões irônicas	44
5.1. ETAPA 1 – Criação de um cabeçalho para cada comentário	45
5.2. ETAPA 2 – Identificação e anotação de opiniões	46
5.2.1. Notepad++	49
5.3. ETAPA 3 – Validação das diretrizes	50
5.4. ETAPA 4 – Anotação de ironia em sentenças opinativa.....	52
5.4.1. Expressões irônicas: oposições.....	53
5.4.2. Pistas: Subcategorias linguísticas	54
Capítulo 6	59
Análise dos Dados	59
6.1. Anotação manual do <i>corpus</i>	60
6.2. Validação da anotação do <i>ComentCorpus</i>	66
6.3. Expressões irônicas e subclasses	66
6.4. Resultados obtidos	71
Conclusão e trabalhos futuros	73
Referências Bibliográficas	75
Anexos	81

Capítulo 1

Introdução

O crescimento da *Web 2.0* colaborou para que grande parte da comunicação diária passasse a ser *on-line* e, conseqüentemente, as chamadas mídias sociais tornaram-se uma fonte valiosa de informações sobre a opinião a respeito de produtos, empresas, políticos, tendências, entre outros (PANG e LEE, 2008). Estes dados opinativos passam a ser interessantes não só pelo seu grande volume, mas pelo constante uso de uma linguagem figurada, em específico, a ironia.

Investigar e descrever a ironia vai além das dificuldades em compreendê-la, é também um desafio definir os limites de seu conceito entre as diferentes abordagens que estão associadas, de uma maneira ou de outra, de um olhar filosófico (KIERKEGAARD, 2017), linguístico (GRICE, 1975), psicolinguístico (SPERBER e WILSON, 1981) ou computacional (CARVALHO et al, 2009; HAO e VEALE, 2009). Tais perspectivas fazem da ironia um fenômeno multifacetado e nos permite interpretá-la de diferentes formas conforme nosso interesse em uma ou outra abordagem. Assumimos aqui o entendimento tradicional, a saber: Ironia é uma “figura de linguagem por meio da qual se passa uma mensagem diferente – em geral contrária – à mensagem literal, geralmente com o objetivo de criticar ou promover humor”. Essa é a definição do Dicionário Houaiss e, no momento, serve aos propósitos deste trabalho. Como veremos mais adiante, há outros aspectos a considerar que serão detalhados ao longo do texto.

A presente dissertação de mestrado é desenvolvida na Linguística Computacional e se desdobra para o campo da Análise de Sentimentos, ou Mineração de Opinião, uma área de estudo que interpreta e analisa computacionalmente opiniões, avaliações, sentimentos e emoções expressas em texto sobre uma determinada entidade e todos seus atributos relacionados (PANG e LEE, 2008; LIU, 2010). Torna-se possível, assim, classificar a polaridade de um determinado texto de acordo com seus pontos favoráveis ou desfavoráveis sobre algum produto do mercado ou até identificar distintas perspectivas ideológicas em um debate político.

A análise de opiniões e sentimentos em língua natural é uma tarefa complexa e exige uma detalhada interpretação das informações explícitas e implícitas na estrutura da língua.

Inferir o significado figurado de um enunciado é, frequentemente, algo não tão simples, mas os seres humanos podem realizar com uma certa facilidade. De fato, a linguagem figurada exige um conhecimento de mundo e familiaridade com o contexto conversacional dos participantes da conversa – informações estas que a máquina não pode acessar facilmente. Desta forma, compreender os mecanismos linguísticos subjacentes que facilitam a comunicação de enunciados não-literais, permite melhorar a compreensão de processos cognitivos, linguísticos e comunicativos (WALLACE, 2015). Além disso, o fato da linguagem figurada estar em constante modificação em função da introdução de mudanças sociais ou mesmo de novos significados que são associados a construções linguísticas já existentes, pode dificultar, por exemplo, o treinamento de algoritmos de Aprendizado de Máquina.

Como em outras categorias de linguagem figurada, o texto irônico não deve ser interpretado em seu sentido literal. Deste modo, um dos problemas na detecção automática de um texto não-literais, em específico a ironia, é o rompimento das expectativas da língua, desde a pronúncia até a escolha lexical, estrutura sintática, semântica e conceitualização. Assim, não é realista buscarmos uma milagrosa solução computacional para linguagem figurada e uma resposta geral para esta questão, dificilmente, será encontrada em alguma técnica ou algoritmo. Em vez disso, parece mais adequado identificar aspectos e formas específicas de linguagem figurada que são suscetíveis de análise computacional e a partir desses tratamentos individuais, tentarmos sintetizar uma solução mais abrangente.

A presença de ironia em um enunciado pode ser responsável por erros em tarefas de identificação e classificação automática da polaridade de um texto (CARVALHO et al., 2009). Neste sentido, a precisão da polaridade de uma sentença pode ser significativamente minada pela presença de ironia, como podemos observar nos exemplos¹ abaixo:

- (1) Ainda bem que este homem mais “honesto” do mundo não vai se candidatar.
- (2) Como é bom ver os políticos brasileiros gastarem o dinheiro público com propina.

Um procedimento básico de Análise de Sentimento, provavelmente, classificaria estas sentenças como positivas, enquanto a intenção é inegavelmente negativa. Em (1) apesar das aspas no adjetivo honesto indicar a presença de ironia é possível dizer que toda a expressão

¹ Todos os exemplos utilizados neste trabalho foram retirados do *ComentCorpus*, com exceção dos identificados no Capítulo 2.

“este homem mais *honesto* do mundo” é uma pista de ironia na qual as aspas são os elementos principais que irão contrapor com toda a sentença e os elementos linguísticos *este homem, mais e do mundo* intensificam o adjetivo. No exemplo (2), a ironia não está marcada em um elemento principal como no exemplo anterior. No entanto, além de compreendermos essa sentença como irônica dado o nosso conhecimento de mundo, pois sabemos que o ato dos políticos brasileiros gastarem o dinheiro público com propina não é, de fato, algo realmente bom. Há um contraste entre *como é bom ver*, um conteúdo positivo e *gastarem o dinheiro público com propina*, como um conteúdo negativo. Logo, é possível compreender a ironia dessa sentença pela oposição entre as sequências de palavras.

Em face às teorias linguísticas e às pesquisas anteriores sobre o tema desta dissertação, buscamos descrever possíveis padrões de ironia. Para o desenvolvimento desta pesquisa, o processo de análise e anotação de opiniões irônicas acontece a partir de duas concepções complementares: da Linguística e da Linguística Computacional. Na primeira concepção, contextualizamos as principais teorias que abarcam, tradicionalmente, a ironia e como estas abordagens são fundamentais no processo de compreensão de um enunciado irônico. Em um segundo momento, propomos um esquema de anotação de *corpus*, que deve arquitetar a ironia através dos princípios estudados pelo viés linguístico para que possamos, em trabalhos futuros, modelá-las para aplicação em Processamento de Língua Natural (PLN).

Para examinar as ocorrências de ironia, partiremos da busca em um *corpus* construído por comentários de notícias e anotado com as expressões irônicas e sarcásticas. A escolha por um *corpus* de comentários retirados da *Web* se dá por acreditarmos que textos em que os usuários se expressam através de impressões, opiniões, sentimentos, emoções e avaliações possuem grande probabilidades de ser, parcialmente, irônicos. Um aspecto importante sobre a delimitação do *corpus* está no fato de ser composto por textos não-jornalísticos em uma linguagem informal com a opinião direta do autor e que se tenha acesso *on-line* destas informações. O *corpus* da pesquisa é construído por 6185 comentários retirados de 90 notícias relacionadas ao período *pré-impeachment* (entre janeiro e julho de 2016) do caderno *Poder* do jornal *Folha de S. Paulo*, totalizando 14 mil sentenças e 207 mil palavras. Para a anotação, foi criado um cabeçalho que identifica o comentário, o autor e a data em que a opinião foi expressa. Além disso, foram criadas etiquetas que indicam a sentença e se ela é uma opinião irônica, não irônica ou contém outro tipo de mecanismo linguístico que interfira em seu sentido literal.

1.1. Objetivos e hipóteses

Considerando que a detecção de ironia é um campo que tem crescido rapidamente nos últimos anos (GHOSH et al., 2015; MAYNARD e GREENWOOD, 2014) e acreditando que a geração de ironia é rigorosamente dependente da cultura e da língua em que é dita, formulamos as seguintes questões que nos guiaram para estabelecer os objetivos e hipóteses desta dissertação:

- Quais são as principais características linguísticas utilizadas para expressar ironia em textos de opinião no domínio político?
- Quais são as classes de ironia mais representativas no *corpus* utilizado?
- Quais são os tipos de ironia mais fáceis de identificar manualmente?

A partir dos direcionamentos acima, levantamos a hipótese de que a presença de elementos lexicais contrastantes em um dado contexto pode ser um indicador de ironia. Desse modo, buscamos não só descrever o fenômeno da ironia em textos em mídias sociais, mas esperamos que esta pesquisa nos ajude a entender melhor a compreensão e desenvolvimento da própria ironia, o que nos levará à descoberta de características e padrões que podem ser compartilhados e confrontados em um momento seguinte com projetos semelhantes.

1.2. Estrutura da dissertação

Esta dissertação está organizada da seguinte forma:

No **Capítulo 2** apresentamos as principais noções da ironia. Primeiramente, é feita uma revisão da ironia através dos tempos e a problemática que envolve os conceitos de ironia sob diferentes perspectivas. Em seguida, é retratada a questão do enunciado irônico através do ponto de vista da pragmática de Grice e do uso ecoico de Sperber e Wilson.

O **Capítulo 3** aborda os principais conceitos que constituem a Análise de Sentimento e seus desafios. Além disso, discorremos sobre as definições e importância da anotação para os estudos descritivos da língua, em específico para a Linguística de Corpus e para o Processamento de Língua Natural. Encerramos o capítulo apresentando os trabalhos relacionados que, de certa forma, se assemelham e compartilham com esta dissertação.

No **Capítulo 4** tratamos das mídias sociais e sua importância para o estudo de opiniões e descrevemos o *corpus* da pesquisa.

O **Capítulo 5** aborda os aspectos metodológicos utilizados nesta pesquisa, isto é, descrevemos todo o processo de anotação e identificação de opiniões irônicas e dos dispositivos identificados e extraídos a partir do *corpus* de estudo.

No **Capítulo 6** expomos os resultados obtidos, entre eles: os dispositivos mais relevantes para a construção do significado irônico no *corpus*, mais precisamente, em comentários do domínio político.

Finalizamos esta dissertação com as **Conclusões e Trabalhos Futuros**.

Capítulo 2

Teorias linguísticas da ironia

O objetivo deste capítulo é apresentar uma síntese dos principais conceitos que abrangem a noção de ironia buscando ressaltar que apesar de existir uma falta de consenso sobre sua conceitualização das diferentes perspectivas teóricas em que a ironia se atribui em um determinado contexto, é possível manter características determinantes que permitem a sua identificação. Entendemos, portanto, que essa reflexão poderá contribuir para melhor compreensão do objeto de estudo dessa pesquisa.

2.1. Contextualização da ironia

O estudo da ironia tem uma longa tradição em filosofia, retórica, crítica literária e linguística. Ao longo do desenvolvimento da sociedade, a ironia adquiriu um *status* categórico diferente: de metáfora no mundo antigo ao modo de pensar e atitude em relação à realidade atual. Segundo Gibbs e Colston (2007), os estudos históricos e contemporâneos da ironia também puderam ser encontrados em áreas tão diversas quanto a antropologia, estudos sociais e até psicologia.

O termo ironia vem do grego *eironeia* (εἰρωνεία) e quer dizer “dissimulação”. Aristóteles reconhecia que a *eironeia* continha diferentes níveis de verdade, sendo que a negação dessa verdade era um dos seus aspectos intrínsecos. O modelo de Aristóteles para a ironia é o da *ironia socrática*, que está de acordo com o método socrático, ou seja, conduzia o indivíduo à consciência do erro. Tal estratégia utilizada sustentava-se na dissimulação de Sócrates, que nada mais era do que uma série de questionamentos com os quais provocava no interlocutor uma confusão de modo a mostrar-lhe as fraquezas de seus argumentos. A ironia socrática criava um mal-estar a partir desse despertar da consciência do indivíduo e, assim, o induzia à sua verdade interior.

A partir de uma perspectiva pragmática, Grice (1975) tenta explicar o funcionamento da ironia focando-se em conceitos que transpassam os níveis tradicionais literários e linguísticos, considerando também os elementos extralinguísticos. Entretanto, anteriormente ao enfoque pragmático, a ironia costumava ser definida através de dicionários e gramáticas. Assim, segundo o dicionário HOUAISS (2009) a ironia pode ser definida como:

ironia s.f. **1** Rubrica: retórica. 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 empr., para definir ou denominar algo [A ironia ressalta do contexto.] **1.1** Rubrica: literatura. esta figura, caracterizada pelo emprego inteligente de contrastes, us. literariamente para criar ou ressaltar certos efeitos humorísticos. **2** m.q. **asteísmo** ('uso sutil'). **3** uso de palavra ou expressão sarcástica; qualquer comentário ou afirmação irônica ou sarcástica. **4** Derivação: sentido figurado. contraste ou incongruência que se afigura como sarcasmo ou troça; acontecimento marcado por esse contraste ou incongruência. Ex: as i. da vida, do destino. (Dicionário eletrônico HOUAISS, 2009)

Ainda em uma perspectiva gramatical, para Cunha e Cintra (1985), a classificação dada aos recursos linguísticos utilizados para tornarem os enunciados mais expressivos encontram-se nas figuras de linguagem. Especificamente, a ironia é uma figura de pensamento, isto é, são mecanismos linguísticos que atuam, geralmente, no campo semântico e permitem que o falante se expresse melhor através de inúmeras combinações de pensamentos que podemos realizar. Neste sentido, a ironia consiste na inversão semântica do sentido e afirmação do contrário daquilo que se pensa. No entanto, apesar da maioria das definições sobre ironia ir para esse caminho, a oposição não prova ser um critério suficiente para explicar como os enunciados irônicos são criados e interpretados.

A intencionalidade é um dos mecanismos mais importantes para diferenciar linguagem literal e figurada. É importante ressaltar que a língua por si só oferece dispositivos linguísticos específicos para expressar deliberadamente diferentes conteúdos figurados: metáfora, alegoria, analogia, prosódia, etc. A figura de linguagem, refere-se, portanto, ao uso de elementos linguísticos que dão a possibilidade de descobrir o significado secundário e depois interpretá-lo dentro de um quadro específico.

As abordagens tradicionais da ironia não explicam o fato da ironia como um fenômeno universal. Curcó (1999) aponta que não existem culturas em que a ironia não é usada ou que precise ser ensinada explicitamente ou aprendida de modo consciente. Pelo contrário, a ironia é uma capacidade que se desenvolve natural e espontaneamente.

2.2. Ironia através da perspectiva pragmática de Grice

O fator que define a ironia é uma relação de oposição entre o significado literal e o que o falante quer comunicar e devido a uma série de inferências que um sujeito faz para compreender um enunciado irônico. Dessa forma, pragmática vai se diferenciar das outras

teorias, tanto retórico-filosóficas quanto linguísticas, pois, procura estudar a língua a partir do uso, das escolhas feitas pelo falante. Esta forma de ver a língua permite dispor de métodos e conceitos que tornam possíveis as análises de elementos que se encontram fora do nível do enunciado que interferem na geração e interpretação do significado implícito.

Para poder abordar as ações através do enunciado e gerar conceitos fundamentais para uma análise mais ampla que a sintática e semântica, Austin (1962) inicia e Searle (1969) desenvolve a Teoria dos Atos de Fala, na qual Austin assume que a comunicação não é composta apenas de palavras e gramática, mas também de ações. Quando um falante diz “Está frio aqui” e há uma janela aberta, ele pode estar querendo muito mais do que constatar o clima daquele lugar, quer pedir ao seu interlocutor que feche a janela. Assim, podemos provocar vários sentidos externos para além daquele que estamos querendo comunicar.

Austin (1962) determina que existem dois tipos de elocuições: as *constativas*, aquelas que descrevem ou relatam algo ou um fato e as *performativas*, aquelas que são, de alguma forma, uma ação. O autor define ainda três atos realizados durante a enunciação, chamados de Atos de Fala: i) *ato locutório*: é o ato de dizer o enunciado, ii) *ato ilocutório*: é o ato de proferir um ato locutório e iii) *ato perlocutório*: é o efeito alcançado do enunciado no ouvinte. É no ato ilocutório que encontramos as elocuições constativas e performativas, dado que é nesse ato que o falante estabelece suas intenções no enunciado. Searle (1979) estende a teoria proposta por Austin e cria uma taxonomia para os atos de fala e estabelece cinco princípios nos quais o falante expressa durante um enunciado: *assertivo*, *compromissivo*, *diretivo*, *declarativo* e *expressivo*.

Dessa forma, não podemos nos prender ao texto como um aglomerado de palavras independentes de toda interação social que o cerca. Entender a ironia sob um viés pragmático é definir que sua identificação só acontece através da percepção da intenção do falante e do contexto em que aquele enunciado é dito. Nas seções seguintes descreveremos o papel de Grice (1975) que complementam a Teoria dos Atos de Fala para a compreensão de um enunciado irônico.

De modo geral, o que a retórica clássica trata como significado figurado, na pragmática de Grice é chamado de implicatura. Segundo o autor, existem dois tipos de implicaturas: as *convencionais* e as *conversacionais*. A primeira refere-se ao significado convencional de um enunciado e a última não está presa ao significado usual e é determinada por princípios básicos de um ato comunicativo, isto é, o que um falante diz difere do que sua frase quer, realmente,

significar. Observemos, inicialmente, a implicatura convencional de acordo com o que foi considerado:

(3) O deputado Eduardo Cunha é um político brasileiro, portanto, é corrupto.

No exemplo (3), está dito que Eduardo Cunha é um político brasileiro e é corrupto e a implicatura convencional através do conectivo *portanto* indica que o *político brasileiro é corrupto*, porém, isso não foi realmente dito. Com base nesse exemplo, é possível entender que a implicatura convencional acontece através da própria força significativa das palavras do enunciado.

As implicaturas conversacionais são, de modo geral, dizer mais do que é realmente dito literalmente. Vejamos o exemplo (4):

(4) a. Eu não acredito mais na política e você?
b. Lula 2018!

Em (4), há a necessidade de fazer inferência para manter o significado do que foi dito, isto é, em (4a) o falante diz não acreditar mais em política, enquanto em (4b) o falante responde “Lula 2018!”, portanto, a compreensão desse enunciado só ocorrerá se inferirmos que ele acredita em política e votará em Lula em 2018.

Um enunciado irônico sempre será implícito e a grande contribuição do autor não está na definição da ironia verbal, mas em introduzir uma explicação sobre o procedimento que um ouvinte segue para reconhecer o enunciado em que o falante tenta comunicar. Em outras palavras, o conceito de implicatura para explicar a ironia é decorrente da relação entre o que é dito e o que está de fato comprometido.

Grice (1975) apresentou um sistema conceitual eficiente para a compreensão de questões que envolvem a significação. Além disso, o autor preocupou-se em descrever e explicar as produções de sentido que vão além do que é dito, ou seja, é possível que um enunciado signifique mais do que seu significado literal. Desta maneira, quando dois indivíduos estão em um diálogo, existem leis implícitas que regem a comunicação. Os interlocutores, mesmo involuntariamente, agem seguindo normas comuns que caracterizam uma cooperação para que a informação possa ser trocada de um modo mais homogêneo possível. Para o autor, estas regras são chamadas de *Princípio da Cooperação* e devem ser seguida em toda a produção de um enunciado respeitando quatro máximas:

- **Máxima de Qualidade:** não diga o que você acredita ser falso; não diga senão aquilo que você possa fortalecer evidência adequada.
- **Máxima de Quantidade:** faça com que sua contribuição não seja tão informativa quanto requerido para o propósito corrente da conversação; não faça a sua contribuição mais informativa do que é requerido.
- **Máxima de Relação:** seja relevante.
- **Máxima de Modo:** seja claro (evite ambiguidades, obscuridade de expressão, seja breve e ordenado).

Como já mencionamos, gerar e processar automaticamente língua natural é algo complicado de formalizar, pois a maioria dos enunciados não segue um padrão e, muitas vezes, seus significados estão estritamente ligados ao contexto. Contudo, existem situações em que nenhuma máxima é violada, ou pelo menos não está claro que qualquer máxima tenha sido violada pelo falante ou existem os casos em que uma máxima é desprezada, propositalmente, como as figuras de linguagem (a metáfora, a hipérbole e a ironia). Nesse contexto, a ideia de ironia para Grice associa-se ao descumprimento da máxima de qualidade “não diga algo que acredita ser falso” - uma das máximas que constituem o Princípio de Cooperação. Em seu exemplo clássico um sujeito X trai uma pessoa A e nesse contexto A diz de X: “X é um grande amigo”. Grice assinala que:

“... é perfeitamente óbvio para A e seu público que o que A disse ou fez como se fosse dizer algo que ele não acredita, e o público sabe que A sabe que isso é óbvio para o público, (...), a menos que o enunciado de A seja inteiramente inútil, deve-se tentar encontrar outra proposição do que a que ele pretende apresentar. Essa deve ser uma proposição obviamente relacionada; a proposição mais obviamente relacionada é a contraditória da qual ele pretende apresentar” (GRICE, 1975, p. 34, trad. nossa)²

O efeito irônico adquirido pela violação da Máxima de Qualidade, no ponto de vista de Grice, está associado à expressão de atitude, sentimento e avaliação dos falantes. Nenhum falante é irônico sem desejar transparecer uma hostilidade, apresentar desprezo ou fazer julgamento depreciativo. Embora a ironia não seja o foco principal nos estudos griceanos, o autor retoma o fenômeno através da pragmática não se distanciando do conceito clássico, ou

² "...it is perfectly obvious to A and his audience that what A has said or has made as if to say is something he does not believe, and the audience knows that A knows that this is obvious to the audience, (...), unless A's utterance is entirely pointless, A must be trying to get across some other proposition than the one he purports to be putting forward. This must be some obviously related proposition; the most obviously related proposition is the contradictory of the one he purports to be putting forward" (GRICE, 1975, p.34)

seja, a ideia de inversão semântica e dissimulação. É importante questionar se o que Grice postula como “perfeitamente óbvio para A e seu público que o que A disse ou fez como se fosse dizer algo que ele não acredita e o público sabe que A sabe que isso é óbvio para o público” é algo que podemos descrever formalmente. Na próxima seção buscamos compreender a ironia além da concepção pragmática, para isso, faremos um percurso sobre um viés cognitivo.

2.3. Perspectiva ecoica de Sperber e Wilson

Apesar dos pensamentos de Grice serem base para os trabalhos de Sperber e Wilson (1981), os autores afirmam que existem vários problemas com a abordagem pragmática, pois, ela não deixa claro porque um ouvinte deve concluir que, ao dizer algo falso, ele comunica o oposto daquilo que realmente pretende. Outro ponto em divergência entre eles é que para Grice, os interlocutores de um enunciado devem estar cientes do contexto, enquanto para Sperber e Wilson o contexto é um conjunto de premissas³ usadas na compreensão de um enunciado.

Os estudos de Wilson e Sperber (2005) consideram como a informação é processada cognitivamente e como é representada através de inferências e se apoia em dois princípios básicos: o lógico e o cognitivo que permanecem constantes de cultura para outra. Explicam também que as abordagens sobre a linguagem figurada (metáfora, eufemismo, ironia) devem ser tomadas com muito cuidado. Embora tenham sido capazes de desenhar uma teoria geral de dispositivos retóricos (SPERBER e WILSON, 1995), a compreensão completa de um texto de humor ou de um enunciado irônico, por exemplo, não é universal. Além disso, postulam a Teoria da Relevância que propõe uma nova perspectiva na compreensão do processamento inferencial e como os processos comunicacionais ocorrem entre os indivíduos.

A afirmação central da Teoria da Relevância é a de que expectativas de relevância geradas por um enunciado são precisas e previsíveis o suficiente para guiar o ouvinte na direção do significado do falante. O objetivo é explicar em termos cognitivamente realísticos a que essas expectativas equivalem e como elas podem contribuir para uma abordagem empiricamente plausível de compreensão (SPERBER e WILSON, 2005, p. 222).

Ademais, postulam que a comunicação humana é resultado de um processo de interação entre falantes e ouvintes e que, ao processarem informações, é possível alterar seus ambientes cognitivos.

³ De acordo com Sperber e Wilson (2005, p. 235), as premissas são "construções de uma hipótese apropriada sobre suposições contextuais pretendidas".

Dito isso, qual é o papel da Teoria da Relevância no processo de compreensão do enunciado irônico? O primeiro elemento da caracterização da ironia são usos particulares de uso interpretativo, no qual um falante não se utiliza de seu enunciado de modo descritivo, mas atribui implicitamente o pensamento que representa seu enunciado para outra pessoa que não seja ele próprio. Um falante pode relatar um pensamento atribuído a outra pessoa e também é capaz de expressar uma própria atitude, simultaneamente, em relação ao conteúdo que relata. Nesses casos, pode-se dizer que o enunciado é ecoico.

De acordo com a explicação proposta pela Teoria da Relevância, a ironia verbal não envolve nenhuma maquinaria especial ou procedimentos que não os já necessários para abordar um uso básico da linguagem, o USO INTERPRETATIVO, e uma forma específica de uso interpretativo, o USO ECOICO. Um enunciado pode ser interpretativamente usado para (meta)representar outro enunciado ou pensamento que se assemelha a ele em conteúdo. O tipo de uso interpretativo mais conhecido é a fala ou pensamento reportado. Um enunciado é ecoico quando ele alcança a maior parte de sua relevância ao expressar a atitude do falante para pontos de vista que ele tacitamente atribui a outrem (WILSON e SPERBER, 2005, p. 246).

Os autores propõem uma explicação para a ironia verbal como um uso ecoico da linguagem: o falante ecoa implicitamente, ou seja, remete um pensamento que atribui a outros pensamentos, sejam eles reais ou não para expressar sua atitude crítica ou ridicularizada dando-a como falsa, irrelevante ou pouco informativa. O conteúdo desse enunciado ecoico se assemelha ao da proposição superficial. O eco pode ocorrer após ao que foi dito, mas também retomar pensamentos reais ou imaginários. Em uma noção mais ampla, o eco tem suas restrições definidoras e, assim, uma representação acessível não pode ser definida como eco, tendo em vista que ele será lembrado a partir da inacessibilidade à representação na qual haverá uma checagem com a relevância do enunciado. Para exemplificar, os autores dão o seguinte exemplo (WILSON e SPERBER, 2005, p. 246):

- (5) Peter: Foi uma festa fantástica.
- (6) Mary: a. [alegremente] Fantástica
 b. [de forma incerta] Fantástica?

 c. [desdenhosamente] Fantástica!

Nos exemplos (5) e (6) dados pelos autores, Mary ecoa um enunciado de Peter indicando em (6a) uma concordância entre eles; em (6b), ela questiona a opinião de e, em (6c) apresenta uma discordância. A fim de reforçar o conceito de eco, damos outro exemplo:

- (7) a. Estou com fome.
b. Está com fome? E o que você acha que eu estou?

Em (7a) a proposição é mencionada implicitamente na resposta (7b). Casos como esse não são usados com o objetivo de informar sobre o conteúdo do enunciado anterior, mas sim para apontar que o enunciado foi recebido e entendido permitindo a expressão imediata do ouvinte.

Nesse contexto, a ironia se enquadra quando o falante menciona uma proposição de modo que a rejeita como sendo falsa, inapropriada ou irrelevante. O reconhecimento de enunciados irônicos como um caso de menção é fundamental para interpretação da ironia. De outro modo, para o ouvinte entender o enunciado, é preciso notar que se trata de um caso de menção e reconhecer as atitudes do falante sobre a proposição mencionada. Os autores afirmam ainda que qualquer objeto no mundo pode ser usado para representar no outro em que ele se pareça. É possível usar um enunciado para representar uma entidade que possua um conteúdo proposicional que tenha uma semelhança a outro enunciado ou oração, pensamento real ou imaginário. Assim, cada enunciado é uma representação pública do pensamento de um falante que possui um conteúdo proposicional semelhante.

Sperber e Wilson (1995) acreditam que os enunciados podem ser considerados interpretações relativamente literais de algum pensamento do falante e, para os autores, é possível representar um enunciado através do pensamento em duas circunstâncias:

- Quando o enunciado representa um pensamento do falante de acordo com o que seu conteúdo proposicional quer de fato representar, descrevendo, verdadeiramente, algum estado de coisa no mundo, isto é, o *uso* de um enunciado faz referência ao que a expressão se refere (referência externa).
- Quando o pensamento ao qual a afirmação representa é uma interpretação de algum outro pensamento a que se assemelha e há uma *menção* de uma sentença referente à própria expressão (metalinguística).

Como resultado, o que a teoria proposta por eles explica é que todos os casos de ironia, padrão ou não e envolvem, geralmente, uma menção a uma proposição anterior. Além disso, apresenta um certo grau de inutilidade. Em outras palavras, a ironia implica a presença de algo

inútil, redundante e totalmente relacionado. Em síntese, postulam que a ironia emerge, como um eco, como resultado da referência ao que foi dito antes, mas em outro contexto. Nesse caso, o falante distingue-se das declarações mencionadas e enfatiza sua atitude negativa.

Através do que apresentamos até o momento, é possível afirmar que a interpretação de um enunciado irônico está intimamente ligada ao contexto. Contudo, Grice acredita que isso acontece quando há a violação da máxima da qualidade, o que resulta em um significado oposto ao que o falante pretende dizer. Para a perspectiva cognitivista de Sperber e Wilson, a ironia é uma menção de outros enunciados existentes que ecoam no que é dito. Desta forma, ao analisarmos as sentenças irônicas no Capítulo 4 será possível observar que existem duas ironias: as explícitas, compostas por contraste entre seus componentes e as implícitas, são irônicas somente por existir um contexto que reiteram outro contexto.

2.4. Ironia *versus* sarcasmo

Muitas vezes a ironia pode ser percebida como a mistura de sarcasmo e sátira, cujo efeito não está baseado apenas em expressar um significado oposto, mas também humor. Embora não haja uma distinção clara sobre os limites de diferenciação entre ironia, sarcasmo, sátira e humor, podemos argumentar que a ironia tende a ser um modo de comunicação mais sofisticado do que o sarcasmo, pois, enquanto a ironia enfatiza uma intensão divertida, o sarcasmo está preocupado com o sentido mordaz e as provocações desordenadas.

Alguns teóricos consideram que a ironia, em específico a verbal, e o sarcasmo são iguais e referem-se a um termo ou outro como sendo o mesmo fenômeno (BURGERS, 2010; CLARK e GERRIG, 1984), enquanto outros afirmam que um é complementar ao outro (ATTARDO, 2000; KREUZ e ROBERTS, 1993). O sarcasmo é visto como uma forma mais agressiva de ironia (ATTARDO, 2000), é dirigido a algo ou alguém (KREUZ e ROBERTS, 1993) e é usado intencionalmente pelo falante (GIBBS et al., 1995). Outra compreensão do sarcasmo é sua utilização para expressar o ridículo (CLIFT, 1999) além de se apresentar através de pistas vocais, como, o tom e a nasalidade da voz (ATTARDO, 2000; HAIMAN, 1998). Embora os pressupostos teóricos apontados acima não apontem nenhuma distinção clara entre a ironia e o sarcasmo, esta pesquisa abordará como sendo fenômenos próximos, porém, distintos.

Outros elementos linguísticos como a sátira e o humor também podem estar relacionados a enunciados irônicos, no entanto, são utilizados para propósitos diferentes. Segundo Singh (2012), o humor consiste na capacidade de reconhecimento de experiências

cognitivas capazes de provocar riso de expressão de algo divertido e inteligente e refere-se, principalmente, no reconhecimento de incongruências ou peculiaridades presentes em uma determinada situação. O autor também esclarece sobre a sátira, afirmando ser uma figura de linguagem que enfatiza a fraqueza e não a pessoa fraca e pode ser usada para expor e criticar algo em um indivíduo ou mesmo a própria sociedade através do humor, da ironia, do exagero ou do ridículo.

Entendemos, portanto, a ironia como uma expressão verbal cujos constituintes formais, ou seja, as palavras, tentam comunicar um significado subjacente oposto ao expresso. Além disso, esta pesquisa faz a diferença entre o *objetivo* e o *efeito*. Ou seja, o objetivo da ironia, de acordo com a definição, é comunicar o oposto do que é literalmente dito. O efeito, entretanto, pode ter uma interpretação sarcástica, satírica ou mesmo divertida que, sem dúvida, introduz conotações negativas. Neste contexto, é conveniente tratar ironia e os dispositivos relacionados como diferentes facetas do mesmo fenômeno. Portanto, dispositivos como sarcasmo, sátira ou humor serão considerados como extensões específicas de um conceito geral e amplo de ironia.

Capítulo 3

Análise de Sentimentos e Recursos Linguísticos

Neste capítulo abordaremos as principais bases teórico-metodológicas voltadas à Linguística Computacional nas quais a atual pesquisa está ancorada. Primeiramente, será apresentada Análise de Sentimentos, uma subárea do PLN que tem ganhado bastante atenção nos últimos anos. Em um segundo momento procura-se descrever a metodologia os recursos linguísticos mais importantes que são utilizados nesse trabalho, concentrando-se recursos utilizados no processo de criação e anotação do *corpus* que está inserido nessa pesquisa.

3.1. Análise de Sentimentos

As opiniões podem influenciar os pensamentos, os comportamentos e o modo de agir de uma pessoa, pois o que cada indivíduo acredita está ligado à maneira de como os outros veem o mundo e com o avanço das tecnologias e a facilidade de comunicação, as pessoas deixaram o “boca a boca” e passaram a expressar suas opiniões e sentimentos em mídias sociais. Tais opiniões, se adequadamente extraídas e analisadas, são capazes não só de explicar fenômenos sociais, mas também podem prevêê-los.

À vista disso, a identificação automática de sentimentos e opiniões em textos extraídos de redes sociais é um recente objeto de estudo para áreas como a Computação e a Linguística, buscando contribuir com o reconhecimento, extração e classificação de opiniões na *Web*. Liu (2012) define o estudo computacional de opiniões, sentimentos, avaliações, atitudes, emoções, subjetividades voltadas a uma entidade, como um produto, pessoa, organização, serviços como Análise de Sentimentos ou Mineração de Opinião. Tal área ainda pode classificar estes dados subjetivos como positivo ou negativo e, em alguns casos, neutro.

Pang e Lee (2008) fazem uma observação sobre a utilização dos termos Análise de Sentimentos e Mineração de Opinião. De acordo com os autores, o termo Mineração de Opinião refere-se às investigações relativas à extração e análise de informações sobre produtos, enquanto Análise de Sentimentos está ligada à classificação e ao tratamento computacional de subjetividade, sentimentos e opiniões em textos. Apesar de afirmarem que ambos os termos denotam o mesmo campo de estudo, sendo considerados uma subárea da Análise da Subjetividade, eles incentivam o uso dos termos Análise de Sentimentos. Dessa forma, em

concordância com os autores, faremos o uso do termo Análise de Sentimentos no decorrer desta pesquisa.

Apresentada a área e suas denominações, nas seções a seguir, descreveremos os principais conceitos e aplicações às quais a Análise de Sentimentos se dedica.

3.1.1. Principais conceitos da Análise de Sentimentos

Alguns conceitos são fundamentais na Análise de Sentimentos. Liu (2012) define que o que os humanos produzem em formato textual pode ser classificados em *fato* ou *opinião*. Os primeiros são, geralmente, as sentenças objetivas desprovidas de sentimentos ou opinião e transmitem alguma informação sobre uma entidade ou alvo. Opiniões são sentenças subjetivas que carregam alguma avaliação, sentimento ou impressão sobre algo. Mais formalmente, para o autor, uma opinião corresponde a uma quintupla: *entidade*, *aspecto* da entidade, *polaridade* da opinião, *autor* da opinião e o *tempo* na qual a opinião foi expressa. *Entidade* pode ser um produto, serviço, pessoa ou organização, por exemplo, uma televisão de uma determinada marca pode ser considerada uma entidade; tamanho, peso, qualidade da imagem são os seus *atributos*. O *autor* da opinião é quem expressou a opinião em um determinado *tempo*. Uma opinião possui uma *polaridade* que define o sentimento do autor em relação a uma entidade, isto é, uma opinião pode ser positiva, negativa ou neutra.

Liu (2012) classifica as opiniões em regulares ou comparativas, diretas ou indiretas e implícitas ou explícitas. Em seguida descreveremos cada um desses conceitos:

- *Opiniões regulares*: o autor da opinião expressa um sentimento sobre um alvo, por exemplo, “O governo atual é muito ruim”.
- *Opiniões comparativas*: a opinião é dita através da relação de similaridade ou diferença entre duas ou mais entidades, por exemplo, “O preço do açúcar é muito melhor que o do álcool”.
- *Opiniões diretas*: é direcionada diretamente a uma entidade, como, “Ela é uma boa presidenta”.
- *Opiniões indiretas*: são ditas indiretamente sobre uma entidade ou aspecto da entidade, por exemplo, “Desde que a presidenta assumiu o governo o país afundou”.
- *Opiniões explícitas*: são as que expressam diretamente a opinião do autor, como “Eu não gosto do atual governo”.

- *Opiniões implícitas*: indicam um sentimento subjacente ao que está sendo dito, por exemplo, “Nada como arrecadar fundos em nome dos pobrezinhos do Brasil ou de qualquer lugar”.

O autor afirma ainda que a detecção do sentimento pode ocorrer em diferentes granularidades que dependem do contexto e da aplicação em que a análise é feita, sendo possível acontecer em três níveis distintos. São eles: i) *nível de documento*: procura-se classificar e tratar um texto opinativo em sua totalidade e pode ser rotulado como positivo, negativo ou neutro; ii) *nível de sentença*: um documento é subdividido e cada sentença é analisada conforme sua polaridade e iii) *nível de aspecto*: são extraídas e agrupadas as opiniões de uma determinada entidade e aspectos e identificados os sentimentos para cada um deles.

3.1.2. Aplicações em Análise de Sentimentos

Como já foi dito, devido ao grande volume de textos que circulam na Web, processá-los manualmente passou a ser uma tarefa inviável e foi necessário desenvolver aplicações capazes de processar dados não estruturados. Tais informações são de extrema importância, pois, podem auxiliar na tomada de decisão de uma pessoa ou dizer sobre a qualidade de um produto. Para Liu (2010), a Análise de Sentimentos pode ser útil, por exemplo, para avaliar produtos ou serviços focados em áreas do *marketing* e relações públicas. De fato, a aplicação mais comum tem como foco revisões de produto e serviços de consumo (FELDMAN, 2013)

Feldman (2013) ainda detalha as aplicações na área em: i) empresas e organizações que buscam por opiniões dos consumidores sobre os produtos que comercializam e/ou os serviços que produzem, ii) indivíduos que tomam decisões para comprar produtos ou contratar um serviço com base em opiniões *on-line*, iii) recuperação de opinião para buscas gerais de opiniões. Já para os consumidores de um produto ou clientes de uma empresa o ponto principal é buscar automaticamente pelas opiniões de outros usuários. Na política, responsáveis por campanhas de partidos ou de um político em si podem extrair, por exemplo, informações importantes sobre opiniões de eleitores em período eleitoral ou dados de outro candidato. O Quadro 1 fornece uma esquematização das aplicações em Análise de Sentimento.

Aplicações em Análise de Sentimentos	
Negócio	Opinião de consumidores
	Reputação de um produto ou empresa
	Publicidade <i>on-line</i>

	Comércio <i>on-line</i>
Política	Aconselhamento em votação
	Esclarecimento das posições de um candidato
	Opinião de um eleitor sobre um candidato
Opinião Pública	Monitoramento de opiniões sobre fenômenos sociais

Quadro 1. Principais aplicações em Análise de Sentimentos

3.1.3. Desafios

O impulso para o desenvolvimento e criação de estudos são os desafios e busca de melhores resultados que a área enfrenta. Buscando apresentar uma solução para este problema, Liu (2012) propõe a criação de sistemas que reconheçam, monitorem, extraiam e processem automaticamente textos opinativos sob o viés da Análise de Sentimentos. Segundo o autor, um dos principais desafios para a área é, certamente, o reconhecimento e classificação automática de opiniões e fatos naturalmente utilizados em língua natural. Embora a diferenciação entre fato e opinião pareça uma questão um tanto quando óbvia, Pang e Lee (2008) acreditam que podem existir em sentenças objetivas opiniões sobrepostas.

Um dos maiores desafios na Análise de Sentimentos, conforme Pang e Lee (2008) indicam, encontra-se na grande diversidade de material textual e seus diferentes estilos, estruturas sintáticas e semânticas, tipos de textos, variações linguísticas e as figuras de linguagem. Nesse quadro podemos acrescentar também os erros de digitação, abreviações, linguagens de Internet e duplicação de letras.

Dessa forma, uma ferramenta de Análise de Sentimentos deve identificar e processar essas adversidades textuais e reconhecer e classificar uma sentença entre positiva, negativa ou neutra, porém, nem sempre é possível ter bons resultados. Com o decorrer das pesquisas e o desenvolvimento de recursos linguísticos para a melhoria da precisão dos métodos em Análise de Sentimentos será possível, futuramente, boas ferramentas capazes de processar opiniões e ter uma boa taxa de acertos e qualidade de análise automática. Na próxima seção apresentaremos um estudo sobre anotação de *corpus*, um dos recursos linguísticos possíveis que, de alguma forma, colaboram para o que foi dito neste parágrafo.

3.2. Anotação de *Corpus*

O conjunto de textos organizados e levantados de acordo com determinados critérios, com o objetivo de permitir o estudo de uma língua de forma representativa é comumente

chamado de *corpus*. Os *corpora*, plural de *corpus*, são grandes coleções de textos que representam uma língua (monolíngue) ou mais línguas (multilíngue). O progresso de conhecimentos computacionais e metodologias voltadas ao desenvolvimento de *corpus* levaram ao crescimento da Linguística de *Corpus*, doravante LC, como seu próprio campo de pesquisa, motivando também campos relacionados com a descrição e investigação linguística, como a Linguística Computacional e o PLN.

Para Santos (1999), existem dois tipos de pesquisadores de *corpus*. O primeiro grupo são os compiladores de *corpora*, ou seja, são aqueles que estão preocupados com questões como criar, estruturar e anotar um *corpus*. O outro grupo são os que estão empenhados em extrair informações a partir de um *corpus* de estudo. Fora estes dois grupos, a autora ainda aponta para um terceiro grupo emergente: os desenvolvedores de ferramentas⁴ para *corpora*. Porém, tal distinção não implica que um pesquisador deva se limitar em apenas um tipo de investigação, aliás, a maioria dos trabalhos recentes pautados pela LC não se satisfazem em apenas um tipo de pesquisa. Em outras palavras, é possível, em um único trabalho, criar um *corpus* e determinar qual o objeto de estudo neste recorte da língua para, então, extrair informações necessárias e, assim, investigá-las visando a criação de algoritmos com a finalidade de aplicação em aprendizado de máquina ou ferramentas de PLN.

Além de abarcar grandes coleções de textos que representem uma língua, um *corpus* também pode possuir informações adicionais, chamada de anotação. A tarefa de anotação pode ser definida como a prática de adicionar informações linguísticas a um *corpus* (LEECH, 1997) inseridas por humanos ou máquinas com um objetivo teórico ou prático (HOVY e LAVID, 2010), explicitando o que antes era implícito na estrutura textual, o que permite que seja possível recuperar e analisar, rapidamente, informações contidas no *corpus* (MCENERY e WILSON, 2001).

Em uma perspectiva linguística, Santos et al. (2015) consideram distintos os interesses entre a linguística e o PLN na anotação de *corpus*, uma vez que a tarefa em larga escala está voltada às necessidades da linguística computacional, enquanto numa anotação que privilegia os estudos linguísticos a exigência de quantidade perde a relevância. Para os autores, a anotação é uma possibilidade de estudar, empiricamente, um *corpus* compilado e enfatizam o processo

⁴ As aplicações em PLN incluem, por exemplo, verificação automática ortográfica e gramática, reconhecimento de fala, sumarização, recuperação de informações e tradução automática. *Corpora* podem ser utilizados também na criação de dicionários e gramáticas de uma língua natural.

de anotação como um procedimento de interpretação, classificação e formalização de um fenômeno linguístico. Em contraponto, a criação do Penn Treebank⁵ (MARCUS et al., 1993), composto por três milhões de palavras, contribuiu com o desenvolvimento de *parsers* de alto desempenho, além disso, serviu de incentivo para a criação de outros grandes *corpora* anotados, como por exemplo, a FrameNet⁶ (BAKER et al., 1998) e PropBank⁷ (PALMER et al., 2005). Assim, entende-se que *corpora* anotados representam um importante recurso tanto para estudos linguísticos como para linguístico-computacionais, posto que anotações acrescentam valor ao *corpus*, permitindo buscas e processamentos mais refinados.

3.2.1. Definições

Leech (1997) define anotação como o processo de “adicionar informações interpretativas e linguísticas para um *corpus* eletrônico de dados de linguagem falada e escrita” é referido como anotação de *corpus*. Anotação, para o autor, é o produto final desse processo de compreensão mental de um texto.

Petrillo e Baycroft (2010) definem a anotação como uma metodologia utilizada para adicionar informações, ou metadados, a um documento a algum nível linguístico. Conforme os autores, a diferença entre uma anotação e qualquer outro tipo de metadado é que a anotação é fundamentada em um conhecimento e nível específico do texto. Além disso, eles classificam e caracterizam a anotação em três tipos. São elas:

- **Anotação manual:** é mais precisa, porém, muito trabalhosa e é, muitas vezes, usada para o treinamento de máquina para a execução de anotações automáticas.
- **Anotação automática:** é menos precisa que a anotação manual, mas pode funcionar em um volume grande de documentos, tarefa na qual o humano realizar razoavelmente.
- **Anotação semiautomática:** as etapas manuais podem vir em várias partes do processo geral. Uma etapa manual inicial pode identificar um conjunto básico de dados ou termos os quais seriam utilizados na criação de uma lista de palavras que uma máquina poderia reconhecer em vários documentos. A anotação manual ainda refinaria o que foi encontrado pelo computador e os resultados seriam desenvolvidos em um processo automático que o tornaria mais preciso.

⁵ Penn Treebank (<http://www.cis.upenn.edu/~treebank/>)

⁶ FrameNet (<http://framenet.icsi.berkeley.edu/>)

⁷ PropBank (<http://verbs.colorado.edu/~mpalmer/projects/ace.html>)

Como as sentenças podem ter múltiplas leituras e interpretações pouco claras, a anotação manual é um trabalho essencial e insubstituível para o PLN. Assim, um anotador tem a árdua tarefa de anotar os recursos linguísticos conforme sua experiência como falante nativo da língua, mas, ao mesmo tempo, precisa seguir um manual, caso contrário as anotações seriam muito desiguais. Por isso, é fundamental ter um manual de anotação que unifique e instrua os anotadores com as mesmas diretrizes.

Hovy e Lavid (2010) denominam anotação como o processo de adicionar novas informações em dados brutos por humanos, isto é, anotadores. Normalmente, as informações são adicionadas por muitas pequenas decisões individuais, em muitos lugares ao longo dos dados. O processo de adição geralmente requer decisões cognitivas que dependem tanto dos textos brutos quanto de alguma teoria ou conhecimento que o anotador tenha internalizado anteriormente. Segundo os autores, para o PLN, a anotação fornece dados para permitir a aprendizagem de máquina de alguma aplicação, transformando o texto “puro” de entrada em texto interpretado e marcado. Para a realização da tarefa é necessário ter vários humanos anotando manualmente os textos com informações, comparar o desempenho e, assim, treinar um algoritmo de aprendizagem para realizar o trabalho automaticamente. No ponto de vista da Linguística, a anotação enriquece o *corpus* com informações linguísticas e permite que o pesquisador descubra fenômenos na língua através da anotação, além de fornecer registros e dados explícitos da análise do pesquisador.

De acordo com os autores, a produção de uma anotação de alta qualidade é essencial para a construção bem-sucedida de processamento automático de texto. Em um esquema de anotação, é importante, primeiramente, construir diretrizes que detalham a tarefa antes mesmo do início da anotação e tais diretrizes devem ser revisadas durante todo o processo de anotação. Entretanto, um dos maiores problemas com o esquema de anotação é o alto custo de treinamento e manutenção dos anotadores.

Para Hovy e Lavid (2010), para que o processo de anotação seja confiável – além de possuir diretrizes que garantam que todos os anotadores realizem de modo uniforme toda a tarefa – é necessário também que a anotação siga um esquema:

1. Definir a tarefa de anotação com base na necessidade da pesquisa.
2. Selecionar os dados a serem anotados.
3. Escrever um conjunto detalhado de diretrizes da anotação.

4. Criar e utilizar boas ferramentas de anotação.
5. Encontrar e treinar anotadores.
6. Anotar o texto:
 - a. Anotar o texto com base nas diretrizes.
 - b. Revisar as diretrizes de anotação, se necessário.
 - c. Monitorar o acordo inter-anotador e retreinar os anotadores.
 - d. Se necessário, modificar a anotação com base nas diretrizes revisadas.
7. Liberar o *corpus* para outras comunidades de pesquisa.

3.2.2. Formatos de anotação

Com a finalidade de satisfazer necessidades de ferramentas de manipulação e anotação de *corpus*, criaram-se formatos próprios para codificação de anotação de *corpus*. Tais formatos podem ser representados, por exemplo, o uso de anotação *stand-off* ou a utilização da linguagem XML – e a escolha de qual formato de anotação depende do recurso linguístico que será identificado e descrito, do projeto e do *corpus*.

Os dados anotados devem ser descritos por uma nomenclatura que siga um padrão e que seja legível para os humanos. Devem permitir também a inserção de outros dados, além de poder ser extraídos ou separados do *corpus*. A anotação *stand-off* permite trocar a anotação de um nível, por exemplo, trocar de etiquetador, sem alterar as outras anotações. Isso ocorre porque cada nível de anotação encontra-se em arquivos separados e o arquivo-fonte não possui anotação. Em oposição, na anotação *in-line* todas as anotações estão no mesmo arquivo-fonte.

O XML (*Extensible Markup Language*) é uma linguagem de editoração que apresenta um formato universal de estruturação de dados. Documentos XML são legíveis por humanos e manipuláveis por máquinas. Os dados são representados de modo estruturado através de um conjunto de *etiquetas*⁸. Por esses conjuntos não serem fixos nem limitados, permite-se que os autores criem suas próprias etiquetas conforme suas necessidades e, por isso o XML é bastante aceito dentro dos padrões de anotação para aplicações em *corpus* baseados em PLN. Essa linguagem permite descrever qualquer espécie de dado, além de ser um padrão aberto para troca de informações. Nesse contexto, por muitas ferramentas reconhecerem esse tipo de padrão, é

⁸ O termo *tag* também pode ser utilizado como sinônimo de *etiqueta*.

realizada a criação de recursos linguísticos padronizados o que favorece a reutilização do *corpus* e ainda facilita a extração de dados uma vez que documentos XML são interpretáveis por humanos e manipuláveis por máquina. Desta maneira, nos utilizamos da linguagem XML como formato de nossa anotação, visto que durante o processo foi necessário criar diversas etiquetas para delimitar e indicar o fenômeno linguístico ali representado.

3.3. Trabalhos Relacionados

Nesta seção, nos concentramos em tarefas de Análise de Sentimentos e, em particular, em projetos de pesquisa que, de alguma forma, precedem a este trabalho, porque dizem respeito ao desenvolvimento de recursos similares aos desenvolvidos nesta dissertação. Apenas alguns recursos foram feitos até agora para o português, em específico para o português do Brasil, enquanto o inglês ainda soma a maioria das pesquisas realizadas até o momento. Entre eles, citamos aqueles que estão de algum modo mais relacionados ao presente estudo.

Em particular, um trabalho recente de Van Hee et al. (2016) aborda a anotação manual de *corpus* e expressões irônicas para o inglês. Para entender como a ironia é compreendida, os autores descrevem a construção de um *corpus* construído por 3 mil mensagens do Twitter com as *hashtags* #irony, #sarcasm e #not e desenvolvem diretrizes para um esquema refinado de anotação, além de explorar recursos lexicais, sintáticos e semânticos capazes de reconhecer automaticamente textos irônicos. Semelhante à anotação desta pesquisa, os textos anotados foram subdivididos em irônicos e não irônicos. Os textos irônicos também foram divididos em outras duas categorias: ironia situacional (situações irônicas) e outros tipos de ironias verbais (não descrevem nem ironia situacional nem choque entre polaridade). O desempenho da classificação quanto às marcas de ironia nos tweets coletados tem 56% de precisão obtida, o que demonstra que o sistema não depende de certas informações, como as *hashtags*. Como resultado, os autores observaram que o sistema apresentava melhor resultado em instâncias que não era necessário o uso de *hashtags*, revelando uma precisão de 83.43% em tweets irônicos que não tinham essa marca. Isso mostra que o conjunto de dados é fortemente lexical, pois, apesar das marcas ajudarem aos humanos a reconhecer a ironia no texto, sua compreensão e reconhecimento está mesmo nos itens lexicais. Através de uma análise qualitativa observaram que o classificador apresentava um melhor desempenho em *tweets* nos quais a ironia era decorrente de um choque de polaridade, isto é, o sentimento expresso era o oposto daquilo que era de fato dito.

O trabalho de Barbieri et al. (2014) descreve um modelo computacional de detecção de sarcasmo no Twitter, além disso procuram criar um *corpus* para pesquisas de sarcasmo e um conjunto de dados capazes de testar o modelo e detecção de sarcasmo. O *corpus* adotado pelos autores contém exemplos positivos marcados como sarcásticos através da *hashtag* #sarcasm e exemplos negativos, marcados por outros tipos de *hashtag*. No total, o *corpus* de pesquisa contém 60 mil *tweets* divididos igualmente em seis categorias: sarcasmo, educação, humor, ironia, política e jornal (10 mil *tweets* dos *The New York Times*, *The Economist* e *The Guardian*). Os outros 50 mil *tweets* foram selecionados automaticamente e anotados com as *hashtags* #humour, #irony, #politics e #sarcasm.

Barbieri et al. (2014) utiliza sete grupos de recursos que representam cada *tweet*: frequência, escrita-fala, intensidade, estrutura, sentimentos, sinônimo e ambiguidade. Para avaliar o sistema, os autores contrastaram cinco conjuntos compostos pelas categorias criadas pelos autores: sarcasmo vs. educação, sarcasmo vs. humor, sarcasmo vs. ironia e sarcasmo vs. jornal, sendo que cada conjunto contém 10 mil *tweets* sarcásticos e outros 10 mil *tweets* de outra categoria não-sarcástica. Através dos resultados das comparações entre *tweets* negativos e sarcásticos e positivos não sarcásticos, foi possível também observar que o sarcasmo usa menos advérbios, porém quando utilizados são mais intensos e como os *tweets* sarcásticos denotam sentimentos mais positivos que a ironia.

Em uma pesquisa similar, González-Ibáñez et al. (2011) apresentam um método de construção de *corpus* de *tweets* sarcásticos e utilizam-se do *corpus* para comparar com outros *tweets* que exprimem um sentimento, positivo ou negativo, sem sarcasmo. O *corpus* consiste em 900 *tweets* classificados em sarcástico, positivo e negativo. Através de uma análise do impacto de itens lexicais e pragmáticos, os autores compararam o desempenho de técnicas de Aprendizagem de Máquina e de anotação manual de *corpus*. Segundo os autores, as principais problemáticas dos resultados não serem tão satisfatório e apresentarem dificuldade no reconhecimento automático se dá pelo tamanho do texto (140 caracteres) e a falta de contexto explícito.

Carvalho et al. (2009) elaboraram algumas pistas para identificar automaticamente frases irônicas, com base em padrões sintáticos e morfológicos específicos, *emoticons*, expressões onomatopaicas, pontuação e aspas. Algumas dessas pistas são específicas para o português (padrões morfológicos), enquanto outras parecem ser independentes da linguagem e estão presentes em todos os lugares nas mídias sociais (*emoticons*). A coleção de textos é

composta por 8.211 notícias e comentários correspondentes postados por leitores on-line. Todos os padrões daquele estudo restringem de alguma forma a polaridade de possíveis sequências de correspondência, uma vez que as pesquisas estavam particularmente interessadas em reconhecer a ironia em frases aparentemente positivas envolvendo entidades nomeadas (NE). No Quadro 2, como é possível observar, os padrões criados pelos autores restringem-se, em sua maioria, em casos de polaridade positiva.

Pistas	Padrões	Exemplo
Pdim	(4-Gram+ NEdim NE 4-Gram+)	"Socratezinho"
Pdem	DEM NE 4-Gram+	"Este Sócrates é muito amigo do Sr. Jack"
Pitj	ITJpos (DEM ADJpos)* NE (! ! ...)	"bravo", "força"
Pverb	NE (tu)* ser2s 4-Gram+	"tu" versus "você"
Pcross	(DEM ART) (ADJpos ADJneut) de NE	"O comunista do ministro"
Ppunct	4-Gram+ (! ! ?! ?)	"!?!?!"
Pquote	"(ADJpos Npos){1,2}"	"Que bom, Sócrates"
Plaugh	(LOL AH EMO+)	"lol" "ah", "eh" and "hi" ":", ";", "-)" ".P"

Quadro 2. Pistas e padrões utilizados por Carvalho et al. (2009)

Como resultado, os padrões mais produtivos estão diretamente relacionados ao uso de sinais de pontuação que são a forma de representar expressões orais ou gestuais no texto escrito. Os padrões mais produtivos envolvem *emoticons* ":" e expressões onomatopeicas como o riso, marcas de pontuação "!?!?", aspas e interjeições positivas. Notavelmente, todos esses padrões estão de alguma forma relacionados à oralidade, o que mostra que construções irônicas são frequentemente sinalizadas por pistas orais. Em termos de distribuição de polaridade, 55.5% das entradas têm polaridade prévia negativa, 21.8% têm polaridade prévia positiva e os restantes 22.7% são considerados neutros, como podemos observar na Tabela 1.

	irônico	não irônico	indecisos	ambíguo
<i>Pitj</i>	44.88%	13.39%	40.94%	0.79%
<i>Ppunct</i>	45.71%	27.53%	26.75%	0.00%
<i>Pquote</i>	68.29%	21.95%	2.73%	7.03%
<i>Plough</i>	85.40%	0.55%	11.13%	2.92%

Tabela 1. Resultados para os padrões usados em experimentos por Carvalho et al. (2009)

Para o português do Brasil, Vanin et al. (2013) apresentam um trabalho inicial para o português do Brasil sobre alguns padrões de detecção de ironia em *tweets*. No artigo, foram desenvolvidos padrões com formas diminutivas, que podem tanto expressar sentimentos positivos, como afeto, ternura e intimidade, mas como também podem apresentar conotações sarcásticas ou irônicas quando a intenção é desvalorizar ou insultar uma determinada entidade com base em uma coleção de 2.780 *tweets* sob o domínio "fim do mundo". O Quadro 8 mostra os quinze padrões implementados e classificadas em sete categorias: listas (C1); expressões (C2); *Part-of-Speech* (C3); *Part-of-Speech* + expressões (C4); *Part-of-Speech* + listas (C5); *Part-of-Speech* + entidades nomeadas (C6); pronomes demonstrativos + entidades nomeadas (C7) e símbolos (C8).

Categorias	Padrões	Expressões
C1	P1	Listas de expressões de riso
	P2	Listas de emoticons
C2	P3	"só que"
	P4	"sim"
	P5	"seria"
	P6	"na boa"
	P7	"medo" "Medo!" "#medo"
	P8	#ironia" #sarcasmo" #joking" #kidding"
C3	P9	ADV + ADV ADJ + ADJ
C4	P10	"tão" + ADJ ou "tão" + ADV
C5	P11	ADJ + Lista de emoticons
C6	P12	DET + ADJ + (PREP+DET) + NE
C7	P13	Pronome Demonstrativo + NE
C8	P14	Expressões
	P15	!* ?* !*?* ?*!*

Quadro 3. Padrões utilizados por Vanin et al. (2013)

Do que foi posto durante esta seção, observamos que a Análise de Sentimentos vem desempenhando um interesse em pesquisa tanto na Computação quanto na Linguística, uma vez que a disponibilidade de mídias sociais permitiu o desenvolvimento de novos recursos para algoritmos de aprendizado de máquinas, por exemplo, a descrição de textos de opinião e avaliação sobre um determinado alvo. Isto justifica a utilização de alguns recursos linguísticos para a realização da tarefa de descrição de um *corpus* de opinião. Apresentados os padrões de detecção de ironia, especificamente, os trabalhos para o português como de Carvalho et al. (2009) e Vanin et al. (2013) buscamos nos basear em algumas pistas propostas pelos autores relacionados nesta seção – como é o caso dos diminutivos, pontuação e aspas – na anotação descrita no Capítulo 5.

Capítulo 4

Mídias sociais e construção do *corpus*

Neste capítulo, apresentamos um dos recursos fundamentais para o desenvolvimento desta dissertação: as mídias sociais. Em particular, procuramos mostrar que esse tipo de mídia e, principalmente, a sua utilidade para a identificação automática de opiniões ao construir *corpus* deste tipo de texto, uma vez que os dados fornecidos são de fácil acessibilidade. Em seguida, baseando-nos na importância das mídias sociais para os estudos da Análise de Sentimentos, descrevemos o *corpus* construído para a realização deste trabalho.

4.1. Mídias sociais como *corpus* para Análise de Sentimentos

As mídias e plataformas sociais como o Twitter, Facebook e Youtube e outros tipos de conteúdo gerado pelo usuário (*user-generated content* ou UGC), são compostas por informações criadas por milhões de pessoas e são acessadas rapidamente e podem afetar a reputação de uma empresa, marca ou uma figura pública, como, os políticos. Em comparação com fontes de texto mais convencionais, como jornais, revistas ou textos literários, as mídias sociais consistem em mensagens mais curtas, como é o caso dos *tweets*, que não podiam exceder 140 caracteres⁹ e outros serviços de redes sociais que oferecem suporte para mensagens mais longas, por exemplo, postagens no Facebook. No entanto, a maioria dos usuários ainda prefere escrever textos mais reduzidos, visto que a linguagem das mídias sociais é, geralmente, mais informal, com numerosos desvios ortográficos (acidentais e/ou deliberados) e inconsistências gramaticais. Como resultado, esses documentos contêm numerosas palavras fora do dicionário e representam desafios para as ferramentas de PLN.

Devido ao crescente interesse de aplicações em Análise de Sentimentos aos dados de redes sociais *on-line*, Pallavicini et al. (2017) acreditam que pesquisa nesta área reconhecem as limitações consequentes ao tratamento de características complexas da língua natural sem considerar os dados coletados através das redes sociais como uma “rede dados”. Contudo, a maior parte do trabalho em Análise de Sentimentos baseia-se em informações textuais expressas em postagens e comentários *on-line* (PANG e LEE, 2008). Pachucki et al. (2010) acreditam que um dos principais problemas com a interpretação de ferramentas de Análise de

⁹ A partir de novembro de 2017, o Twitter atualizou o limite para 280 caracteres.

Sentimentos é porque tais sistemas não consideram dados *on-line* em redes que foram coletados.

Segundo Pozzi et al. (2017), apesar da tendência em pesquisas em Análise de Sentimentos para mídias sociais aplicarem técnicas herdadas desde o início de 2000, é importante considerar a evolução das fontes de informações e opiniões expressas, bem como as estratégias disponíveis atualmente. De acordo com os autores, além das dificuldades já discutidas neste trabalho, a área apresenta novas complexidades, por exemplo, metadados (idade, sexo, localização), conteúdo ruidoso e citam, novamente, mensagens curtas. Assim, para eles, uma vez que as mídias sócias causam um impacto na língua, encontram-se novos desafios em relação à análise de sentimentos, emoções e opiniões, principalmente, por causa do constante desenvolvimento da língua em UGC, isto é, as palavras que nos cercam nos influenciam diariamente nas palavras que utilizamos. Desta maneira, grande parte das palavras que encontramos em nossas redes sociais é uma evolução da nossa interação com a tecnologia. Trazendo essa informação para a nossa pesquisa, por exemplo, o termo “*coxinha*”¹⁰ é facilmente encontrado em plataformas sociais como Facebook e Twitter, mas também em comentários de notícias do domínio político e não é difícil em uma conversa informal ouvirmos alguém dizer, por exemplo, “João é um coxinha!”.

Pozzi et al. (2017) acreditam que isso acontece porque a língua que usamos para nos comunicarmos nas redes sociais tende a ser mais flexível do que a escrita normativa, havendo, então, uma combinação de linguagem informal e pessoal. Desta forma, resolver estes problemas exige um forte empenho linguístico-computacional para desenvolver descrições linguísticas e algoritmos capazes de processar automaticamente tais dados e, conseqüentemente, essa evolução linguística influencia no modo como a ironia e o sarcasmo são compreendidos por estes sistemas. Em conclusão, os autores ainda relacionam a natureza das mídias sociais, que por definição são dinâmicas e heterogêneas, com o desafio diário envolvendo as entidades conectadas entre si. Assim, lidar com esse tipo de ambiente no qual é necessário relacionar informações a todo momento, torna-se obrigatório ir além do conteúdo textual abordado.

¹⁰ O termo “*coxinha*” é uma gíria utilizada para atribuir às pessoas caracterizadas por valores morais em excesso e em conformidade com o que é imposto pela sociedade. Desta forma, o adjetivo “*coxinha*” pode ser utilizado para qualificar uma pessoa que se enquadre em situações de conformismo, seja por suas posições políticas, religiosas ou sociais. O termo também é empregado para se referir àquele que usa roupas e acessórios de grife e por ser influenciado pelas redes sociais e programas de TV.

Focamos nesta pesquisa a identificação de aspectos que podem marcar expressões irônicas em mídias sociais ao visar as personalidades, como os políticos e a criação de um método de mídia social que pode ser usado durante campanhas políticas. Nesta perspectiva, Hubert Massey¹¹, consultor da South University Savannah, afirma que “o uso das mídias sociais na campanha de hoje não é apenas importante, é fundamental”, pois, segundo ele, “milhões de pessoas estão envolvidas no uso diário de redes sociais” e vê como “a oportunidade de estar em contato com um grande número de eleitores de forma rápida, constante e a baixo custo”. Assim, como fornecem uma fonte rica de informações, as mídias sociais têm um enorme impacto em qualquer tipo de eleições, por exemplo. Devido à velocidade das comunicações e ao número de pessoas registradas e envolvidas por hora, esse impacto está ficando ainda mais significativo. Para Ceron et al. (2009), além de ser mais barato e rápido comparado às pesquisas tradicionais, uma análise de mídias sociais permite, por exemplo, o monitoramento de uma campanha eleitoral diariamente. Para os autores, este cenário torna-se bastante fascinante, pois, a previsão de uma eleição é um dos poucos exercícios sobre eventos sociais em que uma medida independente do resultado que um modelo está tentando prever é disponível de forma instantânea.

4.2. Construção do *corpus*

A linguagem utilizada em mídias sociais está baseada na abreviação e simplificação de palavras, permitindo uma comunicação mais rápida, em repetições de vogais e de pontuação que sugere uma reprodução da fala, além de gírias, erros gramaticais e ortográfico e uso de linguagem figurada. Para Bisognin (2009) por se tratar de um léxico bem específico dos meios digitais, a linguagem utilizada em ambientes virtuais possui características da oralidade no modo do usuário expressar suas emoções, opiniões, impressões e avaliações.

Dado que a essência desta pesquisa é a ocorrência de ironia em *corpus* opinativo, procuramos trabalhar com textos nos quais pudéssemos encontrar a opinião direta do autor. Por este motivo, a escolha de comentários dos leitores de um jornal de grande circulação e que tenha grande acesso *on-line* e abre espaço para o público. Além do mais, através deste tipo de mídia social, é possível o leitor colocar suas impressões, opiniões e sentimentos, sendo,

¹¹ Entrevista disponível em: <https://www.southuniversity.edu/whoware/newsroom/blog/political-campaigns-and-social-media-tweeting-their-way-into-office-106986> <Acesso em: 01/02/2018>

supostamente, o material ideal para o desenvolvimento do objetivo da pesquisa, isto é, a elaboração de um esquema de anotação focado nas expressões irônicas. As Figuras 1 e 2 exemplificam a caixa de comentários e o comentário de um leitor, respectivamente.



Figura 1. Interface da caixa de comentários do jornal *Folha de S. Paulo*

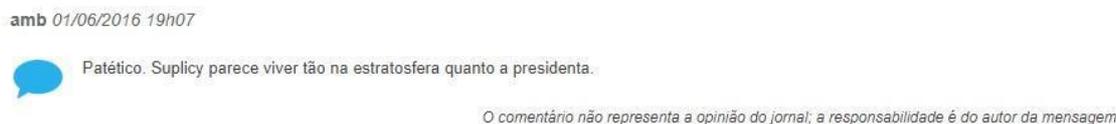


Figura 2. Exemplo de comentário do jornal *Folha de S. Paulo*

O *ComentCorpus* é um *corpus* de comentários de notícias no português do Brasil com anotações semântico-discursivas conforme a intenção de cada sentença identificada e foi criado por extração manual de uma coleção de 90 notícias, retiradas do caderno Poder do jornal *Folha de S. Paulo* no período de janeiro a junho de 2016, relacionadas ao período de impeachment da presidente Dilma Rousseff. A escolha por construir um *corpus* de comentários de notícias se dá ao fato de acreditarmos que esse tipo de texto possui algum grau de subjetividade e opinião do autor e, mesmo apresentando uma estrutura extremamente complexa, acreditamos que por se tratar de textos relativos ao domínio político, é inevitável a presença de sentenças irônicas. O *corpus* é composto por 6.185 comentários, totalizando aproximadamente 14 mil sentenças e 207 mil *tokens*. A Tabela 2 descreve as características do *corpus*.

Notícias	Comentários	Sentenças	Tokens
90	6185	14547	207866

Tabela 2. Características do *ComentCorpus*

A Tabela 3 descreve a descrição dos *subcorpus*¹² relativos aos meses que constituem o *corpus*. Na etapa de compilação atentamos para o seu balanceamento, buscando extrair um número de comentários semelhantes entre os meses de janeiro a julho.

Mês	Tokens	Sentenças	Comentários
Janeiro	26274	1903	798
Fevereiro	24860	1866	798
Março	38033	2649	1136
Abril	47299	3258	1346
Mai	35819	2536	1039
Junho	35581	2527	1068

Tabela 3. Descrição dos meses que compõem o *corpus*

Após a descrição da compilação do *corpus*, a seção seguinte abordará a identificação do comentário, autor, data e sentença, além da anotação de opiniões irônicas e sua intenção – irônica ou não.

¹² Um *subcorpus* possui todas as propriedades de um *corpus*, mas faz parte de um *corpus* maior.

Capítulo 5

Esquema de anotação para expressões irônicas

Este capítulo descreve e propõe um método de anotação para opinião e ironia em *corpus* opinativo. Neste capítulo, descreveremos o processo de anotação do *corpus* descrito. Seguem as seguintes etapas¹³:

1. ETAPA 1: criação de um cabeçalho do comentário que inclui a identificação do comentário, do autor e da data daquele comentário;
2. ETAPA 2: anotação das sentenças opinativas e sua intenção – irônica, não irônica e as que contêm outros dispositivos linguísticos e figurados que interferem na literalidade daquela opinião;
3. ETAPA 3: validação das diretrizes (descritas no Apêndice A) e
4. ETAPA 4: anotação de pistas de ironia em opiniões extraídas a partir da etapa anterior.

Para a realização das etapas acima, contamos com diferentes anotadores em cada uma delas. Na ETAPA 1, um anotador delimitou cada comentário e suas respectivas sentenças. A ETAPA 2 e 3 consiste na anotação da intenção das opiniões. Na ETAPA 2 contamos com dois anotadores, baseados em diretrizes desenvolvidas para este processo de anotação, indicando uma opinião irônica, não irônica ou outro tipo de ironia subjacente ao seu significado. Subsequente, a ETAPA 3 validou duas fases as diretrizes propostas na etapa anterior: na Fase 1 três anotadores marcaram as mesmas classificações de opinião da ETAPA 2 em uma porção geral do *corpus*, adicionando apenas a classificação *não soube responder* para os casos mais difíceis. A Fase 2 seguiu o mesmo modelo da primeira fase, mas aqui contamos com cinco anotadores. Por fim, a ETAPA 4 consiste na identificação e anotação de elementos linguísticos que servem como pistas para a compreensão de um enunciado irônico. Esta etapa foi anotada por um único anotador.

Isto posto, descreveremos nas próximas seções deste capítulo todas as etapas que constituem o processo de anotação realizado nesta pesquisa.

¹³ As quatro etapas que compõem o processo de anotação desta dissertação foram realizadas manualmente pelos anotadores.

5.1. ETAPA 1 – Criação de um cabeçalho para cada comentário

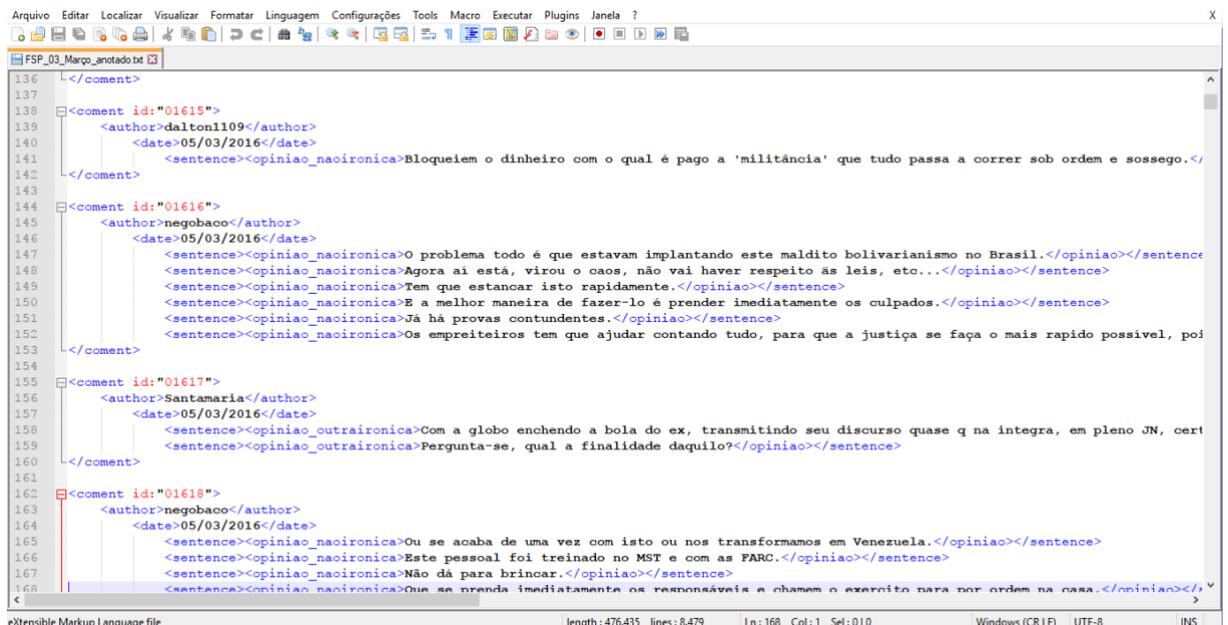
Preliminar à anotação das opiniões, criamos, manualmente, um cabeçalho contendo o número, autor e data de cada comentário analisado. Esta anotação inicial ocorre em duas etapas, ambas anotadas manualmente e por um único anotador, pois entendemos como uma fase menos complexa. A primeira etapa consiste na criação de um cabeçalho e na anotação das opiniões, classificadas em irônicas, não irônicas e as que contém outro tipo de ironia.

Como é possível observar na Figura 3, cada comentário está delimitado com as etiquetas `<coment id:"xxxx">` indicando o início do comentário e `</coment>` indicando o fim de cada comentário do autor. As etiquetas `<author>` e `</author>` marcam o autor do comentário e as etiquetas `<date>` e `</date>` indicam a data em que o comentário foi feito.

```
<coment id:"00018">
  <author>Viva</author>
  <date>02/01/2016</date>
  <sentence>O que o Aécio fez eu não sei, mas a Dilma
    comandou o assalto à Petrobrás.</sentence>
</coment>
```

Figura 3. Exemplo de cabeçalho para cada comentário

Na Figura 4 é possível ter uma visão geral do *corpus* anotado em XML na ferramenta Notepad++.



```
136 </coment>
137
138 <coment id:"01615">
139   <author>dalton1109</author>
140   <date>05/03/2016</date>
141   <sentence><opiniao_naoironica>Bloqueiem o dinheiro com o qual é pago a 'militância' que tudo passa a correr sob ordem e sossego.</opiniao_naoironica></sentence>
142 </coment>
143
144 <coment id:"01616">
145   <author>negobaco</author>
146   <date>05/03/2016</date>
147   <sentence><opiniao_naoironica>O problema todo é que estavam implantando este maldito bolivarianismo no Brasil.</opiniao_naoironica></sentence>
148   <sentence><opiniao_naoironica>Agora ai está, virou o caos, não vai haver respeito às leis, etc...</opiniao_naoironica></sentence>
149   <sentence><opiniao_naoironica>Tem que estancar isto rapidamente.</opiniao_naoironica></sentence>
150   <sentence><opiniao_naoironica>E a melhor maneira de fazer-lo é prender imediatamente os culpados.</opiniao_naoironica></sentence>
151   <sentence><opiniao_naoironica>Já há provas contundentes.</opiniao_naoironica></sentence>
152   <sentence><opiniao_naoironica>Os empreiteiros tem que ajudar contando tudo, para que a justiça se faça o mais rapido possivel, pois
153 </coment>
154
155 <coment id:"01617">
156   <author>Santamaria</author>
157   <date>05/03/2016</date>
158   <sentence><opiniao_outraironica>Com a globo enchendo a bola do ex, transmitindo seu discurso quase q na integra, em pleno JN, cert
159   <sentence><opiniao_outraironica>Pergunta-se, qual a finalidade daquilo?</opiniao_outraironica></sentence>
160 </coment>
161
162 <coment id:"01618">
163   <author>negobaco</author>
164   <date>05/03/2016</date>
165   <sentence><opiniao_naoironica>Ou se acaba de uma vez com isto ou nos transformamos em Venezuela.</opiniao_naoironica></sentence>
166   <sentence><opiniao_naoironica>Este pessoal foi treinado no MST e com as FARC.</opiniao_naoironica></sentence>
167   <sentence><opiniao_naoironica>Não dá para brincar.</opiniao_naoironica></sentence>
168   <sentence><opiniao_naoironica>Que se prenda imediatamente os responsáveis e chameem o exercito para por ordem na casa.</opiniao_naoironica></sentence>
169 </coment>
```

Figura 4. Visualização do *ComentCorpus*

Consideramos sentença a unidade mínima de análise segmentada por um ponto final, exclamação, interrogação ou reticências. Em geral, é necessária a leitura e análise de todo o comentário e, muitas vezes, de comentários anteriores e/ou posteriores para sua identificação:

- (8) A estratégia do governo é dizer que tudo se resume a uma disputa entre Dilma e Cunha: satanizar o Cunha e limpar a Dilma.
- (9) Você diz que o que está acontecendo é golpe, mas fala que a democracia está sendo respeitada ao ter o direito de espremer, digo, judicializar a questão...

No entanto, há sentenças em que a pontuação pode indicar uma hesitação, pausa ou então a reprodução da oralidade na escrita. Sentenças com dois pontos, ponto e vírgula, parênteses e travessões/hífen são considerados como uma única sentença:

- (10) Afinados?... duas pessoas que envergonham a nação!
- (11) É aquela história, quem apoia corrupto, tende a ser também, o Cunha é o pior deles ...

A Figura 5 mostra a anotação das sentenças, delimitada com a etiqueta `<sentence>` indicando o início de uma sentença e `</sentence>` indicando o fim de uma sentença.

```
<coment id:"00018">
  <author>Viva</author>
  <date>02/01/2016</date>
  <sentence>O que o Aécio fez eu não sei, mas a Dilma
    comandou o assalto à Petrobrás.</sentence>
</coment>
```

Figura 5. Exemplo de anotação de sentença

5.2. ETAPA 2 – Identificação e anotação de opiniões

Idealmente, gostaríamos de seguir o esquema “tradicional” apresentado por Hovy e Lavid (2010), no entanto, encontramos vários desafios a algumas características em nosso esquema de anotação e, como resultado, tivemos de fazer algumas adaptações. A princípio, tivemos a dificuldade em encontrar anotadores suficientes para a anotação do *corpus* completo, sendo realizada, inicialmente, por dois anotadores que tiveram de lidar com uma grande quantidade de dados complexos em um tempo limitado.

Como dissemos no Capítulo 3, uma sentença opinativa é aquela que transmite algum conteúdo subjetivo em relação a algo ou alguém e fatos são aquelas sentenças que têm caráter informativo:

- (12) Dilma sofreu impeachment em 2016.

- (13) Lula foi nomeado e destituído no mesmo dia para o cargo de ministro.
- (14) Lula será candidato em 2018.

Os exemplos são claramente uma informação e, dificilmente, poderia ser interpretada como uma opinião. Entretanto, o exemplo é muito semelhante aos anteriores, mas pode ter duas interpretações dependendo do contexto em que é dita. Se um jornalista divulga essa sentença como um título de uma notícia, provavelmente, estamos falando de um fato, mas, se um eleitor do Partido dos Trabalhadores diz que o Lula será candidato em 2018, pode haver o desejo dessa pessoa que Lula se candidate ao cargo de presidente e não, necessariamente, isso seja verdade. Por este motivo e por tratar-se de um *corpus* de comentários, consideramos quase todas as sentenças como opinião, salvo os casos como os exemplos acima.

O passo seguinte após a identificação de uma sentença opinativa foi a anotação da intenção da opinião que pode ser não-irônica, irônica ou conter outro tipo de ironia. Assim, cada sentença identificada como opinião deve ser individualmente indicada pelas etiquetas:

Opinião não irônica: consideramos não irônica as sentenças que não contém mecanismos linguísticos que alternam o seu significado.

- (15) Eu sou a favor da saída da atual Presidente.
- (16) Ela só não disse que convocaria o plebiscito no dia 30 de dezembro de 2018.

A Figura 6 apresenta a anotação de uma opinião não irônica. A etiqueta `<opinio:“naoironica”>` indica o início de uma opinião que não contém mecanismos linguísticos que alternam o seu significado e `</opinio>` indica o fim de uma opinião.

```
<sentence><opinio:“naoironica”>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.</opinio></sentence>
```

Figura 6. Exemplo de cabeçalho para cada comentário

Opinião irônica: o primeiro passo para o modelo foi baseado nas teorias linguísticas, que concordam que esse mecanismo acontece pelo choque entre o significado literal do enunciado e o que se espera do falante, isto é, consideramos ironia apenas a opinião que há a oposição de sentido entre o que é pretendido e o que é realmente dito. Vejamos:

- (17) Como deve ser bom gastar o dinheiro dos contribuintes!!
- (18) Todos os políticos são inocentes!!!!

Na Figura 7 exemplificamos uma opinião irônica, é representada pela etiqueta `<opinio:“ironica”>` indica o início de uma opinião que há inversão no sentido literal da opinião `</opinio>` indica o fim de uma opinião.

```
<sentence><opinio:“ironica”>A Sra. Presidente e tao honesta que vai pro  
ceu..tadinha.. </opinio></sentence>
```

Figura 7. Exemplo de anotação de opinião irônica

Outro tipo de ironia: consideramos outro tipo de ironia a sentença que **NÃO** há oposição entre o sentido literal e significado pretendido, no entanto, o texto ainda carrega algum dispositivo figurado que altera sua literalidade:

- (19) Quanto estão levando pra censurar os comentários contra esse governinho?
- (20) Dimentira não tem um pingão de sentimento e espontaneidade.

No exemplo (19), a sentença pode ser considerada uma espécie de ironia, mas que tem um tom mais agressivo para se referir ao governo do Brasil - que pode ser indicado pelo trecho “esse governinho”. Isso acontece porque o uso de diminutivo, por exemplo, apesar de expressar, em sua maioria, casos de afetividade também pode indicar um sentido de inferioridade ou descaso. Já no exemplo (20), o item lexical “Dimentira” cria uma alcunha com um tom jocoso a ex-presidente do Brasil.

Os casos de outro tipo de ironia podem incluir também o sarcasmo, a sátira e o humor, por exemplo:

Sarcasmo: ironia empregada com o objetivo de atacar ou zombar alguém. Attardo (2000) define sarcasmo como um tipo de ironia abertamente agressiva:

- (21) Já já começará aquela conversinha de que ele é “pelseguido pulíticu” e blá blá blá.

Sátira: é uma maneira de se divertir com as pessoas usando uma linguagem tola ou exagerada. Singh (2012) afirma que a sátira é uma figura de linguagem que enfatiza a fraqueza e não a pessoa fraca. Para o autor, a sátira é usada para expor e criticar algo em um indivíduo ou a própria sociedade através do humor, da ironia, do exagero ou do ridículo:

- (22) O que será que Aécinho Zona Sul fez na noite anterior pra acordar um dia e decidir dar apoio a um golpe perpetrado por uma dupla destas?

Humor: assumimos aqui a posição de Singh (2012) para quem o humor consiste na capacidade de expressão de uma sensação de algo divertido e inteligente. Refere-se, principalmente, no reconhecimento de incongruências ou peculiaridades presentes em uma situação:

(23) O Suplicy não tem amor próprio, parece aqueles balões de plástico infláveis sem vida

(24) Nine Fingers na frente?

Um exemplo de uma sentença com outros tipos de ironia é apresentado no Figura 8 Assim, a etiqueta `<opinio:“outraironica”>` indica o início de uma opinião que não há inversão no sentido literal, no entanto, o texto ainda tem a compreensão literal corrompida.

```
<sentence><opinio:“outraironica”> Já já começará aquela conversinha de ele é "pelsseguido pulíticu" e blá blá blá. </opinio></sentence>
```

Figura 8. Exemplo de anotação com outro tipo de ironia

Com base nos dados anotados descritos nesta seção, a Tabela 4 apresenta a distribuição da anotação por cada anotador, sendo irônica (I), não irônica (NI) e outro tipo de ironia (OI). Vale salientar, que os dados anotados somam a mesma quantidade para cada anotador.

Mês	Anotador 1			Anotador 2			Opiniões
	NI	I	OI	NI	I	OI	
Janeiro	1296	361	246	1321	397	185	1903
Fevereiro	1269	290	307	1310	268	288	1866
Março	1921	508	311	1897	516	327	2740
Abril	1960	702	299	1966	714	281	2961
Mai	1973	362	321	1969	349	338	2656

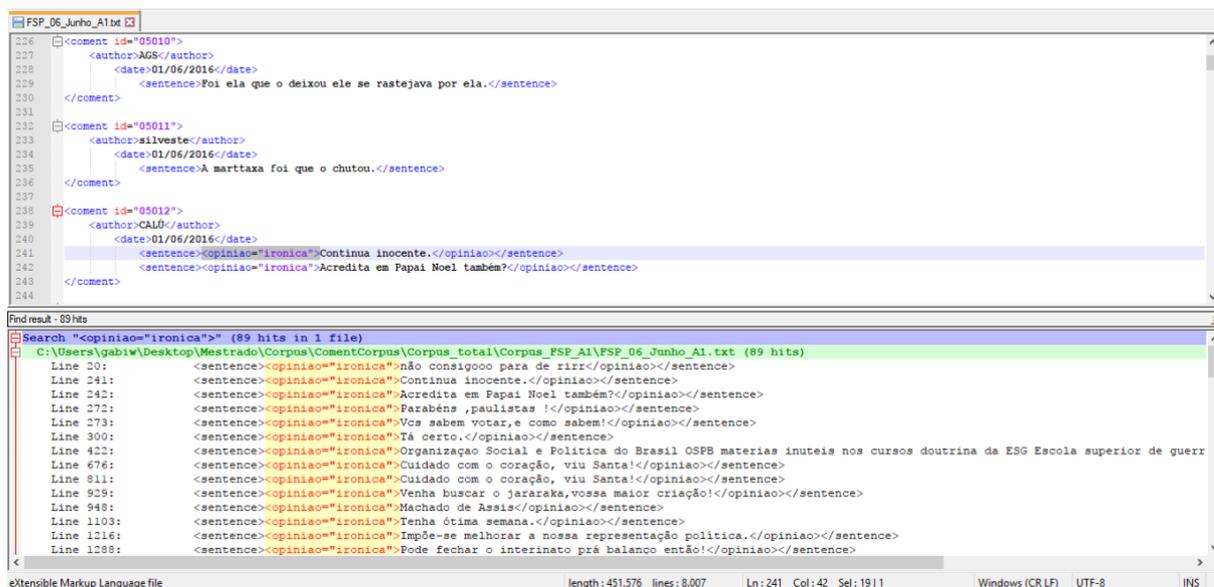
Tabela 4. Dados dos anotadores na ETAPA 2

5.2.1. Notepad++

Neste processo de anotação do *corpus* utilizamos a ferramenta Notepad++, ilustrado através da Figura 4 pode ser usado como bloco de notas, permite editar códigos em diversas linguagens, tais como C, C++, HTML, ASP, Python, JAVA, Pascal, XML, dentre mais de outras quarenta linguagens disponíveis. O Notepad++ também permite a criação de macros.

Basicamente, a macro é uma sequência de ações que você pode ser gravada, salva e reproduzida por meios de códigos criado pelo usuário.

Além disso, a ferramenta ainda permite buscarmos por palavras ou mesmo uma expressão no *corpus*. Assim, para a extração e frequência das ocorrências de opiniões irônicas utilizamos o Notepad++, conforme a Figura 9.



```
FSP_06_Junho_A1.txt [3]
226 <comment id="05010">
227   <author>AGS</author>
228   <date>01/06/2016</date>
229   <sentence>Foi ela que o deixou ele se rastejava por ela.</sentence>
230 </comment>
231
232 <comment id="05011">
233   <author>silveste</author>
234   <date>01/06/2016</date>
235   <sentence>À maritaxa foi que o chutou.</sentence>
236 </comment>
237
238 <comment id="05012">
239   <author>CALB</author>
240   <date>01/06/2016</date>
241   <sentence><opinioao="ironica">Continua inocente.</opinioao></sentence>
242   <sentence><opinioao="ironica">Acredita em Papai Noel também?</opinioao></sentence>
243 </comment>
244

Find result - 89 hits
Search "<opinioao='ironica'" (89 hits in 1 file)
C:\Users\gabin\Desktop\Mestrado\Corpus\ComentCorpus\Corpus_total\Corpus_FSP_A1\FSP_06_Junho_A1.txt (89 hits)
Line 20: <sentence><opinioao="ironica">não consigooo para de rir</opinioao></sentence>
Line 241: <sentence><opinioao="ironica">Continua inocente.</opinioao></sentence>
Line 242: <sentence><opinioao="ironica">Acredita em Papai Noel também?</opinioao></sentence>
Line 272: <sentence><opinioao="ironica">Parabéns , paulistas !</opinioao></sentence>
Line 273: <sentence><opinioao="ironica">Vcs sabem votar,e como sabem!</opinioao></sentence>
Line 300: <sentence><opinioao="ironica">TÁ certo.</opinioao></sentence>
Line 421: <sentence><opinioao="ironica">Organização Social e Política do Brasil OSP materias inuteis nos cursos doutrina da ESG Escola superior de guerr
Line 676: <sentence><opinioao="ironica">Cuidado com o coração, viu Santa!</opinioao></sentence>
Line 811: <sentence><opinioao="ironica">Cuidado com o coração, viu Santa!</opinioao></sentence>
Line 929: <sentence><opinioao="ironica">Venha buscar o jararaka,vossa maior criação!</opinioao></sentence>
Line 948: <sentence><opinioao="ironica">Machado de Assis</opinioao></sentence>
Line 1103: <sentence><opinioao="ironica">Tenha ótima semana.</opinioao></sentence>
Line 1216: <sentence><opinioao="ironica">Impõe-se melhorar a nossa representação política.</opinioao></sentence>
Line 1288: <sentence><opinioao="ironica">Pode fechar o interinato prá balanço então!</opinioao></sentence>

eXtensible Markup Language file length: 451.576 lines: 8.007 Ln: 241 Col: 42 Sel: 19 | 1 Windows (CRLF) UTF-8 INS
```

Figura 9. Busca pela etiqueta <opinioão:“ironica”> no Notepad++

5.3.ETAPA 3 – Validação das diretrizes

Nesta dissertação, nem todas as etapas de uma anotação foram exploradas. Primeiramente, em um trabalho “ideal” de anotação, algumas etapas deveriam ser realizadas, contudo, devido à dificuldade em encontrar anotadores disponíveis e a falta de tempo para calcular a eficácia do modelo proposto entre as anotações, realizamos em uma pequena porção do *corpus* uma metodologia de anotação prototípica. Nesta perspectiva, na ETAPA 3 anotamos uma porção do *corpus* total e visando um melhor resultado na anotação, criamos um método para validar a anotação manual de sentenças irônicas realizada. Tal método de anotação facilita o processo e reduz o tempo de avaliação do anotador. Para este fim, criamos um quadro compartilhado semelhante a um questionário de perguntas fechadas no qual os anotadores deveriam verificar e marcar se uma determinada sentença é:

- a. Irônico;
- b. Não irônico;

- c. Outro tipo de ironia e
- d. Não soube responder.

Como já dissemos, a ETAPA 3 divide-se em duas fases de anotação: a primeira busca treinar os anotadores para a tarefa de anotação. Já a segunda consiste no processo de anotação após discussões e resultados da fase anterior. A princípio, foram extraídas 1904 sentenças a serem classificadas por três anotadores que deveriam classificá-las de acordo com as categorias relacionadas acima. Procuramos nesta etapa não dividir a quantidade de sentenças por anotadores, mas sim a quantidade de *tokens* para cada anotador, permitindo um balanceamento real entre dados anotados, uma vez que algumas sentenças eram curtas e outras extremamente longas, como mostra os exemplos (25) e (26), respectivamente. Assim, em um total de 2482, foram selecionados 5882 *tokens* (extraídos comentários de cinco notícias) para serem anotados igualmente entre os anotadores.

(25) Parabéns!

(26) Mas essa é a verdade esquerda , direita , centro , monarquia , tanto faz o ser humano tem a fome do poder e a fome do dinheiro esse grupo que está no poder a 15 anos , de fato se lambuzou projeto de poder mas principalmente falta de administração só quiseram dinheiro e fama deitaram no berço esplêndido e contaram com a Petrobras para resolver todos os problemas do Brasil qual a grande obra que conseguiram terminar nesses 15 anos ??????

A Tabela 5 expõe os dados anotados na primeira fase da ETAPA 3, nas quais os comentários das notícias selecionadas são representados por N1, N2, N3 e N4 e os anotadores por A1, A2 e A3.

	Irônico				Outro tipo de ironia				Não irônica				Não soube responder			
	N1	N2	N3	N4	N1	N2	N3	N4	N1	N2	N3	N4	N1	N2	N3	N4
A1	9	10	4	39	16	15	37	37	24	16	50	124	0	0	5	16
A2	3	5	3	22	5	9	25	26	41	27	67	167	0	0	1	1
A3	2	4	1	14	18	18	51	67	24	14	43	134	5	5	1	1

Tabela 5. Anotação ETAPA 3: Distribuição de anotação – Três anotadores

Na segunda fase da anotação, 24105 *tokens* (totalizando 1866 sentenças) foram classificadas entre cinco anotadores. Desse conjunto de dados, reservamos 669 *tokens* para serem anotados pelos cinco anotadores, conforme mostra a Tabela 5.3.

	Irônico	Não irônico	Outro tipo de ironia	Não soube responder
Anotador 1	2	43	11	0
Anotador 2	1	25	26	4
Anotador 3	3	35	17	0
Anotador 4	1	46	9	0
Anotador 5	4	35	13	4

Tabela 6. Anotação ETAPA 3: Distribuição de anotação – Cinco Anotadores

5.4.ETAPA 4 – Anotação de ironia em sentenças opinativa

Na ETAPA 4 demos os primeiros passos para uma análise mais profunda do funcionamento da ironia no *corpus*. Apesar das teorias linguísticas serem diversas sobre como definir a ironia, todas elas concordam que isso envolve um conflito entre o significado literal de um enunciado e a intenção do falante. Consideremos os seguintes exemplos:

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

(28) Nossa como ela é legal!

No exemplo (27) infere-se um sentido irônico através do contraste entre “são muito nobres” *versus* “usou dinheiro de corrupção”, ou seja, o autor da sentença utiliza a oposição entre uma sequência positiva e uma negativa para alcançar a ironia, i.e., utiliza-se dispositivos linguísticos que tornam-se pistas para a compreensão não literal da sentença. Nesse caso não é necessário recorrer ao contexto ou às informações extralinguísticas. Já no exemplo (28), quando o autor diz “Nossa como ela é legal!” podemos entender literalmente, mas também é possível que esse enunciado seja irônico. Sua intenção, figurada ou não, está exclusivamente no contexto. Nesse caso, Karoui et al. (2015) mostram que existem duas maneiras de compreender um enunciado irônico:

1. Confiar exclusivamente nas pistas lexicais internas para o enunciado,
2. Combinar essas pistas com um contexto pragmático externo ao enunciado.

Nas seções seguintes, focaremos em explicar as categorias que compõem o esquema de anotação das opiniões irônicas extraídas do *ComentCorpus*.

5.4.1. Expressões irônicas: oposições

Apesar de anotarmos sentenças não irônicas e as que contêm outro tipo de ironia, consideramos nesta etapa de anotação apenas as classificadas como irônicas na ETAPA 3, visto que o foco desta dissertação é compreender e modelar opiniões irônicas. Desse modo, conforme indicado no Capítulo 2, a ironia é frequentemente entendida através da oposição e contradição de elementos, podendo ser oposição entre elementos lexicais da própria sentença, contextual. Por isso, baseamos distinções abaixo no Quadro 9:

OPINIÃO IRÔNICA	
OPOSIÇÃO	
EXPLÍCITA	IMPLÍCITA
A ironia é compreendida através da oposição de pistas presentes no enunciado.	A ironia é compreendida através de pistas presentes e por um contexto pragmático adicional ao enunciado.

Quadro 4. Oposições irônicas – implícitas e explícitas

Embora as opiniões irônicas compostas por uma oposição explícita envolvam um contraste entre elementos que têm polaridades opostas, como o em (29) e (30), elas também existem os casos de contraposição de elementos que não se espera que apareçam juntos, pois, não pertencem a mesma esfera semântica, como no exemplo (31).

(29) As "conquistas" do último governo são as inflações e o desemprego.

(30) Também acho que o Eduardo Cunha seria um homem mais realizado tendo um posto bacana, por exemplo, no quadro do STF. rs

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

No exemplo (29) “*conquistas*”, uma proposição positiva, se opõe a “*as inflações e o desemprego*”, uma proposição negativa. Em (30) “Eduardo Cunha seria um homem mais realizado” e “um posto bacana” são positivas e “no quadro do STF” é negativo”. Já no exemplo (31) além da oposição de polaridade entre as proposições “*muito boa para governar o Brasil*” e “*faliu uma lojinha de 1,99*”, há também um contraste entre o domínio político na qual “*muito boa para governar o Brasil*” se encaixa e “*faliu uma lojinha de 1,99*” que foge desse escopo.

A Figura 10 apresenta a anotação para opiniões irônicas que são construídas a partir de uma contradição explícita. Utilizamos a etiqueta <oposicao:“**explicita**”> para indicar um o tipo de oposição de ironia de uma opinião e </oposicao> para indicar o fim da anotação.

```
<sentence><opinio:"ironica"><oposicao: "explicita">Viva a corrupção, o país de babacas onde vivemos.</oposicao></opinio></sentence>
```

Figura 10. Anotação de oposição explícita

Por outro lado, uma opinião irônica contendo uma contradição implícita descreve um evento ou estado e um contexto pragmático externo ao enunciado. A ironia ocorre porque o autor acredita que o leitor pode detectar o conteúdo irônico com base em conhecimentos de mundo comuns compartilhados com o escritor, como são os casos dos exemplos a seguir:

(32) É incrível o nosso país.

(33) O atual governo parece funcionar muito bem...

Nos exemplos acima é necessário considerar todo o cenário político no qual as notícias e os comentários referentes para inferir o sentido irônico. O Quadro 11 representa a anotação para os casos de ironia implícita, indicada pela etiqueta **<oposicao:"implicita">** e **</oposicao>**.

```
<sentence><opinio:"ironica"><oposicao:"implicita">Agora são trilionários ganharam na mega sena....</oposicao></opinio></sentence>
```

Figura 11. Anotação de oposição implícita

Em suma, existem casos em que a ironia é inferida de várias maneiras dentro de uma sentença. Pode ocorrer que em uma camada superficial, a ironia é explicitamente ativada a partir de palavras lexicalizadas, esses casos são considerados como ironia explícita. Em outro caso, há uma camada “oculta” e mais profunda que pode ser compreendida apenas através de contexto adicional, seja no próprio texto anterior ou posterior ao enunciado irônico ou mesmo um conhecimento pragmático adicional. Neste caso, a ironia deve ser anotada como ironia implícita.

5.4.2. Pistas: Subcategorias linguísticas

As ironias explícitas e implícitas, segundo o esquema descrito na seção anterior, podem ser expressas e alcançadas através de dispositivos e recursos linguísticos e retóricos agrupados em diferentes categorias. Estas categorias não têm seu uso restrito e podem coocorrer na mesma sentença. Nesta seção descreveremos as subcategorias mais relevantes encontradas durante a análise das sentenças identificadas como irônicas. São elas:

1. ENTIDADES NOMEADAS (NE)

Algumas ironias podem ser direcionadas a uma ou mais entidades nomeadas (NE). Quando há uma NE, geralmente, ela é o alvo da ironia e é por esta razão a escolha por indicar sua presença na sentença. Uma NE é marcada pela etiqueta `<ne>` e `</ne>`.

```
<oposicao:"implicita">O <ne>PT</ne> na oposição é nota 100.</oposicao>
```

Figura 12. Anotação de entidades nomeadas (NE)

- (34) Tirar a Dilma e colocar o Cunha ou Temer no lugar, ai você fica feliz né!
- (35) Dinheiro de corrupção para financiar campanha de Dilma não é grave.

2. QUESTÃO RETÓRICA

Questão retórica é um dispositivo linguístico que se utiliza de uma questão para trazer uma opinião e não para obter uma resposta. Em geral, toda questão retórica é considerada por completo como uma opinião irônica, isto é, não existe um elemento linguístico que marque a presença da ironia a não ser ela mesma. Para indicar a presença de uma questão retórica utilizamos a etiqueta `<questao_retorica>` e `</questao_retorica>`.

```
<oposicao:"implicita"><questao_retorica>Acho q eles tão preocupados, milagre né?</questao_retorica></oposicao>
```

Figura 13. Anotação de questão retórica

- (36) Aumentar impostos é fácil, né?
- (37) Devemos ser grato ao P.T e seus eleitores, não é verdade?

3. PONTUAÇÃO

O uso de pontuação como marca de ironia é presença de uma sequência composta de mais de um ponto de exclamação, interrogação ou sequência de vários pontos finais em uma frase também pode ser usada como uma pista para detecção de ironia. Segundo Carvalho et al. (2009) em UGC, a pontuação é usada com o objetivo de verbalizar as emoções e sentimentos do usuário e sinalizar intencionalmente um texto irônico. As etiquetas `<pont>` e `</pont>` foram utilizadas para marcar o uso de uso de pontuação excessiva.

```
<oposicao:"implicita">Longa vida aos iconoclastas  
<pont>!!!!</pont></oposicao>
```

Figura 14. Anotação de pontuação

- (38) E o da campanha do Aécio é limpinho, limpinho, cheirosinho!!!!!!
- (39) Lula, o democrata!!!

4. DIMINUTIVO

O diminutivo é utilizado, muitas vezes, com o propósito de expressar sentimentos positivos, como afeto ou intimidade. No entanto, também podem ser usados para transmitir um sentido irônico ou sarcástico ironicamente. Os casos de diminutivo são delimitados pelas etiquetas **<dim>** e **</dim>**.

```
<oposicao:"implicita">Estão preocupados com o <dim>coitadinho</dim> do  
Cunha.</oposicao>
```

Figura 15. Anotação dos diminutivos

(40) Que perguntinha mais fácil de responder...

(41) Por favor deixem o santinho, o comprador da reeleição, em paz!

No processo de anotação foi observado o uso de “coxinha”, que é usado, frequentemente, de modo pejorativo para referir-se a uma pessoa conservadora e geralmente da classe média. Entretanto, apesar de exprimir um caráter irônico à sentença, considerou-se o termo como um falso positivo, uma vez que não é o sufixo *-inha* que carrega essa ironia, mas sim todo um contexto atual que o envolve.

(42) Golpistas e coxinhas querem que o povo se exploda.

(43) Se for pro pau, os coxinhas fogem.

5. ASPAS

Trabalhos para o português (FREITAS et al., 2014; CARVALHO et al., 2009) entendem as aspas como um possível indicador de ironia. São frequentemente usadas para enfatizar um conteúdo irônico, especialmente, se a sentença tiver uma polaridade e o elemento lexical, por exemplo, um adjetivo positivo que qualifica uma entidade. As aspas são indicadas pelas etiquetas **<aspas>** e **</aspas>**.

```
<oposicao:"explicita">O PMDB é um partido tão <aspas>"pequeno"</aspas> que  
não consegue ser medíocre!</oposicao>
```

Figura 16. Anotação das aspas

(44) Com um "conselheiro" desse, o que se pode esperar.....

(45) Poxa vida, logo agora que o nosso Pravda (FSP) tinha acionado a máquina de propaganda para projetar a imagem do "amigo" J. Wagner.

6. ADJETIVO

Dentro do domínio político alguns adjetivos podem enfatizar uma opinião, principalmente, as irônicas. Muitas vezes estes adjetivos aparecem em oposição a outro de polaridade contrastiva ou mesmo uma sentença inteira. Estes adjetivos são marcados pelas etiquetas `<adj>` e `</adj>`.

```
<oposicao:"explicita">Junto com aquele mais <adj>honesto</adj> do país, que tinham até que canoniza-lo de tão <adj>santo</adj> que é!</oposicao>
```

Figura 17. Anotação de adjetivos

- (46) São muito *nobres*, afinal, a chapa usou dinheiro de corrupção.
- (47) A frase do Sr. Rui Falcão é *fantástica*.
- (48) Nossa, salve o *bastião* do impeachment.

7. DISPARADORES

Alguns elementos lexicais são essenciais para uma sentença e todo o sentido da ironia discorre ao seu redor. Os *disparadores* podem ser definidos como palavras que, geralmente, possuem uma polaridade bem marcada, na maioria dos casos positivas, que disparam a oposição que provoca a ironia. No nosso caso, palavras como *corrupção*, *honesto*, *impeachment*, *imposto*, dentre outras, se opõem a outras palavras ou outro contexto que não está no enunciado, gerando, um enunciado irônico. Utilizamos as etiquetas `<disparador>` e `</disparador>` para delimitar essas ocorrências.

```
<oposicao:"explicita">Dinheiro de <disparador>corrupção</disparador> para financiar <disparador>campanha</disparador> de Dilma ã é grave.</oposicao>
```

Figura 18. Anotação das disparadores

- (49) Aumentar *impostos* é fácil né?
- (50) Aqui pode tudo, só não pode ser *honesto*!
- (51) Todos os meus *subordinados* são *corruptos*, eu não.

8. HIPÉRBOLE

A hipérbole pode ser definida como um exagero deliberado ou exagero de uma característica essencial para um determinado objeto ou fenômeno capazes de gerar uma intenção de exagero, por exemplo, através do uso de superlativos. Os casos de hipérbole são indicados pelas etiquetas `<hiperbole>` e `</hiperbole>`.

```
<oposicao:"explicita"><hiperbole>Belíssimo</hiperbole> modelo de ética e moral dos vermelhos chupins.</oposicao>
```

Figura 19. Anotação de hipérbole

- (52) Ele mudou de nome: ENGANO TODO MUNDO.
- (53) Sob seu comando até o ano de 2050, todos acordos estarão concluídos.

9. INTERJEIÇÃO

Algumas interjeições, principalmente as positivas, podem ser usadas como potenciais pistas para a compreensão de um enunciado irônico. As interjeições são determinadas pelas etiquetas `<itj>` e `</itj>`.

```
<oposicao:"explicita"><itj>Nossa</itj> como ela é legal!</oposicao>
```

Figura 20. Anotação de interjeição

- (54) Claro, deve ser da sua turma e aí podem breçar a Lava-Jato.
- (55) Poxa gente, vamos para de vazar dados sobre os phetistas-pixulequeiros, afinal eles 'lutaram muito' pra conseguir essas bocadas!

10. INTENSIFICADOR/MODIFICADOR

Os intensificadores e modificadores têm a função de reforçar um determinado elemento lexical, uma entidade nomeada ou uma expressão na sentença. As etiquetas `<mod>` e `</mod>` indicam as ocorrências de intensificadores/modificadores na sentença.

```
<oposicao:"implicita">Poxa, a filha dela e <mod>muito</mod> pobre , como filho Lula...</oposicao>
```

Figura 21. Anotação de intensificador/modificador

- (56) Realmente a presidenta é muito boa...tiro o chapéu para ela...simplesmente quebrou o Brasil.
- (57) Tão sábio que prega para o Brasil um governo socialista, mas prefere viver nos USA para usufruir das benesses e oportunidades de um regime capitalista, que prima por valorizar o mérito.
- (58) A Venezuela é absolutamente democrática.

11. REPETIÇÕES

Repetições de uma letra ou palavra pode ter diversos significados dentro de uma sentença, entre eles o de reforçar uma opinião ou mesmo negá-la. As etiquetas **<repet>** e **</repet>** apontam os casos de repetições no texto.

```
<oposicao:"explicita">Antes da Polícia Federal mandar fechar as torneiras o dinheiro vazava pelo ladrão e ninguém viu nada, ninguém sabia de nada, quanta coincidência, foi só fechar as torneiras que aumentaram o fundo partidário, quanta "coincidência", <repet>sei... sei..</repet></oposicao>
```

Figura 22. Anotação de repetições

(59) Eh eh eh... Você é hilário.

(60) Noooooosa, grande vantagem para o cidadão ver a presidenta cair e o vice assumir.

12. RISOS

Comumente em UGC é possível observar que na ironia as expressões de riso são fortes indicadores de opiniões irônicas. As expressões de risos são representadas pelas etiquetas **<riso>** e **</riso>**.

```
<contradicao:"explicita">Acho que eles desconfiam que as empreiteiras estão quebradas. .... <riso>rsrsrsrsrs</riso></contradicao>
```

Figura 23. Anotação de risos

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

(62) Já foi querida hahahaha....

Descrito o processo de anotação das opiniões irônicas, bem como as subcategorias que são consideradas pistas de um enunciado irônico, analisaremos no próximo capítulo os dados extraídos deste processo.

Capítulo 6

Análise dos Dados

Neste capítulo discutimos detalhadamente as experiências e resultados obtidos nesta pesquisa. Embora a identificação da ironia em mídias sociais seja um processo complexo, acreditamos que nosso trabalho possa contribuir para uma melhor compreensão da expressão linguística de ironia para o português do Brasil. Sobretudo, este estudo revela alguns dispositivos linguísticos utilizados na produção de efeito irônico em UGC.

6.1. Anotação manual do *corpus*

Para alcançar os objetivos desta dissertação, propusemos uma anotação manual de *corpus* que apresentou alguns desafios, uma vez que a classificação da intenção da sentença depende da avaliação pessoal de cada anotador. Isso quer dizer que em um processo de anotação de um *corpus* constituído por sentenças irônicas acaba sendo necessário recorrer a intuição do anotador, e apesar da orientação pelas diretrizes, é indiscutível em que ele tenda a colocar seus conhecimentos e experiências na interpretação daquele enunciado, uma vez que expressões figuradas, em particular as irônicas, são conteúdo altamente subjetivos para a compreensão humana. Apesar de dirigirmos nossa análise apenas para as sentenças classificadas como irônicas, alguns casos de outro tipo de ironia tornavam-se confusos e ambíguos para a interpretação de quem estava anotando, uma vez que a distinção de dispositivos figurados não é uma tarefa fácil de ser compreendida e muito menos distinguida.

Embora o *corpus* construído nesta pesquisa tenha sido anotado preliminarmente por dois anotadores na ETAPA 2, consideramos para análise apenas o resultado das anotações realizadas na ETAPA 3, em um primeiro momento por três anotadores e subsequentemente por cinco anotadores. A anotação desta etapa resultou em 198 sentenças irônicas classificadas manualmente, seguindo as diretrizes apresentadas no Apêndice A. Em um total, três anotadores classificaram igualmente 56 sentenças, somando 669 *tokens*, entre irônica, outro tipo de ironia, não irônica e não soube responder. A outra fase de anotação desta etapa consiste em cinco anotadores classificando 405 sentenças, totalizando 5142 *tokens*. A Figura 6 nos mostra a distribuição da classificação por cada anotador. Em ambas as fases de anotação na ETAPA 3 é

possível observar que os casos mais frequentes são os de sentenças não irônica e a variação de anotação como “irônica” é mais consistente na segunda fase.

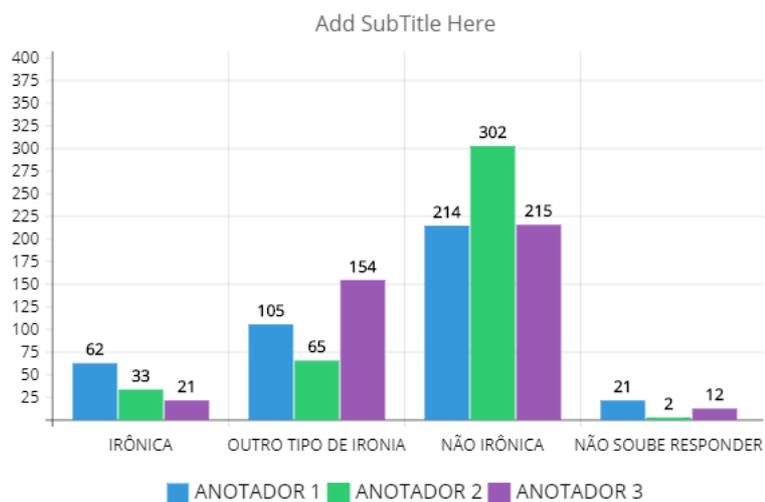


Figura 24. Distribuição das anotações pelos três anotadores na ETAPA 3

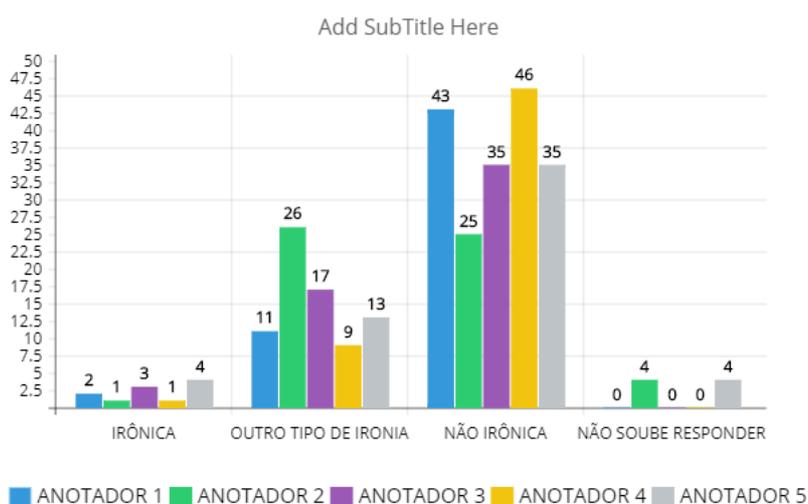


Figura 25. Distribuição das anotações pelos cinco anotadores na ETAPA 3

6.2. Validação da anotação do *ComentCorpus*

Ao fornecer dados para anotação, é necessário avaliar sobre o quão confiável é a anotação e para isso é utilizado um acordo para medir essa confiabilidade do *corpus* anotado: o concordância inter-anotador (*inter-annotator agreement* ou IAA). Em outras palavras, o IAA é uma medida que verifica quanto dois (ou mais) anotadores podem fazer a mesma decisão de anotação para um determinado grupo. Artstein (2017) reconhece que esta é uma prática para comparar a anotação de uma ou mais pessoas em uma única fonte. Segundo o autor, o acordo

pode ser aplicável para validar esquemas e diretrizes de anotação, identificar ambiguidades ou dificuldades na fonte ou avaliar o alcance de interpretações úteis.

A análise realizada nesta seção é baseada no Capítulo 6, *Annotatation and Adjudication*, de Putejovsky e Stubbs (2013). Neste capítulo, os autores descrevem todo o processo de *adjudication* de uma anotação e apresentam e explicam como é feito o cálculo que mede a acurácia da anotação. Deste modo para obter um resultado que compara as anotações, foram desenvolvidas diversas métricas para extrair o acordo inter-anotador, sendo os mais utilizados o kappa de Cohen e de Fleiss e o alpha de Krippendorff. Diferentemente do *kappa* de Cohen, que avalia o acordo de não mais que dois anotadores, o *kappa* de Fleiss é uma medida estatística que avalia o nível de concordância ou reprodutibilidade entre dois ou mais anotadores ao atribuir classificações a um conjunto de dados.

A partir dos dados apresentados no Capítulo 5 (referentes às Tabelas 5 e 6), procuramos nesta seção verificar a confiabilidade das diretrizes desenvolvidas de anotação (disponível no Anexo A) através do *kappa* de Fleiss. A equação de base para o cálculo do coeficiente *kappa* é:

$$k = \frac{P - Pe}{1 - Pe}$$

Na equação acima, *P* é o acordo real e *Pe* é o acordo esperado. Além disso, a tabela usada para representar valores do anotador para o *kappa* de Fleiss possui um eixo para as categorias possíveis que um anotador pode atribuir seus valores e outro eixo para cada um dos anotadores. A Tabela 9 exemplifica o que acabamos de dizer, isto é, as categorias estão no topo, os anotadores estão ao lado e o conteúdo de cada célula representa quantas vezes o anotador atribuiu essa etiqueta ao documento.

	Irônico	Não irônico	Outro tipo	Não soube
Anotador 1	5	34	21	2
Anotador 2	6	42	11	3
Anotador 3	1	27	30	4
Anotador 4	11	34	15	2
TOTAL	23	137	77	11

Tabela 7. Categorias e anotadores para cálculo do coeficiente *kappa*

Para medir o *kappa*, primeiramente, devemos calcular quantas atribuições foram dadas, proporcionalmente, para cada categoria, ou seja, a soma dos valores em sua coluna, dividido

pelo número de anotadores vezes o número de anotações por cada anotador. Esse cálculo é representado pela seguinte equação:

$$P_c = \frac{1}{Aa} \sum_{i=1}^A a_{ic}$$

Onde,

c representa a categoria avaliada,

A é o número de anotadores,

a é o número de anotação por anotadores e

i representa o atual anotador.

Com base nesse cálculo, obtivemos os seguintes resultados, sendo que os TESTES de 1 a 4 são relacionados as anotações com três anotadores e o TESTE 5 na anotação com cinco anotadores:

	Irônica	Não irônico	Outro tipo de ironia	Não soube responder
TESTE 1	.0952	.6054	.2653	.034
TESTE 2	.1544	.4634	.3414	.0406
TESTE 3	.0277	.5555	.3923	.0243
TESTE 4	.1157	.6558	.2006	.0277
TESTE 5	.0392	.6571	.2714	.0285

Tabela 8. Resultados para P_c da anotação

Em seguida, calculamos a P_i , simbolizada pela equação abaixo, que representa o acordo de cada anotador com os outros anotadores em comparação com todos os valores de concordância possíveis. Isso significa que para cada linha, estamos resumindo ao quadrado dos valores em cada coluna e modelando a saída pelo número de anotações totais por cada anotador. O resultado é apresentado na Tabela 11 para as FASES 1 e 2 da anotação.

$$P_i = \frac{(\sum_{c=1}^k a_{ic}^2) - (a)}{a(a-1)}$$

FASE 1				
Anotador 1	.3673	.3292	.4182	.3941
Anotador 2	.7083	.4841	.5618	.6208
Anotador 3	.3741	.317	.4776	.4829
FASE 2				
Anotador 1	.6227			
Anotador 2	.4097			
Anotador 3	.4762			
Anotador 4	.6954			
Anotador 5	.4448			

Tabela 9. Resultado do P_i da anotação

Seguindo, P na equação original de Fleiss é a média dos valores de P_i . Então, calculamos isso somado a coluna P_i dividido pelo número de anotadores. Calculado o P da equação, o próximo passo é determinar Pe somando os quadrados dos valores de P_c . O resultado de P e de Pe é exibido na Tabela 12.

FASE 1		
	P	Pe
Anotação 1	.4832	.4469
Anotação 2	.3767	.3566
Anotação 3	.4858	.4635
Anotação 4	.4992	.4842
FASE 2		
Anotação 5	.5297	.5297

Tabela 10. Resultado para P e Pe da anotação

Finalmente, com base nos resultados, medimos o $kappa$ da anotação proposta e analisada nesta dissertação. A pontuação é descrita na Tabela 13.

FASE 1	
Anotação 1	.0656
Anotação 2	.0312
Anotação 3	.0415
Anotação 4	.0291
FASE 2	
Anotação 5	.0353

Tabela 11. Pontuação $kappa$ para a tarefa de anotação da ETAPA 3

Teoricamente, um escore moderado fica entre 0.41 e 0.60 e um escore perfeito entre 0.80 e 1.0. Contudo, para Pustejovsky e Stubbs (2013), a interpretação do cálculo *kappa* depende muito da complexidade e da objetividade do conteúdo que está sendo anotado. Os autores dão como exemplo em tarefas que envolvem *Part-of-Speech* (PoS) espera-se que a anotação atinja uma pontuação *k* próxima a 1.0 devido ao modo como são definidas as classificações, termos e teorias linguísticas. Em contrapartida, em tarefas de anotação que exigem uma interpretação maior do anotador, como é o caso da ironia, geralmente, não se atinge um padrão alto de anotação. De acordo com essa colocação, a complexidade das sentenças analisadas pelos anotadores nesta pesquisa, mostra que o resultado não se distanciou do que esperávamos obter. Devemos considerar que a maioria das opiniões que compõe o *corpus* está estritamente ligada ao contexto da notícia e ao próprio conhecimento pragmático do anotador, tornando a tarefa demasiadamente subjetiva.

Vale destacar que na maioria das vezes, cada anotador classificou como irônicas as diferentes sentenças, o que nos faz considerar que existem ironias mais ‘prototípicas’ e identificáveis que outras em UGC tão multifacetado como os comentários de notícias *on-line*. O Quadro 5 destaca as ocorrências destes casos.

IRONIAS ‘PROTOTÍPICAS’
Sim ela parece funcionar muito bem...
Quase ninguém passa muita fome ou necessidade lá...
A Sra. Presidente e tao honesta que vai pro ceu..tadinha..
poxa a filha dela e muito pobre ,como filho lula...
ki legal não e mesmo???
Em solidariedade ao Waner que se queixa sobre os vazamentos, poxa gente, vamos para de vazar dados sobre os phetistas-pixulequeiros, afinal eles 'lutaram muito' pra conseguir essas bocadas!
Ô gente, vamos para de vazar dados sobre os phetistas graduados, poxa, afinal eles 'lutaram muito' pra conseguir as bocadas!

Quadro 5. Relação de sentenças irônicas marcadas por todos os anotadores

Por nosso trabalho ser pautado em um *corpus* complexo, cheio de marcas de oralidade, de referências implícitas e outras subjetividades, o processo de interpretação e compreensão torna-se uma tarefa muito mais difícil para os humanos e aspectos como estes devem ser levados em consideração para o desenvolvimento de um método capaz de detectar automaticamente

sentenças irônicas. Todavia, a classificação da intenção de uma opinião em *corpus* em UGC para o português do Brasil abre novos caminhos para a descrição da língua e para novos projetos de anotação e análise de textos opinativos.

6.3. Expressões irônicas e subclasses

De acordo com o esquema de anotação proposto no capítulo anterior, mostramos que uma sentença pode ser classificada de acordo com sua intenção, isto é, entre irônica, não irônica e outro tipo de ironia. A partir desta classificação, as sentenças irônicas foram classificadas em categorias, nomeadas de *ironias com oposições explícitas* e *ironias com oposições implícitas* (Apêndice B). Segundo a Figura 26, das 198 sentenças analisadas, 68% são ironias *implícitas*, ou seja, aquelas que apesar das pistas linguísticas, o sentido irônico só é alcançado através de conhecimentos externos à sentença. Já as opiniões irônicas *explícitas* totalizam 32% das ironias anotadas.

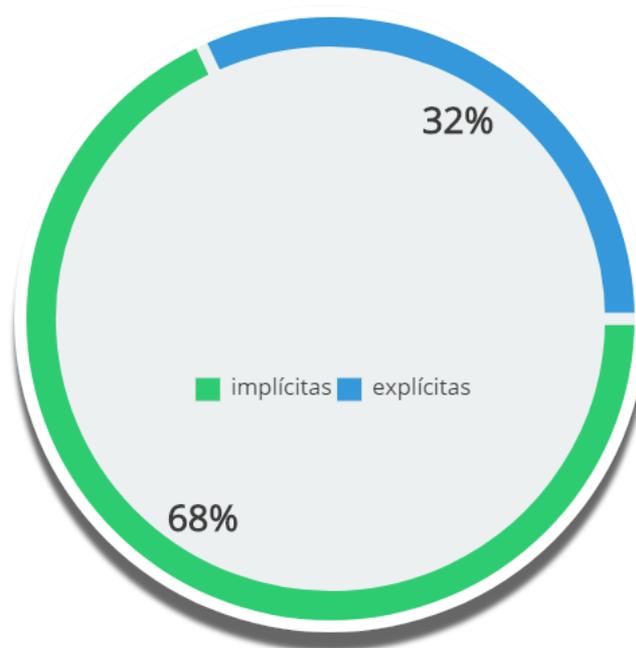


Figura 26. Porcentagem de oposição implícita e explícita

Observamos anteriormente no Capítulo 5, além da oposição entre proposições alguns dispositivos linguísticos podem servir como pistas para a detecção de ironia. De acordo com os resultados apresentados na Tabela 7, as subcategorias mais representativas que criam um efeito irônico em opiniões em mídias sociais são: *disparador* (24,8%), *adjetivo* (21,8%), *entidade nomeada* (13%), *intensificador/modificador* (11,2%), *questão retórica* (8,2%) e *pontuação* (6,8%).

OPINIÕES IRÔNICAS		
	TOTAL	PORCENTAGEM
ENTIDADE NOMEADA (NE)	65	13%
QUESTÃO RETÓRICA	41	8.2%
PONTUAÇÃO	34	6.8%
EMOTICON	1	0.2%
ANALOGIA	3	0.6%
DIMINUTIVO	9	1.8%
ASPAS	26	5.2%
RISOS	2	0.4%
ADJETIVO	109	21.8%
DISPARADOR	85	24.8%
HIPÉRBOLE	4	0.8%
INTERJEIÇÃO	12	2.4%
INTENSIFICADOR/ MODIFICADOR	56	11.2%
EXPRESSÃO	13	2.6%
REPETIÇÃO	2	0.4%

Tabela 12. Distribuição das subcategorias nas sentenças analisadas

A Tabela 7 apresenta a distribuição dos dispositivos entre as ironias implícitas e explícitas. A maioria dos dispositivos marcados como possíveis pistas na detecção de ironia ocorrem nos casos de ironia implícita: *entidade nomeada* (13%) e *questão retórica* (8.2%). Nos exemplos (63) e (64), as entidades nomeadas, *PêeMeDêBê*, *Chavez* e *Maduro* não são os elementos principais, mas são os alvos para qual a ironia do enunciado é direcionada.

(63) O *PêeMeDêBê* é apenas “as meninas” da “casa”, algumas delas mais cobiçadas, recebem mais...

(64) *Chavez* e *Maduro* são democratas.

As *questões retóricas*, em sua maioria, estão relacionadas às ironias implícitas. Isso acontece porque, quando o autor de um comentário se utiliza de *questões retóricas* para ser irônico, ele procura ecoar contextos e situações que não estão, necessariamente, no texto. Um ponto importante a ser destacado é que nos casos das *questões retóricas*, toda a sentença é

considerada como uma pista de ironia. Alguns elementos lexicais e gráficos podem enfatizar a intenção irônica, embora não seja fundamental, como é o caso de *dó* no exemplo (65) e *liderança política* no exemplo (66).

(65) Ninguém tem dó dela não????

(66) Liderança política; ela já teve isso?

Os casos dos *intensificadores/modificadores* (15%) e das *expressões* (4,7%) são mais frequentes nas ironias explícitas. Assim, tanto os intensificadores quanto as expressões marcam de modo mais incisivo uma ironia, como é possível observar nos exemplos (67) e (68). Em (67) o advérbio *absolutamente* intensifica a opinião irônica da Venezuela ser democrática. O mesmo acontece com *vai para o céu*, uma vez que a expressão acentua o *tão honesta*, a principal marca de ironia no enunciado.

(67) A Venezuela é absolutamente democrática.

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

No entanto, de acordo com a Tabela 8, a proporcionalidade dos *adjetivos* (24,6% para as ironias implícitas e 22,4% para as explícitas) e *disparadores* (18,5% para as ironias implícitas e 18,2% para as explícitas) nos mostram que estes dispositivos são fundamentais na construção da ironia em textos opinativos no domínio político.

EXPRESSÕES IRÔNICAS		
	IMPLÍCITA	EXPLÍCITA
ENTIDADE NOMEADA (NE)	16.9%	10.7%
QUESTÃO RETÓRICA	12.9%	4.2%
PONTUAÇÃO	7.3%	7.5%
EMOTICON	0%	0.5%
ANALOGIA	0.4%	0.9%
DIMINUTIVO	1.6%	2.3%
ASPAS	3.6%	7.9%
RISOS	0.4%	0.5%
ADJETIVO	24.6%	22.4%
DISPARADOR	18.5%	18.2%
HIPÉRBOLE	0.8%	0.9%
INTERJEIÇÃO	1.6%	3.7%
INTENSIFICADOR/ MODIFICADOR	9,6%	15%
EXPRESSÃO	1.2%	4.7%
REPETIÇÃO	0.4%	0.5%

Tabela 13. Distribuição das subcategorias entre as oposições implícitas e explícitas

Os resultados mostraram que a porcentagem da subcategoria dos *adjetivos* é a mais alta nas expressões irônicas anotadas na ETAPA 3. Em (69) o adjetivo *fantástica* é o elemento lexical que permite compreendermos o enunciado como uma ironia. Esse tipo de adjetivo possui um alto grau de polaridade positiva (SCOPIIM, 2011) e dependendo do contexto em que ele está inserido (um contexto negativo) é possível alcançar a oposição semântica do enunciado.

(69) A frase do Sr. Rui Falcão é *fantástica*.

ADJETIVOS DE POLARIDADE POSITIVA BEM MARCADA	
belo bom bem excelente extraordinário fantástico feliz glorioso heroico hilário honesto ilustre	incrível legal limpo lindo majestoso maravilhoso nobre ótimo santo satisfeito solícito

Quadro 6. Adjetivos com polaridade bem marcada extraídos de sentenças irônicas

Das ocorrências dos adjetivos positivos (Apêndice B), destacamos no Quadro 24, vinte e três adjetivos sendo de polaridade positiva bem marcada. Vale ressaltar que muitas opiniões irônicas podem conter dois ou mais dispositivos em uma única frase, o que posteriormente causa ambiguidade e imprecisão no tratamento automático. Assim, o uso de adjetivos extremamente positivos, como os listados acima, pode não só representar a subcategoria dos adjetivos como também a subclasse da hipérbole, como no exemplo (70). O exemplo (71) demonstra outra classe que coocorre com a subcategoria dos adjetivos que é o caso das aspas.

(70) Nossos políticos aceitam de bom grado..aliás vivemos um momento *extraordinário!*

(71) Enquanto Temer tenta se safar, Cunha dará andamento ao Impeachment.... a limpeza da casa começa por esse "*pequeno*" detalhe.

Percebemos também que junto aos adjetivos, outras duas subcategorias destacam-se: os *disparadores* e as *entidades nomeadas*. Nesse contexto, as coocorrências destas subcategorias nos dão uma pista de ironia. O exemplo (72) o adjetivo *fugido* qualifica a NE *Tarso Genro* que está ligado a disparador *quebrou*. Contamos também com o advérbio *lindamente* que age como um modificador, isto é, através de *lindamente* é possível alcançar a intenção irônica.

(72) O *fugido* do Rio Grande do Sul *Tarso Genro* que *lindamente* ele *quebrou* deveria ir achar um trabalho...

A partir dos resultados obtidos, observamos que a subcategoria dos *disparadores* é outra classe de dispositivos linguísticos indicativos de ironia. No Capítulo 5 descrevemos as

subcategorias encontradas no processo de anotação da ETAPA 4 e apontamos para o uso de elementos lexicais relativos ao domínio político, como, *corrupção* e *honestidade*.

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

(74) Como é gostoso gastar dinheiro dos *contribuintes*!!!

No exemplo (73), a ironia acontece no contraste entre *muito nobres* e *usou dinheiro de corrupção*. Consideramos *corrupção* um *disparador*, pois, se substituíssemos por outro elemento lexical, como *trabalho* não seria possível criar um enunciado irônico. O mesmo acontece em (74), onde *contribuintes* é considerado um *disparador* se comparada a *gostoso gastar dinheiro*.

No entanto, podemos observar que *os disparadores* não precisam ser necessariamente do domínio político, mas aquela em que é necessária para a construção do sentido irônico. Dessa forma, em (75), o item lexical *parabéns* complementa toda a expressão *meu final de ano me lambuzei de arroz com farofa a carne ficou pra para turma do maiusculo*.

(75) Jaques está de *parabéns*, o meu final de ano me lambuzei de arroz com farofa a carne ficou pra para turma do maiusculo.

Assim, partindo das análises feitas nas seções 6.1, 6.2 e 6.3, o próximo passo foi interpretarmos os resultados obtidos através de uma discussão em que apresentamos a criação de um método de anotação de expressões e ironia.

6.4. Resultados obtidos

Durante todo este trabalho, além de toda elucidação sobre conceitos de ironia e da área Análise de Sentimentos e da importância da anotação de corpus como um recurso linguístico, apresentamos também um método de anotação em um *corpus* opinativo focado em expressões de ironia, a fim de colaborar com futuras pesquisas em PLN voltadas às aplicações e descrições de opiniões irônicas.

Assim, estabelecemos oito etapas resultantes para a anotação do *corpus*, as quais foram:

1. Escolha do domínio político para constituição de um *corpus* e a construção de um corpus opinativo a partir de comentários extraídos de um jornal de grande circulação. No caso desta pesquisa, propomos uma anotação em textos opinativos no domínio político, pois acreditamos possuir um expressivo número de utilização de mecanismos linguísticos para criar diversos efeitos de sentido, entre eles, a ironia.

2. Criação de um cabeçalho para identificar os comentários, os autores, a data e as sentenças do comentário.
3. Criação de diretrizes para o processo de anotação da pesquisa.
4. Anotação das opiniões de acordo com sua intenção.
5. Criação de um método para a validação das diretrizes e da anotação realizada.
6. Cálculo da anotação feita pelos anotadores, medindo a eficácia do método e das diretrizes propostas.
7. Extração e análise das opiniões irônicas extraídas na etapa de validação.
8. Identificação de elementos linguísticos utilizados como pistas na detecção de ironia.

O método descrito acima tem como resultado a relação de ironias com oposição explícitas e implícitas, listas de pistas linguísticas identificadas nestas ironias, além da elaboração de um manual de anotação de corpus voltado à identificação e detecção de opiniões irônicas. Além disso, pudemos verificar particularidades na construção de uma ironia, descritas durante a análise nas seções anteriores.

Conclusão e trabalhos futuros

A presente dissertação procurou estudar e descrever a ironia visando a construção de recursos linguísticos-computacionais para a Análise de Sentimentos. Nosso trabalho descreve um esquema de anotação para opiniões irônicas UGC no domínio político, buscando esclarecer como é o processo de identificação destes casos e se existem dispositivos linguísticos capazes de facilitar sua identificação.

Entendemos que os objetivos propostos foram alcançados, pois pôde-se identificar características de expressão da ironia no domínio escolhido, como o uso de palavras, os disparadores e adjetivos, com polaridade bem marcada que se opõem a outra sequência.

Mesmo que a ironia seja uma figura de retórica conhecida e bastante utilizada desde a antiguidade, a detecção de um enunciado irônico é uma tarefa que exige não só a interpretação da intenção do falante, mas pode envolver também o compartilhamento da cultura, de uma forma de humor e de conhecimento de mundo.

Um dos principais problemas da identificação de ironia em mídias sociais é que fatores linguísticos e sociais intrínsecos em declarações irônicas tornam a tarefa de detecção automática desses mecanismos figurados instâncias bastante complexas, especialmente, devido à falta de informações extralinguísticas.

Nesse contexto, durante o desenvolvimento dessa pesquisa, foi possível observar que o significado irônico deve ser implicado por contextos pragmáticos e, além disso, há a necessidade da realização de processos mentais, como referências a outros enunciados e situações.

No Capítulo 6 foram apresentados os resultados e análise dos dados obtidos. Pudemos demonstrar que a identificação da ironia ainda é uma tarefa que pode ser melhor descrita e analisada a partir da utilização de algumas estratégias linguísticas e de novas pistas para a detecção de ironia ainda puderam ser descritas e descobertas. Como nosso foco era sobretudo o léxico, pudemos constatar que as subclasses que mais colaboram para a identificação do efeito irônico texto do domínio escolhido são os adjetivos (23,6%), os disparadores (18,4%), as entidades nomeadas (14,1%), as questões retóricas (8,9%) e a pontuação (7,4%).

A ironia pode ser considerada um fenômeno generalizado em UGC e as percepções empíricas descritas nesta dissertação podem ser úteis para a melhoria de processamento

automático da linguagem figurada. Sabemos, no entanto, que ainda são necessárias muitas outras pesquisas e novas descrições para atingir esse objetivo. Nosso papel é iniciar uma descrição mais detalhada de mecanismos figurados subjacentes à língua que podem alterar o sentido de uma sentença.

Nesse sentido, um incremento da qualidade de um esquema de anotação de *corpus* de textos opinativos, a melhoria da análise de padrões textuais, bem como a investigação de como funcionam elementos como as perguntas retóricas e os disparadores, podem ajudar a obter uma imagem mais clara de como a ironia funciona nas mídias sociais. Assim, esperamos contribuir para o desenvolvimento de um léxico de sentimentos voltado às mídias sociais, agregando interjeições, adjetivos e advérbios altamente polarizados, expressões cristalizadas mais frequentes em português e disparadores para o domínio político. 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 da Análise de Sentimentos.

Referências Bibliográficas

ALUÍSIO, S.M.; ALMEIDA, G.M.B. O que é e como se constrói um corpus? Lições aprendidas na compilação de vários corpora para a pesquisa linguística. In: **Calidoscópio**. V4, n. 3, setembro/dezembro, p. 156-178, 2006.

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

AUSTIN, J. L. **How to do things with words**. Oxford University Press, London. 1962.

BAKER, C. F.; FILLMORE, C. J.; LOWE, J. B. The berkeley framenet project. In: **Proceedings of the 17th international conference on Computational linguistics-Volume 1**. Association for Computational Linguistics, p. 86-90. 1998.

BALAGE FILHO, P. P.; PARDO, T. A. S.; ALUISIO, S. M. An evaluation of the Brazilian Portuguese LIWC dictionary for sentiment analysis. In: **Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology (STIL)**. p. 215-219. 2013.

BARBIERI, F.; SAGGION, H.; RONZANO, F. Modelling sarcasm in twitter, a novel approach. In: **Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis**. 2014. p. 50-58.

BECKER, K.; TUMITAN, D. **Introdução à mineração de opiniões: conceitos, aplicações e desafios**. Rio Grande do Sul: UFRGS, 2013.

BERBER SARDINHA, T. Linguística de corpus: histórico e problemática. **Delta**, v. 16, n. 2, p. 323-367, 2000.

BISOGNIN, T. R. **Sem medo do internetês**. 1 ed. Porto Alegre: AGE, 2009.

BURGERS, C. **Verbal irony: Use and effects in written discourse**. UB Nijmegen [Host]. 2010.

CARVALHO, P.; SARMENTO, L.; SILVA, M. J.; DE OLIVEIRA, E. Clues for detecting irony in user-generated contents: oh...!! it's so easy;- In **Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion**, p. 53-56, 2009.

CARVALHO, P.; SARMENTO, L.; TEIXEIRA, J.; SILVA, M. J. Liars and saviors in a sentiment annotated corpus of comments to political debates. In: **Proceedings of the 49th**

Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2. Association for Computational Linguistics, p. 564-568. 2011.

CERON, A.; CURINI, L.; IACUS, S. M. Using sentiment analysis to monitor electoral campaigns: Method matters—evidence from the United States and Italy. **Social Science Computer Review**, v. 33, n. 1, p. 3-20, 2015.

CLARK, H. H.; GERRIG, R. J. On the pretense theory of irony. In: **Journal of Experimental Psychology: General**, 113, p. 112 -120. 1984.

CLIFT, R. Irony in conversation. **Language in society**, v. 28, n. 4, p. 523-553, 1999.

CUNHA, C.; CINTRA, L. F. L. **Nova gramática do português contemporâneo.** Rio de Janeiro: Nova Fronteira, 1985.

CURCÓ, C. Irony: Negation, echo and metarepresentation. In: **Lingua**, v. 110, n. 4, p. 257-280, 2000.

DE FREITAS, L. A.; VANIN, A. A.; HOGETOP, D. N.; BOCHERNITSAN, M. N.; VIEIRA, R. Pathways for irony detection in *tweets*. In: **Proceedings of the 29th Annual ACM Symposium on Applied Computing.** ACM, p. 628-633. 2014.

DE OLIVEIRA, L. P. Linguística de Corpus: teoria, interfaces e aplicações. **Matraga-Revista do Programa de Pós-Graduação em Letras da UERJ**, v. 16, n. 24, 2009.

ESULI, A.; SEBASTIANI, F. SENTIWORDNET: A high-coverage lexical resource for opinion mining. **Institute of Information Science and Technologies (ISTI) of the Italian National Research Council (CNR)**, 2006.

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

FREITAS, C.; MOTTA, E.; CESAR, J.; MILIDIÚ, R. L. Vampiro que brilha... rá! Desafios na anotação de opinião em um corpus de resenhas de livros. **ENCONTRO DE LINGUÍSTICA DE CORPUS**, v. 11, p. 1-13, 2012.

GIBBS, R. W. Jr.; COLSTON, H. L. (Eds.). **Irony in language and thought: A cognitive science reader.** Psychology Press, 2007.

- GIBBS, R. W. Jr., RAYMOND W.; O'BRIEN, J. E.; DOOLITTLE, S. Inferring meanings that are not intended: Speakers' intentions and irony comprehension. **Discourse Processes**, v. 20, n. 2, p. 187-203, 1995.
- GIORA, R. On irony and negation. **Discourse processes**, v. 19, n. 2, p. 239-264, 1995.
- GHOSH, A.; LI, G.; VEALE, T.; ROSSO, P.; SHUTOVA, E.; BARNDEN, J.; REYES, A. Semeval-2015 task 11: Sentiment analysis of figurative language in twitter. In: **Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)**. p. 470-478. 2015.
- GONZÁLEZ-IBÁÑEZ, R.; MURESAN, S.; WACHOLDER, N. Identifying sarcasm in Twitter: a closer look. In: **Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers-Volume 2**. Association for Computational Linguistics, p. 581-586. 2011.
- GRICE, H. P. Logic and conversation. In: Cole, P; Morgan, J.L. (Eds.) In: **Syntax and semantics**, Vol. 3: Speech acts. New York, Academic Press, p. 225-242, 1975.
- HAIMAN, J. **Talk is cheap: Sarcasm, alienation, and the evolution of language**. Oxford University Press on Demand, 1998.
- HAO, Y.; VEALE, T. Support structures for linguistic creativity: A computational analysis of creative irony in similes. In: **Proceedings of the Cognitive Science Society**. 2009.
- HOGENBOOM, Alexander et al. Exploiting emoticons in sentiment analysis. In: **Proceedings of the 28th Annual ACM Symposium on Applied Computing**. ACM, 2013. p. 703-710.
- HOUAISS, A.; VILLAR, M. Dicionário Houaiss Eletrônico. **Versão monousuário**, v. 3, 2009.
- HOVY, E.; LAVID, J. 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.
- IDE, N.; PUSTEJOVSKY, J. (Ed.). **Handbook of Linguistic Annotation**. Springer, 2017.
- KENNEDY, G. **An introduction to corpus linguistics**. Series: Studies in language and linguistics. London, England. 2014.
- KIERKEGAARD, S. A. **O conceito de ironia: constantemente referido a Sócrates**. Søren Aabye Kierkegaard; apresentação e tradução, Álvaro Luiz Montenegro Valls. – Petrópolis, RJ. Editora Vozes Limitada, 2017.

- KREUZ, R. J.; CAUCCI, G. M. Lexical influences on the perception of sarcasm. In: **Proceedings of the Workshop on computational approaches to Figurative Language**. Association for Computational Linguistics, p. 1-4. 2007.
- KREUZ, R. J.; GLUCKSBERG, S. How to be sarcastic: The echoic reminder theory of verbal irony. **Journal of experimental psychology: General**, v. 118, n. 4, p. 374, 1989.
- KREUZ, R. J.; ROBERTS, R. M. On satire and parody: The importance of being ironic. **Metaphor and Symbol**, v. 8, n. 2, p. 97-109, 1993.
- KUMON-NAKAMURA, S.; GLUCKSBERG, S.; BROWN, M. How about another piece of pie: The allusional pretense theory of discourse irony. **Journal of Experimental Psychology: General**, v. 124, n. 1, p. 3, 1995.
- LEECH, G. Corpus annotation schemes. In: **Literary and linguistic computing**, v. 8, n. 4, p. 275-281, 1993.
- LEECH, G. Adding Linguistic Annotation. In: Wynne., M. (Eds.) **Developing linguistic corpora: A guide to good practice**. Oxford: Oxbow Books, 2005.
- LIU, B. Sentiment Analysis and Subjectivity. **Handbook of natural language processing**, v. 2, p. 627-666, 2010.
- LIU, B. Sentiment analysis and opinion mining. **Synthesis lectures on human language technologies**, v. 5, n. 1, p. 1-167, 2012.
- MAYNARD, D.; GREENWOOD, M. A. Who cares about Sarcastic *Tweets*? Investigating the Impact of Sarcasm on Sentiment Analysis. In: **Lrec**. p. 4238-4243. 2014.
- MARCUS, M. P.; MARCINKIEWICZ, M. A.; SANTORINI, B. Building a large annotated *corpus* of English: The Penn Treebank. **Computational linguistics**, v. 19, n. 2, p. 313-330, 1993.
- MCENERY, A. M.; WILSON, A. **Corpus linguistics: an introduction**. Edinburgh University Press, 2001.
- KAROUI, J.; BENAMARA, F.; MORICEAU, V.; AUSSÉNAC-GILLES, N.; BELGUITH, L. H. Towards a contextual pragmatic model to detect irony in *tweets*. In: **53rd Annual Meeting of the Association for Computational Linguistics (ACL 2015)**. p. PP. 644-650. 2015.
- KENNEDY, G. **An introduction to corpus linguistics**. Routledge, 2014
- KIERKEGAARD, S. **O Desespero Humano**. São Paulo: Martin Claret, 2006.

- KREUZ, R. J.; ROBERTS, R. M. On satire and parody: The importance of being ironic. **Metaphor and Symbol**, v. 8, n. 2, p. 97-109, 1993.
- PALLAVICINI, F.; CIPRESSO, P.; MANTOVANI, F. Beyond Sentiment: How Social Network Analytics Can Enhance Opinion Mining and Sentiment Analysis. In: **Sentiment Analysis in Social Networks**. p. 13-29. 2017.
- PALMER, M.; GILDEA, D.; KINGSBURY, P. The proposition bank: An annotated corpus of semantic roles. **Computational linguistics**, v. 31, n. 1, p. 71-106, 2005.
- PANG, B.; LEE, L. Opinion mining and sentiment analysis. **Foundations and Trends® in Information Retrieval**, v. 2, n. 1-2, p. 1-135, 2008.
- PAUMIER, Sébastien. Unitex-manuel d'utilisation. 2011.
- PETRILLO, M.; BAYCROFT, J. Introduction to manual annotation. **Fairview Research**, 2010.
- POZZI, F. A.; FERSINI, E.; MESSINA, E.; LIU, B. **Sentiment analysis in social networks**. Morgan Kaufmann, 2016.
- PUSTEJOVSKY, J.; STUBBS, A. **Natural Language Annotation for Machine Learning: A guide to corpus-building for applications**. " O'Reilly Media, Inc.", 2012.
- REYES, A.; ROSSO, P.; BUSCALDI, D. From humor recognition to irony detection: The figurative language of social media. **Data & Knowledge Engineering**, v. 74, p. 1-12, 2012.
- REYES, A.; ROSSO, P.; VEALE, T. A multidimensional approach for detecting irony in twitter. **Language resources and evaluation**, v. 47, n. 1, p. 239-268, 2013.
- SANTOS, D. Disponibilização de corpora de texto através da WWW. In: MARRAFA, P.; MOTA, M. A. (Eds) **Linguística Computacional: Investigação Fundamental e Aplicações Actas do I Workshop sobre Linguística Computacional da Associação Portuguesa de Linguística**. Lisboa: Colibri. Colibri, 1999.
- SANTOS, D.; MARQUES, R.; FREITAS, C.; SIMÕES, A.; MOTA, C. Comparando anotações linguísticas na Gramateca: filosofia, ferramentas e exemplos. **Domínios de Lingu@ gem**, v. 9, n. 2, p. 11-26, 2015.
- SEARLE, J. R. **Speech acts: an essay in the philosophy of language**. Cambridge, Cambridge University Press. 1969.

- SCOPIM, Débora. Estudo de padrões lexicais em textos opinativos. Dissertação de Mestrado. 2012.
- SINGH, R. K. Humour, irony and satire in literature. **International Journal of English and Literature (IJEL)**, v. 3, n. 4, p. 65-72, 2012.
- SPERBER, D; WILSON, D. Irony and the use-mention distinction. **Philosophy**, v. 3, p. 143-184, 1981.
- SPERBER, D; WILSON, D. **Relevance: communication & cognition**. 2ed. Cambridge: Blackwell, 1995.
- SPERBER, D; WILSON, D. As ironias como. In: **Língua e Literatura**, n. 18, p. 81-97, 1990.
- VAN HEE, C.; LEFEVER, E.; HOSTE, V. Monday mornings are my fave:)# not Exploring the Automatic Recognition of Irony in English *tweets*. In: **Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers**. 2016. p. 2730-2739.
- VANIN, A. A.; FREITAS, L. A.; VIEIRA, R.; BOCHERNITSAN, M. Some clues on irony detection in *tweets*. In: **Proceedings of the 22nd International Conference on World Wide Web**. ACM, p. 635-636. 2013.
- VEALE, T.; HAO, Y. Detecting Ironic Intent in Creative Comparisons. In: **ECAI**. 2010. p. 765-770.
- WALLACE, B. C. Computational irony: A survey and new perspectives. In: **Artificial Intelligence Review**, v. 43, n. 4, p. 467-483, 2015.
- WILSON, D. SPERBER, D. Teoria da relevância. In: **Linguagem em (Dis)curso**, v. 5, p. 221-268, 2005.

Anexo A

Manual para anotação de expressão irônica em texto opinativo

Na última década, como aumento ao acesso à Web 2.0 e o avanço de novas tecnologias, textos que carregam uma avaliação ou opinião sobre uma determinada entidade em mídias sociais passaram a ser fontes de informação de opinião a respeito de políticos, empresas, organizações, entre outros.

Estudar expressões irônicas em dados coletados em um corpus constituído por textos opinativos pode aprimorar diversas ferramentas e pesquisas baseadas em Processamento de Língua Natural (PLN). Assim, este estudo tem como objetivo investigar as expressões de ironia para o português do Brasil em conteúdo gerado pelo usuário. Em particular, a pesquisa concentra-se em um tipo específico de ironia, definida como uma maneira de dizer o oposto do que é literalmente dito.

A hipótese desta investigação é de que a ironia poderia ser possivelmente identificada com a ajuda de dispositivos linguísticos que permitem uma compreensão irônica de uma determinada sentença. Como por exemplo,

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

(2) *São muito nobres, afinal, a chapa usou \$\$ de corrupção.*

No exemplo acima, dado o nosso conhecimento de mundo, é possível compreender essa sentença como irônica, pois sabemos que o ato dos políticos brasileiros gastarem o dinheiro público com propina não é algo realmente bom. No entanto, formalmente, a compreensão vai além do contraste entre “dinheiro público” versus “propina”, mas também uma oposição entre esses elementos e a expressão “como é bom”.

Para examinar as ocorrências de ironia e sarcasmo parte-se da busca em corpus – previamente construído. O *ComentCorpus* é um corpus de comentários de notícias no português do Brasil manualmente anotado. Para a criação do *ComentCorpus* foi compilada uma coleção de 90 notícias relacionadas ao período de impeachment da presidente Dilma Rousseff.

Para a escolha das notícias foi realizada uma busca pela palavra *impeachment* no caderno *Poder* do jornal *Folha de S. Paulo* no período de janeiro a junho de 2016. O corpus é composto por 6.185 comentários, totalizando 14 mil sentenças e 207 mil palavras.

As sentenças a serem analisadas são expressões opinativas sobre algum assunto referente ao domínio político. Dessa forma, todo anotador deverá interpretar e marcar um conjunto de comentários e decidir a intenção de cada sentença: se é irônica, não irônica ou contém outro tipo de ironia.

1. Análise no nível da sentença

Considera-se uma sentença a unidade segmentada por um ponto final, exclamação, interrogação ou reticências:

(3) *estratégia do governo é dizer que tudo se resume a uma disputa entre Dilma e Cunha: satanizar o Cunha e limpar a Dilma.*

(4) *Para a minha preocupação, já muito tumulto pode acontecer...*

No entanto, em alguns casos, há sentenças em que a pontuação pode indicar uma hesitação, pausa ou então a reprodução da oralidade na escrita. Sentenças com dois pontos, ponto e vírgula, parênteses e travessões/hífen são considerados como uma única sentença.

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

(6) *É aquela história, quem apoia corrupto, tende a ser também, o Cunha é o pior deles...o mensalão não termina no petrolão...dá-lê Moro, Janot, PF e STF.*

2. Diretrizes para a anotação

1) Sentenças irônicas

Será considerada ironia apenas a sentença de opinião que há a oposição de sentido entre o que o falante pretende dizer e o que é realmente dito. Segundo Colston e Gibbs (2007), a ironia é o uso lúdico da linguagem em que um falante implica o oposto do que é literalmente dito. Já para Giora (1995), a ironia é um tipo de negação indireta que expressa um sentimento em oposição direta ao que realmente se acredita.

(7) *A Sra. Presidente é tão honesta que vai para o céu.*

(8) *Como deve ser bom gastar o dinheiro dos contribuintes!!*

2) Sentenças com outro tipo de ironia

Será considerada outro tipo de ironia a sentença que NÃO há oposição entre o sentido literal e significado pretendido, no entanto, o texto ainda carrega algum dispositivo figurado que alteram sua literalidade.

(9) *Quanto estão levando pra censurar os comentários contra esse governinho?*

(10) *Dilmentira não tem um pingão de sentimento e espontaneidade.*

No exemplo (9), a sentença pode ser considerada uma espécie de ironia, mas que tem um tom mais agressivo para se referir ao governo do Brasil - que pode ser indicado pelo trecho “esse governinho”. Isso acontece porque o uso de diminutivo, por exemplo, apesar de expressar, em sua maioria, casos de afetividade também pode indicar um sentido de inferioridade ou descaso. Já no exemplo (10), o item lexical “Dilmentira” está apelidando com um tom jocoso a ex-presidente do Brasil.

Os casos de outro tipo de ironia podem incluir o sarcasmo, a sátira e o humor, por exemplo:

Sarcasmo: ironia empregada com o objetivo de atacar ou zombar alguém. Attardo (2000) define sarcasmo como um tipo de ironia abertamente agressiva, com pistas alvo mais claros:

(11) *Esse partidinho traidor do Brasil já deveria ter sido varrido.*

Sátira: é uma maneira de se divertir com as pessoas usando uma linguagem tola ou exagerada. Singh (2012) afirma que a sátira é uma figura de linguagem que enfatiza a fraqueza e não a pessoa fraca. Para o autor, a sátira é usada para expor e criticar algo em um indivíduo ou a própria sociedade através do humor, da ironia, do exagero ou do ridículo:

(12) *Já já começará aquela conversinha de que ele é “pelseguido pulíticu” e blá blá blá.*

(13) *O que será que Aécinho Zona Sul fez na noite anterior pra acordar um dia e decidir dar apoio a um golpe perpetrado por uma dupla destas?*

Humor: mensagem que tem o poder de provocar risos através de ingenuidade, habilidade verbal ou incongruência. O humor para Singh (2012) consiste na capacidade de expressão de uma sensação de algo divertido e inteligente. Refere-se, principalmente, no reconhecimento de incongruências ou peculiaridades presentes em uma situação

(14) *O Suplicy não tem amor próprio, parece aqueles balões de plástico infláveis sem vida.*

(15) *Nine Fingers na frente?*

3) Sentenças não irônicas

Será considerada não irônica as sentenças que não contém mecanismos linguísticos que alternam o ser significado.

(16) Eu sou a favor da saída da atual Presidente e sou forte crítico da emenda da reeleição.

(17) Ela só não disse que convocaria o plebiscito no dia 30 de Dezembro de 2018.

4) Caso não tenha certeza se a sentença não se encaixa nas classificações (irônica, outro tipo de ironia e não irônica) marque a quarta coluna: não soube responder.

5) Procure não analisar cada sentença por um longo tempo. Decida com base na sua primeira interpretação.

6) Não hesite em passar para a próxima sentença se não conseguir interpretá-la. Você poderá retornar mais tarde.

7) Faça pausas frequentes! É aconselhável dar um espaço de tempo a cada troca de notícia. Isso deve aumentar sua precisão e evita a desconcentração na anotação.

8) Marcar com a cor vermelha os elementos linguísticos que indicam ironia.

(18) Basta ver as demonstrações de campanha, afinal as "doações" foram todas "legais".

Segundo Burgers et al (2012), marcadores de ironia não são essenciais para tornar um enunciado irônico e, teoricamente, poderiam ser removidos do enunciado sem deixar a expressão perder seu significado irônico, como por exemplo, as aspas. No exemplo (18), é possível observar primeiramente que *doações* e *legais* estão marcadas com aspas, o que já indica uma sentença irônica, no entanto, a intenção irônica da sentença é reforçada pelo contraste entre estes elementos linguísticos e *demonstrações de campanha*.

9) Marcar com a cor azul os elementos linguísticos que indicam outro tipo ironia.

(19) E ao invés de buscar picuinhas para enfatizar esse governozinho, deveriam é reconhecer que eles quebraram.

No exemplo (19) a oposição entre *buscar picuinhas para enfatizar versus reconhecer que eles quebraram* está marcada pela variação *governozinho*, que por sua vez, chama a atenção para a natureza áspera do enunciado através uso do diminutivo.

10) Procure focar na própria sentença como nível de análise. Evite se basear no contexto ou conhecimento de mundo para classificar um enunciado.

Bom trabalho!

Anexo B

Extrações das subcategorias anotadas em sentenças irônicas

ADJETIVOS		
<adj>alérgico</adj>	<adj>fantástica</adj>	<adj>maravilhosa</adj>
<adj>alheio</adj>	<adj>feliz</adj>	<adj>maravilhoso</adj>
<adj>amigo</adj>	<adj>fidalgo</adj>	<adj>meigo</adj>
<adj>anjos</adj>	<adj>fiel</adj>	<adj>melhor</adj>
<adj>apaniguados</adj>	<adj>fugido</adj>	<adj>menor</adj>
<adj>bacana</adj>	<adj>glorioso</adj>	<adj>nobres</adj>
<adj>bastião</adj>	<adj>gostoso</adj>	<adj>novo</adj>
<adj>belíssimo</adj>	<adj>governista</adj>	<adj>ofegante</adj>
<adj>belo</adj>	<adj>grande</adj>	<adj>ótima</adj>
<adj>bem</adj>	<adj>grandes</adj>	<adj>pasma</adj>
<adj>boas</adj>	<adj>grave</adj>	<adj>pequeno</adj>
<adj>bom</adj>	<adj>herói</adj>	<adj>perdida</adj>
<adj>chiquérimaaaas</adj>	<adj>hilário</adj>	<adj>pobre</adj>
<adj>chupins</adj>	<adj>hipócrita</adj>	<adj>pouco</adj>
<adj>cobiçadas</adj>	<adj>honesta</adj>	<adj>preocupada</adj>
<adj>coitada</adj>	<adj>honestidade</adj>	<adj>primeira</adj>
<adj>coitadinho</adj>	<adj>honesto</adj>	<adj>realizado</adj>
<adj>coitado</adj>	<adj>iconoclasias</adj>	<adj>rusa</adj>
<adj>coitados</adj>	<adj>ilustre</adj>	<adj>santo</adj>
<adj>comovido</adj>	<adj>importante</adj>	<adj>satisfeito</adj>
<adj>concluídos</adj>	<adj>incrível</adj>	<adj>saudável</adj>
<adj>confiança</adj>	<adj>injustiçado</adj>	<adj>simples</adj>
<adj>democrata</adj>	<adj>inocentes</adj>	<adj>solícita</adj>
<adj>democratas</adj>	<adj>interessante</adj>	<adj>tadinha</adj>
<adj>desafeto</adj>	<adj>íntimo</adj>	<adj>traidores</adj>
<adj>diferente</adj>	<adj>legal</adj>	<adj>último</adj>
<adj>difícil</adj>	<adj>limpa</adj>	<adj>única</adj>
<adj>emocionante</adj>	<adj>linda</adj>	<adj>velho</adj>
<adj>engraçado</adj>	<adj>longa</adj>	<adj>vilões</adj>
<adj>excelente</adj>	<adj>lord</adj>	<adj>verdadeiro</adj>
<adj>extraordinário</adj>	<adj>majestosa</adj>	
<adj>fácil</adj>	<adj>mal</adj>	
<adj>famosa</adj>	<adj>manipulável</adj>	

ANALOGIA

<analogia>o jogo nunca esteve tão bom e emocionante como agora </analogia>

<analogia>a filha dela e muito pobre, como filho lula</analogia>

<analogia>até jornalista alheio está definindo estratégias de comunicação para ela</analogia>

ASPAS

<aspas>“da casa”</aspas> <aspas>“amigo”</aspas> <aspas>“coincidência”</aspas> <aspas>“coisas menores”</aspas> <aspas>“consciências”</aspas> <aspas>“controlar”</aspas> <aspas>“doações”</aspas> <aspas>“falta de assunto”</aspas> <aspas>“glorioso”</aspas> <aspas>“já mais”</aspas> <aspas>“meninas”</aspas> <aspas>“não sabe”</aspas>	<aspas>“não seja ingênuo”</aspas> <aspas>“pequeno”</aspas> <aspas>“pessoa ilustre”</aspas> <aspas>“programa”</aspas> <aspas>“Quase chego as lágrimas!”</aspas> <aspas>“silêncio das virtudes”</aspas> <aspas>“foda”</aspas> <aspas>“golpe”</aspas> <aspas>“legais”</aspas> <aspas>“chairwoman”</aspas> <aspas>“conselheiro”</aspas> <aspas>“lutaram muito”</aspas>
--	---

DIMINUTIVO

<dim>coitadinho</dim> <dim>perguntinha</dim> <dim>sitiozinhos</dim> <dim>finalzinho</dim> <dim>boquinha</dim>	<dim>Santinho</dim> <dim>limpinho</dim> <dim>cheirosinho</dim> <dim>direitinho</dim> <dim>clubinho</dim>
---	--

EMOTICON

<emo>;-)</emo>

ENTIDADES NOMEADAS (NE)

<ne>Aécio</ne> <ne>André B.T.G</ne> <ne>Carmona, O Breve</ne> <ne>Celso Daniel</ne> <ne>chairwoman</ne> <ne>Chavez</ne> <ne>Color</ne> <ne>comprador de reeleição</ne> <ne>Cunha</ne> <ne>D. Marisa</ne> <ne>Dilma</ne> <ne>Dona Marisa</ne> <ne>Eduardo Cunha</ne> <ne>Eikye Baptista</ne> <ne>Eletrobras</ne> <ne>empreiteiras</ne> <ne>ex</ne> <ne>F. Martins</ne> <ne>Falcão</ne>	<ne>governo</ne> <ne>Hitler</ne> <ne>Jaques</ne> <ne>Johnny Santana</ne> <ne>JS</ne> <ne>JW</ne> <ne>Lula</ne> <ne>Luulla</ne> <ne>Maduro</ne> <ne>Mantega</ne> <ne>Márcio</ne> <ne>Marcola</ne> <ne>Ministro</ne> <ne>P.E.T.R.A.L.H.A.S</ne> <ne>P.T</ne> <ne>PeeMeDeBe</ne> <ne>PêeMeDêBê</ne> <ne>Petrobrás</ne> <ne>PeTe</ne> <ne>PF</ne>	<ne>Pravda</ne> <ne>prefeito de Campinas</ne> <ne>Pres Dilma</ne> <ne>Presidenta</ne> <ne>presidente</ne> <ne>Pro-Exx</ne> <ne>PSDB</ne> <ne>PT</ne> <ne>PTTT</ne> <ne>Rose</ne> <ne>santinho</ne> <ne>Sarney</ne> <ne>Sr. Jaques Wagner</ne> <ne>Sra. Presidente</ne> <ne>Tarso Genro</ne> <ne>Temer</ne> <ne>TFC</ne> <ne>Velhinha de Taubaté</ne>
--	--	---

<ne>filho</ne>	<ne>Pinochet</ne>	<ne>Zé</ne>
<ne>Folha</ne>	<ne>PMDB</ne>	<ne>Zé Honestão</ne>
<ne>Franco</ne>	<ne>ppresidentta</ne>	

EXPRESSÕES
<exp>boca cheia</exp> <exp>casa da mãe Joana</exp> <exp>chupa ovo</exp> <exp>conseguir essas bocadas</exp> <exp>fechar a torneira</exp> <exp>foi pras cucuias</exp> <exp>mesmo saco</exp> <exp>nem que a vaca tussa</exp> <exp>óleo de peroba</exp> <exp> tiro o chapéu</exp> <exp>vai pro céu</exp> <exp>vazar pelo ladrão</exp>

HIPÉRBOLE
<hiperbole>2050</hiperbole> <hiperbole>belíssimo</hiperbole> <hiperbole>chiquérrimaaaas</hiperbole> <hiperbole>trilionários</hiperbole>

INTENSIFICADOR/MODIFICADOR	
<mod>absolutamente</mod> <mod>afinal</mod> <mod>agora</mod> <mod>apesar</mod> <mod>até</mod> <mod>aumento</mod> <mod>como sempre</mod> <mod>demais</mod> <mod>do Brasil</mod> <mod>em paz</mod> <mod>em solidariedade</mod> <mod>extremamente</mod> <mod>ficar longe</mod> <mod>imagine</mod> <mod>incondicionalmente</mod> <mod>lindamente</mod> <mod>logo agora</mod> <mod>mais</mod> <mod>manutenção</mod>	<mod>mas</mod> <mod>menos</mod> <mod>muita</mod> <mod>muito</mod> <mod>não é verdade</mod> <mod>obrigatoriamente</mod> <mod>ou melhor</mod> <mod>pelo menos</mod> <mod>pensando bem</mod> <mod>porque</mod> <mod>quanta</mod> <mod>quão</mod> <mod>sempre tão</mod> <mod>simplesmente</mod> <mod>só</mod> <mod>tanta</mod> <mod>tão</mod> <mod>todos</mod>

INTERJEIÇÃO

<itj>claro</itj>

<itj>Hmmm</itj>

<itj>nossa</itj>

<itj>ô gente</itj>

<itj>oba</itj>

<itj>poxa</itj>

<itj>poxa gente</itj>

<itj>poxa vida</itj>

<itj>uh</itj>

<itj>viva</itj>

DISPARADORES

<disparador>aceitar</disparador>

<disparador>acreditar</disparador>

<disparador>adorar</disparador>

<disparador>alíquota</disparador>

<disparador>alma</disparador>

<disparador>amá-la</disparador>

<disparador>anjos</disparador>

<disparador>boa sorte</disparador>

<disparador>boa viagem</disparador>

<disparador>boa noite</disparador>

<disparador>bocadas</disparador>

<disparador>bondade</disparador>

<disparador>campanha</disparador>

<disparador>canonizá-lo</disparador>

<disparador>canonização</disparador>

<disparador>coincidência</disparador>

<disparador>coisa</disparador>

<disparador>comprar</disparador>

<disparador>contribuintes</disparador>

<disparador>contrição</disparador>

<disparador>corrupção</disparador>

<disparador>corruptos</disparador>

<disparador>cortar gastos</disparador>

<disparador>crescimento</disparador>

<disparador>crise</disparador>

<disparador>culpa</disparador>

<disparador>demitir</disparador>

<disparador>democracia</disparador>

<disparador>desemprego</disparador>

<disparador>desenvolvimento</disparador>

<disparador>desgoverno</disparador>

<disparador>desgraça</disparador>

<disparador>destruído</disparador>

<disparador>desvalorização</disparador>

<disparador>detalhe</disparador>

<disparador>deuses</disparador>

<disparador>dinheirama</disparador>

<disparador>dinheiro</disparador>

<disparador>ditador</disparador>

<disparador>dó</disparador>

<disparador>eleitoreiras</disparador>
<disparador>embuste</disparador>
<disparador>estrigo</disparador>
<disparador>ética</disparador>
<disparador>expertise</disparador>
<disparador>financiar</disparador>
<disparador>fome</disparador>
<disparador>funcionar</disparador>
<disparador>ganhar</disparador>
<disparador>gastar</disparador>
<disparador>gostar</disparador>
<disparador>gostaria</disparador>
<disparador>grato</disparador>
<disparador>honestidade</disparador>
<disparador>ignorância</disparador>
<disparador>impeachment</disparador>
<disparador>inflação</disparador>
<disparador>investimento</disparador>
<disparador>lambuzou</disparador>
<disparador>liderança</disparador>
<disparador>louvor</disparador>
<disparador>megaempreiteiro</disparador>
<disparador>mensalão</disparador>
<disparador>mentiras</disparador>
<disparador>mídia</disparador>
<disparador>milagre</disparador>
<disparador>miséria</disparador>
<disparador>moral</disparador>
<disparador>moralidade</disparador>
<disparador>necessidade</disparador>
<disparador>obrigado</disparador>
<disparador>oposição</disparador>
<disparador>país</disparador>
<disparador>parabéns</disparador>
<disparador>parceria</disparador>
<disparador>partido</disparador>
<disparador>pedaladas</disparador>
<disparador>pena</disparador>
<disparador>política</disparador>
<disparador>político</disparador>
<disparador>prisão</disparador>
<disparador>quadrilha</disparador>
<disparador>quebrou</disparador>
<disparador>reeileição</disparador>
<disparador>refúgio</disparador>
<disparador>rombo</disparador>
<disparador>salve</disparador>
<disparador>santo</disparador>
<disparador>saudade</disparador>
<disparador>sindicalismo</disparador>

<disparador>solidariedade</disparador>
 <disparador>subordinados</disparador>
 <disparador>sugestivo</disparador>
 <disparador>sugestões</disparador>
 <disparador>surpresa</disparador>
 <disparador>tomara</disparador>
 <disparador>tucanas</disparador>
 <disparador>verde-amarelo</disparador>
 <disparador>vermelhos</disparador>
 <disparador>vida</disparador>

PONTUAÇÃO

<pont>!!</pont>	<pont>??</pont>
<pont>!!!</pont>	<pont>??</pont>
<pont>!!!!</pont>	<pont>...</pont>
<pont>!!!?</pont>	<pont>....</pont>
<pont>!!?</pont>	<pont>...</pont>
<pont>?!</pont>	<pont>.....</pont>

REPETIÇÃO

<rep>sei...sei</rep>
 <rep>eh, eh, eh</rep>

RISOS

<risos>Kkkk</risos>
 <risos>kkkk</risos>
 <risos>rs</risos>

QUESTÃO RETÓRICA

<questao_retorica>...aumento da moralidade?</questao_retorica>
 <questao_retorica>...aumento do embuste?</questao_retorica>
 <questao_retorica>...manutenção das mentiras?</questao_retorica>
 <questao_retorica>"já mais"? </questao_retorica>
 <questao_retorica>A culpa desse estado do Brasil não é só dela ? ;-)</questao_retorica>
 <questao_retorica>A presidenta não sabia de nada !!!?</questao_retorica>
 <questao_retorica>Acho q eles tão preocupados, milagre né?</questao_retorica>
 <questao_retorica>Afinal, o PTTT tirou o Color apesar de gostar muito dele, né?</questao_retorica>
 <questao_retorica>Aumentar impostos é fácil né?</questao_retorica>
 <questao_retorica>Bondade empresarial, é ?! </questao_retorica>
 <questao_retorica>Coincidência?</questao_retorica>
 <questao_retorica>Devemos ser grato ao P.T e seus eleitores, não é verdade?</questao_retorica>
 <questao_retorica>É fácil gastar o dinheiro dos outros né ?</questao_retorica>
 <questao_retorica>Eduardo Cunha é o Cara!!?</questao_retorica>
 <questao_retorica>Ela sabe o que é isso?</questao_retorica>

<questao_retorica>Eles são o que, deuses?</questao_retorica>
<questao_retorica>engraçado né?</questao_retorica>
<questao_retorica>Jura que ele eh inocente?</questao_retorica>
<questao_retorica>ki legal não e mesmo??</questao_retorica>
<questao_retorica>Liderança política; ela já teve isso?</questao_retorica>
<questao_retorica>Mas ele pode, né?</questao_retorica>
<questao_retorica>Melhor deixar tudo com está é?</questao_retorica>
<questao_retorica>Nada de irregular? kkkk</questao_retorica>
<questao_retorica>ninguém tem do dela não????..</questao_retorica>
<questao_retorica>O Márcio e o TFC estão falando da única e primeira crise vivida na Era</questao_retorica>
<questao_retorica>Petista, é ?!</questao_retorica>
<questao_retorica>O país esta em excelente fase, assim, se justifica (será que se justificaria em algum momento?) esses valores absurdos...mas recorrer e quem?</questao_retorica>
<questao_retorica>O Sr. quis dizer lambuzou foi a \$\$\$\$ que pegaram ?</questao_retorica>
<questao_retorica>Obrigado ministro, pelo menos, as idéias são pares, ne?</questao_retorica>
<questao_retorica>Ou continuam produzindo domingos sangrentos?</questao_retorica>
<questao_retorica>Ou de todas aquelas que vivemos na Era Tucana ?!</questao_retorica>
<questao_retorica>Ou é?</questao_retorica>
<questao_retorica>ou são funcionários públicos engajados pelas reciprocidades eleitoreiras?</questao_retorica>
<questao_retorica>Ou será sua ekipe os responsáveis por esta maravilhosa idéia de aumentar aalíquota do mais novo velho imposto?</questao_retorica>
<questao_retorica>Performance significa "controlar" a PF?</questao_retorica>
<questao_retorica>Pode ficar feliz a Oposição?!</questao_retorica>
<questao_retorica>Presto, no seu país tem democracia?</questao_retorica>
<questao_retorica>quanta amizade de qualidade vc tem né Lula...ta com medo de que?</questao_retorica>
<questao_retorica>Quem é mesmo esse coitado que tanto tempo fico no ministério ?</questao_retorica>
<questao_retorica>Quer mais ?</questao_retorica>
<questao_retorica>Só agora que você percebeu ?</questao_retorica><questao_retorica>Vai óleo de peroba?</questao_retorica>