



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS
DEPARTAMENTO DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

ARNOLDO NUNES DA SILVA

MODELO DE ANÁLISE DE SENTIMENTOS COM BASE NA ESTRUTURA
LINGÜÍSTICA DA SENTENÇA

FORTALEZA

2021

ARNOLDO NUNES DA SILVA

MODELO DE ANÁLISE DE SENTIMENTOS COM BASE NA ESTRUTURA LINGÜÍSTICA
DA SENTENÇA

Tese apresentada ao Programa de Pós-Graduação em Ciência da Computação do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do título de doutor em Ciência da Computação. Área de Concentração: Sistemas de Informação

Orientador: Prof. Dr. José Neuman de Souza

Coorientador: Prof. Dr. Osvaldo de Souza

FORTALEZA

2021

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- S578m Silva, Arnaldo Nunes da.
Modelo de Análise de Sentimentos com base na Estrutura Lingüística da Sentença / Arnaldo Nunes da Silva. – 2021.
97 f. : il. color.
- Tese (doutorado) – Universidade Federal do Ceará, Centro de Ciências, Programa de Pós-Graduação em Ciência da Computação , Fortaleza, 2021.
Orientação: Prof. Dr. José Neuman de Souza.
Coorientação: Prof. Dr. Osvaldo de Souza.
1. Análise de Sentimentos.. 2. Processamento de Linguagem Natural. 3. Linguística Computacional. I.
Título.

CDD 005

ARNOLDO NUNES DA SILVA

MODELO DE ANÁLISE DE SENTIMENTOS COM BASE NA ESTRUTURA LINGÜÍSTICA
DA SENTENÇA

Tese apresentada ao Programa de Pós-Graduação em Ciência da Computação do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do título de doutor em Ciência da Computação. Área de Concentração: Sistemas de Informação

Aprovada em: 29 de janeiro de 2021

BANCA EXAMINADORA

Prof. Dr. José Neuman de Souza (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Osvaldo de Souza (Coorientador)
Universidade Federal do Ceará (UFC)

Profa. Dra. Marília Soares Mendes
Universidade Federal do Ceará (UFC)

Prof. Dr. Miguel Franklin de Castro
Universidade Federal do Ceará (UFC)

Prof. Dr. João Fernando Lima Alcântara
Universidade Federal do Ceará (UFC)

Prof. Dra. Vlândia Célia Pinheiro
Universidade de Fortaleza (UNIFOR)

Prof. Dr. Ricardo Bezerra de Andrade e Silva
University College London (UCL)

Dedicado à minha família.

AGRADECIMENTOS

Agradeço aos meus orientadores, professores Dr. José Neuman de Souza e Dr. Osvaldo de Souza.

Aos membros da banca examinadora pelas sugestões de melhoria do trabalho.

Aos Departamentos de Ciência da Computação e de Ciências da Informação da Universidade Federal do Ceará.

Ao apoio dos meus familiares, especialmente mãe, pai, irmãos, irmãs, esposa e filha.

RESUMO

A Análise de Sentimentos (SA) é uma área do Processamento de Linguagem Natural que permite detectar a presença de polaridade positiva, negativa ou neutra em um texto. O SA tornou-se um elemento de maior interesse devido ao aumento da geração de dados a partir da internet e aos esforços necessários para processar esse volume de dados. Diante dessa demanda, busca-se obter resultados precisos que permitam a interpretação automática do sentimento inserido no texto. Vários métodos podem ser aplicados nesta tarefa, porém, nesta tese se destacam os aspectos e desafios em explorar a estrutura linguística da frase, para os quais as soluções são direcionadas a regras com as propriedades da gramática que descrevem a linguagem natural. Vale ressaltar que estudos de soluções baseadas em regras também são motivados pela exclusão de custos que envolvem treinamento de dados. Uma revisão do estado da arte, mostra que existem avanços nos estudos da linguística formal que contribuem para a linguística computacional, como regras de uma descrição sintática do português brasileiro já disponível na literatura, e não foram explorados para análise de sentimento. Em particular, trabalhos com modelos restritos de regras e o uso de parsers foram detectados para definir as classes gramaticais, a estrutura da sentença e as relações de dependência. No entanto, nenhuma solução foi encontrada envolvendo uma gramática construída para descrever uma linguagem natural incorporada a um parser que analisa a estrutura da frase caracterizada especificamente pelo sentimento. O principal resultado obtido com esta tese foi um novo modelo de análise de sentimento baseado em uma gramática regular expansível definida por regras de composição semântica. Assim, foi desenvolvido um parser que identifica estruturas de sentenças caracterizadas ou não com polaridade positiva ou negativa. Para atender a essa solução, um conjunto de relações de sentimento entre categorias gramaticais foi estudado e desenvolvido a partir da descrição formal da estrutura da frase. Foi implementado um protótipo para testar a aplicação do modelo em corpora de sentenças e posteriormente para uma avaliação comparativa dos resultados, que apresentaram índices nos mesmos níveis obtidos por outros métodos.

Palavras-chave: Análise de Sentimentos. Processamento de Linguagem Natural. Linguística Computacional

ABSTRACT

Sentiment Analysis (SA) is an area of Natural Language Processing that leads to detect the presence of positive, negative or neutral polarity in a text. SA became an element of greater interest due to the increase of data generation from the internet, and the efforts required to process such volume of data. Due to this demand, there is an endeavour to achieve precise results that will allow the sentiment inserted in the text to be automatically interpreted. Several methods can be applied in this task, however, in this thesis the aspects and challenges in exploring the linguistic structure of the sentence stand out, for which the solutions are directed to rules with the properties of the grammar that describe the natural language. It is worth noting that studies of solutions based on rules are also motivated by excluding costs that involve data training. A review of the state of the art, shows that there are advances in studies of formal linguistics that contribute to computational linguistics, as rules of a syntactic description of Brazilian Portuguese already available in literature, and have not been explored for sentiment analysis. In particular, works with restricted models of rules and the use of parsers were detected to define the parts of speech, sentence structure, and dependency relations. However, no solutions were found involving a grammar constructed to describe a natural language incorporated into a parser that analyzes the sentence structure specifically characterized by sentiment. The main result obtained from this thesis was a new model of sentiment analysis based on an expandable regular grammar defined by rules of semantic composition. Thus, a parser was developed that identifies sentence structures characterized or not with positive or negative polarity. To meet this solution, a set of relations of sentiment between grammatical categories were studied and developed based on the formal description of the sentence structure. A prototype was implemented to test the application of the model in corpora sentences and later for a comparative evaluation of the results, which presented rates at the same levels obtained by other methods.

Keywords: Sentiment Analysis. Natural Language Processing. Computational Linguistics

LISTA DE FIGURAS

Figura 1 – Diagrama de Abordagens para Análise de Sentimentos	22
Figura 2 – Estrutura em Árvore da Teoria X-Barra.	26
Figura 3 – Estrutura do Sintagma Conjuncional na Teoria X-Barra.	35
Figura 4 – Exemplo de estrutura do Sintagma Conjuncional com coordenação múltipla na Teoria X-Barra.	36
Figura 5 – Regras de inferência composicional motivadas pela semântica composicional	38
Figura 6 – Pseudo-código com regras propostas por (ZHANG <i>et al.</i> , 2009)	43
Figura 7 – Regras propostas por (KLENNER <i>et al.</i> , 2009)	44
Figura 8 – Resultados comparativos de desempenho.	44
Figura 9 – Regras trabalhas por (ROMANYSHYN, 2011)	45
Figura 10 – Árvore gerada pelo Pro3Gres	46
Figura 11 – Regra tratada pelo Aleph	46
Figura 12 – Exemplo das dependências tipadas exploradas para a composição	47
Figura 13 – Regras de composição semântica para o sistema MuSES	47
Figura 14 – Gráfico comparativo de desempenho para o Sistema Muses	48
Figura 15 – Regras trabalhas por (KIRITCHENKO; MOHAMMAD, 2016)	49
Figura 16 – Composições trabalhadas em (SANTOS <i>et al.</i> , 2016)	49
Figura 17 – Parte das Regras trabalhas por (TAN <i>et al.</i> , 2016)	50
Figura 18 – Regras trabalhas por (BIDULYA; BRUNOVA, 2017)	50
Figura 19 – Algoritmo com regras trabalhas por (RANI; KUMAR, 2018)	51
Figura 20 – Regras trabalhas por (TOLEDO-RONEN <i>et al.</i> , 2018)	52
Figura 21 – Regras propostas por (RAVISHANKAR; SHRIRAM, 2018)	53
Figura 22 – Diagrama do Sistema com o Parser de Sentimentos acoplado.	55
Figura 23 – Diferentes atributos para o mesmo termo	60
Figura 24 – Amostra do corpus de tuítes referentes a comentários em operadoras de telefonia e TV paga	71
Figura 25 – Amostra do corpus de tuítes dataset 2	72
Figura 26 – Saída para Sentença Simples	73
Figura 27 – Saída para Sentença Múltipla	74
Figura 28 – Saída para sentença com modificador	74
Figura 29 – Saída para sentença com conjunção adversativa	75

Figura 30 – Saída para sentença sarcasmo	76
Figura 31 – Análise Comparativa pelo Gráfico ROC	82
Figura 32 – Resultados de desempenho de dataset2	84

LISTA DE QUADROS

Quadro 1 – Modelos de Regras de Sentenças na Teoria X-Barra	27
Quadro 2 – Regras Gramaticais do Sintagma Flexional	29
Quadro 3 – Regras Gramaticais do Sintagma Complementizador	30
Quadro 4 – Regras Gramaticais do Sintagma Nominal	30
Quadro 5 – Regras Gramaticais do Sintagma Numeral	31
Quadro 6 – Regras Gramaticais do Sintagma Possessivo	31
Quadro 7 – Regras Gramaticais do Sintagma Quantificador	32
Quadro 8 – Regras Gramaticais do Sintagma Determinante	32
Quadro 9 – Regras Gramaticais do Sintagma Verbal	33
Quadro 10 – Regras Gramaticais do Sintagma Preposicional	34
Quadro 11 – Regras Gramaticais do Sintagma Adverbial	34
Quadro 12 – Regras Gramaticais do Sintagma Adjetival	35
Quadro 13 – Padrões de Etiquetagem	54
Quadro 14 – Regras de Exceção	56
Quadro 15 – Exemplos de polaridade e modificador para as Classes Gramaticais	57
Quadro 16 – Possíveis Composições aplicado à estrutura da Teoria X-Barra.	58
Quadro 17 – Regras de Sentimentos do Sintagma Flexional	60
Quadro 18 – Regras de Sentimentos do Sintagma Complementizador	61
Quadro 19 – Regras de Sentimentos do Sintagma Determinante	61
Quadro 20 – Regras de Sentimentos do Sintagma Nominal	61
Quadro 21 – Regras de Sentimentos do Sintagma Adjetival	62
Quadro 22 – Regras de Sentimentos do Sintagma Verbal	62
Quadro 23 – Regras de Sentimentos do Sintagma Preposicional	63
Quadro 24 – Regras de Sentimentos do Sintagma Numeral	63
Quadro 25 – Regras de Sentimentos do Sintagma Quantificador	63
Quadro 26 – Regras de Sentimentos do Sintagma Adverbial	64
Quadro 27 – Regras de Sentimentos do Sintagma Conjuncional	64
Quadro 28 – Regras de Sentimentos do Sintagma Possessivo	65
Quadro 29 – Composição entre elementos da mensagem.	66
Quadro 30 – Regras de composição na presença de conjunção adversativa	67
Quadro 31 – Exemplo de trechos no formato do léxico na gramática do parser	68

Quadro 32 – Interpretação para cada polaridade 79

LISTA DE TABELAS

Tabela 1 – Resultados da busca por trabalhos relacionados	40
Tabela 2 – Comparativo entre os trabalhos relacionados	42
Tabela 3 – Percentagem de cada tipo de erro	78
Tabela 4 – Matriz de confusão para o dataset 1	80
Tabela 5 – Desempenho do Paser de Sentimento para o dataset 1	80
Tabela 6 – Desempenho dos métodos aplicados ao Dataset 1	81
Tabela 7 – Matriz de confusão para o dataset 2	82
Tabela 8 – Métricas de desempenho para o dataset 2	83
Tabela 9 – Desempenho do Parser de Sentimento no Dataset 2	84

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Contextualização e Caracterização do Problema	15
1.2	Hipótese e Questões de Partida	18
1.3	Objetivos Geral e Específicos	18
1.4	Metodologia	19
1.5	Contribuições	19
1.6	Organização da Tese	19
2	FUNDAMENTAÇÃO TEÓRICA	20
2.1	Análise de Sentimento	20
<i>2.1.1</i>	<i>Abordagens de Classificação de Polaridade</i>	21
2.2	Formalismo Gramatical	24
2.3	Teoria X-Barra	26
<i>2.3.1</i>	<i>Regras Sintagmáticas do Português Brasileiro</i>	28
<i>2.3.1.1</i>	<i>Regras Gramaticais de Sentença</i>	29
<i>2.3.1.1.1</i>	<i>Sintagma Flexional</i>	29
<i>2.3.1.1.2</i>	<i>Sintagma Complementizador</i>	29
<i>2.3.1.2</i>	<i>Regras Gramaticais dos Sintagmas para Agrupamentos Nominais</i>	30
<i>2.3.1.2.1</i>	<i>Sintagma Nominal</i>	30
<i>2.3.1.2.2</i>	<i>Sintagma Numeral</i>	31
<i>2.3.1.2.3</i>	<i>Sintagma Possessivo</i>	31
<i>2.3.1.2.4</i>	<i>Sintagma Quantificador</i>	31
<i>2.3.1.2.5</i>	<i>Sintagma Determinante</i>	32
<i>2.3.1.3</i>	<i>Regras Gramaticais dos Sintagmas Verbais</i>	32
<i>2.3.1.4</i>	<i>Regras Gramaticais dos Sintagmas Preposicionais</i>	33
<i>2.3.1.5</i>	<i>Regras Gramaticais do Sintagma Adverbial</i>	34
<i>2.3.1.6</i>	<i>Regras Gramaticais do Sintagma Adjetival</i>	34
<i>2.3.1.7</i>	<i>Sintagma Conjuncional</i>	35
2.4	Conclusão do Capítulo	36
3	TRABALHOS RELACIONADOS	37
3.1	Breve Histórico	37

3.2	Revisão Sistemática	38
3.3	Resumo dos Trabalhos	43
3.4	Conclusão do Capítulo	53
4	PARSER PARA UM SISTEMA DE ANÁLISE DE SENTIMENTOS . .	54
4.1	Regras de Composição Básica	55
4.2	Regras de Exceção	56
4.3	Regras de Produção de Sentimentos	57
4.4	Procedimentos Externo ao Parser	65
4.4.1	<i>Pré-processamento</i>	65
4.4.2	<i>Composição entre Sentenças, Interjeições e Emoticons</i>	66
4.4.3	<i>Composição por Conjunções Adversativas</i>	66
4.5	Léxico de Sentimentos	66
4.6	Conclusão do Capítulo	68
5	EXPERIMENTOS E RESULTADOS	69
5.1	Datasets Utilizados	70
5.2	Casos de Sentenças	72
5.2.1	<i>Sentença simples</i>	72
5.2.2	<i>Múltiplas sentenças</i>	73
5.2.3	<i>Modificadores de Valência</i>	73
5.2.4	<i>Divisão por conjunção</i>	74
5.2.5	<i>Sarcasmo</i>	75
5.2.6	<i>Falsos Resultados</i>	76
5.3	Análise de Desempenho	78
5.3.1	<i>Experimento 1</i>	80
5.3.2	<i>Experimento 2</i>	82
5.4	Conclusão do Capítulo	84
6	CONCLUSÕES	85
6.1	Análise da hipótese e das questões de pesquisa	85
6.2	Resultados alcançados e principais contribuições	87
6.3	Limitações e Trabalhos futuros	88
	REFERÊNCIAS	89

1 INTRODUÇÃO

Esta tese propõe um modelo para análise de sentimentos que explore aspectos linguísticos da estrutura da sentença. Na seção 1.1, são apresentadas a contextualização e a caracterização do problema. A hipótese e as questões da pesquisa que guiaram o desenvolvimento deste trabalho são apresentadas na seção 1.2. Na seção 1.3, os objetivos deste trabalho são listados. Na seção 1.4, são descritas as etapas da metodologia utilizada para a elaboração desta tese. Na seção 1.5, são apresentadas as principais contribuições desta pesquisa. A seção 1.6 finaliza o capítulo apresentando a estrutura do restante desta tese.

1.1 Contextualização e Caracterização do Problema

A internet tem fornecido ferramentas que permitem a produção de conteúdos textuais por seus usuários. A presença de opiniões nestes textos ganhou importância por permitir extrair informações que auxiliem no julgamento da qualidade de uma entidade. Neste contexto, surge a área de análise de sentimentos, ou mineração de opinião, conceituado como o estudo computacional de opiniões, sentimentos e emoções expressas em texto (LIU, 2010). A medida do sentimento é chamada de polaridade ou orientação semântica (TURNEY, 2002) e é determinada com os valores positivo, negativo ou neutro.

Uma das abordagens exploradas envolve o uso de técnicas de Inteligência Artificial através do Aprendizado de Máquina. Este método visa criar classificadores a partir de um conjunto de frases ou documentos que podem estar ou não anotados com atributos de sentimentos. A qualidade dos dados de treinamento é fator fundamental para o desempenho daquela abordagem, este fator é apontado em Taboada *et al.* (2011a), que cita como exemplo os modificadores de valência (negação ou intensificação) de sentimento tratados em Kennedy e Inkpen (2006). A autora argumenta que, para treinar com precisão na aprendizagem de máquina, deve haver exemplos suficientes de todos os tipos de modificadores no corpus de treinamento. Outro fato que pode influenciar na redução do desempenho é a mudança de domínio, onde pode ser necessário um novo processo de treinamento (TABOADA, 2016; ARAÚJO *et al.*, 2014).

Evitar custos com treinamentos de dados é um fator que motiva tratar a análise de sentimentos através de outra abordagem que explore aspectos da estrutura linguística da sentença. Esta vertente é conhecida por análise baseada em léxico contendo palavras e expressões com suas respectivas orientações semânticas. Os dados deste recurso são usados para deduzir o sentimento

do texto em função das polaridades das palavras que o compõem. Esta abordagem também é chamada de análise baseada em regras, pois podem ser aplicadas regras que representam relações entre os constituintes da sentença com base na composição de seus sentidos. O modelo destas relações reflete a característica funcional do Princípio da Composicionalidade (DOWTY *et al.*, 1981), o qual afirma que o significado de um todo é uma função dos significados de suas partes, juntamente com a maneira pela qual essas partes foram combinadas.

O tratamento automatizado da estrutura da sentença está inserido na área da linguística computacional, na qual encontra-se o Processamento de Linguagem Natural (PLN), que se preocupa com o estudo da linguagem voltado para a construção de softwares, aplicativos e sistemas computacionais específicos (OTHERO, 2006). Entre estas aplicações, destaca-se o parser, definido por Satta (2010) como o processo de análise automática de uma sentença, sob a perspectiva de uma sequência de palavras, com o objetivo de determinar suas possíveis estruturas sintáticas. Desta forma, o uso do parser é usado para o processamento sintático e este procedimento, chamado de parsing, corresponde à interpretação automática (ou semi-automática) de sentenças de linguagem natural (OTHERO, 2008). Esta funcionalidade permite analisar uma sentença dizendo se a mesma é gramatical ou agramatical na língua, atribuindo a sua correta estrutura sintática.

O desenvolvimento de parser e suas aplicações nos diversos problemas de PLN requer uma cooperação entre a linguística e a informática, que pode esbarrar em diferentes objetivos. O informata busca atingir a robustez de seu sistema para que atenda textos caracterizados pela espontaneidade, enquanto que o linguista formal está preocupado com a competência linguística e o conhecimento do falante (OTHERO, 2008).

Por outro lado, existem pesquisadores que defendem a importância do formalismo sintático no desenvolvimento da linguística computacional (SIKKEL; NIJHOLT, 1993; BICK, 1996; GARSIDE *et al.*, 1997) e que Othero (2009) resume suas respectivas visões no trecho:

O nível de formalização sintática alcançado pelas pesquisas em Linguística tem proporcionado bons resultados não somente para pesquisas de cunho teórico envolvendo o estudo de línguas naturais, mas também para o desenvolvimento de aplicativos computacionais, como parsers, taggers, corretores ortográficos e gramaticais, tradutores automáticos, etc.

Seguindo os argumentos da importância de se estudar sobre a descrição e o processamento sintático de uma língua, um conjunto de regras formais de descrição sintática do Português Brasileiro (PB) foi proposto por Othero (2009) pensando em sua implementação computacional.

O trabalho do autor atendeu demandas dentro da linguística computacional que, até então, careciam para o PB de modo a permitir uma maior compatibilidade entre os conflitos de objetivos dos linguistas e dos informatas. Para entender como análise de sentimentos tem sido tratado via regras de composição, foi feito um levantamento de trabalhos relacionados, entre os quais foram encontradas propostas de modelos para dedução de polaridades em bigramas (KIRITCHENKO; MOHAMMAD, 2016; SANTOS *et al.*, 2016; TAN *et al.*, 2012; TOLEDO-RONEN *et al.*, 2018; Petrakis, S ; Klenner, 2011). No entanto, o principal foco desta tese está na produção manual de regras e sobre este aspecto foram observados os trabalhos Xie *et al.* (2014), Vilares *et al.* (2017), Neviarouskaya *et al.* (2015), Tan *et al.* (2016), Romanyshyn (2011), Bidulya e Brunova (2017), Rani e Kumar (2018), Zhang *et al.* (2009), Ravishankar e Shriram (2018), Klenner *et al.* (2009). Estas propostas adotam modelos para a execução de heurísticas após a sentença ser aplicada às ferramentas de etiquetagem morfosintática. As soluções apresentadas nestes trabalhos têm em comum o uso de suas regras incorporadas ao código, esta característica torna a adição de novas regras dependente de reimplementação e recompilação. Esta tese considera que a base de conhecimento seja integrada a um componente independente para permitir flexibilidade de expansão. Outro aspecto observado envolve o uso de ferramentas externas para o processo de etiquetagem e estruturação sintagmática. Neste ponto, os modelos de regras dos trabalhos citados devem observar a compatibilidade com o modo de descrição da estrutura da sentença utilizada por outras ferramentas, ficando assim dependentes de seus resultados. A pesquisa tratada nesta tese busca por solução que agregue as funcionalidades referentes ao processo de etiquetagem morfosintática e composição de sentimento, seguindo um modelo padrão de descrição formal da sentença, independente de ferramentas externas.

Os entendimentos e posições epistemológicas delineados até este ponto, e a percepção da oportunidade de uma contribuição para a área, particularmente na disponibilização de uma gramática formalizada para o PB e viável para aplicações em linguística computacional, motivou o desenvolvimento de uma proposta em trabalhar adaptações destas regras para outras linhas de pesquisa em PLN e, mais especificamente nesta tese, a análise de sentimentos baseada na estrutura linguística da sentença.

A tese apresentada neste documento propõe explorar o uso da estrutura linguística da sentença para identificar a sua orientação semântica, através de um conjunto de regras de composição de sentimentos, resultando em uma nova gramática, a qual permite o desenvolvimento de um parser específico para análise de sentimentos. A sua funcionalidade permite identificar o

sentimento da sentença a partir da validação da gramática definida para cada polaridade ou neutralidade. Além disso, é fundamental que o modelo proposto apresente desempenho satisfatório comparado a outras soluções de análise de sentimento.

1.2 Hipótese e Questões de Partida

Esta tese de doutorado busca investigar a seguinte hipótese:

É possível propor e desenvolver um modelo de análise de sentimentos com base em regras de composição semântica, agregadas a um analisador sintático (parser) e desenvolvidas com base nas descrições formais da linguística, apresentando desempenho nos mesmos patamares das taxas obtidas por outros métodos.

Partindo desta hipótese, foram levantadas três questões de pesquisa (QP):

QP1: Quais as demandas para o desenvolvimento de modelo de análise de sentimentos que permita tratar e identificar a estrutura da sentença com característica de sentimento?

QP2: Como adaptar as regras de composição semântica a um modelo de formalismo gramatical a ser agregado a um parser?

QP3: Qual o desempenho de um parser que agrega regras de composição de sentimentos a sua gramática em um processo de análise de sentimentos, comparado a outros métodos quando aplicados a um mesmo corpus?

1.3 Objetivos Geral e Específicos

O objetivo geral deste trabalho concentra-se em propor um modelo de análise de sentimentos utilizando as funcionalidades do analisador sintático automático, agregando regras de composição semântica. Enquanto que os objetivos específicos a serem alcançados ao longo do desenvolvimento são listados abaixo:

1. Identificar características e demandas de um modelo baseado em regras de composição para a análise de sentimentos.
2. Desenvolver um modelo de regras a ser aplicado na definição da composição semântica associada às categorias gramaticais.
3. Implementar protótipo do modelo proposto e aplicar em corpus anotados para análise de sentimentos
4. Avaliar o desempenho do modelo proposto comparado a outros métodos quando aplicados

a um mesmo corpus

1.4 Metodologia

O desenvolvimento do presente trabalho foi realizado de acordo com as seguintes etapas metodológicas:

1. Estudo inicial de conteúdos gerais e específicos, referentes às áreas de Análise de Sentimentos e Formalismo Gramatical;
2. Realização de revisão bibliográfica e levantamento de trabalhos relacionados acerca dos tópicos identificados na etapa anterior;
3. Construção de uma gramática adaptada para sentenças com orientação semântica;
4. Proposta de um modelo para análise de sentimentos com base em regras de composição semântica;
5. Análise de sentenças a partir de corpora com base na gramática e modelo propostos;
6. Implementação de protótipo, testes, aplicação e avaliação dos resultados.

1.5 Contribuições

1. Estudos de padronização de composição entre orientações semânticas de classes gramaticais;
2. Modelo de gramática para um parser sintático de sentimentos;
3. Desenvolvimento e testes de um protótipo de parser sintático para análise de sentimentos.

1.6 Organização da Tese

Este capítulo contextualizou o presente trabalho, incluindo as motivações para o mesmo, a caracterização do problema abordado, os objetivos estabelecidos, as etapas metodológicas procedidas e as contribuições esperadas. O restante desta tese encontra-se organizado da seguinte forma: no Capítulo 2 é apresentada a fundamentação teórica; no Capítulo 3 são abordados os trabalhos relacionados ao escopo da tese; no Capítulo 4, são definidos o modelo para a solução proposta e aspectos de implementação do protótipo; o Capítulo 5 apresenta experimentos e avaliação dos resultados junto a um comparativo com outros métodos e no Capítulo 6, são apresentadas as conclusões sobre este trabalho, bem como limitações observadas, principais resultados alcançados e possíveis trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Análise de Sentimento

A análise do sentimento (também chamada de mineração de opinião) tem a tarefa de detectar, extrair e classificar opiniões, sentimentos e atitudes em relação a diferentes tópicos, conforme expresso em um conteúdo textual (MONTOMOYO *et al.*, 2012). Esta classificação, também denominada de orientação semântica, pode dividir-se em polaridades positiva e negativa ou quando na ausência destas, atribuindo-se o valor neutro. Também podem ser tratados sentimentos mais específicos, como alegria, tristeza, ódio, entre outros que possam caracterizar o sentimento expresso em um conteúdo.

Dependendo do contexto e da aplicação, estas tarefas podem ocorrer em nível de documento, sentença e aspecto (LIU, 2012). No nível de documento, a análise é feita no conteúdo como um todo, composto por sentenças que tratam de uma única entidade. O nível de sentença restringe a analisar uma frase específica, podendo ser dividida em cláusulas. A análise no nível de aspecto e entidade concentra nos elementos que estão sendo julgados, neste caso, a polaridade não se refere a toda uma sentença ou documento, pois cada entidade ou aspecto poderá apresentar classificações diferentes.

A análise de sentimento pode ser dividida em três grandes tarefas (TSYTSAU; PALPANAS, 2012): a) identificar (tópicos, sentenças opinativas); b) classificar a polaridade do sentimento; e c) sumarizar. A identificação consiste na busca por conteúdo subjetivo para posteriormente avaliar sua polaridade. Dependendo do propósito da análise, faz-se necessário identificar entidades que são alvo e seus respectivos aspectos relacionados à subjetividade expressa. O problema de classificação de polaridade visa caracterizar o sentimento, podendo ter divisão binária (positiva ou negativa) ou com uma terceira classe neutra adicionada para casos em que não seja detectado sentimento. A classificação também pode ser subdividida com variações de intensidade entre os extremos das classes binárias. A etapa de sumarização agrupa e representa os resultados obtidos através das análises realizadas anteriormente, ou seja, na presença de um conjunto de opiniões diversificadas, métricas são aplicadas para estabelecer uma opinião média ou prevalente.

No processo de análise, é necessário definir as relações de influência das classes gramaticais na classificação do sentimento. Uma observação relevante é mostrada em Taboada (2016), a qual afirma que a maior parte da pesquisa na análise de sentimentos centrou-se na

natureza avaliativa dos adjetivos. No entanto, trabalhos anteriores, como Benamara *et al.* (2007) e Reforgiato e Subrahmanian (2008), notaram que uma grande quantidade de sentimento é transmitida através de outras classes gramaticais, como substantivos, verbos ou advérbios e frases que contêm essas palavras.

O sentimento transmitido pelas classes gramaticais também pode sofrer modificações de outros termos, como intensificações, atenuações e reversões. Termos que possuem essas características se enquadram na categoria denominada por Polanyi *et al.* (2006) de modificadores de valência. Levar em consideração os modificadores têm demonstrado melhoraria no desempenho dos sistemas de análise de sentimento (ALBORNOZ; PLAZA, 2013) (KENNEDY; INKPEN, 2006), (MORSY; RAFA, 2012), (TABOADA *et al.*, 2011a).

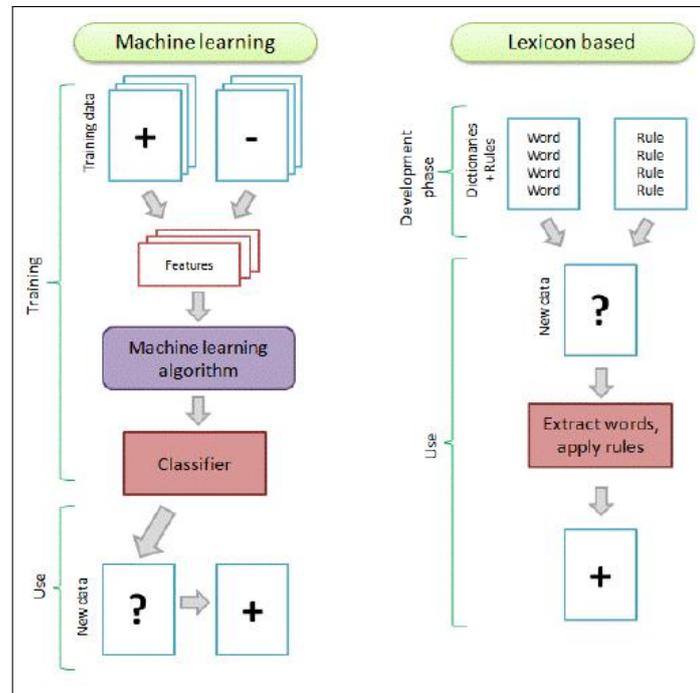
O efeito do modificador de valência pode ser tratado por diferentes abordagens como em Kenedy (2016) e Polanyi *et al.* (2006), que propuseram a sua contabilização usando simples adição e subtração, ou seja, um adjetivo positivo com valor de 2, se intensificado, tornaria-se 3, já em caso de atenuação, reduziria o valor para 1. Em abordagens que trabalham apenas com os extremos da polaridade (positivo e negativo), sem considerar intervalos numéricos, a negação apenas inverte a orientação semântica para a sua oposta. A reversão de polaridade geralmente envolve encontrar uma indicação de negação, como por exemplo em advérbios e locuções adverbiais de negação ("não", "nunca", "jamais", "de jeito nenhum"). Além do advérbio, outras classes gramaticais devem ser consideradas, como nas palavras "falta"(substantivo), "faltar", "curar"(verbo), "sem", "contra"(preposição). As descrições da negação e seu escopo, e como ela pode ser identificada computacionalmente, podem ser encontradas em Sauri (2008), BLANCO e MOLDOVAN (2014).

2.1.1 Abordagens de Classificação de Polaridade

Existem duas principais abordagens para o problema de extração de sentimentos em textos. A primeira baseia-se nos conceitos de aprendizado de máquina e divide-se em supervisionada e não-supervisionada. A segunda abordagem é chamada de baseada em léxico, sendo também referida como baseada em regras. O diagrama mostrado na Figura 1, desenvolvido por (TABOADA, 2016), ilustra os procedimentos de funcionamento das referidas abordagens.

Os algoritmos de Aprendizado de Máquina buscam por padrões dentro de um conjunto de dados para aprender um determinado comportamento. A aplicação desta abordagem ocorre ao considerar a análise de sentimentos como problema de classificação que determina se

Figura 1 – Diagrama de Abordagens para Análise de Sentimentos



Fonte: (TABOADA, 2016)

um sentimento é expresso como positivo ou negativo (SUN *et al.*, 2017a).

A supervisionada busca características que permitam distinguir sentenças com diferentes sentimentos a partir de um processo de treinamento aplicado em um conjunto sentenças previamente rotuladas. Desta forma, obtém-se um modelo a ser usado para identificar o sentimento em novas sentenças. O aprendizado não-supervisionado busca identificar padrões em conjunto de dados não rotulados a partir de algoritmos que fazem o agrupamento das instâncias por meio de medidas de similaridade. Dentre os algoritmos de classificação mais usados nesta área estão Support Vector Machine, Naïve Bayes, Maximum Entropy e algoritmos baseados em redes neurais (YANG; CHEN, 2017; ABIRAMI; GAYATHRI, 2017).

Uma das grandes limitações no uso de aprendizado supervisionado para definição de polaridade é a necessidade de dados rotulados para treino (SUN *et al.*, 2017b). O desempenho destes métodos é afetado não somente pela quantidade, mas igualmente pela qualidade dos dados disponíveis. Outro fator a ser destacado é que dados são treinados para domínios específicos, geralmente sem portabilidade para novos tipos. De acordo com Taboada (2016), a aplicação do modelo a novos contextos e conjuntos de dados normalmente exigirá novos dados de treinamento e, portanto, extensa codificação humana.

A segunda abordagem pode ser chamada de baseada em léxico, sendo também referida como baseada em regras (TABOADA, 2016). Nesta abordagem, um conjunto de regras

é usado para derivar o sentimento do texto a partir das polaridades individuais de suas palavras encontradas em um dicionário de termos rotuladas.

Um léxico de sentimento, em sua estrutura básica, contém a palavra de sentimento e sua respectiva polaridade, expressa como uma categoria ou como um valor em uma escala. No entanto, o léxico pode ser construído com o acréscimo de outros atributos, como: o lema de cada entrada; a categoria gramatical e o alvo do sentimento.

A maioria dos léxicos existentes são dependentes de idioma e foram feitos estritamente para a língua inglesa, como se vê em SentiWordNet (BACCIANELLA *et al.*, 2010) e WordNetAffect (VALITUTTI, 2004). Já para a língua portuguesa estão disponíveis o OpLexicon (SOUZA; VIEIRA, 2011) e o SentiLex-PT (CARVALHO; SILVA, 2015), sendo o primeiro para português do Brasil e o último, para português de Portugal. Outro exemplo é o Linguistic Inquiry and Word Counts (LIWC) (SOUZA *et al.*, 2011), que fornece um software de análise de texto desenvolvido para avaliar os componentes estruturais, cognitivos e emocionais de amostras de texto, sendo que essa análise pode ser feita em vários idiomas disponibilizados na ferramenta.

Uma questão importante, independentemente das partes do discurso que está sendo considerada, é a cobertura dos dicionários do sentimento. É difícil estimar quão grande é o léxico avaliativo de uma língua. Não está claro qual é o tamanho ideal, ou se um idioma pode conter dezenas de milhares de termos avaliativos. Taboada *et al.* (2011a) descobriu que um grande dicionário tende a capturar mais ruído, levando a resultados imprecisos na extração automática do sentimento.

A abordagem baseada em léxico é dividida, conforme Medhat *et al.* (2014), em (i) baseada em dicionário, que busca palavras geradoras de opinião com o auxílio de dicionário de sinônimos e antônimos; (ii) baseada em corpus, que parte de uma lista de palavras sementes com propriedades de sentimento para encontrar em um grande corpus outras palavras com estas propriedades.

Observa-se também que a esquematização apresentada na Figura 1 insere um componente referente a regras na abordagem baseada em léxico, o que leva também a ser chamado de método baseado em regras. Prabowo e Thelwall (2009) definem regra como uma relação entre um antecedente e o seu respectivo conseqüente, normalmente representada por "if-then". O antecedente define a condição e trata-se de um token ou de um conjunto de tokens separados por operador lógico. O conseqüente representa um sentimento que é positivo ou negativo e é o resultado de atender à condição definida pelo antecedente.

Algoritmos que tratam as relações intrínsecas ao modelo de regras, como soluções de parser sintáticos, por exemplo, podem ser utilizados a partir de um modelo de regras associado à gramática. No contexto da linguagem natural, as relações entre as classes gramaticais podem ser trabalhadas e associadas entre si para uma interpretação semântica de toda uma sentença. Esta abordagem é a base do Princípio de Composicionalidade (MONTAGUE, 1970), onde o significado de uma sentença deve ser função do significado de seus constituintes e seus relacionamentos com base em regra sintática. Assim, Dowty (1983) definiu que o significado do todo é resultado da aplicação dos significados dos constituintes imediatos por meio de uma operação semântica que corresponde diretamente à operação sintática relevante.

2.2 Formalismo Gramatical

A gramática de uma língua é um esquema para especificar as sentenças permitidas, indicando a regra sintática para combinar palavras em frases e cláusulas (BARR; FEIGENBAUM, 1981). Em programas de PLN a gramática é utilizada para analisar sentenças e ajudar a determinar seus significados a fim de obter uma resposta apropriada.

O formalismo gramatical pode ser classificado conforme Chomsky (1959) através de 4 níveis hierárquicos de gramáticas. Cada nível é um super conjunto do próximo, ou seja, uma gramática do tipo n é consequentemente uma linguagem do nível $n-1$.

- Tipo 0: Este tipo de gramática utiliza regras irrestritas. Ambos os lados das regras de reescrita podem ter qualquer número de símbolos terminais e não terminais. Estas gramáticas são equivalentes a máquina de Turing em seu poder expressivo.
- Tipo 1: Também conhecida como Gramática Sensível ao Contexto (GSC). A diferença da GSC para a gramática do tipo 0 é que o número de símbolos no lado direito de uma regra na GSC deve ser maior ou igual ao número de símbolos no lado esquerdo da regra e o símbolo inicial não pode aparecer no lado direito de uma regra de produção.
- Tipo 2: Chamada Gramática Livre de Contexto (GLC), é a mais popular para gramáticas de linguagem natural. A restrição é que cada regra deve ter somente um símbolo não terminal no lado esquerdo, ou seja, $S \rightarrow w$, onde S é um símbolo não terminal e w é uma cadeia composta de terminais e não terminais.
- Tipo 3: Conhecida como Gramática Regular (GR), onde toda regra de produção é da forma: $X \rightarrow aY$ ou $X \rightarrow a$. X e Y são símbolos não terminais e a é um terminal.

Uma linguagem gerada por uma GLC é classificada como “livre do contexto”, ou

seja, uma linguagem formal cujas regras de produção podem ser aplicadas independentemente do contexto do símbolo não terminal. Não importa quais símbolos existem na GLC, um único símbolo não terminal existente no lado esquerdo de uma regra pode sempre ser substituído pelo lado direito. Na linguagem natural, os não terminais são representados por constituintes de uma sentença chamados sintagmas ou categorias gramaticais, já os terminais são as palavras da língua em questão. Este modo de análise de sentença é conhecido por Gramáticas de Constituintes Imediato (PSG - Phrase Structure Grammar)(GAZDAR, 1982), cuja estrutura sintática das frases é composta em termos de seus constituintes.

Os trabalhos de Chomsky passaram a ter aplicações não só na linguística, especificamente na descrição da sintaxe das línguas naturais, mas também na ciência da computação, notadamente na construção de compiladores (ALENCAR, 2012). Neste contexto, destaca-se o procedimento de análise sintática (parsing), como sendo o processo de estruturar uma representação linear de acordo com uma dada gramática. O algoritmo, chamado de parser (BARR; FEIGENBAUM, 1981), tem a finalidade de automatizar este processo, utilizando o léxico e a gramática do sistema e, recebendo como entrada a sentença dividida em tokens de palavras. Os parsers podem ser divididos em Constituência e Dependência (VARGAS, 2015). O parser por Constituição consiste na ideia de que grupos de palavras dentro de declarações podem ser representados agindo como unidades individuais. Já o parser por Dependência faz uso da predição de palavras, onde cada palavra na frase tem relação com seus dependentes. Para auxiliar no processo de parsing, é necessário identificar a categoria gramatical de cada palavra, este procedimento é função de uma ferramenta chamada Etiquetador (*tagger*) Morfossintático ou *part-of-speech* (POS) (BICK, 1996).

A tarefa do analisador é essencialmente determinar se e como a entrada pode ser derivada do símbolo inicial da gramática. Isso pode ser feito basicamente de duas maneiras: Análise de cima para baixo (*top-down*), que pode ser vista como uma tentativa de encontrar as derivações mais à esquerda de um fluxo de entrada, procurando por árvores de análise usando uma expansão de cima para baixo das regras gramaticais formais fornecidas. Na análise de baixo para cima (*bottom-up*), um analisador pode começar com a entrada e tentar reescrevê-la no símbolo de início. Intuitivamente, o analisador tenta localizar os elementos mais básicos, depois os elementos que os contêm e assim por diante.

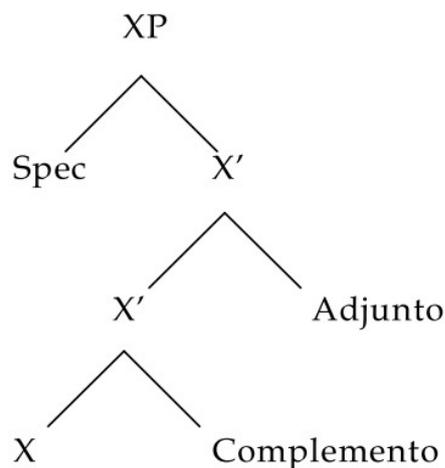
2.3 Teoria X-Barra

Afim de identificar as estruturas semelhantes entre as linguagens, Noam Chomsky propôs a Teoria X-Bar, que trata da organização sintática estrutural de constituintes da sentença em línguas naturais. Conforme Radford (1988), este modelo reduziu a quantidade de regras de estrutura sintagmática em alguns esquemas de aplicação universal.

O modelo adotado para a Teoria X-Barra representa a sentença em uma estrutura hierárquica arbórea, através da posição dos componentes: Projeção máxima de um sintagma (XP), composta pelo nó que representa uma projeção intermediária (X'), podendo ter um nó irmão Spec ocupado pelo sujeito da frase ou algum outro elemento modificador externo. A projeção Intermediária X' poderá ser composta de duas formas: (i) um nó representando o núcleo do sintagma (X), isolado ou acompanhado de um nó irmão na função de Complemento; (ii) uma projeção intermediária com um nó irmão na função de Adjunto.

A Figura 2 ilustra a estrutura arbórea do modelo, considerando a presença de todos os elementos e para refletir que os complementos costumam estar mais próximos do núcleo do que os adjuntos. Sendo que, os nós irmãos de cada par podem apresentar-se em qualquer ordem na construção da sentença. A posição dos elementos permite a compreensão semântica e gramatical conforme a linguagem adotada para o modelo.

Figura 2 – Estrutura em Árvore da Teoria X-Barra.



Fonte: (OTHERO, 2009)

O Quadro 1 mostra modelos de regras, conforme a estrutura da Teoria X-Barra, observando as seguintes possibilidades: a projeção máxima XP pode conter, ou não, um nó na

posições de especificador; a projeção intermediária X' pode apresentar nós que consistem em um núcleo e zero ou mais complementos, e em outro tipo, pode conter X' acompanhado de adjunto. Ou seja, os complementos geralmente estão mais próximos do núcleo do que os adjuntos.

Quadro 1 – Modelos de Regras de Sentenças na Teoria X-Barra

$XP \rightarrow X'$
 $XP \rightarrow \text{Spec } X'$
 $XP \rightarrow X' \text{ Spec}$
 $X' \rightarrow X$
 $X' \rightarrow X' \text{ Adjunto}$
 $X' \rightarrow \text{Adjunto } X'$
 $X' \rightarrow X \text{ Complemento}$
 $X' \rightarrow \text{Complemento } X$

Fonte: do autor

Nesse modelo, as regras de produção podem ter um ou dois símbolos não terminais. Da estrutura do modelo em árvore, observa-se regras com dois símbolos não terminais: $XP \rightarrow \text{Spec } X'$; $X' \rightarrow X' \text{ Comp}$ and $X' \rightarrow X \text{ Adjunto}$. No entanto, regras com apenas um símbolo não terminal podem ocorrer quando um nível intermediário se expande em um núcleo lexical ($X' \rightarrow X$) e a projeção máxima se expande para um nível intermediário sem especificador ($XP \rightarrow X'$) ou dois símbolos não terminais.

A representação dos constituintes na estrutura da Teoria X-barra traz algumas vantagens em relação a um PSG simples, pois conforme afirma Othero (2009), prevê corretamente uma organização hierárquica maior que X e menor que XP e faz uma distinção sintática entre complementos e adjuntos.

A Teoria X-Bar estabelece princípios, listados abaixo, a serem seguidos nos modelos de composição sintática entre os componentes da regra.

- **Endocentricidade:** é o princípio que garante que cada constituinte XP tenha o núcleo X ; todo e qualquer sintagma tem um único núcleo e cada núcleo projeta um sintagma. Todos os sintagmas são encabeçados por um núcleo X de mesma natureza que a projeção máxima.
- **Lexicalidade:** postula que cada elemento não terminal é a projeção de um elemento terminal, garantindo que cada elemento XP ou X' tenha seu núcleo X .
- **Sucessão:** diz respeito à relação de dominância entre os elementos constituintes da frase. O elemento X_n , ou X_{n+1} , pode dominar um outro elemento X_n , podendo ocorrer

em sintagmas verbais com mais de um complemento, ou com mais de um modificador adverbial.

- **Uniformidade:** garante que todas as projeções máximas tenham o mesmo número de barras, neste caso o valor 2. Assim, cada regra de produção poderá ter um ou dois elementos no lado direito.
- **Maximalidade:** postula que todos os elementos funcionais também devem ser projeções máximas;
- **Binaridade:** regula a organização esquemática da Teoria X barra, onde as ramificações estruturais na árvore sintática são sempre binárias.

Othero (2009) propôs um modelo de gramática com regras formais de descrição sintática do português. Apesar de o autor deixar claro que estas "regras não serão capazes de descrever todas as estruturas dos sintagmas em português", seu trabalho mostra capacidade de atender uma ampla gama de sentenças na língua em estudo. Esta gramática não tratou de maneira formal a ocorrência de conjunções, porém, existem formalismos para a consideração de sintagmas conjuncionais (OLIVEIRA, 2011).

A seção a seguir descreve um resumo das regras sintagmáticas tomadas como base para o desenvolvimento da gramática utilizada de um parser de sentimentos.

2.3.1 Regras Sintagmáticas do Português Brasileiro

A estrutura de uma sentença é construída e validada com base em regras gramaticais definidas para a língua em questão. As unidades sintagmáticas e suas relações permitem que a sentença tenha uma boa formação para o estabelecimento da comunicação. Esta seção apresenta um conjunto de regras para o PB modeladas seguindo os princípios e conceitos da Teoria X-barras de modo compatível para a implementação, codificação e execução de um parser automático. A gramática apresentada está baseada no trabalho de (OTHERO, 2009) que foi desenvolvida para abranger uma maior cobertura de sentenças e maior flexibilidade.

A descrição da gramática segue dividida nas seções que abordam as projeções lexicais máximas representadas por sintagmas para agrupamentos nominais, sintagma adjetival, sintagma adverbial e sintagma preposicional. Posteriormente são apresentados as regras para análise da estrutura da sentença representadas pelo sintagma flexional e sintagma complementizador. Por fim, apresenta-se a descrição da solução para trabalhar a ocorrência de conjunções, com a inclusão do sintagma conjuncional (ConjP - Conjunctive Phrase) que ficou ausente na descrição

da gramática original.

2.3.1.1 Regras Gramaticais de Sentença

Esta subseção apresenta a gramática para a descrição sintática da sentença em português brasileiro proposta por (OTHERO, 2009) com base no formalismo linguístico e, ao mesmo tempo, contribuindo para a linguística computacional em seus aspectos de implementação.

2.3.1.1.1 Sintagma Flexional

O Sintagma Flexional (IP - do inglês Inflexion Phrase)(Quadro 2) é a projeção máxima cujo núcleo I ocupando a posição do verbo auxiliar (andar, estar, haver, ir, ser, ter, vir) na ocorrência de sequências verbais: verbo auxiliar + verbo pleno, por exemplo (2.1). Em sentenças com apenas verbos plenos, o núcleo é considerado como elemento vazio: (2.2) e (2.3). Este sintagma tem a função de representar a estrutura principal da sentença, ocupando a raiz da árvore.

(2.1) Ele vai sair.

(2.2) Ele quer um livro.

(2.3) Ele quer sair.

Quadro 2 – Regras Gramaticais do Sintagma Flexional

IP → DP I'
IP → I'
I' → I VP
I' → I IP

Fonte: (OTHERO, 2009)

2.3.1.1.2 Sintagma Complementizador

O Sintagma Complementizador (CP do inglês Complementizer Phrase)(Quadro 3) tem a posição de núcleo C ocupado pelos conectores (que, quando, se) em orações subordinadas, exemplificado nas sentenças: (2.4); (2.5) e (2.6).

(2.4) Ele dirá que chegou.

(2.5) Ele dirá quando chegou.

(2.6) Ele dirá se chegou.

Quadro 3 – Regras Gramaticais do Sintagma Complementizador

CP → C'
C' → C IP

Fonte: (OTHERO, 2009)

2.3.1.2 Regras Gramaticais dos Sintagmas para Agrupamentos Nominais

O agrupamento nominal distribui-se nos sintagmas nominal, determinante, quantificador e numeral. Esta divisão surge para que as regras cumpra os princípios da Teoria X-Barra, pois uma proposta inicial considerava apenas um sintagma nominal para tratar substantivos, que incluía em suas regras os pronomes pessoais, demonstrativos e possessivos, os numerais, artigos e elementos-QU interrogativos, ocupando funções de determinantes e pré-determinantes. As subseções abaixo apresentam a descrição dos referidos sintagmas.

2.3.1.2.1 Sintagma Nominal

O Sintagma Nominal (NP - do inglês *Noun Phrase*)(Quadro 4) tem como núcleo um substantivo. Em suas composições, podem ter o sintagma adjetival como adjunto (2.7) e os sintagmas preposicional e complementizador atuando como adjunto (2.8)(2.9) e também como complemento (2.8)(2.10).

(2.7) NP (N' (**AP**(velho) **N'**(**N'**(carro) **AP**(azul)))).

(2.8) NP (N' (**N'** (**N** (jogador) **PP** (de futebol)) **PP** (do estado))).

(2.9) NP (N' (**N'** (**N**(time)) **CP** (que ganhou))).

(2.10) NP (N' (**N**(time) **CP**(que ele jogou))).

Quadro 4 – Regras Gramaticais do Sintagma Nominal

NP → N'
N' → N
N' → N' AP
N' → AP N'
N' → N' PP
N' → N PP
N' → N' CP
N' → N CP

Fonte: (OTHERO, 2009)

2.3.1.2.2 Sintagma Numeral

O Sintagma Numeral (NumP - do inglês *Numeral Phrase*)(Quadro 5) representa constituintes cujo núcleo é expresso por um numeral, com regras podendo conter sintagmas nominal e preposicional na posição de complementos (2.11).

(2.11) NumP (**Num**(dois) **PP**(dos NumP (**Num**(cinco) **NP**(livros))))).

Quadro 5 – Regras Gramaticais do Sintagma Numeral

NumP → Num' Num' → Num NP Num' → Num PP

Fonte: (OTHERO, 2009)

2.3.1.2.3 Sintagma Possessivo

O Sintagma Possessivo (PossP - do inglês *Possessive Phrase*)(Quadro 6) tem como núcleo o pronome possessivo, com o domínio sobre os sintagmas nominal (2.12) e numeral (2.13) na posição de complementos.

(2.12) PossP (Poss'(**Poss**(meu) **NP**(livro)))

(2.13) PossP (Poss'(**Poss**(meus) **NumP**(dois livros)))

Quadro 6 – Regras Gramaticais do Sintagma Possessivo

PossP → Poss' Poss' → Poss NP Poss' → NP Poss Poss' → Poss NumP
--

Fonte: (OTHERO, 2009)

2.3.1.2.4 Sintagma Quantificador

O Sintagma Quantificador (QP - do inglês *Quantifier Phrase*)(Quadro 7) tem como núcleo um elemento representativo da quantidade. O quantificador é conceituado em (OLIVEIRA *et al.*, 2010) como "um termo cunhado pela semântica formal, a fim de referir-se a um conjunto de itens que expressam contraste ou quantidade como: todo, cada, algum". As regras possuem como complementos os sintagmas: determinante (2.14), preposicional (2.15). Também pode ocorrer sintagma adverbial (2.16) na posição de adjunto.

(2.14) QP(Q'(**Q**(todos) **DP**(os livros)))

(2.15) $QP(Q'(Q(\text{alguns}) PP(\text{dos livros})))$

(2.16) $QP(Q'(AdvP(\text{praticamente}) Q'(\text{todos os alunos chegaram})))$

Quadro 7 – Regras Gramaticais do Sintagma Quantificador

$QP \rightarrow Q'$ $Q' \rightarrow Q DP$ $Q' \rightarrow Q PP$ $Q' \rightarrow DP Q$ $Q' \rightarrow AdvP Q'$
--

Fonte: (OTHERO, 2009)

2.3.1.2.5 Sintagma Determinante

Seguindo os princípios da Teoria X-Barra, o NP passou a ser entendido como uma projeção do determinante. Assim, sua descrição é feita através da projeção máxima do Sintagma Determinante (DP - do inglês *Determiner Phrase*)(Quadro 8), cujo núcleo é um determinante D, que pode ser ocupado por pronomes demonstrativos, pronomes pessoais, artigos, quantificadores, elementos-QU ou por um elemento vazio. A posição de complemento pode ser ocupada pelos sintagmas: nominal (2.17), possessivo (2.18) e numeral (2.19). Também pode conter o sintagma adverbial (2.20) na posição de adjunto.

(2.17) $DP(D'(D(o) NP(\text{livro})))$

(2.18) $DP(D'(D(o) PossP(\text{meu livro})))$

(2.19) $DP(D'(D(os) NumP(\text{dois livros})))$

(2.20) $DP(AdvP(\text{provavelmente}) D'(\text{o carro quebrou}))$

Quadro 8 – Regras Gramaticais do Sintagma Determinante

$DP \rightarrow D'$ $D' \rightarrow D$ $D' \rightarrow D NP$ $D' \rightarrow D PossP$ $D' \rightarrow D NumP$ $D' \rightarrow AdvP D'$

Fonte: (OTHERO, 2009)

2.3.1.3 Regras Gramaticais dos Sintagmas Verbais

O Sintagma Verbal (VP - do inglês *verbal phrase*)(Quadro 9) tem como núcleo um verbo e pode ser composto em sua forma básica por apenas um verbo (2.21). Podem conter adjun-

tos ocupados por sintagmas: adverbiais (2.22)(2.23) e determinantes (2.24). Também apresentam a ocorrência de complementos representados por sintagmas: Adjetivais (2.25); Adverbiais (2.26); Determinante (2.27); Preposicionais (2.28); complementizador (2.29); flexionais (2.30).

(2.21) Ele VP(V'(V(chegou)))

(2.22) Ele VP(V'(AdvP(já) V'(chegou)))

(2.23) Ele VP(V'(V'(chegou) AdvP(bem)))

(2.24) Ele VP(V'(V'(gosta) PP(de livros)))

(2.25) Ele VP(V'(V(chegou) AP(cansado)))

(2.26) Ele VP(V'(V(estava) AdvP(ali)))

(2.27) Ele VP(V'(V(leu) DP(o livro)))

(2.28) Ele VP(V'(V(chegou) PP(em casa)))

(2.29) Ele VP(V'(V(disse) CP(que tentou uma fuga)))

(2.30) Ele VP(V'(V(tentou) IP(chegar)))

Quadro 9 – Regras Gramaticais do Sintagma Verbal

VP → V'
V' → V
V' → AdvP V'
V' → V' AdvP
V' → V' PP
V' → V AP
V' → V AdvP
V' → V DP
V' → V PP
V' → V CP
V' → V IP

Fonte: (OTHERO, 2009)

2.3.1.4 Regras Gramaticais dos Sintagmas Preposicionais

O sintagma preposicional (PP - do inglês *Prepositional Phrase*)(Quadro 10) tem como núcleo a preposição, podendo ter como complemento os sintagmas: determinante (2.31); adverbial (2.32); preposicional (2.33) e complementizador (2.34). Também poderá ocorrer sintagmas adverbiais na posição de adjunto (2.35).

(2.31) Ele está PP(P'(P(em) DP(casa)))

(2.32) Ele está PP(P'(P(por) AdvP(aqui)))

(2.33) Ele foi PP(P'(P(para) PP(fora de casa)))

(2.34) Ele gosta PP(P'(P(de) CP(ler livros)))

(2.35) Isto foi PP(P'(Advp(totalmente) P'(sem sentido)))

Quadro 10 – Regras Gramaticais do Sintagma Preposicional

PP → P'
P' → P
P' → P DP
P' → P AdvP
P' → P CP
P' → P PP
P' → AdvP P'

Fonte: (OTHERO, 2009)

2.3.1.5 Regras Gramaticais do Sintagma Adverbial

O Sintagma Adverbial (AdvP - do inglês *Adverbial Phrase*)(Quadro 11) tem como núcleo o advérbio, podendo este ser intransitivo tendo como adjunto um sintagma adverbial (2.36) ou transitivo, com um complemento ocupado por sintagma preposicional (2.37).

(2.36) AdvP(Adv'(AdvP(bem) Adv'(longe daqui)))

(2.37) AdvP(Adv'(Adv(depois) PP(do jogo)))

Quadro 11 – Regras Gramaticais do Sintagma Adverbial

AdvP → Adv'
Adv' → Adv
AdvP → AdvP Adv'
Adv' → Adv PP

Fonte: (OTHERO, 2009)

2.3.1.6 Regras Gramaticais do Sintagma Adjetival

O Sintagma Adjetival (AP - do inglês *Adjectival Phrase*)(Quadro 12) pode ser constituído apenas por um núcleo adjetival ou acompanhado de modificação adverbial à esquerda (2.39) ou à direita (2.40), ou ainda um preposicional (2.41) na posição de adjunto. Na posição de complemento, também pode conter preposicional (2.42) ou complementizador (2.43).

(2.38) AP(A'(AdvP(Muito) A'(rápido))).

(2.39) AP(A'(A'(Rápido) AdvP(demais))).

(2.40) AP((A'(Amassado) PP(por fora))).

(2.41) AP(A'(A(Sujo) PP(de chocolate))).

(2.42) AP(A'(Confiante) CP(que o time vai vencer)).

Quadro 12 – Regras Gramaticais do Sintagma Adjetival

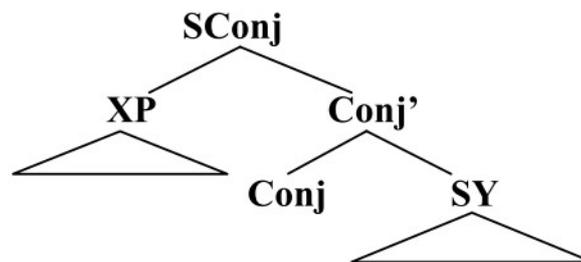
AP → A'
A' → A
A' → AdvP A'
A' → A' AdvP
A' → A' PP
A' → A PP
A' → A CP

Fonte: (OTHERO, 2009)

2.3.1.7 *Sintagma Conjuncional*

Em (OLIVEIRA, 2011) é abordado o modelo da Teoria X-Barra para o Sintagma Conjuncional (SConj ou ConjP - do inglês *Conjunctive Phrase*). O trabalho apresenta a definição de (KAYNE *et al.*, 1994), onde a conjunção surge como o núcleo da estrutura coordenada, na qual os termos coordenados funcionam como Complemento e Especificador do Sintagma Conjuncional. Desta forma, caracteriza-se como uma organização hierárquica para a coordenação, que pode ser representada pela estrutura mostrada na Figura 3.

Figura 3 – Estrutura do Sintagma Conjuncional na Teoria X-Barra.



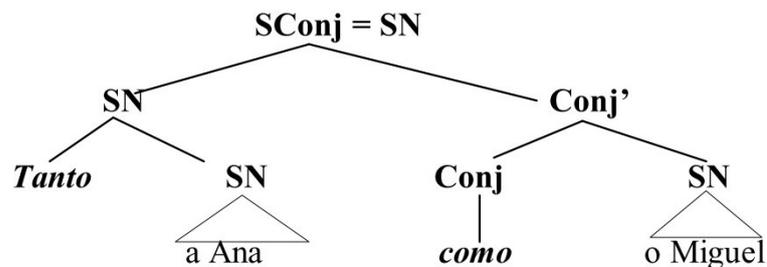
Fonte: (OLIVEIRA, 2011)

Oliveira (2011) defende em sua tese que esta representação se adéqua à estrutura da teoria linguística em questão, conforme pode ser lido abaixo:

Esta hipótese é, do ponto de vista descritivo, bastante útil, pois integra a estrutura da coordenação numa teoria linguística mais geral e unificadora e prevê que as particularidades da coordenação decorram do núcleo Conj e das respectivas conjunções que o preenchem. Alguns fundamentos desta proposta são a ideia que a Conjunção é um núcleo funcional, que seleciona argumentos, sem impor restrições categoriais, não atribuindo papéis temáticos e, nesta medida, os termos coordenados assumem as posições de especificador e de complemento.

Para o caso de coordenação múltipla, (MATOS, 2003) fornece um modelo de estrutura adaptada para que permita a representação da coordenação de dois itens, através de uma expressão descontínua. Isto ocorre em sentenças com o uso, por exemplo, de conectores como: "não só... como (também)", "não só...mas também", "tanto...como", "ou...ou", "nem...nem", etc. A autora ainda destaca que "o núcleo das expressões correlativas se encontra, tipicamente, na conjunção que ocorre entre os termos coordenados e que os elementos adverbiais ou conjuncionais marcam o começo da estrutura coordenada e surgem associados ao primeiro termo da coordenação". A Figura 4 mostra estrutura ilustrada através da aplicação do exemplo de sentença: "Tanto ele como ela sabem o que diz".

Figura 4 – Exemplo de estrutura do Sintagma Conjuncional com coordenação múltipla na Teoria X-Barra.



Fonte: (MATOS, 2003)

2.4 Conclusão do Capítulo

Este capítulo apresentou os conceitos relacionados a Análise de Sentimentos, Formalismo Gramatical e Teoria X-Barra para proporcionar a compreensão dos assuntos abordados nesta tese.

3 TRABALHOS RELACIONADOS

Este capítulo apresenta um levantamento dos trabalhos correlatos aos estudos desta tese para servir como base de referências à pesquisa. Inicialmente é relatado um breve histórico com os trabalhos pioneiros no assunto. Em seguida, é apresentada uma revisão sistemática referente aos anos de 2009 a 2019. O objetivo é fornecer um estado da arte sobre o uso de regras de composição na análise de sentimentos, procurando responder sobre os modos de abordagem deste método e o desempenho alcançado.

3.1 Breve Histórico

Sob o ponto de vista histórico, o trabalho de Hatzivassiloglou e McKeown (1997b) foi pioneiro na tentativa de extrair orientação semântica por associação entre termos, onde conhecendo a polaridade de um adjetivo, deduz-se que um termo da mesma categoria gramatical poderá ter igual polaridade em uma sentença coordenada. Trabalhos como o de Turney (2002) adotaram a mesmo método, mas considerando adjetivos próximos em sentenças não necessariamente coordenadas. Outras categorias gramaticais passaram a ser consideradas por também poder expressar sentimentos, com em Benamara *et al.* (2007) que apresentou estudos envolvendo adjetivos e advérbios, já Reforgiato e Subrahmanian (2008) trabalhou a combinação adjetivo-verbo-advérbio em análise de sentimentos.

Tendo em vista os trabalhos que abordaram a influência das polaridades individuais das palavras junto a sua função gramatical na sentença, Moilanen e Pulman (2007) argumentaram ser possível calcular de forma sistemática as polaridade de constituintes sintáticos em função das polaridades de seus subconstituintes. O modelo proposto trata dois constituintes de entrada de cada vez até gerar o resultado da polaridade global para a composição resultante. Os autores consideram três tipos de situações: (i) propagação de sentimento, onde a polaridade de um constituinte não neutro prevalece sob a polaridade de um neutro; (ii) polaridade reversa, no qual é gerada um valor de polaridade não neutra oposta ao do termo da composição; (iii) conflito de polaridade, onde a composição contém polaridades não-neutras conflitantes. A resolução deste conflito ocorre com base nos pesos atribuídos aos constituintes que decide qual é o mais importante em relação ao sentimento.

Choi e Cardie (2008) explora idéias da semântica composicional através de regras heurísticas incorporando-as abordagem baseada na aprendizagem. Os autores escreveram 8

regras de composição utilizando negação, adjetivo e os sintagmas nominal e verbal, conforme mostra a Figura 5.

Figura 5 – Regras de inferência composicional motivadas pela semântica composicional

	Rules	Examples
1	Polarity(not_[arg1]) = \neg Polarity(arg1)	not [bad] _{arg1} .
2	Polarity([VP]_[NP]) = Compose([VP], [NP])	[destroyed] _{VP} [the terrorism] _{NP} .
3	Polarity([VP1]_to_[VP2]) = Compose([VP1], [VP2])	[refused] _{VP1} to [deceive] _{VP2} the man.
4	Polarity([adj]_to_[VP]) = Compose([adj], [VP])	[unlikely] _{adj} to [destroy] _{VP} the planet.
5	Polarity([NP1]_[IN]_[NP2]) = Compose([NP1], [NP2])	[lack] _{NP1} [of] _{IN} [crime] _{NP2} in rural areas.
6	Polarity([NP]_[VP]) = Compose([VP], [NP])	[pollution] _{NP} [has decreased] _{VP} .
7	Polarity([NP]_be_[adj]) = Compose([adj], [NP])	[harm] _{NP} is [minimal] _{adj} .

Fonte: Choi e Cardie (2008)

3.2 Revisão Sistemática

Com o intuito de identificar os trabalhos relacionados com esta tese de doutorado foi realizada uma revisão sistemática com base nas etapas proposta em (KITCHENHAM, 2004). O objetivo foi identificar os artigos que propõem modelos para análise de sentimentos baseados em regras de composição.

As etapas envolvidas foram definidas de formas diferenciadas para buscas em portais de periódicos e para motor de busca acadêmico.

1. Questões de pesquisa: para atingir o objetivo descrito anteriormente, foram definidas as seguintes questões de pesquisa:

Questão Principal (QP): Quais artigos tratam a análise de sentimentos com modelos de regras de composição?

Questão Secundária 1 (QS1): Como são as abordagens em relação ao uso de regras nos modelos de análise de sentimentos propostos?

Questão Secundária 2 (QS2): Como é aplicada a avaliação comparativa de desempenho?

2. Definição da string de busca: O trabalho opta por utilizar termos que traduzem uma melhor significação das intenções desta pesquisa, assim foi usado o termo "análise de sentimento" sem suas variações. Para maior especificação da relação com este trabalho, foi decidido limitar-se a expressão "composição de sentimentos" e "semântica composicional". Associado a isso, acrescentou a possibilidade de recuperar trabalhos que usassem adicionalmente ou exclusivamente o termo "baseado em regras". Com base nas palavras-chave das questões

de pesquisa, foi elaborada a seguinte string de busca: "sentiment analysis"AND ("sentiment composition"OR "compositional semantic"OR "rule-based"). Apesar do foco da tese ser para o Português Brasileiro, decidiu-se trabalhar com a expressão sem o termo "portuguese", visto que, tentativa de busca com a sua inclusão não retornou resultados.

3. Definição das fontes de pesquisa: foram feitas buscas pelos portais de periódicos IEEE Xplore Digital, ACM Digital Library e Science Direct para o período de 2009 a 2019.

4. Definição dos critérios de inclusão e exclusão: (a) O artigo deve estar relacionado ao tema da pesquisa; (b) O artigo deve estar escrito em português ou inglês.

5. Seleção dos artigos: Para realizar esta etapa foram executadas as seguintes atividades:

a) A string de busca foi executada em cada uma das fontes de pesquisa e seus resultados foram exportados para o formato BibTex.

b) Para selecionar um conjunto inicial de artigos, os títulos e resumos dos resultados das buscas foram lidos e confrontados com os critérios da etapa 4. Durante o processo de busca por referências, verificou-se que o uso do termo "regra" não necessariamente correspondia ao contexto de pesquisa da tese proposta, pois não tratavam de composição entre classes gramaticais. A partir dos resultados obtidos, um processo de seleção de artigos foi aplicado excluindo aqueles que propunham métodos híbridos com foco principal no aprendizado de máquina para a classificação dos sentimentos, no entanto, foram considerados trabalhos que propusessem técnicas de aprendizado de máquina desde que fossem aplicadas para a dedução de regras de composição.

c) Os artigos selecionados no item b foram lidos por completo e novamente confrontados com os critérios definidos na etapa 4, sendo que todos eles estavam de acordo. Os artigos aceitos foram documentados e encaminhados para extração dos dados.

6. Extração dos dados: Os artigos selecionados foram trabalhados para extrair informações relevantes conforme as questões de pesquisa definidas na etapa 1.

Conforme observado na Tabela 1, após a execução da string de busca em cada portal de periódicos, foram selecionados 8 artigos dos 113 recuperados, sendo 4 dos 73 do IEEE, 2 entre os 20 do ACM e 1 dos 20 do ScienceDirect.

Em função da baixa quantidade de artigos selecionados pelos portais escolhidos, optou-se por investigar a possibilidade de adicionar trabalhos encontrados pelo motor de busca acadêmico Google Scholar. Esta segunda fase permitiu a inclusão de mais 7 artigos. Pelo fato

de não ser uma base de periódicos, mas sim um portal de indexação e busca por diversas bases disponíveis na Web, foram trabalhados os seguintes critérios de busca e seleção:

- i) Usar a mesma expressão de busca definida na etapa 2;
- ii) Ordenar os resultados por critério de relevância;
- iii) Considerar as 10 primeiras páginas correspondendo aos 100 primeiros resultados;
- iv) Eliminar Surveys e resultados já encontrados na busca anterior;
- v) Considerar critérios definidos na etapa 4;
- vi) Extrair dados dos artigos selecionados.

Tabela 1 – Resultados da busca por trabalhos relacionados

Portal	Resultado Inicial	Após refinamento
IEEE	73	4
ACM	20	2
Science Direct	20	1
Google Scholar	100	7

Fonte: elaborado pelo autor.

Durante a etapa de seleção de artigos, observou-se referências que não atendiam as relações com a pesquisa desta tese. Foram retornados trabalhos que tratavam de análise de sentimentos aplicando regras para outras finalidades como, por exemplo, em: extração de aspectos (Brunova; Bidulya, 2017; Ruskanda *et al.*, 2018; Fachrina; Widyantoro, 2017; Ruskanda *et al.*, 2019; Bhoir; Kolte, 2015; Rana; Yu-N Cheah, 2015; Akhoundzade; Devin, 2019; Russu *et al.*, 2014; Rao; Shah, 2016; Hamzah; Widyastuti, 2016; Ben-Ami *et al.*, 2015); modelos Fuzzy (Liu; Cocea, 2017; Jefferson *et al.*, 2017; Alshahrani Hasan; Fong, 2018); sistema especialista (Yenkikar *et al.*, 2019); regras baseadas nas ocorrências de palavras com sentimentos (Siddiqua *et al.*, 2016; Nair *et al.*, 2014); operações em valores numéricos de sentimentos (Yan *et al.*, 2018; Li; Yang, 2017; Denecke, 2009; Rane; Khan, 2018), stemização (Sultanova *et al.*, 2019; Rane *et al.*, 2019); conversão de caracteres (Duwairi *et al.*, 2016); algoritmo genético (Iqbal *et al.*, 2019); detecção de negação (Pröllochs *et al.*, 2015; Zirpe; Joglekar, 2017); medidas de centralidade (Jitsakul *et al.*, 2017); sarcasmo (Hiai; Shimada, 2016); geração automática de léxico (Jin *et al.*, 2016); confiabilidade de anotação em corpora (Abdelwahab; Elmaghraby, 2018).

O conjunto final de artigos selecionados corresponde aos requisitos que permitem responder à QP, dado que o escopo desta tese abrange estudos voltados para a classificação de

sentimentos em sentenças com foco da composição de regras a partir de relações linguísticas. Além disso, também foram investigados o modo de avaliação de desempenho nos modelos propostos, uma vez que esta tese busca uma avaliação comparativa a fim de observar se seu desempenho supera ou é próximo ao de outras abordagens.

Para a QS1: Como são as abordagem em relação ao uso de regras nos modelos de análise de sentimentos propostos?

Os trabalhos selecionados mostraram diferentes propostas no modo de aplicar suas regras, onde foram observados as seguintes características em seus modelos: composição em bigramas entre categorias gramaticais utilizadas como treino em processo de aprendizado para dedução de polaridade em novas composições (KIRITCHENKO; MOHAMMAD, 2016; SANTOS *et al.*, 2016; TAN *et al.*, 2012; TOLEDO-RONEN *et al.*, 2018; Petrakis, S ; Klenner, 2011); definição das regras diretamente no código em estruturas de decisão "if-then", executando as regras de composição partindo do nível de palavras até chegar no nível da sentença (XIE *et al.*, 2014; ROMANYSHYN, 2011; TAN *et al.*, 2016; KLENNER *et al.*, 2009; RAVISHANKAR; SHRIRAM, 2018) ou aplicando de operações com valores numéricos atribuídos ao sentimento das palavras (ZHANG *et al.*, 2009; RANI; KUMAR, 2018).

Outro modo de abordar o uso das regras diz respeito às categorias gramaticais trabalhadas nas relações de composição. Os autores buscaram tratar aquelas consideradas como mais influentes no sentimento da sentença e também para não tornar o número de regras inflado. O Quadro 2 exibe as lista de classes gramaticas para cada um dos trabalhos selecionados nesta revisão sistemática, o que permite observar como as regras foram definidas em relação a estas categorias.

No que se refere aos recursos que auxiliam as condições para execução das regras, o mesmo quadro também apresenta as ferramentas externas usadas para identificação de características das palavras na sentença, como no processo de etiquetagem e relação de dependência.

Em relação à QS2, todos os artigos procuram investigar o desempenho de seus modelos, utilizando uma parte ou todas as métricas entre acurácia, precisão, recall, ROC, Medida F e Macro F. Em relação a análise comparativa, foram encontrados trabalhos que apresentavam o desempenho de seus modelos em função das variações de parâmetros internos e outros que confrontavam com uma ou mais abordagens de aprendizado de máquina. A descrição apresentada é sumarizada na Tabela 2.

Considerando os trabalhos que tratam da aplicação de regras de composição em

Tabela 2 – Comparativo entre os trabalhos relacionados

Autor (ano)	Recurso para Etiquetação	Classes Gramaticais nas regras de composição	Abordagem comparativa de desempenho.
(ZHANG <i>et al.</i> , 2009)	Chinese Lexical Analysis System (ICTCLAS)	adjetivos, substantivos, verbos e advérbios	SVM, Naïve Bayes e Árvore de Decisão
(KLENNER <i>et al.</i> , 2009)	TreeTagger	Adjetivo, Substantivo, verbo, advérbio, preposição, reversão e intensificação	Compara com método da classe majoritária.
(ROMANYSHYN, 2011)	UGTag	Adjetivo, Substantivo, verbo, advérbio e preposições	Avaliação interna
(Petrakis, S ; Klenner, 2011)	Pro3Gres	Adjetivo, substantivo, reversão e intensificador	DecisionTable, BayesNet, ADTree, BFTree
(TAN <i>et al.</i> , 2012)	Stanford parser	Adjetivo, Substantivo, verbo, advérbio	Compara seu modelo baseado em Class Sequential Rules com um modelo de heurística com regras de composição
(XIE <i>et al.</i> , 2014)	Stanford POS	Adjetivo, Substantivo, verbo, conjunção adversativa e negação	Compara com a combinações regras + modelo numérica
(NEVIAROUSKAYA <i>et al.</i> , 2015)	Connexor Machine Syntax	Adjetivo, Substantivo, verbo, advérbio e preposições	Compara com arquitetura de Aprendizado SNoW.
(KIRITCHENKO; MOHAMMAD, 2016)	CMU Tweet NLP tool	Adjetivo, Substantivo, verbo, advérbio	Bigramas com técnicas supervisionadas e não supervisionadas.
(SANTOS <i>et al.</i> , 2016)	OpenNLP POS	Adjetivo, Substantivo, verbo, advérbio, preposições, artigos e numerais	Não comparou com outros métodos, apenas internamente.
(TAN <i>et al.</i> , 2016)	Stanford RNN	Adjetivo, Substantivo, verbo, preposição, conjunção adversativa, advérbio de negação	Avaliação interna.
(BIDULYA; BRUNOVA, 2017)	Léxico etiquetado de adjetivos	Adjetivo, conjunção, reversão, intensificação	Naïve Bayes
(VILARES <i>et al.</i> , 2017)	Baseado em Entropia Máxima	Adjetivo e advérbio, intensificadores e reversores, conjunção adversativa	Compara com SentiStrength e SO-CAL.
(RANI; KUMAR, 2018)	NLTK - Pos Tagger (Python)	Adjetivo, advérbio, reversão, intensificação	Avaliação interna
(TOLEDO-RONEN <i>et al.</i> , 2018)	Dataset etiquetado gerado por (KIRITCHENKO; MOHAMMAD, 2016)	Adjetivo, substantivo, verbo, advérbio e reversão	Avaliação interna.
(RAVISHANKAR; SHRIRAM, 2018)	Baseado em Campos Aleatórios Condicionais(CRFs)	Adjetivo, substantivo, verbo, reversão	Compara com variações internas do modelo

Fonte: elaborado pelo autor.

Análise de Sentimentos, a seguinte seção mostra a descrição de cada um dos selecionados.

3.3 Resumo dos Trabalhos

O trabalho de Zhang *et al.* (2009) baseia-se nas pontuações dos termos da sentença contidos em um léxico com palavras equivalentes a adjetivos, substantivos, verbos e advérbios. O procedimento de etiquetagem é feito através de um ferramenta de PLN para a língua chinesa ICTCLAS¹ (Institute of Computing Technology, Chinese Lexical Analysis System). O algoritmo (Figura 6) aplica o cálculo do produto das pontuações, conforme as composições geradas por parser de dependência.

Figura 6 – Pseudo-código com regras propostas por (ZHANG *et al.*, 2009)

```

FOR i = 1 TO |C(w)|
  IF d(w, ci) is VOB, COO, or DE THEN p = WordPolarity
  (ci, s) END IF
  IF d(w, ci) is CMP or MT THEN p = p * I(ci) END IF
  IF d(w, ci) is ADV or ATT THEN Enqueue(Q, ci)
  END IF
END FOR
WHILE(Q(w) ≠ Φ)
  IF(Q(w) ≠ Φ) THEN m1 = Dequeue(Q) END IF
  IF(Q(w) ≠ Φ) THEN m2 = Dequeue(Q) END IF
  IF(m1 is a degree adverb && m2 is a negative word)
  THEN p = -p * I(m1)/2 END IF
  IF(m1 is a negative word && m2 is a degree adverb)
  THEN p = p * I(m1)/2 END IF
  IF(m1 is a degree adverb && m2 is null) THEN p = p *
  I(m1) END IF
  IF(m1 is a negative word && m2 is null) THEN
  p = -p/2 END IF
END WHILE
IF (p = 0) THEN
  FOR i = 1 TO |C(w)|
    p = p + WordPolarity(ci, s)
  END FOR
END IF
RETURN p

```

Fonte: (ZHANG *et al.*, 2009)

O desempenho é avaliado para dois datasets comparando com os métodos de aprendizado (SVM, Naïve Bayes e Decision Tree) com as métricas precisão, revocação, acurácia e

¹ <http://www.ictclas.org/>

Medida F. O trabalho mostra valores favoráveis ao modelo baseado em regras, como por exemplo, obtido na acurácia de 76.33% contra 75.31%, 68.01% e 65.78% para os métodos de aprendizado citados respectivamente.

Em Klenner *et al.* (2009) é apresentada o PolArt, uma abordagem para a composição de sentimentos adotando um número limitado de regras independente de domínio e um léxico de polaridades específico para o domínio. O sistema recebe como entrada um texto sintaticamente fragmentado, onde foi usada a ferramenta TreeTagger (SCHMID, 2013). A partir da estrutura gerada, as regras são executadas iniciando das mais simples, cuja saída é aplicada nas regras mais complexas, passando do nível de palavra para a composição de sentimento no nível de sentença. O modelo foca nas polaridades de substantivos e adjetivos, além de aplicar o modificador que reverte a polaridade. A Figura 7, capturada do seu artigo, mostra um exemplo de regras de composição formadas pelo trígama com uma preposição entre dois sintagmas nominais.

Figura 7 – Regras propostas por (KLENNER *et al.*, 2009)

NP	Prep	NP	→	PP	
POS	to	NEG	→	POS	solution to my problem
POS	for	POS	→	POS	hope for relief
NEG	of	NEG	→	NEG	pain of disappointment
NEG	of	POS	→	NEG	lost of hope

Fonte: (KLENNER *et al.*, 2009)

A avaliação de desempenho foi feita aplicando em um conjunto de 222 sentenças. Os resultados de desempenho foram comparados com um método de classe majoritária que determina a polaridade da sentença a partir do maior número de ocorrências entre palavras positivas e negativas. As medidas mostradas na Figura 8 apontam melhor desempenho para o PolArt.

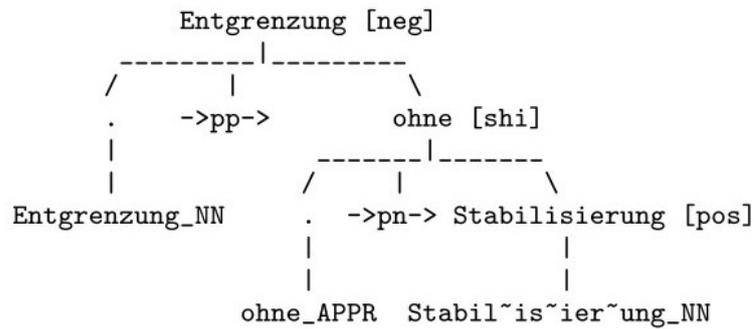
Figura 8 – Resultados comparativos de desempenho.

[Baseline/Polart]	Recall	Precision	F-measure
NEG	0.20/0.59	0.79/0.74	0.33/0.66
POS	0.48/0.4	0.15/0.19	0.23/0.25
NEU	0.56/0.13	0.16/0.11	0.25/0.12

Fonte: (KLENNER *et al.*, 2009)

Romanyshyn (2011) trabalhou regras de composição para a língua Ucrâniana, considerando que os sentimentos de palavras separadas devem ser compostos em uma determinada

Figura 10 – Árvore gerada pelo Pro3Gres



Fonte: (Petrakis, S ; Klenner, 2011)

Figura 11 – Regra tratada pelo Aleph

```

np_pol_neg(A) :-
    depends_BonA(A,B) , has_pol(B,shi) ,
    has_pol(A,pos) .
  
```

Fonte: (Petrakis, S ; Klenner, 2011)

polaridade e detectar o sentimento no nível da frase. Aplicam o uso de regras de seqüência de classes (CSRs) para aprender automaticamente os padrões de dependência tipados, que são obtidos pelo Parser Stanford³. O desempenho é comparado com um método heurístico e os resultados preliminares mostram melhorias no desempenho da classificação, alcançando mais de 80% na medida F nos casos de teste. Além disso, observam as relações mais complexas entre as palavras que podem influenciar a polaridade do sentimento e discutem sobre possíveis abordagens para lidar com seus os efeitos. Um exemplo de regra adotada é apresentada a seguir: “AMOD (ditador, cruel)” poderia indicar uma regra de padrão de polaridade de “AMOD (NN (-), JJ (-)) → (-)” dando uma saída negativa, onde NN (-) e JJ (-) se referem a um termo substantivo negativo e um termo adjetivo negativo, respectivamente. Neste trabalho, foi avaliada apenas a subjetividade das dependências dos tipos: modificador adjetival (AMOD), modificador adverbial (ADVMOD) e modificador de objeto direto, pois são consideradas como contendo a grande maioria de expressões opinativas. Exemplos das dependências tipadas exploradas para a composição são mostrados na Figura 12.

Com base no trabalho de (CHOI; CARDIE, 2008) posteriormente, (XIE *et al.*, 2014) desenvolveu o sistema MuSES para identificação de sentimentos multilinguais adicionando mais 8 novas regras (Figura 13) para acomodar o contexto da mídia social. O MuSES utiliza o Stanford POS (TOUTANOVA *et al.*, 2003) como ferramenta para etiquetagem morfosintática. As

³ <http://nlp.stanford.edu/software/lex-parser.shtml>

Figura 12 – Exemplo das dependências tipadas exploradas para a composição

AMOD(calamity:NN, great:JJ) for “great calamity”
ADVMOD(destroyed:JJ, completely:RB) for “completely destroyed”
DOBJ(killed:VB, innocents:NN) for “killed innocents”

Fonte: (TAN *et al.*, 2012)

regras tratam de composições entre adjetivo com sintagma nominal, negação e os termos “*but* (mas)”, “*despite* (apesar de)” e “*unless* (a menos que)”. Além das regras, os autores também aplicaram uma função de composição mais sofisticada para processar o maior número de regras, cuja saída varia de -2 (fracamente negativo) a +2 (fortemente positivo), sendo que o valor 0 indica neutro. Os experimentos foram aplicados em corpora inglês como idioma original e nas línguas chinesa, alemã e coreana com a aplicação de ferramenta de tradução automática para o inglês. O desempenho foi realizado com base nas métricas de precisão, revocação e Medida F. A avaliação foi aplicada e comparada entre: o modelo de regras semânticas composicionais (CSR); identificação numérica de sentimentos (NSI); combinação entre o método de aprendizado de máquina de melhor desempenho, CSR, NSI (CBM1) e emoticons; combinação entre CSR, NSI, emoticons e palavras-chave específicas de domínio (CBM2) e combinação de todos os anteriores (CBM3). Os resultados mostrados no gráfico da Figura 14 apontam melhor desempenho para CBM3.

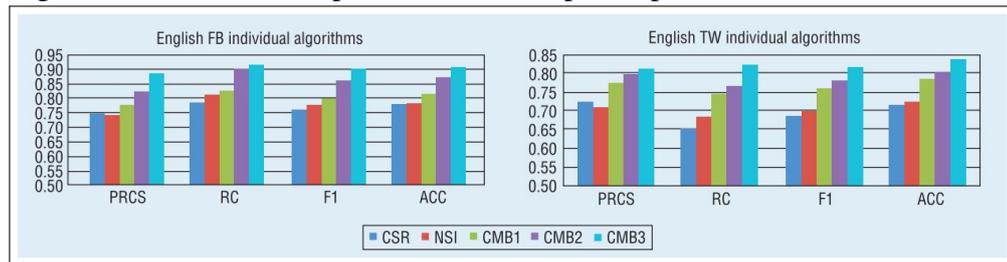
Figura 13 – Regras de composição semântica para o sistema MuSES

Polarity(NP1 of VP1) = Compose(NP1, VP1)	Lack of killing in rural areas.
Polarity(as ADJ as NP) = $\frac{1}{1 + \text{Polarity}(\text{NP} \neq 0)}$ · Polarity(ADJ) + $\frac{1}{1 + \text{Polarity}(\text{NP} \neq 0)}$ · Polarity(NP)	As ugly as a rock.
Polarity(not as ADJ as NP) = -Polarity(ADJ)	That wasn't as bad as the original.
If sentence contains “but,” disregard all previous sentiment and only take the sentiment of the part after “but.”	And I've never liked that director, but I loved this movie.
If sentence contains “despite,” only take the sentiment of the part before “despite.”	I love that movie, despite the fact that I hate that director.
If sentence contains “unless” and “unless” is followed by a negative clause, disregard the “unless” clause.	Everyone likes this video unless he is a sociopath.

Fonte: (XIE *et al.*, 2014)

Neviarouskaya *et al.* (2015) introduzem uma abordagem linguística composicional para reconhecimento de atitudes em texto. O método proposto classifica sentenças usando rótulos de atitude refinada (nove para estados afetivos, dois para julgamento positivo e negativo e dois para avaliação positiva e negativa), em comparação com outros métodos que focalizam principalmente duas categorias de sentimento (positivas e negativas). O trabalho é baseado na análise de relações sintáticas e de dependência entre palavras em uma sentença com o princípio da

Figura 14 – Gráfico comparativo de desempenho para o Sistema Muses



Fonte: (XIE *et al.*, 2014)

composicionalidade. A análise da estrutura sintática e das dependências funcionais da sentença fé feita pelo parser Connexor Machine Syntax ⁴. Também é aplicada uma nova abordagem linguística baseada nas regras elaboradas para classes verbais semanticamente distintas e um método considerando a hierarquia de conceitos. A análise comparativa de desempenho é feita com os resultados em (ALM *et al.*, 2005), que aplica a arquitetura de aprendizado SNoW em sentenças no domínio narrativas infantis de contos de fadas. Os resultados apresentaram cerca de 2% de ganho de precisão: 80.8% vs. 79%.

Kiritchenko e Mohammad (2016) explora a composição do sentimento em bigramas com uma palavra positiva e outra negativa entre os pares de categorias gramaticais, conforme mostra a Figura 15, como por exemplo: "*happy accidents* (acidentes felizes)". Usando um conjunto de dados, foram identificados padrões linguísticos com polaridade opostas. Técnicas não supervisionadas e supervisionada (Suport Vector Machine) foram aplicadas para investigar a sua eficácia. O melhor resultado ocorre com o uso de técnica superviosada, incorporando informações sobre os constituintes da frase e pontuações associadas ao sentimento, obtendo uma precisão de mais de 80%. A etiquetagem morfosintática é feita através da ferramenta CMU Tweet NLP (GIMPEL *et al.*, 2010).

O trabalho de Santos *et al.* (2016) explora o princípio da composicionalidade semântica, porém as regras a serem aplicadas são deduzidas automaticamente tomando como base um corpus anotado que relaciona o sentimento de frases com as polaridades de seus constituintes. Os autores aplicam a probabilidade condicional como uma medida de confiabilidade das regras de composição. Não há comparações com outros métodos, mas o autor apresenta o desempenho com a métrica de confiabilidade das regras aprendidas que variou de 60,44% para frases verbo-substantivo a 100% para adjetivo-adjetivo. O corpus de bigramas foi construído a partir da coleta de um conjunto de frases de manchetes em português. Posteriormente, foi aplicada a

⁴ <http://www.connexor.eu/technology/machine/machinese-syntax/>

Figura 15 – Regras trabalhas por (KIRITCHENKO; MOHAMMAD, 2016)

SCP	Occ.	# phrases
▽adj. + △adj. → △phrase	0.76	17
▽adj. + △noun → ▽phrase	0.59	68
△adj. + ▽noun → ▽phrase	0.53	73
△adverb + ▽adj. → ▽phrase	0.89	18
△adverb + ▽verb → ▽phrase	0.91	11
▽noun + △noun → △phrase	0.60	10
△noun + ▽noun → ▽phrase	0.52	25
▽verb + det. + △noun → ▽phrase	0.65	17
▽verb + △noun → ▽phrase	0.82	17

Fonte: (KIRITCHENKO; MOHAMMAD, 2016)

identificação da classe gramatical, através da ferramenta OpenNLP POS Tagger⁵. A extração dos bigramas foi feita com base nos padrões de composição mostrados na Tabela 16.

Figura 16 – Composições trabalhadas em (SANTOS *et al.*, 2016)

Pattern	#phrases	Sem. Orientation		
		-	0	+
ADJ N	433	71	300	62
ADV ADJ	38	9	13	16
ADJ ADJ	56	4	50	2
N ADJ	1,753	496	860	397
ADV VRB	240	37	183	20
N prp N	2,112	703	961	448
VRB (pr art num)? N	3,147	999	1,151	997
Total:	7,779	2,319	3,518	1,942

Fonte: (SANTOS *et al.*, 2016)

Tan *et al.* (2016) apresenta um algoritmo de análise de sentimentos baseado em regras para classificação de polaridade de artigos de notícias financeiras. Dados sobre as classes gramaticais das palavras que compõem a sentença são extraídos usando o parser Stanford RNN (KONG; SMITH, 2014). As regras de composição de sentimentos são usadas para determinar a polaridade de cada frase, enquanto a razão Positividade/Negatividade (relação P/N) é usada para calcular os valores de sentimento do conteúdo geral. O desempenho foi avaliado comparando às anotações manuais. O resultado foi considerado satisfatório pelo valor da Medida F, obtendo 75.6% para as classificações positiva e negativa.

O trabalho de (BIDULYA; BRUNOVA, 2017) teve como objetivo desenvolver um classificador baseado em regras para a análise de sentimento no domínio da qualidade do serviço

⁵ <http://opennlp.sourceforge.net/>

Figura 17 – Parte das Regras trabalhas por (TAN *et al.*, 2016)

Rules	POS Combination(JJ-VB/ JJ-VP/ ADJP-VB/ ADJP-VP)	→	Output (VP)
VP1	(NEG)(POS)	→	NEG
VP2	(NEG)(NEG)	→	NEG
VP3	(NEG)(NEU)	→	NEG
VP4	(POS)(POS)	→	POS
VP5	(POS)(NEG)	→	NEG
VP6	(POS)(NEU)	→	POS
VP7	(NEU)(POS)	→	POS
VP8	(NEU)(NEG)	→	NEG

Fonte: (TAN *et al.*, 2016)

bancário. O léxico de sentimento incluía 286 adjetivos positivos e 385 negativos. Para testar o algoritmo, 200 resenhas em russo foram analisadas. O experimento demonstra que a eficiência do classificador baseado em regras é maior comparado ao Naïve Bayes, sendo verificado uma melhor detecção para as críticas positivas.

Figura 18 – Regras trabalhas por (BIDULYA; BRUNOVA, 2017)

<p>Rule 1.1. <ALT>(<POS>){n}→<NEG>{n} Доверия нет! (There is no trust!), count=-1.</p> <p>Rule 1.2. <ALT>(<NEG>){n}→<POS>{n} Сами работники банка не грубые. (The bank officers are not rude.), count=1.</p> <p>Rule 2.1. (<INC><NEG>)→<NEG><NEG> Сотрудники крайне некомпетентны. (The employees are extremely incompetent.), count=-2.</p> <p>Rule 2.2. (<INC><POS>)→<POS><POS> Большое спасибо сотрудникам банка (Many thanks to the bank officers!), count=2.</p> <p>Rule 3.1. (<TH><ALT><NEG>)→<NEG> Меня еще никогда так не обманывали. (I was never cheated like this.), count=-2.</p> <p>Rule 3.2. (<TH><ALT><POS>)→<POS> Такой оперативности я еще не видел. (I have not met such an efficient response.), count=1.</p> <p>Rule 4. <POS>{1}<QM>→<NEG> Просто? (Easy?), count=-1.</p>	<p>Rule 5. <QM>→<NEG> Что теперь ждать-мо? (What shall I wait for?), count=-1.</p> <p>Rule 6. <Q><POS><Q>→<NEG> «Надежный» банк. (A "reliable" bank), count=-1.</p> <p>Rule 7.1. <WT>(<INC>){0}<POS>→<NEG> Мое обращение оставили без внимания. (My application was ignored), count=-1.</p> <p>Rule 7.2. <WT>(<INC>){0}<NEG>→<POS> Это банк без неприятных сюрпризов. (This is a bank without unpleasant surprises.), count=1.</p> <p>Rule 8.1. <POS>→<POS><POS> Спасибо Сбербанку! (Thanks to Sberbank!) count=2</p> <p>Rule 8.2. <NEG>→<NEG><NEG> Я в шоке! (I am shocked!), count=-2.</p> <p>Rule 9.1. <POS><CAP>→<POS><POS> ХОРОШО, когда есть такие сотрудники. (It is GOOD when there are such employees.), count=2.</p> <p>Rule 9.2. <NEG><CAP>→<NEG><NEG> ОТВРАТИТЕЛЬНЫЙ сервис. (It is an AWFUL service.), count=-2.</p>
---	---

Fonte: (BIDULYA; BRUNOVA, 2017)

Vilares *et al.* (2017) trabalham a análise de sentimentos multilíngües orientada por regras baseadas em sintaxe de composição. Os autores implementam e avaliam um modelo para análise de sentimentos universal, conduzida por um conjunto de regras sintáticas que envolvem o uso de modificadores e intensificadores de sentimentos. As regras utilizam etiquetagem, classes gramaticais e uma lista de palavras de negação, intensificadores, conjunções adversativas e palavras que introduzem condicionais (como o inglês “if” ou “would”). As regras são aplicadas em uma estrutura gerada por um parser de dependência através da ferramenta baseada no método

de Entropia Máxima trabalhada por (TOUTANOVA; MANNING, 2000). O desempenho do algoritmo, através da acurácia, foi comparado com dois sistemas não supervisionados: SO-CAL e SentiStrength. Para comparação com a abordagem supervisionada, foi considerada a rede neural recursiva profunda apresentada por (SOCHER *et al.*, 2013).

Rani e Kumar (2018) propuseram um sistema para identificar o sentimento de tweets com base em pontuações associadas a adjetivos e advérbios contidos no léxico usado. A identificação das classes gramaticais foi feita pela ferramenta NLTK⁶ da linguagem Python. O tipo de cálculo é aplicado conforme a composição entre adjetivo e os tipo do advérbio (afirmação; dúvida e de intensidades fraca e forte). O resultado de desempenho, em relação às anotações manuais, atingiu uma precisão de 86% no teste em 500 tweets e 94% para 200 frases em inglês.

Uma pontuação de -5 a 5 foi atribuída a todos os adjetivos, onde 5 é o adjetivo positivo mais forte e -5 é o negativo mais intensidade. O algoritmo (Figura 19) trabalha estes valores junto a advérbios para a definição da pontuação final na combinação Advérbio-Adjetivo.

Figura 19 – Algoritmo com regras trabalhas por (RANI; KUMAR, 2018)

Algorithm 1 Variable Scoring Algorithm

```

1: if  $adv \in AFF \cup STRONG$  then
2:   if  $scr(adj) > 0$  then
3:      $f_{VS}(adv, adj) = scr(adj) + (5 -$ 
        $scr(adj)) * scr(adv).$ 
4:   if  $score(adj) < 0$  then
5:      $f_{VS}(adv, adj) = scr(adj) - (5 -$ 
        $scr(adj)) * scr(adv).$ 
6: if  $adv \in WEAK \cup DOUBT$  then
7:   if  $scr(adj) > 0$  then
8:      $f_{VS}(adv, adj) = scr(adj) - (5 -$ 
        $scr(adj)) * scr(adv).$ 
9:   if  $scr(adj) < 0$  then
10:     $f_{VS}(adv, adj) = scr(adj) + (5 -$ 
       $scr(adj)) * scr(adv).$ 

```

Fonte: (RANI; KUMAR, 2018)

Toledo-Ronen *et al.* (2018) propõe abordagem para aprender a composição de sentimentos a partir de um grande corpus não rotulado e um léxico de sentimento no nível

⁶ <https://www.nltk.org/>

da palavra para a supervisão. Os dados referentes à etiquetagem foram baseados no dataset etiquetado gerado por (KIRITCHENKO; MOHAMMAD, 2016). Foi definido um conjunto de classes de composição de sentimento que tratam de regras de reversão, propagação, dominação e relações com adjetivos, cobrindo uma variedade de processos e a partir desses recursos o método faz aquisição automática de léxicos para cada uma dessas classes. O desempenho é avaliado por meio da acurácia e precisão, não havendo comparação com outros métodos.

Figura 20 – Regras trabalhadas por (TOLEDO-RONEN *et al.*, 2018)

#	Class	Condition		Predicted Bigram Polarity	
		UG1	UG2		
Composition Classes					Sample match
1	REV \oplus	w	\ominus	\oplus	combat loneliness
2	REV \ominus	w	\oplus	\ominus	lacked courage
3	PROP \oplus	$w \wedge \ominus$	\oplus	\oplus	overwhelming success
4	PROP \ominus	$w \wedge \oplus$	\ominus	\ominus	fresh trouble
5	DOM \oplus	w		\oplus	improving nutrition
6	DOM \ominus	w		\ominus	reckless decision
Adjective Classes					Sample match for (high,low)
7	ADJ(a_1, a_2) $\oplus\ominus$	$E(a_1)$	w	\oplus	high morale
8	ADJ(a_1, a_2) $\oplus\ominus$	$E(a_2)$	w	\ominus	low morale
9	ADJ(a_1, a_2) $\ominus\oplus$	$E(a_1)$	w	\ominus	higher inflation
10	ADJ(a_1, a_2) $\ominus\oplus$	$E(a_2)$	w	\oplus	lower inflation

Fonte: (TOLEDO-RONEN *et al.*, 2018)

Ravishankar e Shriram (2018) desenvolveram uma ferramenta de categorização de sentimentos baseada em regras gramaticais para tweets em Tâmil (uma língua do sul da Índia), referentes a comentários sobre filmes. A fase de etiquetagem é feita com base no modelo probabilístico de Campos Aleatórios Condicionais (LAFFERTY *et al.*, 2001) aplicado a Tâmil por Pandian e Geetha (2009). A gramática tâmil classifica as palavras em oito categorias de classes: o verbo, o substantivo, o pronome, o adjetivo, o advérbio, a posposição, a conjunção e a exclamação. Os modelos de linguagem podem ser formados usando uma sequência de padrões de classes gramaticais, como um adjetivo seguido por um substantivo ou verbo, ou de categoria sintática, como substantivos e/ou sintagmas verbais. A gramática proposta concentra-se apenas nos seguintes aspectos que são relevantes para os tweets; regras de adjetivos e regras de negação, que são derivadas de verbos/advérbios são usados para inverter o significado de uma frase (Figura 21).

O trabalho utiliza o TF-IDF (frequency-inverse document frequency) como base para avaliação comparativa de desempenho da abordagem proposta. TF-IDF é baseado no número de ocorrências de palavras-chave no conjunto de dados. Para cada gênero de filmes Tamil, o

Figura 21 – Regras propostas por (RAVISHANKAR; SHRIRAM, 2018)

<i>Patterns of Adjectives</i>	<i>Example</i>
Adjective → Noun → Verb (ANV)	அதிரடி படமா வந்திருக்கிறது (Athiradi padama vanthirukkirathu)
Verb → Noun → Adjective (VNA)	பார்க்க படம் அருமை (parka padam arumai)
Noun → Verb → Adjective (NVA)	திரையில் பார்த்தேன் மகிழ்ச்சி (Thiraiyil parthen magilchi)
Verb → Adjective → Noun (VAN)	பார்த்து ரசித்த படம் (Parthu rasitha padam)
Adjective → Verb → Noun (AVN)	மகிழ்ச்சியாக போகிறது படம் (Magilchiyaga pogirathu padam)
Noun → Adjective → Verb (NAV)	படத்தை ரசித்து பாருங்கள் (padathai rasithu parungal)
Noun → Adjective → Noun (NAN)	அஜீதின் நல்ல படம் (Ajithin nalla padam)

Fonte: (RAVISHANKAR; SHRIRAM, 2018)

TF-IDF calcula a pontuação e indica a importância de uma palavra-chave para as críticas de um filme no corpus. O modelo semântico baseado em regras de gramática produziu uma melhor acurácia média de 64,72% em relação ao TF-IDF.

3.4 Conclusão do Capítulo

Este capítulo apresentou uma revisão sistemática com o objetivo de explorar a produção acadêmica em relação aos assuntos de pesquisa desta tese, mais especificamente, como foram tratadas soluções para análise de sentimentos com base em regras de composição que sejam capazes de representar a estrutura da sentença. Foram observados que os trabalhos procuravam limitar a quantidade de regras, argumentando desvantagens com o esforço da produção manual. Este argumento também levou pesquisadores a abordarem métodos para dedução de regras a partir de técnicas de aprendizado automático. Apesar destes trabalhos fazerem uso de recursos como etiquetadores e parser sintáticos, os algoritmos propostos não exploravam soluções que tratassem da possibilidade de expansão das regras ao mesmo tempo que estas fossem modeladas com base na descrição da linguística formal. Dentro desse contexto, também foi observado a carência de trabalhos para o Português Brasileiro. Além disso e pelo fato desta tese procurar confrontar o desempenho do modelo proposto em relação a outras abordagens de análise de desempenho, esta revisão procurou investigar como os trabalhos trataram deste procedimento, mostrando desde avaliações limitadas a variações de parâmetros do próprio modelo, como também comparações com diferentes abordagens.

4 PARSER PARA UM SISTEMA DE ANÁLISE DE SENTIMENTOS

Este capítulo apresenta a proposta de abordagem com base em uma nova gramática, que permite o desenvolvimento de parser específico para análise de sentimentos. A gramática foi modelada com um conjunto de regras de composição de sentimentos a partir da descrição sintática formal da estrutura da sentença para o português brasileiro.

A definição das etiquetas segue um padrão de formato na especificação do sentimento com base em cada componente da estrutura da Teoria X-Barra. O Quadro 13 mostra os padrões, onde X representa a categoria sintática.

Quadro 13 – Padrões de Etiquetagem

Etiqueta	Descrição
X, Xpos, Xneg, Xrev X', X'pos, X'neg, X'rev	Núcleo Neutro, Núcleo Positivo, Núcleo Negativo, Núcleo Reversor Projeção Intermediária Neutra, Projeção Intermediária Positiva, Projeção Intermediária Negativa, Projeção Intermediária Reversora,
XP, XPpos, XPneg, XPrev	Sintagma Neutro, Sintagma Positivo, Sintagma Negativo, Sintagma Reversor

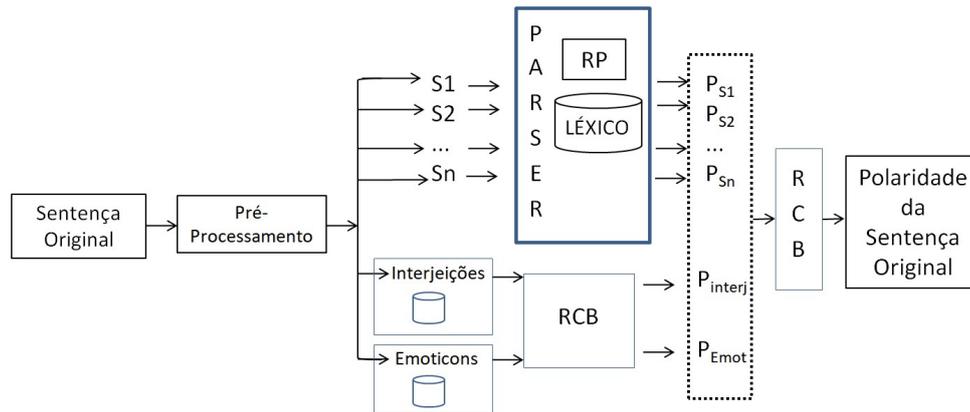
Fonte: elaborado pelo autor.

No âmbito de sua funcionalidade, foi associada uma gramática para cada uma das três categorias de sentimentos (positiva, negativa e neutra), nas quais o parser verifica a qual estrutura pertence a sentença. A diferença está apenas pela ordem de verificação das regras de sentença, onde primeira regra representa a raiz da estrutura com a projeção máxima do sintagma flexional do sentimento. Ou seja, as gramáticas positiva, negativa e neutra iniciam respectivamente por IPpos, IPneg e IP.

O Parser de sentimentos é composto por uma gramática de regras de composição de sentimentos definidas a partir de um conjunto de regras formais de descrição sintática do PB. Desta forma, o Parser analisa a sentença verificando com qual sentimento a sua estrutura está caracterizado. Quando a mensagem é composta por múltiplas sentenças, o parser analisa cada uma delas e posteriormente aplica-se a composição de sentimentos entre polaridade das sentenças. Este mesmo procedimento externo também é aplicado na presença de interjeições e emoticons. A próximas seções mostram detalhes de cada um dos procedimentos e recursos necessários, tanto para as execuções internas e externas ao parser. A Figura 22 mostra uma visão geral do sistema no qual o parser está inserido.

As regras de produção da gramática sintática são estendidas com composição de sentimentos para reconhecer uma estrutura sentencial dotada de orientação semântica, originando

Figura 22 – Diagrama do Sistema com o Parser de Sentimentos acoplado.



Fonte: do autor.

assim, as Regras de Produção de Sentimentos (RP). Ou seja, o modelo permite gerar uma especificação para um Parser de Sentimentos, que tem a função de informar se uma sentença atende a uma das estruturas caracterizadas pelas possíveis polaridades.

Para atender esta solução, a gramática é definida a partir de regras de composição entre pares de categorias gramaticais com um dos atributos (positivo, negativo, neutro, reversor). O léxico fornece um agrupamento de palavras que possui estes atributos em comum.

As regras da gramática de sentimento são definidas a partir das Regras de Composição Básica (RCB) e nos casos em que estas não são seguidas, é definida em outra categoria chamada Regras de Exceção (RE).

4.1 Regras de Composição Básica

As Regras de Composição Básica representam uma relação lógica entre pares de componentes em função de suas polaridades e/ou seus atributos modificadores. As RB correspondem a composições generalizadas que serão a base da representação do modelo de regras de produção. A formação lógica entre os elementos das RCB são definidas a partir de uma interpretação semântica quando ocorridas na sentença e foram baseadas em Moilanen e Pulman (2007). Estas regras também são aplicadas nas relações de composição de sentimentos entre sentenças e entre os segmentos de sentenças unidas por conjunções adversativas. As RCB seguem os seguintes princípios:

1. Relações entre polaridades:

- a) a predominância da polaridade de um constituinte não neutro sobre outro neutro. Este princípio baseia-se na fato de um termo neutro normalmente não interfere no

- resultado da composição. A relação em questão refere-se a uma palavra positiva ou negativa composta com outra neutra e sem atributo de modificador de valência. Nesta situação, o resultado da composição será ou positivo ou negativo, respectivamente;
- b) a composição entre constituintes com a mesma polaridade resulta nesta polaridade. Considera-se que a relação entre duas palavras, onde ambas são positivas ou negativas ou neutras, a composição gera um resultado positivo, negativo ou neutro, respectivamente;
- c) Dominância do elemento de polaridade negativa sobre o positivo. Este caso reflete o problema do conflito de polaridades, no entanto, foi considerado este domínio por não serem levados em conta variadas pontuações de sentimentos nas palavras para o tratamento deste conflito.

2. Relações com modificadores de valência:

- a) Quando um dos termos é um reversor, a regra resulta em uma polaridade oposta a do outro termo, se este for não-neutro e igual se for neutro;

4.2 Regras de Exceção

As RCB adotam um modelo generalizado a ser seguido primordialmente na expansão das regras de produção sintática para o modelo de orientação semântica. No entanto, ocorrem exceções cujos sentimento resultante do par da relação não correspondem à composição definida nas RCB, definido assim as Regras de Exceção (RE) (Quadro 14). Para estas situações, deve ser verificado a existência de padrões nas expressões geradas pelo par de termos. Como exemplo, considere a expressão "sem internet", a qual pode ser usada como negativa em um comentário sobre o serviço de uma provedora de internet. Neste caso, o a preposição "sem", que tem atributo reversor, atua sobre um elemento neutro, que pelas RCB geraria uma orientação semântica neutra. Desta forma, o modelo de composição com os atributos destas categorias gramaticais terá sua regras de sentimento negativa. O quadro mostra as regras em os respectivos exemplos com a ocorrência de exceções.

Quadro 14 – Regras de Exceção

Regra de Exceção	Exemplo
Q'pos -> AdvPpos Qlneg	"Felizmente todos os inimigos"
D'pos -> AdvPpos Dlneg	"Felizmente os inimigos "
P'neg -> Prev DP	"sem internet"

Fonte: do autor.

4.3 Regras de Produção de Sentimentos

As RP integram as GLCs para o processo de análise sintática e são expressas por um conjunto de símbolos não terminais e um conjunto de símbolos terminais. As RP são definidas a partir das RCB e RE.

O modelo proposto é baseado nas regras de uma gramática de descrição sintática da estrutura da sentença, a partir das quais aplicam-se as relações de composição de sentimento gerando um conjunto de novas regras e conseqüentemente, uma nova gramática que representa a estrutura de sentimento da sentença. Assim, o parser verifica se a frase está de acordo com uma estrutura gramatical para os casos de polaridades positiva, negativa ou neutra.

O núcleo da estrutura sintática pode ter polaridade positiva ou negativa se a categoria gramatical a que pertence transmitir sentimento, ou seja, adjetivos, advérbios, substantivos e verbos. Todas as projeções máximas (XP) e intermediárias (X') podem ter polaridades positivas e negativas, uma vez que todas as regras de produção gramaticais têm pelo menos um componente com essas polaridades.

Para a categoria reversora, o núcleo deve corresponder a uma classe gramatical que tenha a característica de inverter a polaridade de seu par. Isto aplica-se a advérbios, adjetivos, substantivos, verbos, conjunções, preposições. O Quadro 15 mostra exemplos em que há ocorrência dos atributos apresentados.

Quadro 15 – Exemplos de polaridade e modificador para as Classes Gramaticais

Classe Gramatical	Positivo	Negativo	Reversor
Verbo	Gostei desse trabalho.	Odeio o calor.	Curou a doença.
Substantivo	Encontrou a felicidade .	Declarou guerra .	O contrário da derrota.
Adjetivo	Ficou feliz .	Ficou infeliz .	Livre do vírus.
Advérbio	Felizmente deu certo	Jogou mal	Não adoeceu.
Preposição	-	-	Ele está sem a doença.
Pronome Possessivo	-	-	-
Quantificador	-	-	Nenhuma doença
Determinante	-	-	Ninguém gostou.
Conjunção	-	-	Nem andou, nem ganhou

Fonte: do autor.

Com base nos padrões apresentados, as regras de composição de sentimento foram aplicadas aos componentes do modelo da Teoria X-Bar. A composição do sentimento para regras com um símbolo não terminal: $XP \rightarrow X'$ e $X' \rightarrow X$ depende apenas da polaridade da projeção intermediária e do núcleo, respectivamente. Os símbolos terminais representam palavras e expressões com atributos referentes à categoria gramatical e sua polaridade a priori.

Para composições binárias, o Quadro 16 mostra as combinações possíveis com as atribuições de polaridades e modificador de valência para a estrutura de formação da Teoria X-Barra. No modelo de regra ($XP \rightarrow \text{Spec } X'$), por exemplo, a projeção máxima representada pelo Sintagma Inflexional $IP \rightarrow DP \ I'$ gera $IP_{\text{pos}} \rightarrow DP \ I'_{\text{pos}}$ (positivo) e $IP_{\text{neg}} \rightarrow DP \ I'_{\text{neg}}$ (negativo).

Quadro 16 – Possíveis Composições aplicado à estrutura da Teoria X-Barra.

XP X' X'	Spec X' X	X' Adjunto Complemento
POSITIVO	POS NEU POS REV NEG	POS POS NEU NEG REV
NEGATIVO	NEG NEU NEG NEG POS REV POS	NEG NEG NEU POS NEG POS REV
NEUTRO	NEU	NEU
REVERSOR	REV NEU	NEU REV

Fonte: do autor.

Considerando as possíveis combinações nas RCB e as polaridades e atributos que cada elemento recebe, defini-se um conjunto de regras para o sentimento da estrutura da sentença a partir da expansão das regras sintagmáticas.

Aplicar todas combinações às regras da gramática pode gerar regras de sentimentos desnecessárias, pois existem classes gramaticais que não contém polaridade a priori. Conforme visto, os sentimentos podem ser encontrados em palavras para adjetivos, substantivos, advérbios, verbos. No entanto, todos os sintagmas apresentam estas propriedades, pois em todos os casos haverá pelo menos um dos componentes com polaridade de sentimento. Os exemplos de bigramas abaixo mostram estes casos para DP (4.1), QP (4.2), PossP (4.3), PP (4.4), IP (4.5) e CP (4.6).

(4.1) $DP_{\text{neg}} \rightarrow D' (D (A) NP_{\text{neg}} (N'_{\text{neg}} (N_{\text{neg}} (\text{guerra}))))$

(4.2) $QP_{\text{neg}} \rightarrow Q'_{\text{neg}} (Q (\text{Todas}) DP_{\text{neg}} (\text{as guerras}))$

(4.3) $PossP_{\text{pos}} \rightarrow Poss'_{\text{pos}} (Poss (\text{Minha}) NP_{\text{pos}} (\text{felicidade}))$

(4.4) $PP_{\text{pos}} \rightarrow P'_{\text{pos}} (P (\text{por}) DP_{\text{pos}} (\text{amor}))$

(4.5) IPpos -> I'pos(I(**está**) VP(**ganhando**))

(4.6) CPpos -> C'pos(C(**que**) IPpos(**está ganhando**))

Os exemplos acima mostraram também que a projeção intermediária X' tem polaridade, pois em todos os casos, poderá ser projeção de um par de elementos que contém ou o adjunto ou o complemento, mesmo que os seus respectivos irmãos X' e X sejam neutros.

As propriedades modificadoras de valência na gramática de sentimentos, da mesma forma, dependerá de suas projeções. No modelo proposto, está sendo considerado apenas o reversor de polaridade, uma vez que os intensificadores e atenuadores influenciam em valores numéricos atribuídos ao sentimento, sem alterar a orientação semântica. Para um componente ser reversor, é necessário que esteja associado a uma das classes gramaticais que pode ser caracterizada por este atributo. Exemplos de sentenças com reversor são mostrados abaixo, através da ocorrência em: núcleo e projeção intermediária para verbos (V'rev e Vrev)(4.7); substantivo (Nrev) (4.8); preposição (Pprev) (4.9); adjetivo (Adjrev) (4.10) e o exemplo para este atributo em um sintagma XPrev (4.11) quando a sentença está na voz passiva com a composição, onde o verbo principal é reversor, resultando em um sintagma flexional reversor (IPrev).

(4.7) VPpos -> V'rev(Vrev(**faltou**) DPpos(**amor**))

(4.8) (NPpos(Nlpos(Nrev **cura**)(PPneg(Plneg **de**)(DPneg (Dlneg (D **a**) (NPneg (Nlneg (Nneg **doença**))))))

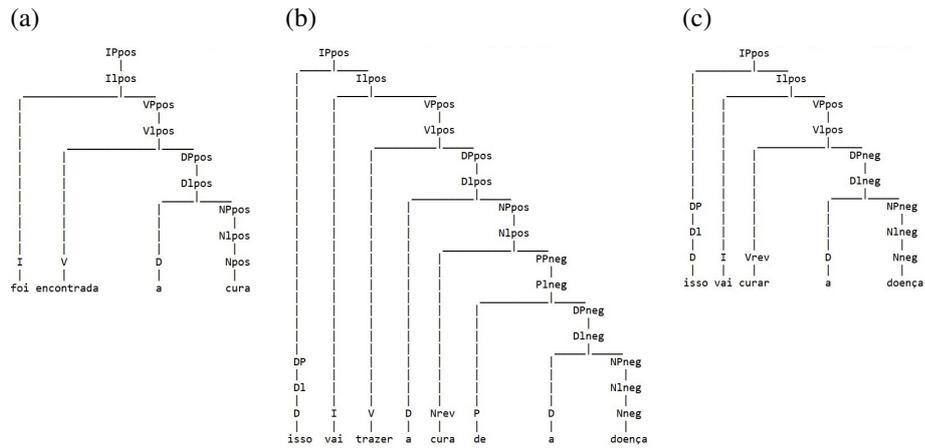
(4.9) (PPpos (Plpos (Pprev **contra**) (DPneg (Dlneg (D **o**) (NPneg (Nlneg (Nneg **vírus**))))))

(4.10) (APpos (Adjlpos (Adjrev **livre**) (PPneg (Plneg (P **de**) (DPneg (Dlneg (D **a**) (NPneg (Nlneg (Nneg **doença**))))))

(4.11) (IPpos (DPneg (Dlneg (D **a**) (NPneg (Nlneg (Nneg **doença**)))))(Ilrev(I **havia**) (IPrev (Ilrev (I **sido**) (VPrev (Vlrev (Vrev **curada**))))))

Vale salientar que um mesmo termo pode pertencer a mais de um grupo de atributo, por exemplo, o termo "cura", pode atuar ora como substantivo positivo ("foi encontrada a cura") (Figura 23a), ora como substantivo reversor ("isso vai trazer a cura da doença") (Figura 23b), ora como verbo reversor ("isso vai curar a doença") (Figura 23c). Observa-se que nestas ocorrências, o parser identifica a regra correta por conta das diferentes composições: D'pos → D NPpos; N'pos → Nrev PPneg e V'pos → Vrev DPneg, respectivamente.

Figura 23 – Diferentes atributos para o mesmo termo



Fonte: do autor

Com o levantamento das possíveis situações de ocorrência das propriedades de sentimento e modificação, os quadros abaixo apresentam as regras distribuídas para cada sintagma a partir das combinações trabalhadas em suas interpretações semânticas.

Quadro 17 – Regras de Sentimentos do Sintagma Flexional

POSITIVO	
IPpos →	I'pos DP I'pos DPpos I' DPpos I'pos ConjP I'pos ConjPpos I' ConjPpos I'pos
I'pos →	I VPpos I IPpos
NEGATIVO	
IPneg →	I'neg DP I'neg DPpos I'neg DPneg I' DPneg I'neg DPneg I'pos ConjP I'neg ConjPneg I' ConjPneg I'pos ConjPneg I'neg
I'pos →	I VPpos I IPpos
NEUTRO	
IP →	I'
I' →	DP I' DPrev I' ConjP I'
I' →	I VP I IP
REVERSOR	
IPrev →	I'rev DP I'rev
I'rev →	I VPrev I IPrev

Fonte: do autor.

Quadro 18 – Regras de Sentimentos do Sintagma Complementizador

POSITIVO	
CPpos →	C'pos
C'pos →	C IPpos
NEGATIVO	
CPneg →	C'neg
C'neg →	C IPneg
NEUTRO	
CP →	C'
C' →	C IP

Fonte: do autor.

Quadro 19 – Regras de Sentimentos do Sintagma Determinante

POSITIVO	
DPpos →	D'pos
D'pos →	D NPpos D PossPpos D NumPpos AdvPpos D'pos AdvPpos D' AdvP D'pos
NEGATIVO	
DPneg →	D'neg
D'neg →	D NPneg D PossPneg D NumPneg AdvPneg D'neg AdvPneg D' AdvP D'neg AdvPneg D'pos AdvPpos D'neg
NEUTRO	
DP →	D'
D' →	D D NP D PossP D NumP AdvP D'
REVERSOR	
DP →	D'rev
D'rev →	D NPrev D NumPrev AdvPrev D'

Fonte: do autor.

Quadro 20 – Regras de Sentimentos do Sintagma Nominal

POSITIVO	
NPpos →	N'pos
N'pos →	Npos N'pos AP APpos N' N'pos PP Npos PP N'pos CP Npos CP N' APpos AP N'pos APint N'pos N' PPpos N PPpos N' CPpos N CPpos N'pos APpos APpos N'pos N'pos PPpos Npos PPpos N'pos CPpos Npos CPpos N'rev APneg APrev N'neg N'rev PPneg Nrev PPneg N'rev CPneg Nrev CPneg
NEGATIVO	
NPneg →	N'neg
N'neg →	Nneg N'neg AP APneg N' N'neg PP Nneg PP N'neg CP Nneg CP N' APneg AP N'neg APint N'neg N' PPneg N PPneg N' CPneg N CPneg N'neg APneg APneg N'neg N'neg PPneg Nneg PPneg N'neg CPneg Nneg CPneg Nlneg APpos APneg Nlpos Nlneg PPpos Nneg PPpos Nlneg CPpos Nneg CPpos Nlpos APneg APpos Nlneg Nlpos PPneg Npos PPneg Nlpos CPneg Npos CPneg N'rev APpos APrev N'pos N'rev PPpos Nrev PPpos N'rev CPpos Nrev CPpos
NEUTRO	
NP →	N'
N' →	N N' AP AP N' N' PP N PP N' CP N CP
REVERSOR	
NPrev →	N'rev
N'rev →	Nrev N'rev AP AP N'rev N'rev PP Nrev PP N'rev CP Nrev CP

Fonte: do autor.

Quadro 21 – Regras de Sentimentos do Sintagma Adjetival

POSITIVO	
APpos →	A'pos
A'pos →	Apos AdvPpos A'pos A'pos AdvPpos A'pos PPpos Apos PPpos Apos CPpos AdvP A'pos A' AdvPpos A' PPpos A PPpos A CPpos AdvPpos A' A'pos AdvP A'pos PP Apos PP Apos CP AdvPrev A'neg A'rev AdvPneg A'rev PPneg Arev PPneg Arev CPneg
NEGATIVO	
A'neg →	Aneg AdvPneg A'neg A'neg AdvPneg A'neg PPneg Aneg PPneg Aneg CPneg AdvP A'neg A' AdvPneg A' PPneg A PPneg A CPneg AdvPneg A' A'neg AdvP A'neg PP Aneg PP Aneg CP AdvPpos A'neg A'pos AdvPneg A'pos PPneg Apos PPneg Apos CPneg AdvPneg A'pos A'neg AdvPpos A'neg PPpos Aneg PPpos Aneg CPpos AdvPrev A'neg A'rev AdvPneg A'rev PPneg Arev PPneg Arev CPneg
NEUTRO	
AP →	A'
A' →	A AdvP A' A' AdvP A' PP A PP A CP
REVERSOR	
APrev →	A'rev
A'rev →	Arev AdvPrev A' A'rev AdvP A'rev PP Arev PP Arev CP AdvP A'rev A' AdvPrev A' PPrev A PPrev A CPrev

Fonte: do autor.

Quadro 22 – Regras de Sentimentos do Sintagma Verbal

POSITIVO	
VPpos →	V'pos
V'pos →	V'pos AdvP Vpos Vpos DP Vpos PP Vpos CP Vpos AP Vpos AdvP V'pos PP Vpos IP V' AdvPpos V DPpos V PPpos V CPpos V APpos V AdvPpos V' PPpos V IPpos V'rev AdvPneg Vrev DPneg Vrev PPneg Vrev CPneg Vrev APneg Vrev AdvPneg V'rev PPneg Vrev IPneg
NEGATIVO	
VPpos →	V'pos
V'pos →	V'pos AdvP Vpos Vpos DP Vpos PP Vpos CP Vpos AP Vpos AdvP V'pos PP Vpos IP V' AdvPpos V DPpos V PPpos V CPpos V APpos V AdvPpos V' PPpos V IPpos V'rev AdvPneg Vrev DPneg Vrev PPneg Vrev CPneg Vrev APneg Vrev AdvPneg V'rev PPneg Vrev IPneg
NEUTRO	
VP →	V'
V' →	V' AdvP V V DP V PP V CP V AP V AdvP V' PP V IP
REVERSOR	
VPrev →	V'rev
V'rev →	Vrev

Fonte: do autor.

Quadro 23 – Regras de Sentimentos do Sintagma Preposicional

POSITIVO	
PPpos →	P'pos
P'pos →	AdvPpos P' P DPpos P AdvPpos P CPpos P PPpos
NEGATIVO	
PPneg →	P'neg
P'neg →	AdvPneg P' P DPneg P AdvPneg P CPneg P PPneg P
NEUTRO	
PP →	P'
P' →	AdvP P' P DP P AdvP P CP P PP P
REVERSOR	
PPrev →	P'rev
P'rev →	Prev

Fonte: do autor.

Quadro 24 – Regras de Sentimentos do Sintagma Numeral

POSITIVO	
NumPpos →	Numlpos
Numlpos →	Num NPpos Num PPpos
NEGATIVO	
NumPneg →	Numlneg
Numlneg →	Num NPneg Num PPneg Numneg
NEUTRO	
NumP →	Numl
Numl →	Num NP Num PP Num
REVERSOR	
NumPrev →	Numlrev
Numlrev →	Num NPrev Num PPrev

Fonte: do autor.

Quadro 25 – Regras de Sentimentos do Sintagma Quantificador

POSITIVO	
QPpos →	Q'pos
Qlpos →	Q DPpos Q PPpos DPpos Q AdvPpos Q' AdvP Q'pos AdvPpos Q'pos AdvPrev Q'neg
NEGATIVO	
QPneg →	Q'neg
Q'neg →	Q DPneg Q PPneg DPneg Q AdvPneg Q' AdvPneg Q'neg AdvP Q'neg AdvPneg Q' AdvPpos Q'neg AdvPneg Q'pos AdvPrev Q'pos
NEUTRO	
QP →	Q'
Q' →	Q DP Q PP DP Q Q AdvP Q'
REVERSOR	
QPrev →	Q'rev
Q'rev →	Q DPrev Q PPrev DPrev Q AdvPrev Q'

Fonte: do autor.

Quadro 26 – Regras de Sentimentos do Sintagma Adverbial

POSITIVO	
AdvPpos →	Advlpos
Advlpos →	AdvPpos Advl Advpos PP AdvP Advlpos Adv PPpos AdvPpos Advlpos Advpos Advpos PPpos AdvPrev Advlneg Advrev PPneg
NEGATIVO	
AdvPneg →	Advlneg
Advlneg →	AdvPneg Advl Advneg PP AdvP Advlneg Adv PPneg AdvPneg Advlneg Advneg Advneg PPneg AdvPrev Advlpos Advrev PPpos
NEUTRO	
AdvP →	Advl
Advl →	AdvP Advl Adv Adv PP
REVERSOR	
AdvPrev →	Advlrev
Advlrev →	Advrev

Fonte: do autor.

Quadro 27 – Regras de Sentimentos do Sintagma Conjuncional

POSITIVO	
ConjPpos →	Conjlpos IPpos Conjlpos DPpos Conjlpos AdvPpos Conjlpos APpos Conjlpos QPpos Conjlpos PossPpos Conjlpos PPpos Conjlpos VPpos Conjlpos IP Conjlpos DP Conjlpos AdvP Conjlpos AP Conjlpos QP Conjlpos PossP Conjlpos PP Conjlpos VP Conjlpos IPpos Conjl DPpos Conjl AdvPpos Conjl APpos Conjl QPpos Conjl PossPpos Conjl PPpos Conjl VPpos Conjl
Conjlpos →	IPpos Conjl DPpos Conjl AdvPpos Conjl APpos Conjl QPpos Conjl PossPpos Conjl PPpos Conjl VPpos Conjl
NEGATIVO	
ConjPneg →	Conjlneg IPneg Conjlneg DPneg Conjlneg AdvPneg Conjlneg APneg Conjlneg QPneg Conjlneg PossPneg Conjlneg PPneg Conjlneg VPneg Conjlneg IP Conjlneg DP Conjlneg AdvP Conjlneg AP Conjlneg QP Conjlneg PossP Conjlneg PP Conjlneg VP Conjlneg IPneg Conjl DPneg Conjl AdvPneg Conjl APneg Conjl QPneg Conjl PossPneg Conjl PPneg Conjl VPneg Conjl IPneg Conjlpos DPneg Conjlpos AdvPneg Conjlpos APneg Conjlpos QPneg Conjlpos PossPneg Conjlpos PPneg Conjlpos VPneg Conjlpos IPpos Conjlneg DPpos Conjlneg AdvPpos Conjlneg APpos Conjlneg QPpos Conjlneg PossPpos Conjlneg PPpos Conjlneg VPpos Conjlneg
Conjlneg →	Conj DPneg Conj IPneg Conj AdvPneg Conj APneg Conj QPneg Conj PossPneg Conj PPneg Conj VPneg
NEUTRO	
ConjP →	Conjl IP Conjl DP Conjl AdvP Conjl AP Conjl QP Conjl PossP Conjl PP Conjl VP Conjl
Conjl →	Conj DP Conj IP Conj AdvP Conj AP Conj QP Conj PossP Conj PP Conj VP
REVERSOR	
ConjPprev →	Conjlrev IP Conjlrev DP Conjlrev AdvP Conjlrev AP Conjlrev QP Conjlrev PossP Conjlrev PP Conjlrev
Conjlrev →	Conjrev DP Conjrev IP Conjrev AdvP Conjrev AP Conjrev QP Conjrev PossP Conjrev PP

Fonte: do autor.

Quadro 28 – Regras de Sentimentos do Sintagma Possessivo

POSITIVO	
PossPpos →	Poss'pos
Poss'pos →	Poss NPpos NPpos Poss Poss NumPpos
NEGATIVO	
PossPneg →	Poss'neg
Poss'neg →	Poss NPneg NPneg Poss Poss NumPneg
NEUTRO	
PossP →	Poss'
Poss' →	Poss NP NP Poss Poss NumP

Fonte: do autor.

4.4 Procedimentos Externo ao Parser

Conforme citado no início do capítulo, após a etapa de pré-processamento, a sequência de entrada para análise pode ser composta de múltiplas sentenças (S) ou ainda alguns elementos que não são tratados nas regras da gramática, mas que podem ser decisivos na identificação do sentimento, como interjeições(I) e emoticons(E).

4.4.1 Pré-processamento

Esta seção descreve a etapa de pré-processamento que envolve o preparo dos dados antes do envio ao parser. Em qualquer sistema, os procedimentos desta fase dependerão da forma como o algoritmo trabalha. Para a abordagem tratada nesta tese, o parser necessita das palavras que compõem a estrutura da sentença e em sua forma original, pois a análise é feita com base nas relações entre seus pares. Portanto, não se aplica procedimentos que envolvam padronização com quebra de termos, como *stemização* e *lematização*. Outro processo dispensado refere-se a retirada de *stopwords*, pois os termos considerados com esta característica também são fundamentais para a interpretação final da orientação semântica. Considera-se então, neste trabalho, que o pré-processamento se volte para a limpeza e fragmentação da sentença, envolvendo os seguintes procedimentos:

- Redução de múltiplos espaços em branco entre os termos;
- Extração de interjeições e emoticons para tratamento externo;
- Retirada de URLs;
- Retirada de caracteres diferentes de: letras do alfabeto; dígitos numéricos; símbolos de pontuação (',', ';', ',', '!', '?');
- Fragmentação em múltiplas sentenças, identificado pela presença dos sinais de pontuação que não são excluídos na limpeza;

- Separação de preposições contraídas (exemplo: 'da' - 'de' 'a', 'na' - 'em' 'a');
- Divisão de cada sentença em tokens.

Após estes procedimentos, intervenções poderão ser necessárias caso haja ocorrência de termos não reconhecidos pelo parser, o que implica em sua inclusão no léxico.

4.4.2 *Composição entre Sentenças, Interjeições e Emoticons*

A sequência de palavras pode ser agrupadas em sentenças, emoticons e interjeições. Assim, considere M uma mensagem que pode ser formado por um ou mais elementos pertencentes a qualquer dos conjuntos S, I e E. Quando há ocorrência de mais de um destes elementos, há um processo de composição de sentimentos externo ao parser e que tem como base os princípios que definem as RCB, conforme mostra o Quadro 29.

Quadro 29 – Composição entre elementos da mensagem.

Polaridade da Mensagem	Polaridade do elementos
POSITIVO	Todos elementos positivos elementos positivos e neutros
NEGATIVO	Pelo menos um elemento negativo Todos os elementos negativos
NEUTRO	Todos os elementos neutros

Fonte: do autor.

4.4.3 *Composição por Conjunções Adversativas*

Também são verificadas, em cada grupo, a existência de conjunção adversativa para uma análise a partir de regras de composição entre as partes da sentença dividida pela conjunção. Os fragmentos da sentença são analisados pelo parser e suas polaridades aplicadas às regras de composição mostradas no Quadro 30. Este resultado parcial é guardado para compor com as polaridades detectadas nos demais segmentos da mensagem original.

4.5 **Léxico de Sentimentos**

O modo de funcionalidade do parser trabalha com uma estrutura de GLC, cujas regras equivalem aos símbolos não terminais. Os elementos terminais atribuem as palavras aos símbolos com suas propriedades.

O léxico deve conter palavras e expressões com as respectivas propriedades a serem

Quadro 30 – Regras de composição na presença de conjunção adversativa

Saída	S1	Conjunção Adversativa	S2
POS	POS		POS
	NEG		POS
	NEU		POS
	POS		NEU
NEG	POS		NEG
	NEG		NEG
	NEG		NEU
	NEU		NEG
NEU	NEU		NEU

Fonte: do autor.

tratadas pelo modelo adotado. Nesta proposta, considera-se os atributos referentes a polaridades (positiva, negativa e neutra), o modificador de valência (reversor) e a propriedade de adversidade que tem um conjunto próprio de regras. Existe uma série de léxicos propostos e disponíveis variando conforme os tipos de valores que são atribuídos aos seus elementos. Para o modelo proposto, é necessário apenas a indicação de polaridade e modificador de valência para um cada palavra ou expressão. O léxico do parser é composto por uma lista de palavras agrupadas de acordo com a categoria gramatical (X) e o atributo de polaridade ou de modificador, de acordo com o formato: Xatributo → 'palavra_1' | 'palavra_2' | 'palavra_3' | ... | 'palavra_n'. O exemplo de um trecho neste formato é mostrado no Quadro 31.

A associação de cada palavra ao seu grupo de atributos tem como base dois léxicos de sentimentos: OpLexicon e SentiWordNet. Observa-se que há uma certa dificuldade nas decisões de escolha de léxicos devido à subjetividade no processo de atribuir sentimentos às palavras e também pela suas dependências de contexto, podendo gerar problemas com ambiguidade. Apesar disso, optou-se pela escolha destes léxicos por suas amplas coberturas de palavras e por serem consolidados na literatura. O procedimento de consulta aos léxicos se dá verificando inicialmente se a palavra está contida no OpLexicon, caso não seja encontrada, será consultada no SentiWordNet com sua correspondente tradução. Quando nenhum dos léxicos atende a demanda, a decisão é tomada conforme o significado da palavra com sua posterior inclusão no léxico. Este fato pode ocorrer com termos regionais, neologismos, palavrões ou expressões.

Com o propósito em atender o objetivo da pesquisa referente a funcionalidade do parser em tratar regras de composição de sentimentos e que os experimentos possibilitassem explorar uma vasta abrangência de diferentes estruturas de sentenças, levou-se em conta para sua execução as seguintes considerações em relação aos termos das sentenças:

Quadro 31 – Exemplo de trechos no formato do léxico na gramática do parser

POSITIVO		
Npos	→	'felicidade' 'alegria' 'saúde'
Advpos	→	'felizmente' 'bem'
Adjpos	→	'agradável' 'bonito' 'saudável'
Vpos	→	'gostar' 'amar' 'agradar'
NEGATIVO		
Nneg	→	'tristeza' 'doença' 'roubo'
Advneg	→	'infelizmente' 'mal'
Adjneg	→	'infeliz' 'mau' 'doente'
Vneg	→	'roubar' 'matar' 'cair' 'enganar'
NEUTRO		
N	→	'consórcio' 'serviços' 'download'
Adv	→	'só' 'lá' 'ontem' 'breve' 'através' 'depois' 'antes'
Adj	→	'online' 'pago' 'meio' 'brasileiro' 'acostumado' 'normal' 'neutro'
REVERSOR		
Nrev	→	'cura' 'falta' 'ausência'
Advrev	→	'nem' 'nunca' 'não'
Adjrev	→	'livre'
Vrev	→	'salvei' 'vencer' 'faltar' 'curar' 'negar' 'derrubado' 'estornar'
Prev	→	'sem'
Qrev	→	'nenhum'
Conjrev	→	'nem'

Fonte: do autor.

- Presença de todas as palavras no léxico. Este fator é possível com um processo de extração das palavras não repetidas dos corpora trabalhados, suas correções ortográficas e posteriormente a identificação de suas categorias gramaticais e sentimentos;
- Possibilidade de uma mesma palavra está contida em mais de um grupo. Algumas palavras podem estar associadas a mais de uma categoria gramatical, onde o parser identifica seu uso conforme a regra de composição;
- Possibilidade de sintagma com núcleo vazio. Por exemplo, na ausência de determinantes, verbos, núcleo complementizador e núcleo flexional, este último quando não há verbo composto na sentença.

4.6 Conclusão do Capítulo

Este capítulo apresentou o modelo de análise de sentimentos proposto a partir da definição de um conjunto de regras de composição de sentimentos que formam a gramática. Mostrou-se uma nova abordagem baseada em regras, considerando um parser que identifica se a estrutura da sentença está de acordo com a gramática específica de cada orientação semântica. Foram mostrados os possíveis modelos de relação entre os componentes que estruturam a Teoria X-Barra, as características do léxico e das funcionalidades tratadas externamente ao parser.

5 EXPERIMENTOS E RESULTADOS

O presente capítulo apresenta o processo de aplicação do modelo proposto. Para dar suporte aos experimentos, um protótipo foi implementado em Python com o uso de sua API para processamento de linguagem natural NLTK (Natural Language Toolkit). Inicialmente, são apresentados os datasets utilizados nos experimentos e seu processo de busca e escolha. Em seguida, são mostrados exemplos de resultados em diferentes estruturas de sentenças. Por fim, o processo de avaliação de desempenho é apresentado com os resultados das métricas utilizadas. A comparação dos resultados de desempenho foi feita em relação a trabalhos que apresentaram suas medidas de desempenho a partir da aplicação de outros métodos nos mesmos datasets, sendo que, não houve reimplementação de outras abordagens.

A escolha pela utilização de resultados publicados para fins de comparação se deu no decorrer da pesquisa bibliográfica, onde iniciou-se estudos para primeiro contato com o objeto da pesquisa. Na sequência, foi adotada a revisão sistemática para dar corpo e um contato mais sólido com os desdobramentos dos achados. Como parte dos resultados da revisão sistemática, destacou-se a inexistência de trabalho, em português brasileiro, com abordagens baseadas na estrutura linguística, dificultando futuras comparações com o método proposto nesta tese. Desta forma, foi aplicado também um procedimento de busca mais específico, para este idioma, que levassem a trabalhos que apresentassem seus resultados de desempenho e ao mesmo tempo disponibilizassem os datasets usados. Assim, restaram trabalhos em português brasileiro apenas baseados em outras áreas da inteligência artificial e com uso de um conjunto de ferramentas de análise de sentimentos, os quais foram então selecionados para serem utilizados nas comparações. Na avaliação de como seria o procedimento no restante da pesquisa, os autores de trabalhos correlatos foram contatados para solicitar acesso aos dados e métodos utilizados. Todavia, as respostas obtidas indicaram que tais autores não tinham mais os dados, ou os métodos, ou não poderiam fornecê-los. Diante desta dificuldade, ficou completamente fora do escopo desta tese implementar diversos métodos, o que por si só implicaria em maiores desafios. Na busca por uma solução, optou-se por comparar os resultados do presente método aos já previamente aprovados pela comunidade científica, publicados e achados na revisão sistemática. De outra forma e ainda que possível, reproduzir diversos trabalhos não traria solidez a presente tese, visto que já teriam sido feitos pelo mesmo pesquisador em seus trabalhos.

5.1 Datasets Utilizados

As buscas por referências levaram em conta aquelas que apresentassem resultados de suas propostas e ao mesmo tempo disponibilizassem os seus datasets anotados. Desta forma, foram estabelecidos os seguintes critérios para consideração da escolha do conjunto de dados de trabalho:

- Sentenças curtas:

Neste caso, com tamanho máximo equivalente ao das mensagens do Microblog Twitter

- Anotações de sentimentos (positivo, negativo e neutro) obtidos por um processo manual: Optou-se por esta preferência para que a análise de desempenho fosse em relação à visão humana de sentimento da sentença. Portanto, corpus gerado por coleta que fizesse julgamento automático em função da presença de alguma hashtag ou emoticons indicadores de sentimentos foram descartados.

- Disponibilização para download direto e pronto para uso:

Neste requisito, o dataset deve estar em local de acesso público e liberado para baixar, sem necessitar de instalações ou configurações adicionais. O conteúdo deve estar legível, sem truncamentos e completo conforme as especificações apontadas pelo artigo, além de compatível com leitores de texto.

- Ter sido usado em outro trabalho:

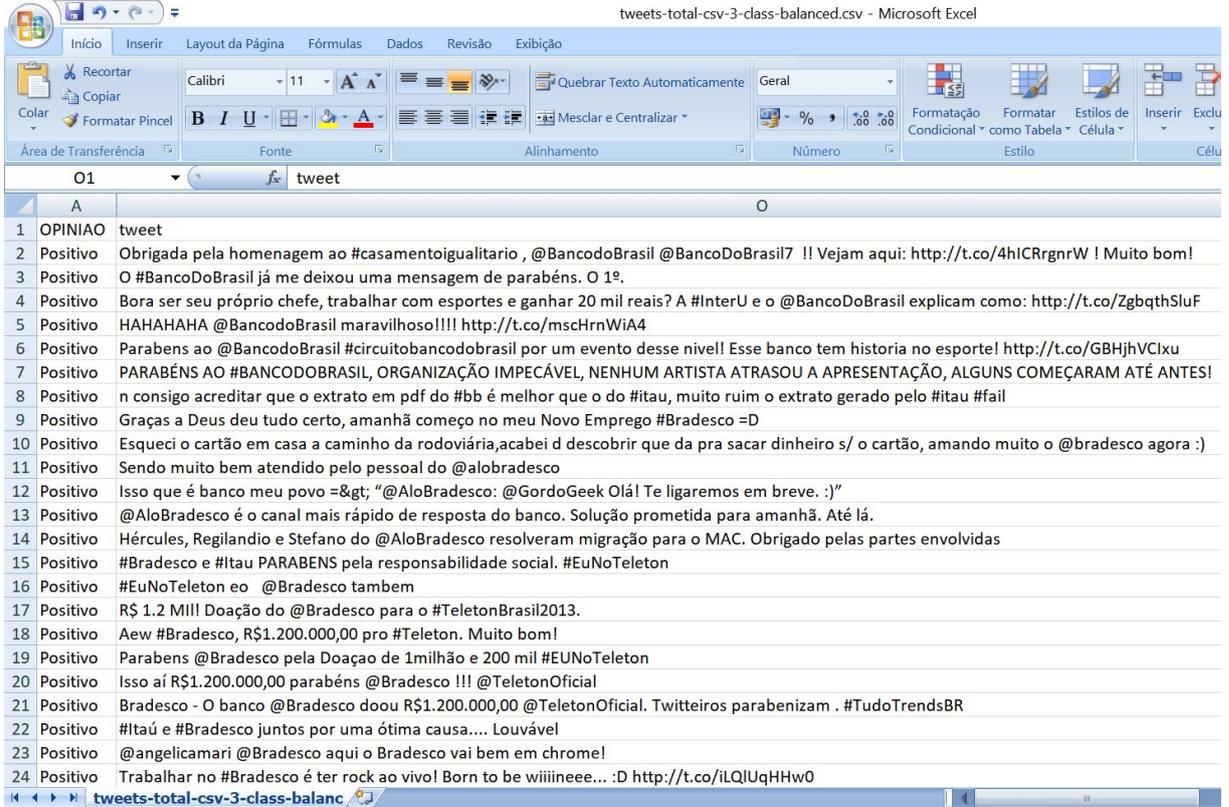
Neste item, o trabalho deve ser uma publicação com resultados abordando aplicação em métodos de aprendizado de máquina ou de modelos baseados em léxico e/ou regras. Além disso, disponibilizando resultados de desempenho dos mesmos.

Em uma busca no Google Scholar, com a string: "análise de sentimentos"AND desempenho AND corpus AND português, foram retornados 113 resultados, dos quais foram encontrados 2 trabalhos (AGUIAR *et al.*, 2018; REIS *et al.*, 2015) com referência a dataset que atendiam os requisitos desejados.

O primeiro dataset é composto por comentários em operadoras de telefonia e TV paga extraído do Twitter. Este corpus foi gerado e disponibilizado pelo grupo MiningBR e contém 2516 tuítes anotados manualmente (332 positivos, 1465 negativos e 719 neutros). A Figura 24 exibe uma amostra deste tuítes com anotação de polaridade positiva.

O trabalho de Aguiar *et al.* (2018) utilizou este dataset para análise de sentimentos com os métodos de aprendizado de máquina (Naive Bayes, Suporte Vector Machine, Decision Tree, Random Forest, Logistic Regression e Comite). A avaliação de desempenho utilizou as

Figura 24 – Amostra do corpus de tuítes referentes a comentários em operadoras de telefonia e TV paga



O1	tweet
1	OPINIAO tweet
2	Positivo Obrigada pela homenagem ao #casamentoigualitario , @BancodoBrasil @BancoDoBrasil7 !! Vejam aqui: http://t.co/4hICRgrnrW ! Muito bom!
3	Positivo O #BancoDoBrasil já me deixou uma mensagem de parabéns. O 1º.
4	Positivo Bora ser seu próprio chefe, trabalhar com esportes e ganhar 20 mil reais? A #InterU e o @BancoDoBrasil explicam como: http://t.co/ZgbqthSluF
5	Positivo HAHAAHAHA @BancodoBrasil maravilhoso!!!! http://t.co/mscHrnWiA4
6	Positivo Parabens ao @BancodoBrasil #circuitobancodobrasil por um evento desse nível! Esse banco tem historia no esporte! http://t.co/GBHjhVCixu
7	Positivo PARABÊNS AO #BANCODOBRASIL, ORGANIZAÇÃO IMPECÁVEL, NENHUM ARTISTA ATRASOU A APRESENTAÇÃO, ALGUNS COMEÇARAM ATÉ ANTES!
8	Positivo n consigo acreditar que o extrato em pdf do #bb é melhor que o do #itau, muito ruim o extrato gerado pelo #itau #fail
9	Positivo Graças a Deus deu tudo certo, amanhã começo no meu Novo Emprego #Bradesco =D
10	Positivo Esqueci o cartão em casa a caminho da rodoviária,acabei d descobrir que da pra sacar dinheiro s/ o cartão, amando muito o @bradesco agora :)
11	Positivo Sendo muito bem atendido pelo pessoal do @alobradesco
12	Positivo Isso que é banco meu povo => @AloBradesco: @GordoGeek Olá! Te ligaremos em breve. :)"
13	Positivo @AloBradesco é o canal mais rápido de resposta do banco. Solução prometida para amanhã. Até lá.
14	Positivo Hércules, Regilandio e Stefano do @AloBradesco resolveram migração para o MAC. Obrigado pelas partes envolvidas
15	Positivo #Bradesco e #Itau PARABENS pela responsabilidade social. #EuNoTeleton
16	Positivo #EuNoTeleton eo @Bradesco tambem
17	Positivo R\$ 1.2 Mil! Doação do @Bradesco para o #TeletonBrasil2013.
18	Positivo Aew #Bradesco, R\$1.200.000,00 pro #Teleton. Muito bom!
19	Positivo Parabens @Bradesco pela Doacao de 1milhão e 200 mil #EUNoTeleton
20	Positivo Isso aí R\$1.200.000,00 parabéns @Bradesco !!! @TeletonOficial
21	Positivo Bradesco - O banco @Bradesco doou R\$1.200.000,00 @TeletonOficial. Twitteiros parabenizam . #TudoTrendsBR
22	Positivo #Itaú e #Bradesco juntos por uma ótima causa.... Louvável
23	Positivo @angelicamari @Bradesco aqui o Bradesco vai bem em chrome!
24	Positivo Trabalhar no #Bradesco é ter rock ao vivo! Born to be wiiiiineee... :D http://t.co/iLQIUqHHw0

Fonte: Imagem capturada pelo autor.

métricas Acurácia, Revocação, Precisão e Medida F. Para avaliar os classificadores, foi utilizada a curva Receiver Operating Characteristic (ROC). Nesta proposta de tese, foram medidos o desempenho do modelo proposto nas classificações de casos positivos, negativos, neutros, de forma isolada, e também abrangendo todos os caso juntos. Outras medidas consideradas para fins comparativos foram as taxas de verdadeiros positivos e de falsos positivos.

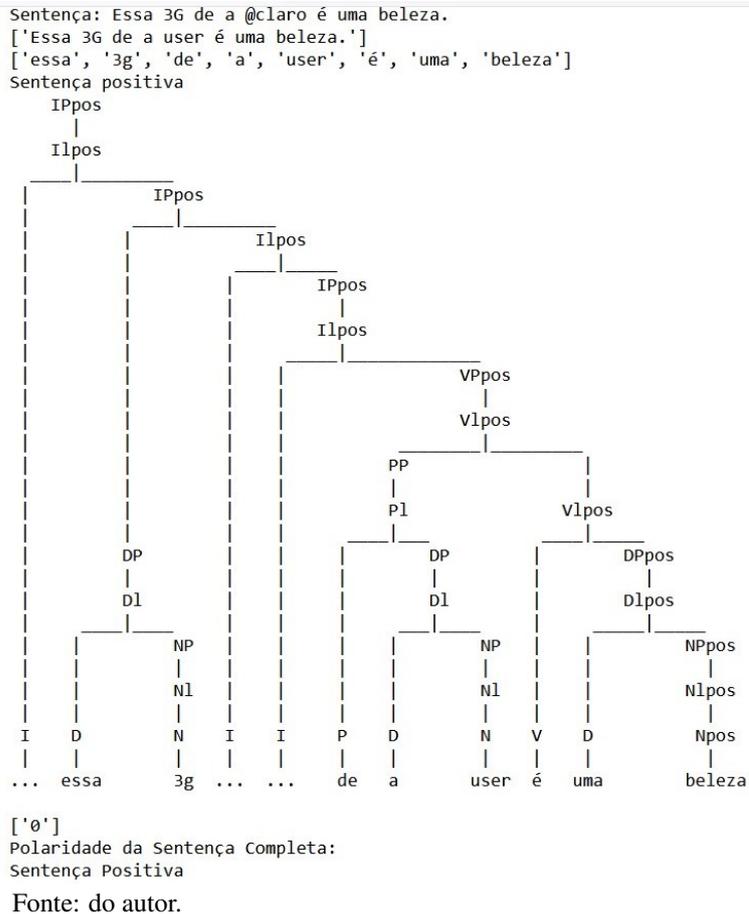
A execução do protótipo inicia com a fase de pré-processamento: limpeza de dados referentes a menções a usuário, símbolos de hashtag, urls e excesso de espaço em branco entre as palavras. Este procedimento facilita a fase seguinte correspondente à detecção de múltiplas sentenças reconhecidas pela presença de pontuações: ponto-final; exclamação e interrogação. Além disso, as interjeições e emoticons são separados para análises isoladas. Posteriormente os símbolos de pontuações são retirados dos grupos de sequências obtidas.

O segundo dataset¹ é composto de tuítes de conteúdos aleatórios em língua portuguesa. O corpus possui 774 tuítes manualmente anotados, sendo 297 positivos, 213 negativos e 264 neutros. A Figura 25 mostra uma parte do conteúdo e seu formato com as sentenças e seus respectivos valores de polaridade.

¹ <http://www.dcc.ufmg.br/fabricio>

neutros, o algoritmo de parser verificou que esta sentença pertence a uma gramática para uma estrutura de sentença positiva.

Figura 26 – Saída para Sentença Simples



5.2.2 Múltiplas sentenças

Uma mensagem com múltiplas sentenças é composta por mais de uma sentença simples. O processo de análise é feito através da composição das polaridades em cada sentença interna. O exemplo deste caso é a mensagem positiva "*Sempre MUITO bem atendida pela @Itaucard ! Muito obrigada!*", a Figura 28 mostra, na saída dos resultados, a divisão entra as duas partes, onde foram julgadas como duas sentenças simples positivas e gerando em sua composição uma sentença final positiva.

5.2.3 Modificadores de Valência

A presença do modificador de valência ocorre na sentença negativa "*Falta de respeito com o cliente! Bradesco Procon <http://t.co/mD2tdAY453>*". Neste exemplo o termo "*falta*" atua

O exemplo expõe um comentário irônico. O expressão interrogativa expressa em "*o que está acontecendo*" causa uma reversão no aparente elogio final, dando em seu todo o real sentido negativo da mensagem. O modelo proposto não tem regras que tratam o aspecto semântico da composição e características dos termos da expressão em questão.

b) Negativo julgado como neutro:

"@oi eu quero a velocidade da minha internet de volta "

Não houve indícios de sentimentos negativos conforme visto pelo modelo proposto. No entanto, é possível verificar um modo de transmitir insatisfação com o serviço, que pode ser encontrado com os termos "*eu quero*" e "*de volta*".

c) Positivo julgado como neutro:

"Fiz meu primeiro saque na agência totalmente digital do #Bradesco - agência do futuro."

Ao ler a mensagem, depara-se com o modo que o transmissor adotou para elogiar o alvo do comentário. O termo "*do futuro*" caracteriza-se como locução adjetiva e tem um contexto positivo, mas nem sempre causa esta orientação semântica, como por exemplo em: "*carro do futuro*". Desta forma, o modelo não apresenta regras para tratar esta exceção.

d) Positivo julgado como negativo:

"Que demora para o #Santander aprovar um compra online se fosse #Itaucard era em minutos... #tenso"

Os termos "*demora*" e "*tenso*" têm sentimentos a priori negativos, no entanto, a anotação manual caracterizou toda a mensagem como positiva. Isso pode ter ocorrido por conta da direções opostas apresentadas na mesma mensagem, porém para alvos diferentes e que foi dominado pelo alvo positivo. No entanto, o modelo proposto considerou o domínio do sintagma negativo.

e) Neutro julgado como negativo:

"Dificuldade de mandar UM sms. A #claro tá com algum problema será?"

Esta mensagem apresenta duas sentenças, sendo que a primeira direciona para um sentimento negativo por conta do termo "*dificuldade*". Juntamente com a segunda mensagem julgada como neutra pelo modelo. A composição de sentenças resultou na polaridade negativa ao aplicar as regras externas ao parser. A anotação manual foi dada como neutra para toda a mensagem, provavelmente, pelo seu caráter interrogativo.

f) Neutro julgado como positivo:

"Um bom Rock and Roll! Capital Inicial #festa #Bradesco <http://t.co/qiJ6uL7OEx>"

Este é um caso de falso positivo, em função de não haver comentário para nenhum dos alvos do domínio, porém julgado pelo parser de sentimento como positivo sob a influência do termo "bom".

A Tabela 3 apresenta a taxa de ocorrência para cada tipo de erro, definido conforme a sua causa.

Tabela 3 – Percentagem de cada tipo de erro

Anotado	Obtido	Causa	Taxa(%)
Positivo	Negativo	Presença de termos negativos	6,2
	Neutro	Termos positivos ou negativos não encontrados	17,8
Negativo	Positivo	Sarcasmo	2,7
		Presença de termos positivo	5,3
	Neutro	Termos positivos ou negativos não encontrados	42,0
Neutro	Positivo	Presença de termos positivos sem alvo específico	17,8
	Negativo	Presença de termos negativos sem alvo específico	8,2

Fonte: do autor.

5.3 Análise de Desempenho

A análise de desempenho tratada nesta seção tem o objetivo de avaliar o classificador em relação aos resultados encontrados para cada polaridade, através das métricas: acurácia; precisão; revocação e medida F. O modelo de parser de sentimento foi avaliado para os dois datasets e posteriormente comparado com resultados disponíveis na literatura que trabalharam outros métodos nestes mesmos conjunto de dados. A interpretação dos parâmetros no contexto de classificação de polaridade é mostrado no Quadro 32.

Em relação aos valores das métricas, é importante observar que aplica-se o seu cálculo individualmente em cada polaridade e no contexto geral.

Qualquer mensagem classificadas erroneamente como neutra estará ocultando as reais demandas, que corresponde na necessidade em saber se uma entidade está sendo bem ou mal avaliada. Classificar um mensagem positiva com outro sentimento, implica em falsos negativos, isso significa que está sendo descartada uma opinião favorável, no entanto é importante verificar para qual sentimento este falso negativo está sendo direcionado. Mensagens positivas classificadas como negativas poderá implicar em falsos alertas de que uma entidade está mal avaliada quando na verdade não está. Cabe ressaltar que as consequências deste fato são relativas, pois do ponto vista da exposição destes valores, haverá uma falsa percepção ruim de público.

Quadro 32 – Interpretação para cada polaridade

POSITIVA	Polaridade Real da Sentença	Julgada pelo Classificador
Verdadeiro Positivo	Positivo	Positivo
Falso Positivo	Negativa ou Neutra	Positivo
Verdadeiro Negativo	Negativa ou Neutra	Negativa ou Neutra
Falso Negativo	Positivo	Negativa ou Neutra
NEGATIVA	Polaridade Real da Sentença	Julgada pelo Classificador
Verdadeiro Positivo	Negativa	Negativa
Falso Positivo	Positivo ou Neutra	Negativa
Verdadeiro Negativo	Positivo ou Neutra	Positivo ou Neutra
Falso Negativo	Negativa	Positivo ou Neutra
NEUTRA	Polaridade Real da Sentença	Julgada pelo Classificador
Verdadeiro Positivo	Neutra	Neutra
Falso Positivo	Negativa ou Positivo	Neutra
Verdadeiro Negativo	Negativa ou Positivo	Negativa ou Positivo
Falso Negativo	Neutra	Negativa ou Positivo

Fonte: do autor.

Em relação ao fornecedor, poderá acontecer aplicação de mudanças em algo que na verdade está tendo bom retorno.

A classificação de mensagens negativas como positivas pode ser considerada uma falha grave, em função de que, avaliações desfavoráveis a entidade estão sendo interpretadas como bem vista pelo usuário, esta condição pode afetar estratégias de tomada de decisão do não investimento urgente na melhoria do produto ou serviço que está sendo oferecido para evitar prejuízos. A partir destas interpretações, observa-se que é muito importante que o classificador minimize a taxa de falsos positivos e maximize a taxa de verdadeiros negativos. Por outro lado, também é fundamental manter baixa as taxa de falsos positivos para evitar mudanças equivocadas de estratégias e propaganda negativa.

As interpretações acima são consideradas com base nos valores das métricas trabalhadas. A precisão para uma certa polaridade representa o grau de acerto em relação aos que foram classificadas com esse valor. A revocação indica a frequência com que o classificador encontra exemplos de uma classe, ou seja, o quanto realmente apresenta aquele sentimento.

A medida F combina precisão e revocação de modo a trazer um número único que indique a qualidade geral do modelo e trabalha bem até com conjuntos de dados que possuem classes desproporcionais.

As métricas utilizadas para avaliar os métodos basearam-se na precisão e revocação para obter o F1 por classe (positivo e negativo). A medida F1 é a média harmônica entre a precisão e a recuperação.

5.3.1 Experimento 1

O primeiro experimento é aplicado sobre o dataset 1, verificando a polaridade gerada pelo modelo proposto em cada sentença. Inicialmente, foram gerados os valores da matriz de confusão (Tabela 4) a serem usados no cálculo das demais métricas. A avaliação de desempenho com base nas métricas pré-estabelecidas apresentou os resultados exibidos na Tabela 5.

A polaridade negativa apresentou o melhor resultado de precisão, ou seja, entre os comentários negativos sobre o produto, 95,2% foram classificados corretamente. Em termos práticos, é importante que esse valor seja alto, pois a classificação incorreta de comentários negativos pode ocultar ocorrências de problemas relatados pelos usuários.

Em relação à classificação positiva, o resultado da revocação (80,7%) mostrou-se maior que a precisão (73%), porém, independente desta comparação, os resultados permitem indicar que o classificador comportou-se nesta amostra com baixo valor de falsos negativos, ao mesmo tempo que a precisão mostrou baixo valor de falsos positivos, este quadro é favorável considerando que comentários classificados incorretamente com polaridade positiva causariam uma falsa impressão de boa qualidade do produto. Além de observar o equilíbrio entre precisão e revocação, o valor da medida F (76,7%) permitir afirmar que o valor da acurácia é confiável

Tabela 4 – Matriz de confusão para o dataset 1

	POS	NEG	NEU	TOTAL
POS	268	16	48	332
NEG	45	1196	224	1465
NEU	51	44	624	719

Fonte: do autor.

Tabela 5 – Desempenho do Paser de Sentimento para o dataset 1

	Geral	Positivo	Negativo	Neutro
Acuracia	88,0	93,5	85,3	85,3
Precisão	82,8	73,0	95,2	69,4
Revocação	81,4	80,7	79,0	86,8
Medida F	82,1	76,7	86,4	77,1

Fonte: do autor.

No contexto da avaliação comparativa, os resultados são mostrados na Tabela 6, juntamente com os valores de desempenho obtidos pelos métodos trabalhados por (AGUIAR

et al., 2018). Em termos de acurácia, o parser de sentimento proposto apresentou valor 88%, indicando ter maior taxa de acertos para a base utilizada em relação aos demais métodos.

Tabela 6 – Desempenho dos métodos aplicados ao Dataset 1

Método	Acurácia	Precisão	Revocação	Medida F
Parser de Sentimentos	0.88	0.82	0.81	0.82
Naive Bayes	0.810	0.814	0.810	0.812
SVM	0.841	0.846	0.841	0.844
Decision Tree	0.800	0.827	0.800	0.815
Radon Forest	0.856	0.862	0.856	0.859
Logistic Regression	0.842	0.845	0.842	0.842
Comite	0.865	0.866	0.864	0.865

Fonte: adaptado de (AGUIAR *et al.*, 2018).

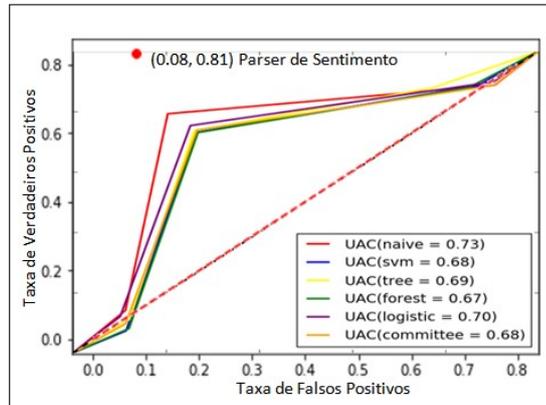
Outra ferramenta usada para avaliação é o gráfico ROC (Receiver Operating Characteristics). Um modelo de classificação é representado por um ponto, com valores entre 0 e 1, no espaço ROC que representa a taxa de verdadeiros positivos e a taxa de falsos positivos. Desta forma, um conjunto de predições permite formar uma curva, cuja área abaixo desta é chamada de (AUC), em inglês *Area Under Curve* e seu valor é usado como índice, que varia entre 0 e 1. Quanto mais próximo do valor 1, melhor será o classificador. Nesse gráfico é definida uma linha limite, que indica qual o valor mínimo de AUC deve ser considerado para que o mesmo ainda seja viável para a aplicação.

Alguns classificadores produzem naturalmente uma probabilidade ou pontuação de instância, um valor numérico que representa o grau em que uma instância é membro de uma classe (FAWCETT, 2006). Valores diferentes podem ser calculados e usados como limiares para produzir um conjunto distinto de pontos n ROC. A maneira mais eficiente de gerar essa curva é ordenar todos os casos de teste de acordo com o valor contínuo predito pelo modelo. Quanto mais longe a curva estiver da diagonal principal, melhor será o desempenho do sistema de aprendizado para esse domínio. Esta mesma interpretação pode ser verificada visualmente, por exemplo, um ponto no espaço ROC é melhor que outro se estiver a noroeste do primeiro, isto implica que a taxa verdadeiros positivos é maior e a taxa de falsos positivos é menor.

Neste contexto, a Figura 31 exhibe o gráfico ROC com os resultados de Aguiar *et al.* (2018), que exhibe o valor de AUC para comparar o desempenho dos métodos de aprendizado de máquina, com o melhor resultado para Naïve Bayes. Para efeito comparativo, o par (0,81, 0,08) gerado pelo parser de sentimentos proposto nesta tese também foi plotado no mesmo espaço. Sua posição possibilitando identificar condições favoráveis para o classificador, conforme descrito

no início da seção, caracterizado pela necessidade de manter baixa as taxa de falsos positivos e maximizar as taxas de verdadeiros positivos.

Figura 31 – Análise Comparativa pelo Gráfico ROC



Fonte: adaptado de (AGUIAR *et al.*, 2018).

5.3.2 Experimento 2

A segunda parte da avaliação aplicou o modelo no dataset 2, sendo verificadas as classificações corretas e os valores referentes aos erros de classificação. Após a contabilização dos resultados, construiu-se a matriz de confusão mostrada na Tabela 7 para dar suporte ao cálculo das métricas a serem trabalhadas.

Tabela 7 – Matriz de confusão para o dataset 2

	POS	NEG	NEU	TOTAL
POS	214	22	57	293
NEG	5	188	20	213
NEU	53	4	207	264

Fonte: do autor.

A Tabela 8 apresenta os resultados das métricas de desempenho individualmente para cada polaridade e seu valor geral. Assim como no primeiro dataset, os resultados mostram valores maiores para a polaridade positiva (82,2%), ficando pouco abaixo dos neutros (82,6%), refletindo a importância do aumento destes resultados para a sua interpretação prática. O mesmo pode-se afirmar para os resultados na polaridade negativa, conforma as justificativas e interpretações já tratadas e explicadas nesta seção.

Em função do objetivo em comparar com resultados de desempenho obtidos por

Tabela 8 – Métricas de desempenho para o dataset 2

	Geral	Positivo	Negativo	Neutro
Acuracia	77,8	82,2	70,8	82,6
Precisão	79,1	78,7	87,9	72,9
Revocação	59,9	73,0	41,0	78,4
Medida F	68,2	75,8	55,9	75,5

Fonte: do autor.

outras abordagens. Foram observados os valores medidos em Araújo *et al.* (2016), que aplicou o mesmo corpus nas seguintes 23 ferramentas de análise de sentimentos: Emoticons (GONÇALVES *et al.*, 2013a); Opinion Lexicon (HATZIVASSILOGLOU; MCKEOWN, 1997a); Opinion Finder (MPQA) (WILSON *et al.*, 2005); SentiWordNet (ESULI; SEBASTIANI, 2006); LIWC (TAUSCZIK; PENNEBAKER, 2010); Sentiment140 (GO *et al.*, 2009); SenticNet (CAMBRIA *et al.*, 2014); AFINN (NIELSEN, 2011); SO-CAL (TABOADA *et al.*, 2011b); Emoticons DS (Distant Supervision) (HANNAK *et al.*, 2012); NRC Hashtag (MOHAMMAD, 2012); Pattern.en (SMEDT; DAELEMANS, 2012); SASA (WANG *et al.*, 2012); PANAS-t (GONÇALVES *et al.*, 2013b); Emolex (MOHAMMAD; TURNEY, 2013); USent (PAPPAS *et al.*, 2013); Sentiment140-Lexicon (MOHAMMAD *et al.*, 2013); SentiStrength (SOCHER *et al.*, 2013); Stanford Recursive Deep Model (THELWALL, 2014); Umigon (LEVALLOIS, 2013); ANEW_SUB (WARRINER *et al.*, 2013); VADER (HUTTO; GILBERT, 2014); Semantria (LEXALYTICS, 2015).

As métricas com os valores disponibilizados pelos autores envolveram a Medida F para as polaridades positivas e negativas. Em função do conjunto de dados apresentar desequilíbrio, foram calculados os valores de Macro F obtidos a partir da média entre as Medidas F para as duas polaridades. Além disso, também foi considerada a cobertura (Cov) que indica o número de sentenças totais menos as indefinidas, tudo isso dividido pelo total de sentenças. Os resultados deste trabalho são mostrados na Figura 32 capturada do artigo referido, especificamente para o corpus no idioma português.

Considerando as métricas utilizadas para critérios de comparação e analisando os valores obtidos pelo parser de sentimentos (Tabela 9) em relação as outras ferramentas, observou-se que o valor da métrica Macro F posiciona o modelo proposto nesta tese entre os 4 primeiros de maior valor, ficando atrás de SentiStrength, ML SentiStrength e Emoticons.

Figura 32 – Resultados de desempenho de dataset2

Method	Portuguese			
	F1(+)	F1(-)	Macro-F1	Cov
AFINN	0.86	0.79	0.82	0.51
Emolex	0.66	0.69	0.68	0.53
Emoticons	0.93	0.86	0.90	0.06
Emoticons DS	0.74	0.00	0.37	0.98
Happinnes Index	0.80	0.54	0.67	0.57
LIWC	0.79	0.65	0.72	0.61
NRC Hashtag	0.50	0.66	0.58	0.65
Opinion Finder	0.77	0.78	0.77	0.42
Opinion Lexicon	0.84	0.79	0.81	0.44
PANAS-t	0.77	0.72	0.74	0.08
Pattern.en	0.86	0.74	0.80	0.72
SANN	0.82	0.75	0.78	0.46
SASA	0.77	0.61	0.69	0.58
SenticNet	0.78	0.54	0.66	0.91
Sentiment140	0.75	0.68	0.71	0.96
SentiStrength	0.94	0.91	0.93	0.31
SentiWordNet	0.76	0.59	0.67	0.89
SO-CAL	0.85	0.81	0.83	0.66
Stanford Deep Mode	0.51	0.67	0.59	0.86
Umigon	0.87	0.82	0.84	0.56
Vader	0.87	0.80	0.83	0.79
ML SentiStrength	0.95	0.83	0.89	0.05
Semantria	0.87	0.82	0.85	0.56

Fonte: (ARAÚJO *et al.*, 2016)

Tabela 9 – Desempenho do Parser de Sentimento no Dataset 2

	Positive	Negative
Precisão	97,7	89,5
Revocação	72,0	88,2
Medida F	82,9	88,8
Macro F		0,86
Cobertura		0,84

Fonte: do autor.

5.4 Conclusão do Capítulo

Este capítulo apresentou a avaliação do desempenho do modelo proposto aplicado em dois datasets e a comparação dos resultados com outros métodos usados nos mesmos conjunto de dados. O objetivo fez parte do processo de validação da hipótese, mostrando melhores valores referentes ao gráfico ROC comparados aos métodos de aprendizado de máquina. Um segundo experimento apresentou resultados entre os quatro melhores entre outras 23 ferramentas de análise de sentimentos.

6 CONCLUSÕES

Este capítulo é dedicado às considerações finais desta tese de doutorado. A hipótese e as questões de pesquisa são analisadas na seção 6.1. A seção 6.2 apresenta as principais contribuições e os resultados alcançados com esta tese. Por fim, a seção 6.3 é dedicada aos possíveis trabalhos futuros.

6.1 Análise da hipótese e das questões de pesquisa

A hipótese e as questões de pesquisa que nortearam o desenvolvimento desta tese de doutorado são analisadas nesta seção. As questões de pesquisa (QP) são discutidas a seguir:

Q1: Quais as demandas para o desenvolvimento de modelo de análise de sentimentos que permita tratar e identificar a estrutura da sentença com característica de sentimento?

A revisão sistemática mostrou uma variedade de metodologias que abordam regras de composição em análise de sentimentos, entre os quais relacionavam pares de classes gramaticais com suas respectivas orientações semânticas. Cada abordagem tratava seus modelos a partir de um subconjunto de categorias gramaticais.

No contexto algorítmico, os trabalhos faziam uso de parser sintático em módulos funcionais separados a partir de ferramentas de terceiros.

Trabalhos que focavam em características sintáticas para definir relações entre termos da sentença propunham regras que não cobriam todas as classes gramaticais. Também não foram encontradas propostas que adotassem a estrutura da sentença com base em formalismo linguístico, assim como, abordagens detecção de sentimentos agregados à gramática do parser, de modo que o sentimento fosse tratado ao mesmo tempo que a análise sintática.

Desta forma, a resposta para esta questão apontou: necessidade de maior abrangência de relações entre classes gramaticais que possa alcançar maior cobertura de tipos de estruturas sentenciais; gramática sintagmática com regras que representem a descrição sintática da sentença com base em um formalismo linguístico e que possam ser computacionalmente tratadas; flexibilidade na expansão da gramática para aumentar o poder de solução, de modo que evite custos adicionais de processamento, como por exemplo, retreinamentos.

Q2: Como as regras de composição devem ser modeladas para a representação da estrutura da sentença?

A Seção 2.3 apresentou a gramática de Othero (2009), que propunha "descrever a

estrutura interna da sentença em português brasileiro com base em um formalismo linguístico (...) nos moldes propostos pela teoria X-barra". Ao mesmo tempo, esta gramática deveria ser passível de implementação computacional. Com base nesta gramática, as regras de composição de sentimentos puderam ser modeladas, estendendo com as relações entre polaridades de categorias gramaticais com base nos princípios de composição tratados em Moilanen e Pulman (2007). Um protótipo foi desenvolvido e testado inicialmente para a gramática original, a qual apresentou resultados coerentes com a proposta de seu desenvolvimento, posteriormente, foi implementado com as regras de sentimentos. Algumas questões de codificação tiveram que ser tratadas para ampliar a abrangência de sentenças. Além disso, foi necessário adicionar novas regras não trabalhadas, como por exemplo na ocorrência de sentenças com conjunções adversativas.

Q3: Como adaptar as regras de composição semântica a um modelo de formalismo gramatical a ser agregado a um parser?

As seções 4.1 e 4.2 apresentaram um conjunto de regras básicas a serem usadas para a composição de sentimentos. Este modelo inicial apresentou de forma generalizada quais as polaridades resultantes na relação entre dois elementos com os possíveis valores positivo, negativo, neutro e reversor. Estas definições foram estabelecidas a partir de interpretações semânticas das relações. A seção 4.3 abordou a aplicação das regras básica na estrutura da Teoria X-barra e as composições que poderão ser geradas entre os constituintes da sentença. Em seguida, foram apresentadas todas as regras adaptadas para cada sintagma descritas na gramática da seção 2.3.

Q4: Qual o desempenho de um parser que agrega regras de composição de sentimentos à sua gramática em um processo de análise de sentimentos comparado a outros métodos quando aplicados a um mesmo corpus? A resposta desta questão, trabalhada na seção 5.2, finalizou a pesquisa ao mostrar o desempenho do modelo proposto aplicado a dois conjuntos de dados de diferentes domínios. Estes corpora foram usados em outros trabalhos através de métodos de aprendizado supervisionado e ferramentas de análise de sentimentos. O desempenho do parser de sentimentos apresentou resultados satisfatórios de precisão e revocação em relação os valores anotados. Além disso, o modelo proposto foi comparado com o desempenho de outros métodos e ferramentas propostas na literatura quando aplicados ao mesmo conjunto de dados.

Hipótese: É possível propor e desenvolver um modelo de análise de sentimentos com base em regras de composição semântica de sentimentos, agregadas a um analisador sintático (parser) e desenvolvidas com base nas descrições formais da linguística, apresentando

desempenho nos mesmos patamares das taxas obtidas por outros métodos.

Resultado: Validada. Os resultados dos experimentos do capítulo 5 mostraram que o modelo proposto apresentou parâmetros que o qualifica positivamente como classificador de sentimentos. Foram usados dois conjuntos de dados com uma extensão de conteúdos que permitiram colocar o modelo diante de variados tipos de estruturas de sentenças. Desta forma, a gramática modelada seguindo princípios do formalismo linguístico e utilizada como base para as regras de composição de sentimentos atenderam as demandas diversificadas.

Em termos numéricos, o parser de sentimento proposto, ao ser aplicado ao primeiro dataset, apresentou equilíbrio entre valores de precisão e revocação. Os resultados indicaram respectivamente (73% e 80.7%) para sentenças com polaridades positivas e (95.2% e 79%) para negativas. Estes valores apontaram maiores possibilidade do classificador não mascarar comentários desfavoráveis, que acarretaria em uma falsa interpretação de que a entidade alvo estaria sendo bem avaliada. Por outro lado, estes valores mostraram que o classificador tende a não passar a falsa impressão de que o produto alvo esteja bem avaliado, quando na verdade vendo sendo mal visto pelos seus consumidores.

Outro parâmetro verificado foi a relação entre a taxa de falsos positivos e a taxa de verdadeiros positivos usados para o gráfico ROC. O valor do ponto (0.08, 0.81) no espaço do gráfico ROC, juntamente com os valores gerados para outros métodos baseados em aprendizado supervisionado, mostrou o classificador melhor que os demais pela sua posição nas coordenadas que indica maior número de resultados corretos sobre a ocorrência e não ocorrência de determinada polaridade. Com relação ao segundo dataset, os resultados para sentenças positivas e negativas foram comparados através do Macro F, onde foi obtido o valor 0.86, que o posicionou entre os 4 melhores comparados a outras 23 ferramentas de análise de sentimentos.

Além disso, os resultados mostraram equilíbrio entre os dois experimentos com o mesmo conjunto de regra e léxico, porém aplicados datasets com domínios diferentes.

6.2 Resultados alcançados e principais contribuições

As principais contribuições desta tese com base nos resultados alcançados na pesquisa foram:

Um conjunto de regras de composição de sentimentos baseados em uma gramática descrita conforme o formalismo linguístico, que foi fruto de um estudo de padronização de composição entre orientações semânticas em classes gramaticais;

Modelo de classificação de sentimentos cuja funcionalidade baseia-se em um parser onde a sua gramática permite identificar se a sentença pertence ou não a uma estrutura com uma das características: positiva, negativa ou neutra. Esta abordagem tem a vantagem de ser expandida com maior flexibilidade para adição de novas regras e termos na gramática, afim de aumentar seu poder de solução, além de ser uma abordagem dispensa a necessidade de treinamentos.

O desenvolvimento e testes de protótipo permitiu mostrar a possibilidade de implementação computacional da gramática sintática utilizada e conseqüentemente a sua adaptação para um parser com a gramática de regras de composição de sentimentos.

A modelagem e os resultados obtidos foram publicados em:
da Silva, A. N.; de Souza, O.; de Souza, J. N. Sentiment parser based on x-bar theory to brazilian portuguese. In: **2020 International Conference on Computing, Electronics Communications Engineering (iCCECE)**. [S. l.: s. n.], 2020. p. 166–171.

Vale salientar que esta publicação passa a ser a primeira referência encontrada¹ no portal IEEE quando se adiciona o termo "portuguese" à expressão de busca usada na revisão sistemática. Outros resultados semelhantes ocorrem também quando são utilizados termos como: **"X-bar theory" AND "sentiment analysis"**. Usando a expressão: **"sentiment analysis" AND portuguese**, o mesmo portal o classifica na sétima posição entre 15 trabalhos retornados. Este fato permite mostrar a importância e o pioneirismo da pesquisa, podendo vir a ser usada como referência inicial para futuras trabalhos envolvendo os mesmos assuntos.

6.3 Limitações e Trabalhos futuros

Ao longo da pesquisa e do desenvolvimento desta tese, foram observadas propostas de possíveis melhorias para ampliar os recursos fornecidos pelo parser de sentimento. A seguir, são listados perspectivas de trabalhos futuros, que está diretamente associada as limitações do trabalho: a adição de atributos numéricos que permitam aplicar intervalo de valores aos sentimentos; aplicação de padrões propostos na literatura na detecção de sarcasmos para melhoria dos resultados diante desse tipo de sentença; exploração do o uso de relações de dependência nas composições; adaptação do modelo de regras de composição para maior especificação do tipo de sentimento, como alegria, raiva, tristeza, etc.; estudo de modelos padrões de sentenças pseudo-subjetivas.

¹ Buscas feitas em 26 de janeiro de 2021

REFERÊNCIAS

- Abdelwahab, O.; Elmaghraby, A. Deep learning based vs. markov chain based text generation for cross domain adaptation for sentiment classification. In: **2018 IEEE International Conference on Information Reuse and Integration (IRI)**. [S. l.: s. n.], 2018. p. 252–255.
- ABIRAMI, A.; GAYATHRI, V. A survey on sentiment analysis methods and approach. In: **IEEE. 2016 Eighth International Conference on Advanced Computing (ICoAC)**. [S. l.], 2017. p. 72–76.
- AGUIAR, D.; FAIC, B. S.; SILVA, G. C. Análise de Sentimento em Redes Sociais para a Língua Portuguesa Utilizando Algoritmos de Classificação. 2018.
- Akhoundzade, R.; Devin, K. H. Persian sentiment lexicon expansion using unsupervised learning methods. In: **2019 9th International Conference on Computer and Knowledge Engineering (ICCKE)**. [S. l.: s. n.], 2019. p. 461–465. ISSN 2643-279X.
- ALBORNOZ, J. Carrillo-de; PLAZA, L. An emotion-based model of negation, intensifiers, and modality for polarity and intensity classification. **Journal of the American Society for Information Science and Technology**, Wiley Online Library, v. 64, n. 8, p. 1618–1633, 2013.
- ALENCAR, L. F. DONATUS: UMA INTERFACE AMIGÁVEL PARA O ESTUDO DA SINTAXE FORMAL UTILIZANDO A BIBLIOTECA EM PYTHON DO NLTK. *Alfa*, 2012. ISSN 00025216.
- ALM, C. O.; ROTH, D.; SPROAT, R. Emotions from text: Machine learning for text-based emotion prediction. **HLT/EMNLP 2005 - Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference**, n. October, p. 579–586, 2005.
- Alshahrani Hasan, A.; Fong, A. C. Sentiment analysis based fuzzy decision platform for the saudi stock market. In: **2018 IEEE International Conference on Electro/Information Technology (EIT)**. [S. l.: s. n.], 2018. p. 0023–0029. ISSN 2154-0373.
- ARAÚJO, M.; GONCALVES, P.; CHA, M.; BENEVENUTO, F. iFeel: A System That Compares and Combines Sentiment Analysis Methods. **Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion**, p. 75–78, 2014. Disponível em: <http://dx.doi.org/10.1145/2567948.2577013>.
- ARAÚJO, M.; PEREIRA, A. C. M.; HORIZONTE, B.; REIS, J. C. S. An Evaluation of Machine Translation for Multilingual Sentence-level Sentiment Analysis. 2016.
- BACCIANELLA, S.; ESULI, A.; SEBASTIANI, F. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In: CHAIR), N. C. C.; CHOUKRI, K.; MAEGAARD, B.; MARIANI, J.; ODIJK, J.; PIPERIDIS, S.; ROSNER, M.; TAPIAS, D. (Ed.). **Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)**. Valletta, Malta: European Language Resources Association (ELRA), 2010. ISBN 2-9517408-6-7.
- BARR, A.; FEIGENBAUM, E. A. The handbook of artificial intelligence. william kaufmann. **Inc., Los Altos, CA**, p. 163–171, 1981.

- Ben-Ami, Z.; Feldman, R.; Rosenfeld, B. Exploiting the focus of the document for enhanced entities' sentiment relevance detection. In: **2015 IEEE International Conference on Data Mining Workshop (ICDMW)**. [S. l.: s. n.], 2015. p. 1284–1293. ISSN 2375-9259.
- BENAMARA, F.; CESARANO, C.; PICARIELLO, A.; REFORGIATO, D.; SUBRAHMANNIAN, V. S. Sentiment analysis: Adjectives and adverbs are better than adjectives alone. In: **Proceedings of the International Conference on Weblogs and Social Media (ICWSM)**. [S. l.: s. n.], 2007. Short paper.
- Bhoir, P.; Kolte, S. Sentiment analysis of movie reviews using lexicon approach. In: **2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCI)**. [S. l.: s. n.], 2015. p. 1–6.
- BICK, E. Automatic parsing of portuguese. In: **Proceedings of the Second Workshop on Computational Processing of Written Portuguese, Curitiba**. [S. l.: s. n.], 1996. p. 91–100.
- BIDULYA, Y.; BRUNOVA, E. Sentiment analysis for bank service quality: A rule-based classifier. **Application of Information and Communication Technologies, AICT 2016 - Conference Proceedings**, 2017.
- BLANCO, E.; MOLDOVAN, D. Retrieving implicit positive meaning from negated statements. **Natural Language Engineering**, Cambridge University Press, v. 20, n. 4, p. 501–535, 2014.
- Brunova, E.; Bidulya, Y. Aspect extraction and sentiment analysis in user reviews in russian about bank service quality. In: **2017 IEEE 11th International Conference on Application of Information and Communication Technologies (AICT)**. [S. l.: s. n.], 2017. p. 1–4. ISSN 2472-8586.
- CAMBRIA, E.; OLSHER, D.; RAJAGOPAL, D. Senticnet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis. In: **Proceedings of the AAAI Conference on Artificial Intelligence**. [S. l.: s. n.], 2014. v. 28, n. 1.
- CARVALHO, P.; SILVA, M. J. . Sentilex-Pt: Principais Características E Potencialidades. v. 7, n. 1, 2015.
- CHOI, Y.; CARDIE, C. Learning with compositional semantics as structural inference for subsentential sentiment analysis. **Proceedings of the Conference on Empirical Methods in Natural Language Processing - EMNLP '08**, n. October, p. 793, 2008. Disponível em: <http://portal.acm.org/citation.cfm?doid=1613715.1613816>.
- CHOMSKY, N. On certain formal properties of grammars. **Information and control**, Elsevier, v. 2, n. 2, p. 137–167, 1959.
- Denecke, K. Are sentiwordnet scores suited for multi-domain sentiment classification? In: **2009 Fourth International Conference on Digital Information Management**. [S. l.: s. n.], 2009. p. 1–6.
- DOWTY, D.; DOWTY, D.; WALL, R.; PETERS, S. **Introduction to Montague Semantics**. Springer Netherlands, 1981. (Studies in Linguistics and Philosophy). ISBN 9789027711410. Disponível em: <https://books.google.com.br/books?id=PMFOow9IYYC>.
- DOWTY, D. R. Word meaning and montague grammar. the semantics of verbs and times in generative semantics and in montague's ptq. **Journal of Symbolic Logic**, Association for Symbolic Logic, v. 48, n. 2, p. 501–502, 1983.

Duwairi, R. M.; Alfaqeh, M.; Wardat, M.; Alrabadi, A. Sentiment analysis for arabizi text. In: **2016 7th International Conference on Information and Communication Systems (ICICS)**. [S. l.: s. n.], 2016. p. 127–132.

ESULI, A.; SEBASTIANI, F. Sentiwordnet: A publicly available lexical resource for opinion mining. In: CITESEER. **LREC**. [S. l.], 2006. v. 6, p. 417–422.

Fachrina, Z.; Widyantoro, D. H. Aspect-sentiment classification in opinion mining using the combination of rule-based and machine learning. In: **2017 International Conference on Data and Software Engineering (ICoDSE)**. [S. l.: s. n.], 2017. p. 1–6.

FAWCETT, T. An introduction to ROC analysis. **Pattern Recognition Letters**, v. 27, n. 8, p. 861–874, 2006. ISSN 01678655.

GARSDIE, R.; LEECH, G. N.; MCENERY, T. **Corpus annotation: linguistic information from computer text corpora**. [S. l.]: Taylor & Francis, 1997.

GAZDAR, G. Phrase structure grammar. In: **The nature of syntactic representation**. [S. l.]: Springer, 1982. p. 131–186.

GIMPEL, K.; SCHNEIDER, N.; O’CONNOR, B.; DAS, D.; MILLS, D.; EISENSTEIN, J.; HEILMAN, M.; YOGATAMA, D.; FLANIGAN, J.; SMITH, N. A. **Part-of-speech tagging for twitter: Annotation, features, and experiments**. [S. l.], 2010.

GO, A.; BHAYANI, R.; HUANG, L. Twitter sentiment classification using distant supervision. **CS224N project report, Stanford**, v. 1, n. 12, p. 2009, 2009.

GONÇALVES, P.; ARAÚJO, M.; BENEVENUTO, F.; CHA, M. Comparing and combining sentiment analysis methods. In: **Proceedings of the first ACM conference on Online social networks**. [S. l.: s. n.], 2013. p. 27–38.

GONÇALVES, P.; BENEVENUTO, F.; CHA, M. Panas-t: A psychometric scale for measuring sentiments on twitter. **arXiv preprint arXiv:1308.1857**, 2013.

Hamzah, A.; Widyastuti, N. Opinion extracting and classification from questionnaire comments using hmm-pos tagger and machine learning techniques. In: **2016 International Conference on Data and Software Engineering (ICoDSE)**. [S. l.: s. n.], 2016. p. 1–6.

HANNAK, A.; ANDERSON, E.; BARRETT, L. F.; LEHMANN, S.; MISLOVE, A.; RIEDEWALD, M. Tweetin’ in the rain: Exploring societal-scale effects of weather on mood. In: **Proceedings of the International AAAI Conference on Web and Social Media**. [S. l.: s. n.], 2012. v. 6, n. 1.

HATZIVASSILOGLOU, V.; MCKEOWN, K. Predicting the semantic orientation of adjectives. In: **35th annual meeting of the association for computational linguistics and 8th conference of the european chapter of the association for computational linguistics**. [S. l.: s. n.], 1997. p. 174–181.

HATZIVASSILOGLOU, V.; MCKEOWN, K. R. Predicting the semantic orientation of adjectives. In: **Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics**. Stroudsburg, PA, USA: Association for Computational Linguistics, 1997. (ACL ’98), p. 174–181. Disponível em: <https://doi.org/10.3115/976909.979640>.

Hiai, S.; Shimada, K. A sarcasm extraction method based on patterns of evaluation expressions. In: **2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)**. [S. l.: s. n.], 2016. p. 31–36.

HUTTO, C.; GILBERT, E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: **Proceedings of the International AAAI Conference on Web and Social Media**. [S. l.: s. n.], 2014. v. 8, n. 1.

Iqbal, F.; Hashmi, J. M.; Fung, B. C. M.; Batool, R.; Khattak, A. M.; Aleem, S.; Hung, P. C. K. A hybrid framework for sentiment analysis using genetic algorithm based feature reduction. **IEEE Access**, v. 7, p. 14637–14652, 2019. ISSN 2169-3536.

Jefferson, C.; Liu, H.; Cocea, M. Fuzzy approach for sentiment analysis. In: **2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)**. [S. l.: s. n.], 2017. p. 1–6. ISSN 1558-4739.

Jin, Z.; Yang, Y.; Bao, X.; Huang, B. Combining user-based and global lexicon features for sentiment analysis in twitter. In: **2016 International Joint Conference on Neural Networks (IJCNN)**. [S. l.: s. n.], 2016. p. 4525–4532. ISSN 2161-4407.

Jitsakul, W.; Meesad, P.; Sodsee, S. Enhancing comment feedback classification using text classifiers with word centrality measures. In: **2017 2nd International Conference on Information Technology (INCIT)**. [S. l.: s. n.], 2017. p. 1–5.

KAYNE, R.; KAYNE, P.; KAYNE, R. **The Antisymmetry of Syntax**. London, 1994. (Linguistic inquiry monographs). ISBN 9780262611077. Disponível em: <https://books.google.com.br/books?id=tnXJVbGpMfEC>.

KENEDY, E. Possíveis contribuições da linguística gerativa à formação do professor de língua portuguesa. v. 1, n. 4, p. 72–79, 2016.

KENNEDY, A.; INKPEN, D. Sentiment classification of movie reviews using contextual valence shifters. **Computational Intelligence**, v. 22, n. 2, p. 110–125, 2006. ISSN 08247935.

KIRITCHENKO, S.; MOHAMMAD, S. M. Sentiment Composition of Words with Opposing Polarities. **Naacl**, p. 1102–1108, 2016. Disponível em: <http://www.saifmohammad.com/WebDocs/OPP-naacl2016.pdf>.

KITCHENHAM, B. Procedures for performing systematic reviews. **Keele, UK, Keele University**, v. 33, n. 2004, p. 1–26, 2004.

KLENNER, M.; PETRAKIS, S.; FAHRNI, A. Robust compositional polarity classification. **Proceedings of RANLP**, p. 180–184, 2009. ISSN 13138502. Disponível em: <http://acl.eldoc.ub.rug.nl/mirror/R/R09/R09-1034.pdf>.

KONG, L.; SMITH, N. A. An empirical comparison of parsing methods for stanford dependencies. **arXiv preprint arXiv:1404.4314**, 2014.

LAFFERTY, J. D.; MCCALLUM, A.; PEREIRA, F. C. N. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: **Proceedings of the Eighteenth International Conference on Machine Learning**. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001. (ICML '01), p. 282–289. ISBN 1558607781.

- LEVALLOIS, C. Umigon: sentiment analysis for tweets based on lexicons and heuristics. 2013.
- LEXALYTICS. **Sentiment extraction - measuring the emotional tone of content**. [S. l.], 2015.
- Li, J.; Yang, L. A rule-based chinese sentiment mining system with self-expanding dictionary - taking tripadvisor as an example. In: **2017 IEEE 14th International Conference on e-Business Engineering (ICEBE)**. [S. l.: s. n.], 2017. p. 238–242.
- LIU, B. Sentiment analysis and subjectivity. In: INDURKHYA, N.; DAMERAU, F. J. (Ed.). **Handbook of Natural Language Processing, Second Edition**. Boca Raton, FL: CRC Press, Taylor and Francis Group, 2010. ISBN 978-1420085921.
- LIU, B. Sentiment Analysis and Opinion Mining. **Synthesis Lectures on Human Language Technologies**, v. 5, n. May, p. 1–108, 2012. ISSN 1947-4040.
- Liu, H.; Cocea, M. Fuzzy rule based systems for interpretable sentiment analysis. In: **2017 Ninth International Conference on Advanced Computational Intelligence (ICACI)**. [S. l.: s. n.], 2017. p. 129–136.
- MATOS, G. Gramática da língua portuguesa. In: _____. [S. l.]: Ines Duarte, Hub Faria, Mira Mateus, 2003. cap. Estruturas de Coordenação, p. 549–592.
- MEDHAT, W.; HASSAN, A.; KORASHY, H. Sentiment analysis algorithms and applications: A survey. **Ain Shams Engineering Journal**, Faculty of Engineering, Ain Shams University, v. 5, n. 4, p. 1093–1113, 2014. ISSN 20904479. Disponível em: <http://dx.doi.org/10.1016/j.asej.2014.04.011>.
- MOHAMMAD, S. # emotional tweets. In: * **SEM 2012: The First Joint Conference on Lexical and Computational Semantics–Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)**. [S. l.: s. n.], 2012. p. 246–255.
- MOHAMMAD, S. M.; KIRITCHENKO, S.; ZHU, X. Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. **arXiv preprint arXiv:1308.6242**, 2013.
- MOHAMMAD, S. M.; TURNEY, P. D. Crowdsourcing a word–emotion association lexicon. **Computational intelligence**, Wiley Online Library, v. 29, n. 3, p. 436–465, 2013.
- MOILANEN, K.; PULMAN, S. Sentiment Composition. **Proceedings of the Fourth International Conference on Recent Advances in Natural Language Processing (RANLP 2007)**, n. M, p. 378–382, 2007. ISSN 13138502. Disponível em: http://www.clg.ox.ac.uk/{_}media/people:karo:sentcompranlp07final.
- MONTAGUE, R. English as a formal language. In: VISENTINI, B. (Ed.). **Linguaggi nella società e nella tecnica**. [S. l.]: Edizioni di Comunità, 1970. p. 188–221.
- MONTOYO, A.; MARTÍNEZ-BARCO, P.; BALAHUR, A. Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments. **Decision Support Systems**, v. 53, n. 4, p. 675 – 679, 2012. ISSN 0167-9236. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0167923612001339>.

MORSY, S. A.; RAFEA, A. Improving document-level sentiment classification using contextual valence shifters. In: SPRINGER. **International conference on application of natural language to information systems**. [S. l.], 2012. p. 253–258.

Nair, D. S.; Jayan, J. P.; Rajeev, R. R.; Sherly, E. Sentima - sentiment extraction for malayalam. In: **2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)**. [S. l.: s. n.], 2014. p. 1719–1723.

NEVIAROUSKAYA, A.; PRENDINGER, H.; ISHIZUKA, M. Attitude sensing in text based on a compositional linguistic approach. **Computational Intelligence**, v. 31, n. 2, p. 256–300, May 2015. ISSN 0824-7935.

NIELSEN, F. Å. A new anew: Evaluation of a word list for sentiment analysis in microblogs. **arXiv preprint arXiv:1103.2903**, 2011.

OLIVEIRA, M. d. C. P. **A sintaxe da coordenação e os conectores conclusivos**. 344 p. Tese (Doutorado) – Faculdade de Letras da Universidade do Porto, 2011.

OLIVEIRA, M. de; SANTOS, E. dos; CAMPOS, E. **Análise sintática do português falado no Brasil**. Editora Multifoco, Luminária Academia, 2010. (Análise sintática do português falado no Brasil, v. 1). ISBN 9788579610981. Disponível em: <https://books.google.com.br/books?id=VK0BywAACAAJ>.

OTHERO, G. d. Á. Lingüística Computacional: uma breve introdução. **Letras de hoje**, p. 341–351, 2006. Disponível em: file:///tmp/othero{_}06{_}Lingstica-Compu.

OTHERO, G. d. Á. LINGÜISTA “PURO” VS. LINGÜISTA “COMPUTACIONAL”: REVISITANDO A DISTINÇÃO ENTRE “LINGÜISTA DE POLTRONA” E “LINGÜISTA APLICADO” Gabriel de Ávila Othero *. **Domínios de Lingu@gem**, p. 1–13, 2008.

OTHERO, G. d. Á. **a Gramática Da Frase Em Português**. [S. n.], 2009. ISBN 9788574308548. Disponível em: <http://www.pucrs.br/orgaos/edipucrs/gramaticadafrase.pdf>.

PANDIAN, S. L.; GEETHA, T. Crf models for tamil part of speech tagging and chunking. In: SPRINGER. **International Conference on Computer Processing of Oriental Languages**. [S. l.], 2009. p. 11–22.

PAPPAS, N.; KATSIMPRAS, G.; STAMATATOS, E. Distinguishing the popularity between topics: a system for up-to-date opinion retrieval and mining in the web. In: SPRINGER. **International conference on intelligent text processing and computational linguistics**. [S. l.], 2013. p. 197–209.

Petrakis, S ; Klenner, M. Learning Theories for Noun-Phrase Sentiment Composition. n. August, p. 179–188, 2011.

POLANYI, L.; ZAENEN, A.; ALTO, P. Contextual valence shifters. **Computing attitude and affect in text: Theory and applications**, p. 1–10, 2006.

PRABOWO, R.; THELWALL, M. Sentiment analysis: A combined approach. **Journal of Informetrics**, v. 3, n. 2, p. 143 – 157, 2009. ISSN 1751-1577. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1751157709000108>.

- Pröllochs, N.; Feuerriegel, S.; Neumann, D. Enhancing sentiment analysis of financial news by detecting negation scopes. In: **2015 48th Hawaii International Conference on System Sciences**. [S. l.: s. n.], 2015. p. 959–968. ISSN 1530-1605.
- RADFORD, A. **Transformational Grammar: A First Course**. Cambridge University Press, 1988. (Cambridge Textbooks in linguistics, v. 2). Disponível em: <https://books.google.com.br/books?id=et2rngEACAAJ>.
- Rana, T. A.; Yu-N Cheah. Hybrid rule-based approach for aspect extraction and categorization from customer reviews. In: **2015 9th International Conference on IT in Asia (CITA)**. [S. l.: s. n.], 2015. p. 1–5.
- Rane, F.; Kauthankar, G.; Naik, A.; Gawas, S. Online product review classification. In: **2019 1st International Conference on Advances in Information Technology (ICAIT)**. [S. l.: s. n.], 2019. p. 228–232.
- Rane, P. S.; Khan, R. A. Ranked rule based approach for sentiment analysis. In: **2018 International Conference on Recent Innovations in Electrical, Electronics Communication Engineering (ICRIEECE)**. [S. l.: s. n.], 2018. p. 2318–2324.
- RANI, S.; KUMAR, P. Rule based sentiment analysis system for analyzing tweets. **2017 International Conference on Infocom Technologies and Unmanned Systems: Trends and Future Directions, ICTUS 2017**, v. 2018-Janua, p. 503–507, 2018.
- Rao, A.; Shah, K. An optimized rule based approach to extract relevant features for sentiment mining. In: **2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)**. [S. l.: s. n.], 2016. p. 2330–2336.
- RAVISHANKAR, N.; SHRIRAM, R. Grammar rule-based sentiment categorisation model for classification of Tamil tweets. **International Journal of Intelligent Systems Technologies and Applications**, v. 17, n. 1-2, p. 89–97, 2018. ISSN 17408873.
- REFORGIATO, D.; SUBRAHMANIAN, V. Ava: Adjective-verb-adverb combinations for sentiment analysis. **IEEE Intelligent Systems**, v. 23, p. 43–50, 07 2008. ISSN 1541-1672. Disponível em: doi.ieeecomputersociety.org/10.1109/MIS.2008.57.
- REIS, J.; GONÇALVES, P.; ARAÚJO, M.; PEREIRA, A. C.; BENEVENUTO, F. Uma Abordagem Multilíngue para Análise de Sentimentos. 2015.
- ROMANYSHYN, M. Rule-based sentiment analysis of Ukrainian reviews. **International Journal of Artificial Intelligence & Applications (IJAIA)**, v. 2, n. 4, p. 91–103, 2011.
- Ruskanda, F. Z.; Widyantoro, D. H.; Purwarianti, A. Comparative study on language rule based methods for aspect extraction in sentiment analysis. In: **2018 International Conference on Asian Language Processing (IALP)**. [S. l.: s. n.], 2018. p. 56–61.
- Ruskanda, F. Z.; Widyantoro, D. H.; Purwarianti, A. Sequential covering rule learning for language rule-based aspect extraction. In: **2019 International Conference on Advanced Computer Science and Information Systems (ICACSIS)**. [S. l.: s. n.], 2019. p. 229–234. ISSN 2330-4588.
- Russu, R. M.; Dinsoreanu, M.; Vlad, O. L.; Potolea, R. An opinion mining approach for romanian language. In: **2014 IEEE 10th International Conference on Intelligent Computer Communication and Processing (ICCP)**. [S. l.: s. n.], 2014. p. 43–46.

SANTOS, A. P.; RAMOS, C.; MARQUES, N. Finding Compositional Rules for Determining the Semantic Orientation of Phrases. 2016.

SATTA, M.-J. N.-G. **Theory of Parsing**. [S. l.]: Clark-Fox-Lappin, 2010.

SAURI, R. **A Factuality Profiler for Eventualities in Text**. Tese (Doutorado), USA, 2008. AAI3304029.

SCHMID, H. Probabilistic part-of-speech tagging using decision trees. In: **New methods in language processing**. [S. l.: s. n.], 2013. p. 154.

SENNRICH, R.; SCHNEIDER, G.; VOLK, M.; WARIN, M. A new hybrid dependency parser for german. **Proceedings of the German Society for Computational Linguistics and Language Technology**, p. 115–124, 2009.

Siddiqua, U. A.; Ahsan, T.; Chy, A. N. Combining a rule-based classifier with ensemble of feature sets and machine learning techniques for sentiment analysis on microblog. In: **2016 19th International Conference on Computer and Information Technology (ICCIT)**. [S. l.: s. n.], 2016. p. 304–309.

SIKKEL, K.; NIJHOLT, A. (Ed.). **Natural Language Parsing: Methods and Formalisms - ACL/SIGPARSE Workshop: Proceedings of the sixth Twente Workshop on Language Technology**. Netherlands: University of Twente, 1993. (Twente Workshop on Language Technology). 6th Twente Workshop on Language Technology, TWLT 6, TWLT ; Conference date: 16-12-1993 Through 17-12-1993.

SMEDT, T. D.; DAELEMANS, W. Pattern for python. **The Journal of Machine Learning Research**, JMLR. org, v. 13, n. 1, p. 2063–2067, 2012.

SOCHER, R.; PERELYGIN, A.; WU, J.; CHUANG, J.; MANNING, C. D.; NG, A.; POTTS, C. Recursive deep models for semantic compositionality over a sentiment treebank. In: **Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing**. Seattle, Washington, USA: Association for Computational Linguistics, 2013. p. 1631–1642. Disponível em: <http://www.aclweb.org/anthology/D13-1170>.

SOUZA, M.; VIEIRA, R. Construction of a portuguese opinion lexicon from multiple resources. **Anais do Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana, 2011, Brasil**, 2011.

SOUZA, M.; VIEIRA, R.; BUSETTI, D.; CHISHMAN, R.; ALVES, I. M.; UNISINOS, F. D. L. Construction of a portuguese opinion lexicon from multiple resources. In: **In 8th Brazilian Symposium in Information and Human Language Technology - STIL, Mato Grosso**. [S. l.: s. n.], 2011.

SRINIVASAN, A. **The aleph manual**. 2001.

Sultanova, N.; Kozhakhmet, K.; Jantayev, R.; Botbayeva, A. Stemming algorithm for kazakh language using rule-based approach. In: **2019 15th International Conference on Electronics, Computer and Computation (ICECCO)**. [S. l.: s. n.], 2019. p. 1–4.

SUN, S.; LUO, C.; CHEN, J. A review of natural language processing techniques for opinion mining systems. **Information Fusion**, Elsevier B.V., v. 36, p. 10–25, 2017. ISSN 15662535. Disponível em: <http://dx.doi.org/10.1016/j.inffus.2016.10.004>.

- VALITUTTI, R. Wordnet-affect: an affective extension of wordnet. In: **In Proceedings of the 4th International Conference on Language Resources and Evaluation**. [S. l.: s. n.], 2004. p. 1083–1086.
- VARGAS, N. L. Analisadores sintáticos de linguagens naturais: uma comparação do estado da arte. Universidade Federal do Pampa, 2015.
- VILARES, D.; GÓMEZ-RODRÍGUEZ, C.; ALONSO, M. A. Universal, unsupervised (rule-based), uncovered sentiment analysis. **Knowledge-Based Systems**, Elsevier, v. 118, p. 45–55, 2017.
- WANG, H.; CAN, D.; KAZEMZADEH, A.; BAR, F.; NARAYANAN, S. A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In: **Proceedings of the ACL 2012 system demonstrations**. [S. l.: s. n.], 2012. p. 115–120.
- WARRINER, A. B.; KUPERMAN, V.; BRYLSBAERT, M. Norms of valence, arousal, and dominance for 13,915 english lemmas. **Behavior research methods**, Springer, v. 45, n. 4, p. 1191–1207, 2013.
- WILSON, T.; HOFFMANN, P.; SOMASUNDARAN, S.; KESSLER, J.; WIEBE, J.; CHOI, Y.; CARDIE, C.; RILOFF, E.; PATWARDHAN, S. Opinionfinder: A system for subjectivity analysis. In: **Proceedings of HLT/EMNLP 2005 Interactive Demonstrations**. [S. l.: s. n.], 2005. p. 34–35.
- XIE, Y.; CHEN, Z.; ZHANG, K.; CHENG, Y.; HONBO, D. K.; AGRAWAL, A.; CHOUDHARY, A. N. MuSES: Multilingual sentiment elicitation system for social media data. **IEEE Intelligent Systems**, v. 29, n. 4, p. 34–42, 2014. ISSN 15411672.
- Yan, D.; Hu, B.; Qin, J. Sentiment analysis for microblog related to finance based on rules and classification. In: **2018 IEEE International Conference on Big Data and Smart Computing (BigComp)**. [S. l.: s. n.], 2018. p. 119–126. ISSN 2375-9356.
- YANG, P.; CHEN, Y. A survey on sentiment analysis by using machine learning methods. In: **IEEE. 2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)**. [S. l.], 2017. p. 117–121.
- Yenkikar, A.; Babu, N.; Sangve, S. R-sa: A rule-based expert system for sentiment analysis. In: **2019 IEEE Pune Section International Conference (PuneCon)**. [S. l.: s. n.], 2019. p. 1–7.
- ZHANG, C.; ZENG, D.; LI, J.; WANG, F.-Y.; ZUO, W. Sentiment analysis of chinese documents: From sentence to document level. **J. Am. Soc. Inf. Sci. Technol.**, John Wiley; Sons, Inc., USA, v. 60, n. 12, p. 2474–2487, dez. 2009. ISSN 1532-2882.
- Zirpe, S.; Joglekar, B. Negation handling using stacking ensemble method. In: **2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA)**. [S. l.: s. n.], 2017. p. 1–5.