



UMA ANÁLISE COMPARATIVA ENTRE LÉXICOS AFETIVOS EM PORTUGUÊS

Rodolpho da Silva Nascimento

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação, do Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, CEFET/RJ, como parte dos requisitos necessários à obtenção do título de mestre.

Orientador(a): Gustavo Paiva Guedes e Silva

Rio de Janeiro,
abril 2021

UMA ANÁLISE COMPARATIVA ENTRE LÉXICOS AFETIVOS EM PORTUGUÊS

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação, do Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, CEFET/RJ, como parte dos requisitos necessários à obtenção do título de mestre.

Rodolpho da Silva Nascimento

Banca Examinadora:

Presidente, Professor D.Sc. Gustavo Paiva Guedes e Silva (CEFET/RJ) (Orientador(a))

Professora D.Sc. Kele Teixeira Belloze (CEFET/RJ)

Professor D.Sc. Raimundo Santos Moura (UFPI)

Rio de Janeiro,

abril 2021

Ficha catalográfica elaborada pela Biblioteca Central do CEFET/RJ

N244 Nascimento, Rodolpho da Silva
Uma análise comparativa entre léxicos afetivos em português /
Rodolpho da Silva Nascimento – 2021.
98f + apêndice : il. (algumas color.), enc.

Dissertação (Mestrado) Centro Federal de Educação
Tecnológica Celso Suckow da Fonseca, 2021.

Bibliografia : f. 85-98

Orientador: Gustavo Paiva Guedes e Silva

1. Computação. 2. Mineração de dados (Computação). 3.
Processamento de linguagem natural (Computação). 4. Sistemas
especialistas (Computação). I. Silva, Gustavo Paiva Guedes e
(Orient.). II. Título.

CDD 004

DEDICATÓRIA

Esta pesquisa é dedicada a Deus, e a pessoa mais especial da minha vida, minha esposa Vanessa. Nesses 22 anos juntos, além de me dar uma linda família, ela contribui para que cada dia meu seja sempre melhor do que ontem.

AGRADECIMENTOS

À Deus, por escrever uma história incrível na minha vida.

À minha esposa Vanessa, por ser o pilar de tudo que tenho e tudo que sou.

À minha filha Rafaely, por ser o meu presente imensurável que Deus me deu.

À Celinha, Relphy, Viviane, Pedro e Raquel, pelo apoio mesmo que distante.

Ao meu professor e orientador Gustavo Guedes, por acreditar em mim e me direcionar em um caminho justo e reto, contribuindo para a materialização de um sonho.

Ao professor Diego Haddad, por me ensinar a compreender um teste qui-quadrado.

À professora Kele Belloze, por me contar a história da travessia do rio.

Ao professor Eduardo Bezerra, por participar da banca de qualificação além de me explicar o funcionamento de um neurônio artificial.

Ao professor Eduardo Ogasawara, pelo grande empenho na coordenação do programa e pelas aulas em inglês. Foram excepcionais.

À professora Rafaelli Coutinho, por me ensinar a diferenciar um *cluster* de um *grid*.

Ao professor Fellipe Duarte (UFRRJ), por aceitar participar da banca de qualificação.

Ao professor Raimundo Moura (UFPI), por aceitar participar da banca de defesa e pelo breve e saudável bate papo pós-congresso, em Belém-PA.

Ao meu amigo Ribamar Matias, pelo apoio incontestável nesta jornada.

Ao meu amigo Antônio Castro (presidente), pelas nossas conversas no trem sobre o curso.

Ao meu amigo Leonardo Ferreira, pela participação muito próxima deste projeto.

Ao meu amigo Flávio Damasceno, pelo grande apoio e por dividir comigo um peixe *filhote* em Belém-PA.

Ao meu amigo Gabriel Nascimento, apoio incontestável ao fornecer códigos em Python.

Ao meu amigo Alexandre Cunha, pela ajuda e nossas risadas em Belém-PA.

Ao meu amigo Rafael Guimarães, por me ajudar a ingressar no curso.

Aos amigos que fiz no CEFET: Adalberto, Alan, Arthur, Augusto, Carlos, Cedric, Daniel, Edson, Fernando, Gustavo Alexandre, Gustavo Pacheco, Jefferson, Heraldo, Ricardo, Jomar, Raphael Abreu, Raphael Fialho, Rafaela, Thiago, e Yukio.

À Rodolfo e Ricardo, pelo grande apoio e incentivo neste projeto.

À todos os citados, expresso aqui o meu sincero MUITO OBRIGADO.

RESUMO

Uma Análise Comparativa Entre Léxicos Afetivos em Português

A Computação Afetiva é uma área de pesquisa que estuda o uso de técnicas que objetivam habilitar um computador a reconhecer emoções por meio da análise e interpretação de dados. Dentro dessa área destaca-se a Análise de Sentimentos (AS), que corresponde a um campo de pesquisa que emprega ferramentas automáticas capazes de extrair informações subjetivas em textos, como opiniões e sentimentos. Existem trabalhos em AS que utilizam Léxicos Afetivos (LA) atuando como apoio aos algoritmos de aprendizagem de máquina, ou como componente principal. Um LA é uma coleção de palavras que capturam o conhecimento que falantes e ouvintes têm sobre expressões lexicais básicas, formando um conjunto de palavras pertencentes ao contexto dos sentimentos humanos. Todavia, poucos trabalhos abordam a utilização de LA na língua portuguesa. Esta pesquisa tem como objetivo realizar uma investigação dos LA disponíveis em português e avaliar suas contribuições em AS. Nesta vertente, pretende-se fornecer recursos para que trabalhos futuros sejam capazes de elaborar novos LA em português ou aprimorar os já existentes. Também pretende-se, com esta pesquisa, ser um referencial para melhor escolha de LA em português, auxiliando trabalhos relacionados e contribuindo para melhores resultados em tarefas de AS em português.

Palavras-chave: Léxicos Afetivos, Computação Afetiva, Análise de Sentimento

ABSTRACT

A Comparative Analysis Between Affective Lexicons in Portuguese

Affective Computing is a research area that studies the use of techniques that aim to enable a computer to recognize emotions through data analysis and interpretation. Sentiment Analysis (SA) is a relevant search field in this area, which corresponds to uses automatic tools capable of extracting subjective information in texts, such as opinions and feelings. There are works in SA that use Affective Lexicons (AL) to support automatic machine learning algorithms or as a main component. An AL is a set of words that capture the knowledge that speakers and listeners have about basic lexical expressions, forming a set of words that belong in human feelings context. However, few studies apply AL in the Portuguese language. This research aims to investigate available AL in Portuguese and evaluate their contributions in SA. In this regard, it is intended to provide resources for future work to develop new LA in Portuguese or improve existing ones. This research is also intended to be a reference for a better choice of AL in Portuguese, helping related works and contributing to better results in SA Portuguese tasks.

Keywords: Affective Lexicons, Affective Computing, Sentiment Analysis

LISTA DE ILUSTRAÇÕES

- Figura 1 – Modelo proposto para reconhecimento afetivo. Os dados são coletados, preparados e seguidamente, enviados à avaliação e identificação de padrões, conseqüentemente possibilitando o reconhecimento da afetividade. Fonte: Elaborada pelo autor. 20
- Figura 2 – Exemplo de utilização do Fleiss' Kappa para a sentença S1. Juizes julgam o sentimento evocado e o Fleiss' Kappa ajuda a compreender a concordância entre si. Fonte: Elaborada pelo autor. 26
- Figura 3 – Exemplo de utilização do Fleiss' Kappa para todas as cinco sentenças. Fonte: Elaborada pelo autor. 27
- Figura 4 – Exemplo de matriz de confusão. A tabela indica a frequência de erros/acertos em um modelo que detecta spam. Fonte: Elaborada pelo autor. 31
- Figura 5 – Exemplo da palavra “triste” estruturada no LIWC2007-PT-br. Fonte: Elaborada pelo autor. 38
- Figura 6 – Exemplo da palavra “triste” estruturada no LIWC2015-PT-br. Fonte: Elaborada pelo autor. 39
- Figura 7 – Visão geral do *framework* elaborado para a execução dos experimentos. Na etapa 1 ocorre a coleta dos dados. Na etapa 2 ocorre as atividades de pré-processamento. Na etapa 3 ocorre a análise de sentimento com apoio da etapa 4 (normalização das valências). Na etapa 5 ocorre a aplicação das medidas de avaliação. Fonte: Elaborada pelo autor. 60

- Figura 8 – Limiar da pontuação para detecção de sentimento. Se o valor da pontuação P for maior ou igual a $0,003$, define-se o sentimento como **positivo**. Se o valor da pontuação P for menor ou igual a $-0,003$, define-se o sentimento como **negativo**. Se o valor da pontuação P estiver entre $-0,003$ e $0,003$, define-se o sentimento como **neutro**. Fonte: Elaborada pelo autor. 62
- Figura 9 – Medida F1 no conjunto de dados MQD. O léxico AffectPT-br apresenta o melhor resultado entre os demais na análise de sentimento. Fonte: Elaborada pelo autor. 69
- Figura 10 – Medida F1 no conjunto de dados SADT. O léxico AffectPT-br apresenta o melhor resultado entre os demais na análise de sentimento. Fonte: Elaborada pelo autor. 69
- Figura 11 – Medida F1 no conjunto de dados TOPIE. O léxico SentiStrength apresenta o melhor resultado entre os demais na análise de sentimento. Fonte: Elaborada pelo autor. 70
- Figura 12 – Medida F1 no conjunto de dados TweetSentBR. O léxico LIWC2015-PT-br apresenta o melhor resultado entre os demais na análise de sentimento. Fonte: Elaborada pelo autor. 70
- Figura 13 – Matriz de confusão - MQD. O eixo x (Predito) representa os valores preditos pelo léxico e o eixo y (Atual) representa os valores rotulados. O sentimento *positivo* é caracterizado nos eixos pelo número 1, seguidamente de -1 para *negativo*, e 0 para *neutro*. Fonte: Elaborada pelo autor. 73
- Figura 14 – Níveis de concordância afetiva entre os léxicos. Nota-se que os léxicos apresentam baixa concordância afetiva, indicando grande divergência entre si ao detectar sentimentos. Fonte: Elaborada pelo autor. 74
- Figura 15 – Fleiss' Kappa dos Léxicos Afetivos mais influentes. Fonte: Elaborada pelo autor. 75

- Figura 16 – Concordância entre palavras do léxico AffectPT-br. X_1 representa palavras equivalentes entre os léxicos. X_2 representa palavras de valência afetiva equivalentes. Observa-se que para o léxico AffectPT-br há maior concordância com o léxico LIWC2015-PT-br. Fonte: Elaborada pelo autor. 76
- Figura 17 – Concordância entre palavras do léxico LIWC2015-PT-br. X_1 representa palavras equivalentes entre os léxicos. X_2 representa palavras de valência afetiva equivalentes. Observa-se que para o léxico LIWC2015-PT-br há maior concordância com o léxico LIWC2007-PT-br. Fonte: Elaborada pelo autor. 76
- Figura 18 – Concordância entre palavras do léxico SentiStrength. X_1 representa palavras equivalentes entre os léxicos. X_2 representa palavras de valência afetiva equivalentes. Observa-se que para o léxico SentiStrength há maior concordância com os léxicos MPQA-PT e Polyglot. Fonte: Elaborada pelo autor. 77
- Figura 19 – Conjunto de dados - MQD. Fonte: Elaborada pelo autor. 103
- Figura 20 – Conjunto de dados - SADT. Fonte: Elaborada pelo autor. 106
- Figura 21 – Conjunto de dados - TOPIE. Fonte: Elaborada pelo autor. 109
- Figura 22 – Conjunto de dados - TweetSentBR. Fonte: Elaborada pelo autor 112
- Figura 23 – Gráfico de concordância entre palavras. X_1 representa palavras equivalentes entre os léxicos. X_2 representa palavras de valência afetiva equivalentes. Fonte: Elaborada pelo autor. 116

LISTA DE TABELAS

Tabela 1	– Exemplo da utilização do Fleiss' Kappa em Análise de Sentimento. Cinco juízes votam em cinco sentenças, classificando-as nas categorias: positivo, negativo ou neutro. A coluna p_i indica a concordância entre os juízes para cada sentença, e a linha p_j indica a concordância entre os juízes para cada categoria.	25
Tabela 2	– Escala de concordância para interpretação do Fleiss' Kappa.	27
Tabela 3	– Medidas de avaliação calculadas para o exemplo apresentado da Matriz de Confusão.	30
Tabela 4	– Strings de busca utilizadas nas pesquisas.	33
Tabela 5	– Léxicos Afetivos em português encontrados na literatura.	34
Tabela 6	– Linhas iniciais do léxico AffectPT-br. 30 representa a categoria <i>affect</i> , 31 a categoria <i>posemo</i> , 32 a categoria <i>negemo</i> e 35 a categoria <i>sad</i> . Em destaque a categoria que indica a valência afetiva da palavra “abandon*” (negativa). O asterisco (*) indica que o léxico desconsidera caracteres subsequentes, abrangendo variações da palavra.	35
Tabela 7	– Linhas iniciais do léxico ANEW-Br. Sua estrutura é composta por valores de <i>valência</i> e <i>alerta</i> , seguidos do desvio padrão (DP). Em destaque a dimensão que indica a valência afetiva da palavra “abandonado” (negativa).	36
Tabela 8	– Linhas iniciais do léxico BabelSenticnet. Sua estrutura é composta por valores de <i>agradabilidade</i> , <i>atenção</i> , <i>sensibilidade</i> , <i>aptidão</i> e <i>polaridade</i> . Em destaque a categoria que indica a valência afetiva da expressão “a_alucinar” (negativa).	37

- Tabela 9 – Linhas iniciais do léxico LIWC2007-PT-br. 146 representa a categoria *bio*, 125 a categoria *affect*, 127 a categoria *negemo* e 129 a categoria *anger*. Em destaque a categoria que indica a valência afetiva da palavra “*abafad**” (negativa). O asterisco (*) indica que o léxico desconsidera caracteres subsequentes, abrangendo variações da palavra. 38
- Tabela 10 – Linhas iniciais do léxico LIWC2015-PT-br. 70 representa a categoria *bio*, 72 a categoria *health*, 20 a categoria *verb*, 115 a categoria *death*, 113 a categoria *money*, 30 a categoria *affect*, 32 a categoria *negemo* e 35 a categoria *sad*. Em destaque a categoria que indica a valência afetiva da palavra “*abatid**” (negativa). O asterisco (*) indica que o léxico desconsidera caracteres subsequentes, abrangendo variações da palavra. 40
- Tabela 11 – Linhas iniciais do léxico MPQA-PT. A letra “A” representa a classificação gramatical da palavra (adjetivo), os termos *weaksbj/strongsubj* representam o “peso” da subjetividade (forte/fraca), atuando como intensificadores emocionais. A valência afetiva é representada pelos termos *positive/positive*. Em destaque a valência afetiva para a palavra “*aborrecer*” (negativa). 41
- Tabela 12 – Linhas iniciais do léxico OpLexicon. O termo *adj* representa um adjetivo, *vb* representa um verbo. Em destaque a valência afetiva da palavra “*acreditado*” (positiva). 42
- Tabela 13 – Linhas iniciais do léxico Polyglot. O número 1 representa um sentimento positivo, -1 representa um sentimento negativo. Em destaque a valência afetiva da palavra “*aberração*” (negativa). 43
- Tabela 14 – Linhas iniciais do léxico ReLi-Lex. O número 1 representa um sentimento positivo, -1 representa um sentimento negativo. Em destaque a valência afetiva da palavra “*amar*” (positiva). 43

Tabela 15 – Linhas iniciais do léxico Sentilex-PT. A categoria Pos=Adj representa um adjetivo, a categoria TG=HUM:N0 representa um nome de tipo humano, a categoria POL:N0=1 representa a polaridade (1 positiva, -1 negativa) e a categoria ANOT=JALC representa que a polaridade foi anotada automaticamente. Em destaque a valência afetiva da palavra “ <i>aceitável</i> ” (positiva).	44
Tabela 16 – Linhas iniciais do léxico SentiStrength. Em destaque a valência afetiva da palavra “ <i>abraços</i> ” (positiva).	45
Tabela 17 – Linhas iniciais do léxico WordnetAffectBR. O símbolo “+” representa um sentimento positivo e “-” representa um sentimento negativo. Em destaque a valência afetiva da palavra “ <i>abrandamento</i> ” (positiva).	46
Tabela 18 – Linhas iniciais do léxico WordnetAffectBR (adapt). O símbolo “+” representa um sentimento positivo e “-” representa um sentimento negativo. Em destaque a valência afetiva da palavra “ <i>agraciado</i> ” (positiva).	47
Tabela 19 – Exemplo de desempenho dos Léxicos Afetivos na utilização de Tradução Automática de Texto (TAT) pelo iFeel. Com a aplicação de TAT, menos léxicos afetivos conseguiram detectar corretamente o sentimento da sentença (5 léxicos). Sem a aplicação de TAT (em destaque), mais léxicos acertaram corretamente o sentimento da sentença (13 léxicos).	49
Tabela 20 – Exemplo de dados do MQD.	56
Tabela 21 – Exemplo de dados do SADT.	57
Tabela 22 – Exemplo de dados do TOPIE.	58
Tabela 23 – Exemplo de dados do TweetSentBR.	59
Tabela 24 – Exemplo de múltiplos tipos de representações dos VA nas palavras.	63
Tabela 25 – Resumo da normalização efetuada nos Lexicos Afetivos. Pode-se notar que alguns lexicos não são normalizados (e.g., Polyglot).	64
Tabela 26 – Tabela dos melhores léxicos classificados pelo teste estatístico de Friedman.	71

Tabela 27 – Medida F1 dos três melhores léxicos. Percebe-se que o léxico AffectPT-br ocupa o 1º lugar nos conjuntos de dados MQD e SADT, ficando em 2º lugar no TweetSentBR e 3º no TOPIE. O léxico LIWC2015-PT-br ocupa o 1º lugar no conjunto de dados TweetSentBR, ficando em 2º no MQD e TOPIE e na 3ª posição no SADT. O léxico SentiStrength ocupa o 1ª lugar no conjunto de dados TOPIE, 2ª lugar no SADT e 3º lugar nos conjuntos de dados MQD e TweetSentBR.	72
Tabela 28 – Escala para interpretação de valores do método Fleiss' Kappa.	74
Tabela 29 – Exemplo de anotações divergentes no corpus.	80
Tabela 30 – Resultados da Análise de Sentimento - MQD.	100
Tabela 31 – Resultados da Análise de Sentimento - SADT.	100
Tabela 32 – Resultados da Análise de Sentimento - TOPIE.	101
Tabela 33 – Resultados da Análise de Sentimento - TweetSentBR.	101

LISTA DE ALGORITMOS

- Algoritmo 1 – Pseudocódigo do método tradicional para detectar sentimento, utilizado pela Abordagem Baseada em Dicionário 24
- Algoritmo 2 – Pseudocódigo para detectar sentimento utilizando a técnica de Abordagem Baseada em Dicionário 61

LISTA DE ABREVIATURAS E SIGLAS

ABD	Abordagem Baseada Em Dicionário
AS	Análise De Sentimentos
CA	Computação Afetiva
LA	Léxico Afetivo
MC	Matriz De Confusão
MQD	Meu Querido Diário - Dataset
PLN	Processamento De Linguagem Natural
RS	Redes Sociais
SADT	Sentiment Analysis Dataset Twitter-pt-br
TAT	Tradução Automática De Texto
VA	Valência Afetiva

SUMÁRIO

Introdução	14
1 Referencial Teórico	19
1.1 Computação Afetiva	19
1.2 Análise de Sentimento	21
1.3 Léxicos Afetivos	22
1.4 Léxicos Afetivos e Análise de Sentimento	23
1.5 Método Fleiss' Kappa	24
1.6 Teste de Friedman	27
1.7 Medidas de avaliação	28
2 Léxicos Afetivos em português	32
2.1 AffectPT-br	33
2.2 ANEW-Br	35
2.3 BabelSenticnet	36
2.4 LIWC2007-PT-br	37
2.5 LIWC2015-PT-br	39
2.6 MPQA-PT	40
2.7 OpLexicon	41
2.8 Polyglot	42
2.9 ReLi-Lex	43
2.10 Sentilex-PT	44
2.11 SentiStrength	45
2.12 WordnetAffectBR	45
2.13 WordnetAffectBR (adapt)	46
3 Trabalhos relacionados	48

3.1	iFeel	48
3.2	Comparação entre Léxicos Afetivos para análise de sentimento no Twitter	50
3.3	Revisão Sistemática sobre Léxicos Afetivos para o Português do Brasil	51
3.4	Um levantamento da análise de sentimento na língua portuguesa	51
3.5	Comparação sistemática de ferramentas automatizadas de análise de conteúdo	52
4	Metodologia	55
4.1	Linguagem de programação	55
4.2	Conjunto de dados	55
4.2.1	MQD	56
4.2.2	SADT	57
4.2.3	TOPIE	57
4.2.4	TweetSentBR	58
4.3	Framework para Análise de Sentimento	59
4.4	Método Utilizado para Análise de Sentimento	60
4.5	Normalização das valências dos Léxicos Afetivos	62
5	Resultados	68
5.1	Avaliação dos Léxicos Afetivos	68
5.2	Seleção dos melhores léxicos para apresentação dos resultados	70
5.3	Matriz de Confusão	72
5.4	Concordância Afetiva nas sentenças	73
5.5	Concordância nas palavras entre os Léxicos Afetivos	75
6	Discussão dos Resultados	78
6.1	Precisão, revocação, Acurácia e F1	78
6.2	Utilização de conjunto de dados de diferentes domínios	79
6.3	Anotações incorretas no corpus	80
6.4	Divergências entre os Léxicos Afetivos	81
6.5	Melhores Léxicos Afetivos	82
7	Conclusão	83
	Referências	84
A	Apêndice: Resultados de acurácia, revocação, precisão e F1	99

B	Apêndice: Matriz de Confusão	102
C	Apêndice: Concordância nas palavras entre os Léxicos Afetivos	115

Introdução

A Internet tem se mostrado ser de grande relevância para a sociedade moderna, mudando a forma de como os indivíduos se comunicam. Por meio da Internet, um simples compartilhamento de mensagens facilita atividades rotineiras da sociedade, como por exemplo, uma negociação de um bem, produto ou serviço. Neste contexto, os aplicativos e plataformas *online* de Redes Sociais (RS) se destacam no processo de comunicação entre os indivíduos. As RS proporcionam o relacionamento em espaços virtuais criados pelos próprios usuários, destinados a assuntos de interesses mútuos, resultando em processos de difusão da informação [Araya et al., 2019].

No Brasil, o uso da Internet vem ganhando grande adesão pelos indivíduos, tornando-se cada vez mais popular entre os brasileiros. Uma pesquisa realizada pelo IBGE¹ indicou que o uso da Internet em domicílios brasileiros tem crescido de forma rápida, dado que em 2018 a Internet era utilizada em 79,1% dos domicílios permanentes do país e no ano seguinte (2019), este número aumentou para 82,6% [PNAD, 2019].

Ainda segundo o IBGE, cerca de 143 milhões de pessoas no Brasil utilizam a Internet e desse número, 95% afirmam enviar ou receber mensagens que não sejam *e-mails* (e.g. redes sociais) [PNAD, 2019]. Sendo assim, o uso das RS tem contribuído para que o mundo virtual se torne cada vez mais uma extensão da vida de seus usuários [Althoff et al., 2017; Sajadi et al., 2018].

Uma área que vem estudando a vasta quantidade de dados produzidos pelas RS e tem buscado compreender seus aspectos é a Computação Afetiva (CA) [Zhao et al., 2019]. A CA é uma área de pesquisa que estuda o uso de técnicas que permitem que um computador seja capaz de reconhecer, emular ou influenciar emoções por meio da análise e interpretação de dados [Picard, 2000], provenientes de imagens, sons, sensores ou textos. Para atingir seus objetivos, a CA depende de contribuições de áreas distintas da computação, como psicologia, fisiologia, engenharia, sociologia, matemática, educação e linguística, expondo então a complexidade em descrever, compreender e emular a dinâmica do sentimento [Daily et al., 2017].

Nesse cenário, a Análise de Sentimentos (AS) destaca-se em desempenhar um

¹<https://www.ibge.gov.br/>

papel indispensável na transformação dos atuais sistemas de Inteligência Artificial, para a próxima geração de dispositivos emocionais inteligentes [Picard, 2000; Minsky, 2007]. A AS (também conhecida como, Mineração de Opinião, Análise de Opinião, Extração de Opinião, Mineração de Sentimento, Análise de Subjetividade, Análise Afetiva, Análise de Emoção e Mineração de Revisão [Liu, 2017]), é um campo de pesquisa que tem como objetivo definir ferramentas automáticas que sejam capazes de extrair informações subjetivas em textos, como opiniões e sentimentos, de modo a criar um conhecimento estruturado e acionável, podendo ser utilizado por qualquer sistema de apoio à tomada de decisão [Pozzi et al., 2017].

Em processos de tomada de decisão, um importante item a ser considerado é “saber o que as pessoas pensam” [Pang and Lee, 2008]. Obter previamente uma opinião pública sobre um produto ou serviço pode oferecer vantagens relevantes à uma organização, permitindo que a mesma molde seu produto ou serviço possibilitando então maior aceitação pública. Isso ocorre de forma que a empresa se adapte às necessidades daqueles que expressaram suas opiniões, atendendo seus anseios e mitigando desperdícios econômicos, causados por uma possível rejeição.

A Internet aparece nesse contexto, permitindo que uma grande quantidade de opiniões e experiências sejam expostas por pessoas que não conhecemos e não são críticos profissionais [Pang and Lee, 2008]. Essas experiências, por muitas vezes, podem auxiliar outros indivíduos no processo de tomada de decisão sobre a aquisição de um produto ou serviço. Por exemplo, o site ReclameAqui² concentra opiniões públicas sobre o consumo de produtos ou serviços, classificando estas opiniões como negativa ou positiva, e conseqüentemente gerando potencial para influenciar indivíduos.

A análise de opiniões expostas em RS por seus usuários constituem uma tendência que vem despertando diversas pesquisas sobre AS [Wang et al., 2016; Shahare, 2017; Wang and Castanon, 2015; Sanguansat, 2016; Cirqueira et al., 2018]. Estudos relacionados a AS, abordam como técnica principal os algoritmos de aprendizagem de máquina [Rumelli et al., 2019; Zhang and Zheng, 2016; Poornima and Priya, 2020; S. and Ramathika, 2019] ou até mesmo a utilização das redes neurais profundas [Goularas and Kamis, 2019; Quan et al., 2018; Wang et al., 2019; Cheng and Tsai, 2019]. Todavia, há estudos que adotam como técnica principal o uso exclusivo de Léxicos Afetivos [Sabra et al., 2017; Rabeya et al., 2019; Thavareesan and Mahesan, 2020; Tanantong et al., 2020].

²<https://www.reclameaqui.com.br/>

Um Léxico Afetivo (LA) consiste em um conjunto de palavras e/ou frases comumente utilizadas em contextos afetivos [Liu, 2017], selecionadas por técnicas de linguística [Clore et al., 1987]. Tais frases ou palavras, são associadas à uma Valência Afetiva (VA) (também chamada de polaridade afetiva ou orientação semântica) que indica se uma palavra/frase transmite um conteúdo **positivo** ou **negativo** [Bestgen et al., 2008]. Utilizar um LA em tarefas de AS, possibilita que um sentimento expresso em um texto seja identificado.

A abordagem da utilização de LAs em tarefas de AS, pretende aplicar uma cobertura mais abrangente em detectar sentimentos [Ahire, 2014]. Desta maneira, um LA objetiva ser capaz de classificar corretamente grandes quantidades de textos não vistos no processo de elaboração do léxico, atuando de maneira diferente dos algoritmos de aprendizagem de máquina. Estes algoritmos (e.g. Naive Bayes, SVM, Decision Tree) são treinados utilizando um conjunto limitado de exemplos, e também são dependentes de uma correta anotação no corpus [Pustejovsky and Stubbs, 2012]. Embora esta abordagem produza bons resultados, a mesma pode não produzir acurácias semelhantes, quando o algoritmo for aplicado em corpus diferentes daquele utilizado em seu treinamento e validação [Guellil et al., 2018].

A AS com base no uso de LAs vem despertando muitas pesquisas relacionadas. Dentre essas pesquisas, a maioria utiliza LAs construídos para língua inglesa [Liu, 2017; Agarwal et al., 2016; Imran et al., 2020; Gupta et al., 2020]. Pesquisas semelhantes que abordem a criação e avaliação de LAs construídos em outras línguas não são comuns [Asghar et al., 2019]. Neste sentido, poucos trabalhos abordam a construção de LAs na língua portuguesa [Avanco and Nunes, 2014], o que torna essa tarefa um desafio e potencialmente, objeto de investigação nesta área.

Dentre as pesquisas que se propõem a realizar AS baseadas em LAs utilizando conjunto de dados na língua portuguesa, algumas empregam uma técnica conhecida como Tradução Automática de Texto (TAT) [Gelbukh et al., 2004], que compreende na conversão automática de um texto, mudando-o de uma língua para outra e preservando sua semântica. Santos et al. [2014] elaborou um estudo que converteu textos do português para o inglês, possibilitando a utilização de um LA disponível em inglês para AS, e destacou ainda a capacidade desta técnica em suportar multi-idiomas.

No entanto, nem sempre a utilização de TAT produz um resultado consistente. Estudos indicam que a aplicação de TATs não são absolutamente fidedignas, demandando

um processo de intervenção humana [Rodrigues et al., 2020], sendo este um problema recorrente [Eskander and Rambow, 2015]. Desta forma, enfatiza-se o conceito de que a aplicação de TAT não é suficiente para tarefas de AS utilizando LAs em línguas não nativas [Basiri and Kabiri, 2017], indicando a importância da disponibilização de LAs em línguas nativas para AS.

Além disso, pode-se destacar dois grandes problemas que são destaques em trabalhos relacionados à AS abordando a utilização de LAs: 1) *“Como identificar e lidar com palavras e frases que têm sentimento dependente de domínio ou de contexto (e.g. não gosto)”*; e 2) *“Como identificar palavras e frases factuais que impliquem sentimento (e.g. minha nossa)”* [Liu, 2017]. Ainda segundo o autor, para o primeiro problema quase sempre há algumas palavras ou frases em um domínio que expressam orientações diferentes em alguns contextos, distanciando-se das suas orientações padrões de um LA para uso geral, podendo potencialmente ocasionar baixo desempenho na AS. O segundo problema é ainda mais difícil de resolver, porque existem muitas possibilidades e muitas vezes é necessário um senso comum e conhecimento prévio de domínio para melhor compreensão do sentimento. Há pesquisas que objetivam resolver tais problemas elaborando novos LAs [Huang et al., 2014] ou ampliando-os [Lu et al., 2011]. Todavia, poucas pesquisas foram feitas para resolver esses problemas [Liu, 2017].

Segundo González-Bailón and Paltoglou [2015], a grande variedade de LAs disponíveis é resultado de sua capacidade em oferecer um método direto para classificar documentos em tarefas de AS. Ainda segundo os autores, os LAs simplesmente precisam de uma contagem de frequência de palavras e alguma lógica para ponderar e agregar as pontuações de sentimento associada às palavras que fazem parte de sua estrutura .

Em contraste com a abordagem de aprendizagem de máquina (ML), os métodos baseados em léxicos são métodos independentes de domínio, segundo Basiri and Kabiri [2018]. Eles ainda destacam que os LAs não precisam de um grande corpus de treinamento anotado e portanto, são mais rápidos, o que torna a abordagem baseada em LA prevalente na comunidade de AS .

É relevante destacar que com a disponibilização de muitos LAs, inconsistências e erros são inevitáveis [Liu, 2017]. Consequentemente, emerge a demanda por estudos que objetivem avaliar a estrutura de tais LAs, checando a consistência afetiva das palavras que compõem estes léxicos, objetivando compreender melhor a distribuição de emoções evocadas por termos em diferentes partes de discurso [Mohammad and Turney, 2010].

É importante reconhecer a relevância das pesquisas de AS em textos, e dos benefícios que tais trabalhos proporcionam quando aplicados à sociedade. Também destaca-se os problemas existentes para os LAs citados previamente (e.g. como lidar com palavras com sentimento dependente de domínio, problemas na utilização de TAT, consistência afetiva entre os LAs), acrescido da baixa quantidade de estudos que abordem LAs em português e a ausência de trabalhos investigativos sobre o desempenho de LAs em português. Sendo assim, torna-se muito importante investigar e avaliar a contribuição que os LAs disponíveis na língua portuguesa trazem quando utilizados em tarefas de AS.

Esta pesquisa tem como objetivo fazer um levantamento dos LAs existentes na língua portuguesa, realizar experimentos que avaliem o desempenho dos léxicos, em destaque a aplicação dos LAs em diferentes domínios, e evidenciar os melhores para auxiliar no desenvolvimento de trabalhos com foco em métodos baseados em LAs para efetuar AS em português.

Enfatiza-se que esta pesquisa, além de fornecer dados para avaliação de LAs em português, pretende ser um referencial para estudos que objetivem elaborar ou aprimorar o desempenho de LAs disponíveis em português. Para que os resultados se apresentem de maneira generalizada (diferentes domínios), considera-se indispensável a utilização de diferentes conjuntos de dados nos experimentos. Para esta pesquisa, foi utilizado conjunto de dados estruturados em sentenças extraídos de RS brasileiras.

A estrutura desta pesquisa está dividida na seguinte forma: no Capítulo 1 é apresentado o **Referencial Teórico**, contendo os conceitos necessários para compreensão das abordagens utilizadas nesta pesquisa. No Capítulo 2 são descritos os **LAs em português**, bem como o processo de seleção e descrição das características de cada um. No Capítulo 3 encontra-se os **Trabalhos Relacionados**, debatendo seus pontos fracos e a correlação de contribuição desta pesquisa. No Capítulo 4 descreve-se as características da **Metodologia** utilizada nesta pesquisa para obter seus resultados, como linguagem de programação utilizada, conjunto de dados selecionados e desenvolvimento de *framework* para apoio. No Capítulo 5 são apresentados os **Resultados** obtidos pelos experimentos executados nesta pesquisa, bem como tabelas e gráficos para visualização dos dados. No Capítulo 6 apresenta-se a **Discussão dos Resultados** e no Capítulo 7, a **Conclusão** de toda a investigação aplicada nesta pesquisa.

1- Referencial Teórico

Este capítulo apresenta os conceitos fundamentais para a compreensão da proposta desta pesquisa, começando pelos princípios de Computação Afetiva e Análise de Sentimento. Em seguida, aborda-se os conceitos teóricos de Léxicos Afetivos bem como sua relação com Análise de Sentimento. Posteriormente apresenta-se conceito das medidas estatísticas Fleiss' Kappa e Teste de Friedman e por fim, conceitos teóricos referente às medidas de avaliação de modelos preditivos.

1.1- Computação Afetiva

A CA é um campo de estudo que objetiva habilitar os computadores para que eles sejam capazes de detectar ou simular sentimentos, de forma muito próxima à dos seres humanos [Picard, 2000]. Por meio da leitura de dados, que podem ser provenientes de imagens, sons, sensores ou textos, um computador pode identificar padrões e detectar o estado emocional de um indivíduo, detectando uma **emoção** (*tristeza, felicidade, ansiedade, raiva, etc...*), ou um **sentimento** (*positivo, negativo, neutro*), interligando a afetividade humana com uma máquina. A Figura 1 exemplifica um modelo proposto de reconhecimento afetivo por um computador.

Capacitar um computador para reconhecer ou simular emoções, não só o torna mais inteligente, mas também traz benefícios em setores importantes da sociedade [Picard, 2000]. Por exemplo, saber se um programa de computador dispõe de recursos que tornem o seu uso agradável ao usuário, pode ser aferido por meio da leitura de expressões faciais deste usuário, utilizando uma câmera para capturar essas expressões [Medjden et al., 2020]; um filme pode ser classificado como bom, ruim, ou mediano, ao analisar comentários expostos em blogs, fóruns ou sites do gênero [Patel et al., 2021]; um programa de TV pode ser avaliado em tempo real, analisando emoções expressas em conteúdos de RS como Twitter [Yamauchi et al., 2013]; uma empresa pode avaliar a aceitação popular de um potencial produto ou serviço, por meio de opiniões expostas

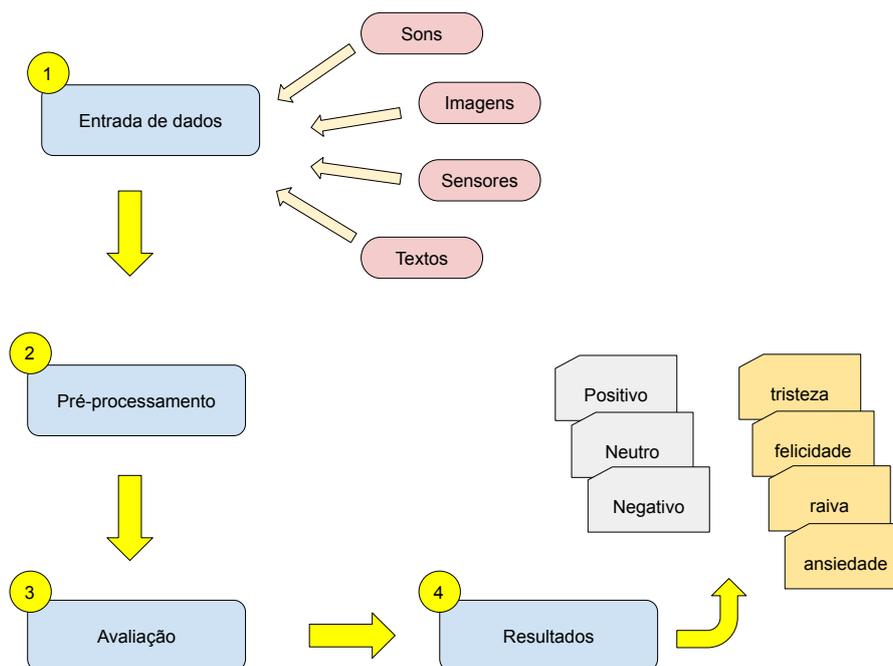


Figura 1 – Modelo proposto para reconhecimento afetivo. Os dados são coletados, preparados e seguidamente, enviados à avaliação e identificação de padrões, consequentemente possibilitando o reconhecimento da afetividade. Fonte: Elaborada pelo autor.

em RS [Nascimento et al., 2020]; sintomas de manifestações depressivas podem ser detectados em comentários expostos em RS [da Silva Nascimento et al., 2019]. Tais exemplos demonstram a relevância da CA aplicada à sociedade.

Um computador que identifica o estado emocional de um indivíduo, observando e compreendendo o comportamento humano, pode tornar mais eficiente um processo de tomada de decisão [Picard, 2000]. Ao lidar melhor com problemas que envolvam inspeção afetiva em larga escala, ele facilita a análise de comportamento social ao observar grande quantidade de dados, tarefa esta que seria extremamente onerosa e bastante suscetível a erros, se realizada por indivíduos. Neste sentido, o processo de tomada de decisão torna-se mais consistente, ao considerar uma quantidade maior de exemplos analisando-os automaticamente. Sendo assim, é muito apropriado examinar como os sentimentos podem ser incorporados nos modelos de inteligência computacional, e suas interações com humanos [Picard, 2000].

O uso da CA para reconhecimento de emoções e sentimentos em indivíduos é de grande importância para compreensão de comportamento humano, podendo ser aplicada em larga escala e de maneira automatizada por computadores. A CA não só contribui

para uma qualidade de interação mais rica entre homem x máquina, mas ela também afeta diretamente a capacidade de um sistema interagir de forma mais inteligente [Picard, 2000].

1.2- Análise de Sentimento

A AS, também conhecida como Mineração de Opinião [Pang and Lee, 2008], Análise de Opinião, Extração de Opinião, Mineração de Sentimento, Análise de Subjetividade, Análise de Afeto, Análise de Emoção e Mineração de Revisão, é um campo de estudo na computação que objetiva detectar emoções, opiniões, sentimentos e atitudes dos indivíduos expressos em textos [Liu, 2017]. Engaja-se como objetivo principal, na construção de ferramentas automáticas capazes de extrair informações subjetivas de textos, como opiniões e sentimentos, de modo a criar um conhecimento estruturado e acionável para ser usado por qualquer sistema de suporte à decisão, ou um tomador de decisão [Pozzi et al., 2017].

A AS em documento é potencialmente o tópico mais estudado ultimamente [Liu, 2017]. Consiste em analisar um documento, extrair o sentimento expresso e determinar a sua orientação emocional. Dentre os tipos de tarefas de AS, destaca-se a detecção de **sentimento**, que é a orientação afetiva em um texto, que pode ser caracterizada da seguinte maneira: **positiva**, **negativa** ou **neutra**. Exemplificando, a frase *“Hoje eu estou muito feliz, pois ganhei algo muito desejado!”* representa a expressão de um sentimento **positivo**. Em outro exemplo, a frase *“Odeio quando me acordam cedo...”*, representa a expressão de um sentimento **negativo**. Já a frase *“A lua gira em torno da Terra.”*, representa uma expressão de sentimento **neutro**.

Em AS existem três diferentes níveis de abordagem para análise de textos: **nível de documento**, **nível de aspecto** e **nível de sentença** [Liu, 2017]. Para o **nível de documento**, a detecção do sentimento consiste na análise de todo conteúdo apresentado, considerando todas as sentenças dispostas no documento.

Para o **nível de aspecto**, a detecção do sentimento consiste em uma análise de fina granularidade. Neste nível, o objetivo é descobrir o sentimento sobre entidades e/ou seus aspectos. Em um documento ou sentença, diversas opiniões podem estar presentes.

Para detectar o sentimento do texto, é preciso ter uma perspectiva diferente, analisando o contexto. Por exemplo, a frase “*O smartphone pode ser bom, mas a bateria tem pouca duração ou o preço é caro*”, observando-a de forma literal em destaque a expressão “pode ser bom”, remete a ideia de um **sentimento positivo**. Entretanto, ao observar o contexto da frase é evidente que o sentimento presente no texto é um **sentimento negativo**.

Para o **nível de sentença**, a detecção do sentimento consiste apenas na análise de uma frase apresentada, independente de esta frase estar contida em um documento ou não. Esta abordagem torna mais próximo a AS da vida real [Liu, 2017] e portanto, será adotada no processo de AS previsto nesta pesquisa.

1.3- Léxicos Afetivos

Em termos gerais, um léxico é uma coleção de palavras que capturam o conhecimento que falantes e ouvintes têm sobre expressões lexicais básicas em uma língua [Pustejovsky and Batiukova, 2019]. Um LA é um conjunto de palavras ou frases (e.g. bom, incrível, ruim, horrível) que as pessoas costumam expressar opiniões positivas ou negativas [Liu, 2017]. A palavra é o elemento mais significativo para detectar um sentimento presente em um texto, sendo caracterizada neste contexto como **palavras de opinião** [Liu, 2017]. Por exemplo, a palavra *feliz* evoca um sentimento *positivo*, ao contrário da palavra *triste*, que evoca um sentimento *negativo*.

Os LAs são elaborados utilizando técnicas psicolinguísticas, selecionando palavras comumente utilizadas em expressões de sentimento [Clore et al., 1987]. A finalidade é reservar-se exclusivamente de palavras que representem o domínio dos sentimentos humanos [Zhang and Singh, 2014]. Estas palavras normalmente são associadas à uma VA, indicando se a palavra evoca um sentimento **negativo** (e.g. ruim, péssimo), **positivo** (e.g. feliz, alegre) ou **neutro** (e.g. calor, chuva). Assim, dada a palavra p , seu valor v pode ser representado por um valor numérico de -1 , 1 ou 0 . A equação 1 demonstra a representação numérica destes valores afetivos. Palavras associadas com valor “1” indicam sentimento positivo, “-1” sentimento negativo e “0” sentimento neutro.

$$p(v) = \begin{cases} 1 & \rightarrow \textit{positivo} \\ -1 & \rightarrow \textit{negativo} \\ 0 & \rightarrow \textit{neutro} \end{cases} \quad (1)$$

1.4- Léxicos Afetivos e Análise de Sentimento

Em tarefas de AS, os LAs comumente são utilizados como componente principal, ou como componente de apoio aos *frameworks* de aprendizagem de máquina. Nas abordagens que utilizam algoritmos de aprendizagem de máquina (e.g. Naive Bayes, SVM, MLP, Linear Regression, etc...), os LAs apoiam os algoritmos melhorando seus resultados [Mohamad Beigi and Moattar, 2021]. Nas abordagens que utilizam os LAs como componente principal, aplica-se uma técnica que consiste em extrair os valores associados às palavras do LA, e estes valores são computados gerando uma pontuação que possibilita detectar o sentimento expresso no texto [Taboada et al., 2011]. Esta técnica é conhecida como Abordagem Baseada em Dicionário (ABD) [Liu and Zhang, 2012].

A ABD geralmente atua aplicando um método tradicional, que consiste em selecionar palavras equivalentes entre o LA e a sentença [Park and Kim, 2016]. Cada palavra do LA possui uma VA, e os valores associados em cada palavra da sentença viabiliza a geração de uma pontuação que conseqüentemente, possibilita a detecção de sentimento expresso na sentença [Hu and Liu, 2004]. O Algoritmo 1 apresenta um pseudocódigo da aplicação desta abordagem.

Um exemplo está na sentença “*Você me faz bem e eu sinto que sou amada verdadeiramente*”. Selecionando palavras e suas respectivas VAs do léxico, teríamos: “bem”(+1) e “amada”(+1), e que a soma dos valores ($1 + 1 = 2$) indicariam um sentimento **positivo** (i.e. maior que 1). A sentença “*Detesto este tipo de comida*”, é um exemplo de sentimento **negativo**, dado que a palavra selecionada pelo léxico seria somente “detesto”(−1). Já a sentença “*Achei legal você não ter destruído meu trabalho*”, indicaria um exemplo de sentimento **neutro**, dado que “legal”(+1) e “destruído”(−1), possui um

Algoritmo 1 – Pseudocódigo do método tradicional para detectar sentimento, utilizado pela Abordagem Baseada em Dicionário

```

1  Entrada
2      sentença  $S$ 
3      léxico  $L$ 
4  Saída
5      sentimento  $R$ 
6
7  Procedimento DetectarSentimento ( $S, L$ )
8      pontuação  $P \leftarrow 0$ 
9      valência afetiva  $v \leftarrow 0$ 
10
11     para cada palavra  $p$  em  $S$  faça:
12         para cada palavra  $q$  em  $L$  faça:
13             se  $p = q$  então:
14                  $v \leftarrow \text{valenc}(q, L)$ 
15                 ( $\text{positivo} = 1; \text{negativo} = -1; \text{neutro} = 0$ )
16                  $P \leftarrow P + v$ 
17             fim-se
18         fim-para
19     fim-para
20     se  $P \geq 1$  então:
21          $R \leftarrow \text{positivo}$ 
22     senão se  $P \leq -1$  então:
23          $R \leftarrow \text{negativo}$ 
24     senão:
25          $R \leftarrow \text{neutro}$ 
26     fim-se
27
28     retorna  $R$ 

```

resultado igual a: $+1 - 1 = 0$.

1.5- Método Fleiss' Kappa

O Fleiss' Kappa [Fleiss, 1971] é um teste de concordância que objetiva identificar um ponto de equilíbrio em sistemas de votação que utilizem variáveis categóricas ou numéricas, determinando o grau de concordância entre dois ou mais juízes. O Fleiss' Kappa é definido pela equação 2, sendo o denominador k indicando o grau de concordância.

$$k = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e} \quad (2)$$

Podemos exemplificar a utilização do Fleiss' Kappa em um caso de AS, ao apresentar um conjunto de sentenças à um grupo “juízes” que devem “votar” o sentimento evocado em um conjunto de sentenças. Cinco juízes (n) atribuem em conjunto de cinco sentenças (N) as três (k) possíveis categorias (positivo, negativo, neutro) conforme exemplo exposto na Tabela 1. O valor da coluna p_i para a primeira sentença (S1) indica a concordância na votação entre os demais juízes, variando em uma escala de 0 – 1 a qual 1 indica concordância plena e 0 nenhuma concordância (Figura 2). Portanto, o valor de p_i é calculado como:

$$p_i = \frac{1}{n(n-1)} \left(\sum_{j=1}^k n_{ij}^2 j - n \right) \quad (3)$$

ou

$$1/(5 \cdot (5 - 1)) \cdot ((3^2 + 1^2 + 1^2) - 5) = 0,05 \cdot 11 = 0,55 \quad (4)$$

Sentenças	Categorias			p_i
	positivo	negativo	neutro	
S1	3	1	1	0,55
S2	0	5	0	1,00
S3	1	0	4	0,60
S4	2	1	2	0,20
S5	3	0	2	0,40
Total de votos	9	7	9	
p_j	0,36	0,28	0,36	

Tabela 1 – Exemplo da utilização do Fleiss' Kappa em Análise de Sentimento. Cinco juízes votam em cinco sentenças, classificando-as nas categorias: positivo, negativo ou neutro. A coluna p_i indica a concordância entre os juízes para cada sentença, e a linha p_j indica a concordância entre os juízes para cada categoria.

O valor da linha p_j indica a concordância entre os juízes para cada categoria disposta. Neste exemplo, a categoria **positivo** proporcionalmente às demais categorias

tem o seu valor de p_j calculado como:

$$p_j = \frac{1}{N(n)} \sum_{i=1}^N n_{ij} \quad (5)$$

ou

$$1 / 25 \cdot 9 = 0,36 \quad (6)$$

Calculando \bar{P} e \bar{P}_e , temos:

$$\bar{P} = \frac{1}{N} \sum_{i=1}^N p_i = (0,55 + 1,00 + 0,60 + 0,20 + 0,40) / 5 = 0,55 \quad (7)$$

$$\bar{P}_e = \sum_{j=1}^N p_j^2 = 0,36^2 + 0,28^2 + 0,36^2 = 0,3376 \quad (8)$$

O Fleiss' Kappa para as cinco sentenças (Figura 3) então é calculado como:

$$Kappa = (0,55 - 0,3376) / (1 - 0,3376) = 0,320 \quad (9)$$

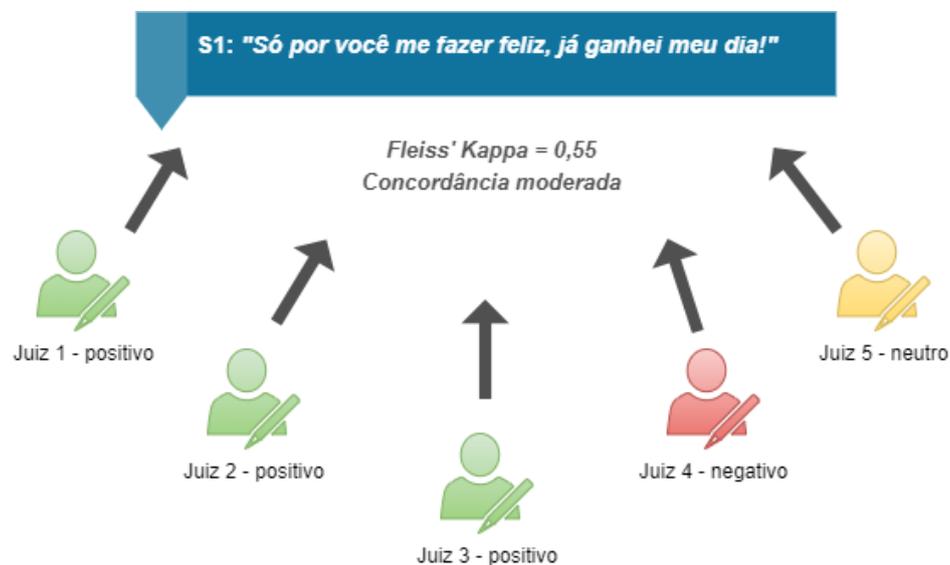


Figura 2 – Exemplo de utilização do Fleiss' Kappa para a sentença S1. Juízes julgam o sentimento evocado e o Fleiss' Kappa ajuda a compreender a concordância entre si. Fonte: Elaborada pelo autor.

Uma tabela auxilia na interpretação do método Fleiss' Kappa (Tabela 2) facilitando a compreensão de como os juízes concordam entre si em sistemas de votação [Landis

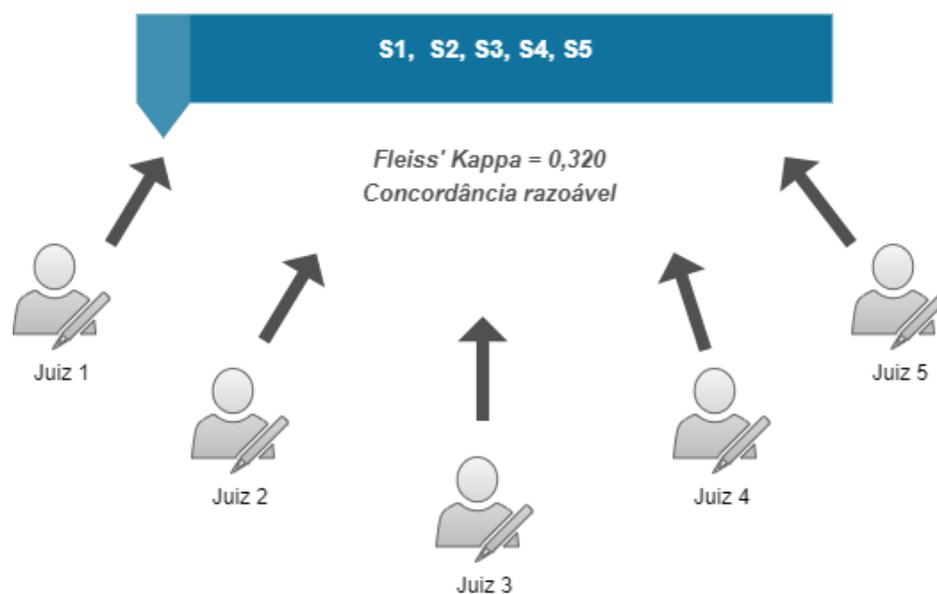


Figura 3 – Exemplo de utilização do Fleiss' Kappa para todas as cinco sentenças. Fonte: Elaborada pelo autor.

and Koch, 1977]. A utilização do Fleiss' Kappa é bem recorrente em trabalhos de AS [Al-Thubaity et al., 2018; Roy et al., 2019; Bian et al., 2018].

Tabela 2 – Escala de concordância para interpretação do Fleiss' Kappa.

Valor	Interpretação
< 0,00	Nenhuma ou quase nenhuma concordância
0,00–0,20	Breve concordância
0,21–0,40	Concordância razoável
0,41–0,60	Concordância moderada
0,61–0,80	Concordância considerável
0,81–1,00	Concordância plena

1.6- Teste de Friedman

O Teste de Friedman [Friedman, 1937] é um teste estatístico não-paramétrico, utilizado para comparar dois ou mais grupos de mesmo tamanho, em observância à uma

variável numérica qualitativa (*ranked data*). A hipótese nula (H_0) afirma que K variáveis numéricas procedem da mesma população ou de populações com a mesma mediana.

O procedimento consiste primeiramente em ordenar os dados de cada grupo numa tabela. Em seguida, um teste é aplicado para avaliar se as diferentes colunas da tabela ordenada (variável numérica K) pode-se supor que todas procedem do mesmo universo. Este teste é feito pelo cálculo das classificações médias para as K colunas de um teste estatístico X^2 , que tende a ser distribuída de acordo com a distribuição X^2 usual, quando a classificação é, de fato, aleatória ou seja, quando o fator testado não tem influência.

1.7- Medidas de avaliação

Algumas medidas auxiliam na compreensão dos resultados, quando se pretende conhecer o quão bom é um modelo de predição [Arendt et al., 2020]. Quantificar o desempenho de um modelo preditivo sabendo o quanto ele erra ou acerta, auxilia na comparação entre modelos proporcionando uma melhor escolha [Han et al., 2011].

Para melhor compreensão, existem quatro termos que representam elementos utilizados para computar medidas de avaliação em modelos preditivos [Han et al., 2011]. É importante entendê-los pois desta maneira, auxiliarão na compreensão das medidas de avaliação. São eles:

- **Verdadeiros positivos (VP)**: Representa o número de exemplos verdadeiros que o modelo acertou ao detectar no conjunto de dados observado (e.g. sentimento positivo? → sim, é verdadeiro; valor previsto pelo modelo → sim, é verdadeiro).
- **Verdadeiros negativos (VN)**: Representa o número de exemplos que não são verdadeiros e que o modelo acertou ao detectar no conjunto de dados observado (e.g. sentimento positivo? → não, é falso; valor previsto pelo modelo → não, é falso).
- **Falsos positivos (FP)**: Representa o número de exemplos positivos que o modelo errou ao detectar no conjunto de dados observado (e.g. sentimento positivo? → não, é falso; valor previsto pelo modelo → sim, é verdadeiro).

- **Falsos negativos (FN):** Representa o número de exemplos negativos que o modelo errou ao detectar no conjunto de dados observado (e.g. sentimento positivo? → sim, é verdadeiro; valor previsto pelo modelo → não, é falso).

Nesta vertente, destacamos algumas medidas que são bastante utilizadas para avaliação de modelos preditivos [Olson and Delen, 2008] e também em pesquisas relacionadas à AS [Ribeiro et al., 2016]. São elas: precisão, revocação, acurácia, e F1.

A **precisão** é a medida que avalia o quão bom foi o modelo de predição em relação aos exemplos VP selecionados, considerando o que ele acertou como VP e o que ele errou ao desconsiderar VP [Han et al., 2011]. Observando os exemplos de VP selecionados pelo modelo, avalia-se o quão desses exemplos são de fato VP. Podemos exemplificar em um modelo que tenta identificar cães em um banco de imagens com diversos animais. Neste banco de imagens dos 10 exemplos disponíveis, 4 são cães e 6 são outros animais. O modelo então prediz 5 exemplos como cães. Ao checar os exemplos identificados pelo modelo, dentre estes, 3 são de fato cães (acertou, VP) e 2 são outros animais (errou, FP). Neste caso, sua precisão é de $\frac{3}{3+2} = 0.60$. A Equação 10 exhibe o modelo matemático para o cálculo da precisão.

$$Precisão = \frac{VP}{VP + FP} \quad (10)$$

A **revocação** é a medida a qual objetiva-se quantificar os elementos relevantes extraídos pelo modelo, considerando os elementos VP [Han et al., 2011]. A finalidade é avaliar o quão todos os exemplos de VP o modelo consegue abranger corretamente. Utilizando o exemplo citado no parágrafo anterior, ao checar os 5 exemplos identificados pelo modelo, 3 são de fato cães (acertou, VP), 1 é um outro animal (errou em detectar que era um cão, FP) e 1 era de fato um cão, mas o modelo não o considerou (errou, FN). Neste caso, a revocação é de $\frac{3}{3+1} = 0.75$. A Equação 11 exhibe o modelo matemático para o cálculo da revocação.

$$Revocação = \frac{VP}{VP + FN} \quad (11)$$

A **acurácia** é a medida a qual objetiva-se identificar dentre erros e acertos, o quanto o modelo acertou das previsões possíveis [Han et al., 2011]. Busca evidenciar dentre todos os exemplos de VP e FP, o quão o modelo consegue acertar. Ainda utilizando o exemplo do banco de imagens (10 animais, sendo 4 cães e 6 outros animais), a acurácia

seria $\frac{3+5}{3+2+5+1} = 0.72$, pois dos 5 exemplos identificados o modelo acertou em detectar 3 cães (VP), acertou em desconsiderar como cães 5 exemplos (VN), errou em detectar 2 exemplos como cães (FP) e errou em deixar de detectar corretamente 1 cão (FN). O modelo matemático para o cálculo da acurácia é demonstrado na Equação 12.

$$Acurácia = \frac{VP + VN}{VP + FP + VN + FN} \quad (12)$$

O F1-score ou simplesmente **F1**, é a medida harmônica entre a precisão e a revocação. Tem como objetivo buscar um equilíbrio entre o que o modelo preditivo acertou (precisão) e o que ele recuperou de relevante (revocação). Utilizando o exemplo citado previamente sendo precisão 0.60 e revocação 0.75, então o F1 é de $2 \cdot \frac{0.60 \cdot 0.75}{0.60 + 0.75} = 0.67$. A Equação 13 exhibe o modelo matemático para o cálculo de F1.

$$F1 = 2 \cdot \frac{precisão * revocação}{precisão + revocação} \quad (13)$$

Uma técnica bastante utilizada para compreensão detalhada de modelos preditivos é conhecida como Matriz de Confusão (MC). Esta técnica consiste em elaborar uma tabela que exhibe a frequência de erros/acertos para cada classe prevista no modelo [Han et al., 2011]. Os valores computados em uma MC são: VP, FP, VN e FN.

A Figura 4 apresenta um exemplo de MC. Neste exemplo, dentro de uma coleção de e-mails o modelo tenta detectar o resultado de duas classes definidas, sendo elas **spam** e **não spam**. O modelo previu corretamente 3 e-mails que **eram spam** (VP) e 4 e-mails que **não eram spam** (VN). Na predição incorreta, o modelo errou em prever 2 e-mails como **spam, quando não eram** (FP), e também errou ao detectar 1 e-mail como **não spam, em que na verdade eram um spam** (FN). A Tabela 3 apresenta os valores das medidas calculadas para este exemplo.

Tabela 3 – Medidas de avaliação calculadas para o exemplo apresentado da Matriz de Confusão.

Medida de avaliação	Valor
Precisão	0,60
Revocação	0,75
Acurácia	0,70
F1	0,66

		Valores preditos	
		Spam	Não spam
Valores reais	Spam	VP	FN
	Não spam	FP	VN

		Valores preditos	
		Spam	Não spam
Valores reais	Spam	3	1
	Não spam	2	4

Figura 4 – Exemplo de matriz de confusão. A tabela indica a frequência de erros/acertos em um modelo que detecta spam. Fonte: Elaborada pelo autor.

2- Léxicos Afetivos em português

Há uma quantidade muito maior de pesquisas em AS que utilizam LAs disponíveis em inglês, quando comparado com pesquisas do mesmo gênero que utilizam LAs em outras linguagens [Chang et al., 2019] e dentre elas, inclui-se a língua portuguesa. Nesta vertente, torna-se importante explorar a disponibilidade de LAs para língua portuguesa. Para tal realização, recorre-se à buscas na literatura.

Nas pesquisas em banco (ou base) de dados de acesso à literatura acadêmica / científica, considera-se importante definir termos que extraíam documentos relevantes para a pesquisa. Para definir tais termos, adotou-se termos relevantes em inglês/português comumente aplicados em estudos acadêmicos como artigos científicos, dissertações e teses. Termos como “*sentiment*”, “*sentimento*”, “*lexicon*” e “*léxico*”, fornecem um exemplo de elementos aplicados à busca.

Existem diversas bases de conhecimento direcionadas à indexação de artigos científicos. Dentre elas, podemos destacar Scopus¹ e Google Scholar², que agem como fontes concentradoras de artigos indexados em outras bases (e.g. ACM³, IEEEExplorer⁴). Assim, o Scopus e Google Scholar foram as de conhecimento utilizadas para busca de LAs em português. A Tabela 4 exhibe a expressão de busca utilizada nas respectivas bases de dados, bem como o filtro por período e a quantidade de documentos retornados para análise posterior.

Os documentos inicialmente recuperados foram avaliados individualmente, com o objetivo de selecionar pesquisas que desenvolvam LAs em português. Também foi aplicada a técnica de *snowballing*, que possui duas subdivisões: *forward snowballing* que busca estudos **que citaram** um artigo específico; *backward snowballing*, que busca estudos **que são citados** por um artigo específico. Após esta dinâmica, 13 LAs em português foram selecionados para esta pesquisa conforme apresentado na Tabela 5. É importante ressaltar que nesta pesquisa, não há distinção de variantes da língua portuguesa abordada nos LAs (e.g. português Brasil, português Portugal). Tal fato dá-se

¹<https://www.scopus.com>

²<https://scholar.google.com.br/>

³<https://dl.acm.org>

⁴<https://ieeexplore.ieee.org/>

Tabela 4 – Strings de busca utilizadas nas pesquisas.

Base	String de busca	Data limite	Qtd.
Scopus	(“sentiment lexicon” OR “sentiment dictionary” OR “affective lexicon” OR “affective dictionary”) AND (“sentiment analysis”) OR (“affective computing”, “portuguese”, “brazilian”)	até 2015	857
Google Scholar	(“sentiment lexicon” OR “sentiment dictionary” OR “affective lexicon” OR “affective dictionary”) AND (“sentiment analysis”) OR (“affective computing” “portuguese” “brazilian”)	até 2015	180
Google Scholar	(“léxico afetivo” OR “dicionário afetivo” OR “léxico de sentimento”) AND (“análise de sentimento”) OR (“computação afetiva”, “português”, “brasil”)	até 2015	39

Total: 1.076

motivado pela subjetividade da maioria dos autores das pesquisa selecionadas, sem afirmações categóricas sobre a variante do português utilizada nos léxicos. Embora seja possível encontrar outras categorias em alguns léxicos, esse trabalho se limita a utilizar as categorias de sentimentos/emoção. As subseções a seguir descrevem os 13 LAs em português encontrados na literatura.

2.1- AffectPT-br

O AffectPT-br [Carvalho et al., 2018] foi construído utilizando como base outro LA, o LIWC em inglês versão padrão 2015 [Pennebaker et al., 2015]. Seu desenvolvimento deu-se por meio da seleção da categoria de afeto mais ampla no LIWC (i.e. “*affect*”) e suas sub-categorias “*positive emotion*” e “*negative emotion*”. Para cada palavra associada à tais categorias, houve o processo de tradução do inglês para o português.

O AffectPT-br é composto por 1.139 palavras, em que 479 palavras são associadas à categoria “*posemo*”, que representa palavras de sentimento positivo e 661 palavras são associadas à categoria “*negemo*”, que representa palavras de sentimento negativo. Palavras com asterisco (*) indicam a posição inicial que o léxico passa a desconsiderar

Tabela 5 – Léxicos Afetivos em português encontrados na literatura.

	Léxico Afetivo	Qtd. palavras	Traduzido	Polaridades
1	AffectPT-br	1.139	sim	positiva/negativa
2	ANEW-BR	1.046	sim	positiva/negativa/neutra
3	BabelSenticNet	22.597	sim	positiva/negativa
4	LIWC2007-PT-br	127.161	sim	positiva/negativa
5	LIWC2015-PT-br	14.459	não	positiva/negativa
6	MPQA-PT	9.945	sim	positiva/negativa/neutra
7	OpLexicon	32.191	sim	positiva/negativa/neutra
8	Polyglot	3.953	sim	positiva/negativa
9	ReLi-Lex	609	não	positiva/negativa
10	Sentilex-PT	7.014	não	positiva/negativa/neutra
11	SentiStrength	2.112	sim	positiva/negativa/neutra
12	WordnetAffect-BR	289	sim	positiva/negativa
13	WordnetAffect-BR (adapt)	1.023	não	positiva/negativa

caracteres subsequentes, abrangendo variações da palavra (e.g. abandon* = abandonado, abandonada, abandonados, abandonadas). Para esta pesquisa serão utilizados apenas as categorias “31-*posemo*” e “32-*negemo*” para extrair a VA da palavra. A Tabela 6 apresenta uma breve visualização do léxico.

Tabela 6 – Linhas iniciais do léxico AffectPT-br. 30 representa a categoria *affect*, 31 a categoria *posemo*, 32 a categoria *negemo* e 35 a categoria *sad*. Em destaque a categoria que indica a valência afetiva da palavra “abandon*” (negativa). O asterisco (*) indica que o léxico desconsidera caracteres subsequentes, abrangendo variações da palavra.

AffectPT-br	
abaix*	30 32 35
abandon*	30 32 35
abatid*	30 32 35
abenço*	30 31

2.2- ANEW-Br

O ANEW-Br [Kristensen et al., 2011] foi traduzido do original em inglês [Bradley and Lang, 1999] adaptado e normatizado para o português. O estudo foi dividido em duas grandes etapas, sendo a primeira uma tradução direta do ANEW (do inglês para o português brasileiro), produzindo o ANEW-Br. Na segunda etapa, um grupo de 755 voluntários atuaram de forma independente, agindo como juízes na tarefa de avaliação das VA das palavras traduzida no ANEW-Br, possibilitando ajustes no léxico.

O ANEW-Br é composto por 1.046 palavras com medidas para duas dimensões emocionais e os respectivos desvios padrão (DP): 1) “*valência*”, que consiste na avaliação do quão agradável ou desagradável um estímulo é percebido, variando em uma escala de 1 a 9 sendo 1 palavra muito desagradável (negativo) e 9 palavra muito agradável (positivo). Palavras de valência neutra concentram-se na faixa de valores entre 4 a 6, conforme indicado pelos autores; 2) *alerta*, que consiste na avaliação do quão estimulado ou relaxado um estímulo nos deixa, variando em uma escala de 1 a 9 sendo 1 muito estimulado e 9 muito relaxado.

Um exemplo está na palavra “**abandonado**” que apresenta valência=1,85, DP=1,70, alerta=5,48 e DP=2,49. Para esta pesquisa será utilizado apenas a dimensão *valência* para extrair a VA da palavra. A Tabela 7 apresenta uma breve visualização do léxico.

Tabela 7 – Linhas iniciais do léxico ANEW-Br. Sua estrutura é composta por valores de *valência* e *alerta*, seguidos do desvio padrão (DP). Em destaque a dimensão que indica a valência afetiva da palavra “abandonado” (negativa).

ANEW-Br
abalado;2,58;1,81;5,11;2,82
abandonado;1,85;1,70;5,48;2,49
abdução;5,00;1,85;3,36;2,14
abraço;8,77;0,75;3,38;3,00

2.3- BabelSenticnet

O BabelSenticnet [Vilares et al., 2018] foi criado a partir do léxico Senticnet [Cambria et al., 2020], com suporte a multilínguas. É disposto em 40 línguas (Árabe, Bósnio, Búlgaro, Catalão, Chinês, Croata, Tcheco, Dinamarquês, Holandês, Inglês, Estoniano, Finlandês, Francês, Alemão, Grego, Haitiano, Hebraico, Hindi, Hmong, Húngaro, Indonésio, Italiano, Japonês, Coreano, Letão, Lituano, Malaio, Maltês, Português, Romeno, Russo, Sérvio, Singlish, Eslovaco, Esloveno, Espanhol, Tailandês, Turco, Ucraniano e Urdu), e foi inicialmente traduzido do inglês para as respectivas línguas utilizando um tradutor automático (i.e. Bing ⁵).

Sua versão em português é composta por 22.597 termos e expressões. As palavras estão estruturadas em categorias afetivas que indicam níveis de “*agradabilidade*”, “*atenção*”, “*sensibilidade*”, “*aptidão*” e “*polaridade*”. Um exemplo está na palavra “*doença*”, que apresenta *agradabilidade* = -0,963, *atenção* = 0; *sensibilidade* = 0.064; *aptidão* = 0,960 *polaridade* = -0.662. A categoria “*polaridade*” tem sua VA representada de forma numérica, variando em uma escala de “-1” (muito negativo) a “1” (muito positivo). Um exemplo está na palavra “*aceitar*”, que apresenta VA de 0.284 (positivo).

⁵<https://www.microsoft.com/en-us/translator/business/translator-api/>

As expressões são representadas de forma única, colocando todas as palavras juntas e não em termos individuais (e.g. “*a brincar=a.brincar*”), atribuindo uma VA única para todas as palavras caracterizando a afetividade da expressão. Tomemos como exemplo a palavra “*morte*”, que tem sua VA =-0.089 entretanto, a expressão “*morte_de_estrangulamento*” que também possui a palavra “*morte*”, tem sua VA definida como -0.137.

Para contribuir com mais qualidade semântica nas palavras traduzidas automaticamente, uma avaliação realizada por humanos por meio de anotação da concordância dos termos foi exaustivamente abordada neste léxico. Para esta pesquisa será utilizada apenas a categoria “*polaridade*” para extrair a VA da palavra. A Tabela 8 apresenta uma breve visualização do léxico.

Tabela 8 – Linhas iniciais do léxico BabelSenticnet. Sua estrutura é composta por valores de *agradabilidade*, *atenção*, *sensibilidade*, *aptidão* e *polaridade*. Em destaque a categoria que indica a valência afetiva da expressão “*a_alucinar*” (negativa).

BabelSenticnet
32_dentes;0.965; 0.775;0; 0.968; 0.903
a_alucinar;0;0; -0.83; -0.861; -0.564;
a_brincar;-0.141; 0.098; 0.662; -0.656; -0.454;
a_cheirar;0;0; -0.894; -0.993; -0.629;

2.4- LIWC2007-PT-br

O *Linguistic Inquiry and Word Count* (LIWC) [Pennebaker et al., 2001] é um LA disponível na língua portuguesa [Balage Filho et al., 2013], composto por 127.161 palavras e 64 categorias. Cada palavra está associada a uma ou mais categorias, que possibilita extrair informações para análise de aspectos emocionais, estruturais e cognitivos. Os autores compararam o LIWC2007-PT-br contra dois outros LAs disponíveis para o português: OpLexicon e o SentiLex, objetivando fornecer novos *insights* sobre como esses léxicos seriam úteis na análise de sentimento e suas principais características.

As categorias que indicam VA associados à palavra são: “*126-posemo (Positive*

Emotions) e “127-negemo (*Negative Emotions*). Podemos citar como exemplo a palavra *triste*, que está associada à categoria “127-negemo, e categorias adicionais (“124-humans, 125-affect, 130-sad, 140-percept, 141-see, 143-feel), indicando o aspecto estrutural linguístico. A Figura 5 ilustra o exemplo estruturado da palavra.

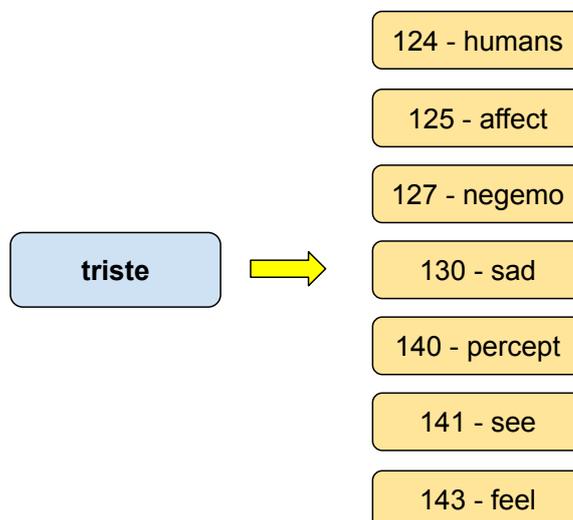


Figura 5 – Exemplo da palavra “*triste*” estruturada no LIWC2007-PT-br. Fonte: Elaborada pelo autor.

Palavras com asterisco (*) indicam a posição inicial que o léxico passa a desconsiderar caracteres subsequentes, abrangendo variações da palavra (e.g. *abafad** = *abafado*, *abafada*, *abafados*, *abafadas*). Para esta pesquisa serão utilizados apenas as categorias “126-*posemo*” e “127-*negemo*” para extrair a VA da palavra. A Tabela 9 apresenta uma breve visualização do léxico.

Tabela 9 – Linhas iniciais do léxico LIWC2007-PT-br. 146 representa a categoria *bio*, 125 a categoria *affect*, 127 a categoria *negemo* e 129 a categoria *anger*. Em destaque a categoria que indica a valência afetiva da palavra “*abafad**” (negativa). O asterisco (*) indica que o léxico desconsidera caracteres subsequentes, abrangendo variações da palavra.

LIWC2007-PT-br	
aba	146
abafa	125 127 129
abafad*	125 127 129
abafai	125 127 129

2.5- LIWC2015-PT-br

O LIWC2015-PT-br é um LA disponível na língua portuguesa, sendo a versão mais nova do LIWC em relação ao apresentado na seção anterior e que tem apresentado melhores resultados [Carvalho et al., 2019]. É composto por 14.459 palavras e 73 categorias e subcategorias. Cada palavra está associada a uma ou mais categorias/subcategorias, que possibilita extrair informações para análise de aspectos emocionais, estruturais e cognitivos.

A categoria mais ampla que identifica aspectos afetivos em uma palavra é a categoria “30-Affect”. As subcategorias que indicam VA associadas à palavra são “31-*posemo*” (*Positive Emotions*) e “32-*negemo*” (*Negative Emotions*). Podemos citar como exemplo, com a palavra *triste* que está associada às subcategorias “32-*negemo*” e “35-*sad*”, e com as categorias “30-*affect*” e “21-*adj*”, indicando o aspecto estrutural linguístico. A Figura 6 ilustra um exemplo da palavra estruturada.

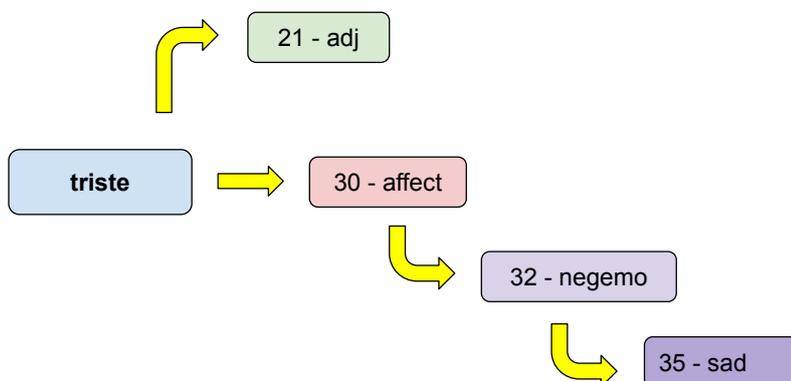


Figura 6 – Exemplo da palavra “*triste*” estruturada no LIWC2015-PT-br. Fonte: Elaborada pelo autor.

Palavras com asterisco (*) indicam a posição inicial que o léxico passa a desconsiderar caracteres subsequentes, abrangendo variações da palavra (e.g. *abus** = *abuso*, *abusivo*, *abusada*, *abusado*). Para esta pesquisa serão utilizados apenas as categorias “31-*posemo*” e “32-*negemo*” para extrair a VA da palavra. A Tabela 10 apresenta uma breve visualização do léxico.

Tabela 10 – Linhas iniciais do léxico LIWC2015-PT-br. 70 representa a categoria *bio*, 72 a categoria *health*, 20 a categoria *verb*, 115 a categoria *death*, 113 a categoria *money*, 30 a categoria *affect*, 32 a categoria *negemo* e 35 a categoria *sad*. Em destaque a categoria que indica a valência afetiva da palavra “*abatid**” (negativa). O asterisco (*) indica que o léxico desconsidera caracteres subsequentes, abrangendo variações da palavra.

LIWC2015-PT-br
abatacepte 70 72
abater 20 115
abatid* 30 32 35 115
abatimento* 113

2.6- MPQA-PT

O MPQA-PT (Multi-Perspective Question Answering) [Almeida et al., 2015] é um LA em português que foi traduzido do inglês [Wiebe et al., 2005a]. Para validar o léxico, os autores utilizaram um corpus manualmente anotado, contendo 80 tópicos de notícias com 10 documentos cada, extraídos de jornais, sites pessoais, sites de TV e rádio, nos períodos de 2010-2011 e 2012-2013. A anotação foi realizada por três linguistas, respeitando as premissas preconizadas pelo MPQA original em inglês [Wiebe et al., 2005a] que inclui informações de intensidade e polaridade. Cada documento foi anotado por dois dos três linguistas e depois revisado pelo terceiro linguista. Em caso de divergências, o terceiro linguista debatia com os demais anotadores para atingir um consenso final.

É composto por 9.945 palavras e expressões. Cada palavra está estruturada em 3 categorias: 1) “*classificação gramatical da palavra*” (V=verbo, A=adjetivo, etc...); 2) “*classificação de subjetividade*” (*weaksubj*=fraca subjetividade, *strongsubj*=forte subjetividade); 3) *valência* (*negative*=negativo, *positive*=positivo, *neutral*=neutro). Para esta pesquisa será utilizada apenas a categoria “*valência*” para extrair a VA da palavra. A Tabela 11 apresenta uma breve visualização do léxico.

Tabela 11 – Linhas iniciais do léxico MPQA-PT. A letra “A” representa a classificação gramatical da palavra (adjetivo), os termos *weaksbj/strongsubj* representam o “peso” da subjetividade (forte/fraca), atuando como intensificadores emocionais. A valência afetiva é representada pelos termos *positive/positive*. Em destaque a valência afetiva para a palavra “aborrecer” (negativa).

MPQA-PT
abordável A weaksubj positive
aborrecer V strongsubj negative
aborrecido A strongsubj negative
aborrecimento N strongsubj negative

2.7- OpLexicon

O OpLexicon (ou Opinion Lexicon) [Souza et al., 2011] é um LA disponível em português traduzido do inglês [Hu and Liu, 2004]. Foi elaborado a partir da integração de 3 diferentes métodos previstos na literatura: abordagem baseada em corpus, abordagem baseada em léxico (ou dicionário) e abordagem baseada em tradução. Para abordagem baseada em corpus, os autores utilizaram 346 documentos referentes à críticas de filmes, extraídos dos sites CinePlayers⁶, e Cinema com Rapadura⁷, além de 970 textos jornalísticos sobre diversos temas extraídos do corpus de categorias PLN-Br [Bruckschen et al., 2008]. O resultado final contém 1.317 documentos e cerca de um milhão de palavras.

Para abordagem baseada em léxicos, os autores utilizaram o dicionário TEP thesaurus [Maziero et al., 2008] contendo 44.077 palavras anotadas. Para abordagem baseada em tradução, os autores utilizaram o dicionário em inglês Opinion Lexicon [Hu and Liu, 2004] traduzindo-o para o português por meio da ferramenta de apoio Google Translate⁸. Na etapa final unificaram as 3 abordagens com interseções, considerando sua concordância e VA selecionando então as palavras de opinião que compõem a versão inicial do léxico.

Em sua última versão disponível (3.0), o léxico é composto por 32.191 palavras.

⁶<https://www.cineplayers.com/>

⁷<https://cinemacomrapadura.com.br/>

⁸<https://translate.google.com/>

Estas palavras estão estruturadas nas seguintes categorias: “*categoria morfológica*” (vb=verbo, adj=adjetivo, etc...) e *valência* (escala de –1 para “*muito negativo*” e 1 para “*muito positivo*”). Para esta pesquisa será utilizada apenas a categoria “*valência*” para extrair a VA da palavra. A Tabela 12 apresenta uma breve visualização do léxico.

Tabela 12 – Linhas iniciais do léxico OpLexicon. O termo *adj* representa um adjetivo, *vb* representa um verbo. Em destaque a valência afetiva da palavra “*acreditado*” (positiva).

OpLexicon
acreditadas,adj,1
acreditado,adj,1
acreditados,adj,1
acreditar,vb,0

2.8- Polyglot

O Polyglot [Chen and Skiena, 2014] é um LA disposto em mais de 136 línguas, dentre elas o português. Elaborado a partir do trabalho de Al-Rfou’ et al. [2013], foi traduzido para o português utilizando como ferramenta de apoio o Google Translate⁹, validando o processo de anotação de sentimento utilizando humanos. Segundo os autores, o Polyglot atinge uma concordância afetiva de 95,7% com outros léxicos publicados, além de uma cobertura de 45,2%.

É composto por 3.953 palavras estruturadas em uma única categoria. Esta categoria representa a VA da palavra, limitando-se a dois valores: “–1” representa um sentimento negativo e “1” representa um sentimento positivo. A Tabela 13 apresenta uma breve visualização do léxico.

⁹<https://translate.google.com/>

Tabela 13 – Linhas iniciais do léxico Polyglot. O número 1 representa um sentimento positivo, -1 representa um sentimento negativo. Em destaque a valência afetiva da palavra “*aberração*” (negativa).

Polyglot
abençoar [1]
aberração [-1]
abertura [1]
aborrecido [-1]

2.9- ReLi-Lex

O ReLi-Lex [Freitas, 2013] é um LA disponível em português elaborado a partir de um corpus composto por resenhas de livros publicados na Internet. Foi previamente anotados com informação de opinião, e criado com o propósito de atender dois objetivos: fornecer subsídios para aplicação de Processamento de Linguagem Natural (PLN) e contribuir para a caracterização do vocabulário afetivo da língua portuguesa. Vale destacar que possui ênfase em contextos informais

O ReLi-Lex é composto por 609 palavras e expressões estruturadas em uma única categoria. Esta categoria representa a VA da palavra limitando-se a dois valores: “-1” representa um sentimento negativo e “1” representa um sentimento positivo. A Tabela 14 apresenta uma breve visualização do léxico.

Tabela 14 – Linhas iniciais do léxico ReLi-Lex. O número 1 representa um sentimento positivo, -1 representa um sentimento negativo. Em destaque a valência afetiva da palavra “*amar*” (positiva).

ReLi-Lex
amadurecer;1
amar;1
amor à primeira vista;1
anacrônico;-1

2.10- Sentilex-PT

O Sentilex-PT é um LA especificamente concebido para a análise de sentimento e opinião sobre entidades humanas em textos redigidos em português [Carvalho and Silva, 2015]. As palavras deste léxico foram organizadas e classificadas de modo semi-automático, utilizando abordagens linguísticas (extraindo palavras candidatas a partir de um corpus) e abordagens de aprendizagem de máquina (filtrando palavras de uma lista de candidatas). A VA aplicada às palavras foi determinada com base em dicionários de sinônimos disponíveis em português. O objetivo era identificar palavras que apresentam polaridades distintas em função de diferentes contextos em que se encontram, e consequentemente remover possíveis ruídos. Por exemplo, a palavra “*quente*” pode evocar um sentimento positivo em um contexto (e.g. “*a sopa ainda está quente*”) porém, a mesma palavra pode evocar um sentimento negativo (e.g. “*o champanhe está quente*”).

O Sentilex-PT é composto por 7.014 palavras e expressões. Estas palavras/expressões estão estruturadas nas seguintes categorias: 1) “*PoS (gramatical)*” (ADJ=adjetivo, N=nome, V=verbo, IDIOM=idioma), 2) “*polaridade (POL)*” (1 positiva, -1 negativa, 0 neutra), 3) “*alvo da polaridade (TG)*” (corresponde a um nome de tipo humano “HUM”, com função de sujeito “N0” e/ou complemento “N1”, 4) “*anotação de polaridade (ANOT)*” (MAN=manualmente anotada, JALC=atribuída automaticamente). Um exemplo está na palavra “*aberração*” que apresenta PoS=N (nome), TG=HUM (humano), POL= -1 (negativa) e ANOT=MAN (manual). Para esta pesquisa será utilizada apenas a categoria “*POL*” para extrair a VA da palavra. A Tabela 15 apresenta uma breve visualização do léxico.

Tabela 15 – Linhas iniciais do léxico Sentilex-PT. A categoria Pos=Adj representa um adjetivo, a categoria TG=HUM:N0 representa um nome de tipo humano, a categoria POL:N0=1 representa a polaridade (1 positiva, -1 negativa) e a categoria ANOT=JALC representa que a polaridade foi anotada automaticamente. Em destaque a valência afetiva da palavra “*aceitável*” (positiva).

Sentilex-PT
aceitável.PoS=Adj;TG=HUM:N0;POL:N0=1;ANOT=JALC
aceite.PoS=Adj;TG=HUM:N0;POL:N0=0;ANOT=JALC
aceito.PoS=Adj;TG=HUM:N0;POL:N0=0;ANOT=JALC
acelerado.PoS=Adj;TG=HUM:N0;POL:N0=0;ANOT=MAN

2.11- SentiStrength

O SentiStrength [Thelwall et al., 2010], é um LA que tem como objetivo principal a AS em textos curtos, abrangendo linguagem formal e informal. Esta disponível nas línguas Finlandês, Alemão, Espanhol, Holandês, Russo, Português, Francês, Árabe, Polonês, Persa, Sueco, Grego, Galês, Italiano e Turco. É composto por 9 arquivos, sendo selecionados para esta pesquisa apenas 3 arquivos principais, sendo: palavras de associação emocional (EmotionLookupTable), *emoticons* (EmoticonLookupTable) e palavras de intensidade (BoosterWordList). A motivação para descartar o uso dos demais arquivos deu-se pela baixa quantidade de palavras, sendo SlangLookupTable (1 palavra), EnglishWordList (1 palavra), IdiomLookupTable (1 palavra), IronyTerms (6 palavras), NegatingWordList (4 palavras) e QuestionWords (5 palavras).

Os 3 arquivos selecionado compõem 2.112 palavras estruturadas em uma única categoria. Esta categoria representa a VA da palavra em uma escala valores de “-5” (extremamente negativo) a “5” (extremamente positivo). A faixa [-1 .. 1] representa um sentimento neutro. A Tabela 16 apresenta uma breve visualização do léxico.

Tabela 16 – Linhas iniciais do léxico SentiStrength. Em destaque a valência afetiva da palavra “abraços” (positiva).

SentiStrength	
abraço	3
abraços	3
absurdo	-2
abuso	-2

2.12- WordnetAffectBR

O WordnetAffectBR [Pasqualotti and Vieira, 2008] foi elaborado a partir do original em inglês Wordnet [Miller, 1995]. Os autores adotaram uma “teoria de emoções” para

identificar “palavras” que tenham significado no domínio das emoções e, consequentemente, gerar uma base lexical. A partir desta base lexical, apoiaram-se em outros recursos que permitissem estender a base do WordnetAffectBR (e.g. um base contendo substantivos e respectivos adjetivos possibilitaram a criação de 42 palavras), ampliando a cobertura do domínio das emoções com o máximo possível de palavras. A base inicial do WordnetAffectBR foi composta inicialmente por palavras extraídas do Wordnet atribuídas à categoria (CAT=emo), que representa palavras do estado afetivo das emoções. Seguidamente, ocorreu a tradução do inglês para o português e sua validação, proporcionando a versão final do WordnetAffectBR.

O WordnetAffectBR é composto por 289 palavras e expressões estruturadas em uma única categoria. Esta categoria representa a VA da palavra em um símbolo de “+”, indicando um sentimento positivo e “-”, indicando um sentimento negativo. A Tabela 17 apresenta uma breve visualização do léxico.

Tabela 17 – Linhas iniciais do léxico WordnetAffectBR. O símbolo “+” representa um sentimento positivo e “-” representa um sentimento negativo. Em destaque a valência afetiva da palavra “*abrandamento*” (positiva).

WordnetAffectBR
abominação;-
abrandamento;+
acanhado;-
admiração;+

2.13- WordnetAffectBR (adapt)

O WordnetAffectBR Adapt [Longhi, 2011], é um LA elaborado por uma tese de doutorado. A pesquisa contribuiu com a ampliação do WordnetAffectBR [Pasqualotti and Vieira, 2008], cuja base inicial apresentava uma estrutura de 289 palavras e expressões. O WordnetAffectBR (adapt) abrange registros de léxico afetivo, *emoticons* (e.g. “:”) e “:(”), interjeições, artigos, preposições, pronomes e alguns verbos e advérbios.

O WordnetAffectBR (adapt) é composto por 1.023 palavras e expressões. Estas

palavras/expressões estão estruturadas nas seguintes categorias: 1) “*código*” (código da palavra), 2) “*imagem*” (código da imagem associada à emoção), 3) “*quadrante*” (intensificador afetivo, em uma escala de valor 1 a 4), 4) “*valência*” (+ positiva, – negativa). Para esta pesquisa serão utilizadas apenas as categorias “*valência*” e “*quadrante*” para extrair a VA da palavra. A Tabela 18 apresenta uma breve visualização do léxico.

Tabela 18 – Linhas iniciais do léxico WordnetAffectBR (adapt). O símbolo “+” representa um sentimento positivo e “–” representa um sentimento negativo. Em destaque a valência afetiva da palavra “*agraciado*” (positiva).

WordnetAffectBR (adapt)
(57 ,'agonia' , '121', 3 , '-')
(60 ,' agraciado ' , , 1 , '+')
(61 ,'agradável' , '236', 1 , '+')
(64 ,'agressão' , , 2 , '-')

3- Trabalhos relacionados

Neste capítulo, são abordados alguns trabalhos com temáticas relacionadas ao presente estudo. Inicialmente, debate-se sobre um sistema web que possibilita AS em 61 línguas (por meio da aplicação de TAT) utilizando 16 LAs. Em seguida, apresenta-se um estudo comparativo entre quatro LAs para tarefas de AS. Por último, debate-se uma revisão sistemática sobre LAs em português do Brasil.

3.1- iFeel

O iFeel¹ é um sistema web, inicialmente proposto por Araújo et al. [2014] (versão 1.0). Tem como finalidade ser uma ferramenta que possibilita AS utilizando LAs. Em sua versão inicial, utilizou como apoio 7 LAs disponíveis na língua inglesa: PANAS-T [Gonçalves et al., 2013b], Emoticons [Gonçalves et al., 2013a], Sentiwordnet [Baccianella et al., 2010], Happiness Index [Dodds and Danforth, 2010], SentiStrength [Thelwall, 2017], SASA [Wang et al., 2012] e Senticnet [Cambria et al., 2014]. Deste modo, a AS restringia-se à textos na língua inglesa.

Seu modelo básico de funcionamento, consiste em receber arquivos enviados por usuários para que sejam submetidos à avaliação dos LAs. Objetivando a AS, o iFeel utilizou aquilo que os autores nomearam de Método Combinado (*Combined Method*), que consiste em uma “votação” dos LAs para cada documento, adotando como “sentimento do documento” aquele identificado em maior consenso entre os LAs. Após a AS do texto, o sistema disponibiliza um arquivo contendo o resultado possibilitando o *download* pelo usuário.

A versão seguinte (2.0) [Araujo et al., 2016], oferece suporte para AS em multi-línguas. Além do inglês, 60 novas línguas foram adicionadas e dentre elas, o português. Os autores acrescentaram mais 12 léxicos afetivos: Opinion Lexicon [Hu and Liu, 2004], Opinion Finder [Wilson et al., 2005], AFINN [Nielsen, 2011], SO-CAL [Taboada et al.,

¹<http://www.ifeel.dcc.ufmg.br/>

2011], Emoticons Distant Supervision [Hannak et al., 2012], NRC Hashtag [Mohammad, 2012], EmoLex [Mohammad and Turney, 2013], SANN [Pappas and Popescu-Belis, 2013], Sentiment140 Lexicon [Mohammad et al., 2013], Stanford Recursive Deep Model [Socher et al., 2013], Umigon [Levallois, 2013] e Vader [Gilbert and Hutto, 2014].

Apesar de ser multi-idiomas, todos os léxico dispostos no iFeel versão 2.0 suportam somente a língua inglesa ou seja, não há versões exclusivas em outras línguas. Para realizar a AS, o sistema aplica o uso de TAT, traduzindo textos da língua original para língua inglesa. A ferramenta de apoio utilizada para TAT foi a Yandex Translation API².

Conforme é mencionado, em AS, a utilização de LA que não sejam na língua nativa do texto à ser avaliado pode não apresentar resultados adequados. Um dos fatores que podem influenciar na inconsistência destes resultados são as variações de estilos linguísticos, comumente presentes na língua. Tomemos um exemplo de inconsistência apresentada pelo iFeel, um experimento realizado nesta pesquisa. Submeteu-se ao sistema uma sentença em português (que evoca um sentimento negativo) para a aplicação de TAT: “*Enjoado das minhas músicas mas com preguiça de procurar novas*”. Seguidamente, a mesma sentença foi submetida ao sistema sem a aplicação de TAT (convertida naturalmente para o inglês sem tradutor automático: “*Sick of my music but too lazy to look for new ones*”). Coletou-se ambos os resultados e ao avaliá-los, percebeu-se que a aplicação de TAT degradou o desempenho dos LAs prejudicando a detecção do sentimento. A Tabela 19 apresenta os resultados apurados.

Tabela 19 – Exemplo de desempenho dos Léxicos Afetivos na utilização de Tradução Automática de Texto (TAT) pelo iFeel. **Com** a aplicação de TAT, menos léxicos afetivos conseguiram detectar corretamente o sentimento da sentença (5 léxicos). **Sem** a aplicação de TAT (em destaque), mais léxicos **acertaram** corretamente o sentimento da sentença (13 léxicos).

Classe	Com aplicação de TAT	Sem aplicação de TAT
Negativo	5	13
Positivo	3	1
Neutro	8	2

²<https://yandex.com/dev/translate/>

3.2- Comparação entre Léxicos Afetivos para análise de sentimento no Twitter

Avaliar a eficácia de LAs nas abordagens de AS baseadas em léxicos foi objeto de investigação de Musto et al. [2014]. Os autores pesquisaram a eficácia de quatro LAs disponíveis em inglês para tarefa de AS utilizando dados da RS Twitter. São eles: SentiWordNet [Baccianella et al., 2010], WordNet-Affect [Strapparava et al., 2004], MPQA [Wiebe et al., 2005b] e SenticNet [Cambria et al., 2014].

O trabalho citado adotou uma “técnica base” para AS, aquilo que os autores chamaram de “micro-frases”. Esta técnica consistiu em fragmentar comentários do Twitter adotando como “*pontos de divisões*” as pontuações, advérbios e conjunções, gerando sub-sentenças do texto. Dada tal fragmentação, o sentimento presente na sentença foi definido por meio da soma das polaridades veiculadas em cada uma das micro-frases que compõem a sentença. As VAs dos termos foi obtida extraíndo recursos dos léxicos utilizados no estudo.

Os autores definiram quatro implementações diferentes para abordagem de micro-frases: 1) **básica**, consiste na detecção do sentimento por meio da soma da polaridade extraída do primeiro elemento de cada micro-frase; 2) **normalizada**, consiste na detecção do sentimento por meio da soma da polaridade de todos os elementos na micro-frase; 3) **enfaticada**, consiste na extensão da formulação **básica** entretanto, fornecendo um peso maior à polaridade quando o termo do léxico for pertencente à categorias específicas que indicam intensificação de afetividade; 4) **enfaticada-normalizada**, que consiste na unificação do segundo e terceiro método.

Os experimentos foram aplicados explorando dois conjuntos de dados: SemEval-2013 [Rosenthal et al., 2017] (14.435 Tweets) e Stanford Twitter Sentiment-STS (1.600.000 Tweets) [Go et al., 2009], ambos contento Tweets rotulados nas classes *positivo*, *negativo* e *neutro*.

Os resultados indicaram que no conjunto de dados SemEval-2013, o SentiWordNet apresentou um desempenho superior ao SenticNet e WordNet-Affect, empatando estatisticamente com o MPQA. No conjunto de dados STS, o SenticNet apresentou o melhor resultado, seguido do SentiWordNet, MPQA e WordNet-Affect. Referente à abordagem das micro-frases, os autores citaram não haver diferenças estatísticas significantes. Os autores ainda destacam os LAs MPQA e SentiWordNet como os de melhor

desempenho sobre os conjuntos de dados utilizados. Embora o trabalho citado apresente um estudo comparativo de LAs, a quantidade de léxicos utilizados é baixa e o estudo abordou somente léxicos disponíveis em inglês.

3.3- Revisão Sistemática sobre Léxicos Afetivos para o Português do Brasil

A importância da utilização de LA aplicados ao estudo das emoções se apresenta como destaque na pesquisa de Cruz et al. [2017]. Os autores salientaram a utilização de LA para identificação de características emocionais em textos escritos por indivíduos. Tal atividade pode contribuir em importantes áreas da sociedade, como inferir fatores de personalidade [Golbeck et al., 2011], o desenvolvimento de robôs emocionais [Aly and Tapus, 2013], e até detectar sintomas de manifestações depressivas em redes sociais [De Choudhury et al., 2013].

Ainda segundo os autores, poucos trabalhos na literatura disponibilizam LA em português do Brasil. Nesta vertente, realizaram como contribuição uma revisão sistemática de LA em português do Brasil, objetivando motivar novos estudos para capacitar detecção de emoções em textos na língua portuguesa.

Os LA abordados na pesquisa foram: ANEW-Br [Kristensen et al., 2011], LIWC [Balage Filho et al., 2013], Opinion Lexicon [Souza et al., 2011], Personalitatem Lexicon [Machado et al., 2015], ReLi-Lex [Freitas, 2013], WordnetAffectBR [Pasqualotti and Vieira, 2008] e WordnetAffectBR (adapt) [Longhi, 2011]. Todavia, a pesquisa não sucedeu nenhum tipo de avaliação de desempenho em tarefas de AS utilizando os LA abordados. Neste aspecto, se suscita o desenvolvimento de pesquisas que abordem avaliações de desempenho dos LA em tarefas de AS, provendo informações que capacitem estudos futuros objetivando a elaboração de novos léxicos com qualidade superior.

3.4- Um levantamento da análise de sentimento na língua portuguesa

Pereira [2021] forneceu uma visão geral em pesquisas publicadas de AS exclusiva-

mente na língua portuguesa. O autor apresenta um levantamento de estudos referentes à AS em português, abordando os recursos utilizados nestes estudos como processamento de linguagem natural, tradução automática de textos, léxicos afetivos e conjunto de dados. O estudo faz revisão do estado da arte, discutindo desafios relacionados à área de AS.

Referente a utilização de léxicos afetivos para AS em português, o autor destaca seis léxicos encontrados: OpLexicon [Souza et al., 2011], LIWC2007 [Balage Filho et al., 2013], Onto.PT [Gonçalo Oliveira, 2018], ReLi-Lex [Freitas, 2013], SentiLex [Carvalho and Silva, 2015], SentiStrength [Thelwall et al., 2010]. O autor ainda destaca os esforços aplicados nas pesquisa de AS para português durante os últimos anos, proporcionando novas abordagens e novos métodos.

Entretanto, trata-se apenas de uma revisão sistemática de recursos utilizados em pesquisa, não executando nenhum tipo de estudo comparativo entre léxicos afetivos em AS. Vale ressaltar que o léxico Onto.PT é um léxico que extrai a polaridade das palavras de outro léxico, o SentiLex, sendo utilizado em uma abordagem diferente da proposta nesta pesquisa (Seção 4.4).

3.5- Comparação sistemática de ferramentas automatizadas de análise de conteúdo

Avaliar a eficácia de léxicos afetivos em AS utilizando diferentes ambientes de mídias sociais foi destaque na pesquisa elaborada por González-Bailón and Paltoglou [2015], objetivando fornecer orientações para a escolha do método mais adequado, dadas as diferenças nas fontes de dados, especialmente em conteúdo online gerado pelo usuário.

Os dados utilizados neste estudo foram extraídos de quatro mídias sociais: BBC (comentários sobre notícias), Twitter, Digg (comentários sobre assuntos diversos), Youtube e comentários de blogs relacionados à política. Os léxicos afetivos utilizados foram: LexiconBased [Paltoglou and Thelwall, 2012], SentiStrength [Thelwall et al., 2010], ANEW [Bradley and Cuthbert, 1999], labMT [Dodds et al., 2011], e Lexicoder [Young and Soroka, 2012].

Os autores mediram a concordância entre os léxicos nas tarefas de AS (teste

de confiabilidade), seguidamente da comparação das tarefa de classificação (teste de validade). Segundo os autores, os resultados indicaram que a validade e a confiabilidade variam com a formalidade e a diversidade do texto. A concordância entre os léxicos (teste de confiabilidade) apresentou-se baixa, indicando léxicos que não são adequados para classificação ternária (ou seja, positivo, negativo e neutro), e a tarefa de classificação (teste de validade) tende a apresentar baixo desempenho, caso o conteúdo analisado for composto por domínios específicos que não são bem representados por léxicos afetivos de uso geral.

Apesar de importantes, os trabalhos supracitados apresentam lacunas inexploradas no campo da AS baseada em LAs. Para melhor compreender seus aspectos positivos e negativos, a seguir apresentam-se de maneira sintetizadas as características predominantes destes trabalhos.

1. iFeel

- **Aspectos positivos:** Possibilita análise de sentimento em uma grande quantidade de línguas; grande quantidade de LAs utilizados; apoia-se na opinião de maior consenso entre LAs para detectar o sentimento.
- **Aspectos negativos:** Utilização de LAs somente em inglês; a aplicação de Tradução Automática em Textos (TAT) não apresenta resultados consistentes na análise de sentimento.

2. Comparação entre Léxicos Afetivos para análise de sentimento no Twitter

- **Aspectos positivos:** Busca compreender o desempenho de quatro LAs em análise de sentimento e conseqüentemente, apresenta os melhores resultados possibilitando uma boa escolha para trabalhos relacionados.
- **Aspectos negativos:** Utilização de LAs somente em inglês; estudo com baixa quantidade de LAs.

3. Revisão Sistemática sobre Léxicos Afetivos para o Português do Brasil

- **Aspectos positivos:** LAs em português do Brasil; boa quantidade de LAs encontrado na literatura.
- **Aspectos negativos:** Não realizaram avaliações experimentais dos LAs, o que dificulta a compreensão da real eficácia dos léxicos na análise de sentimento.

4. Um levantamento da análise de sentimento na língua portuguesa

- **Aspectos positivos:** Revisão sistemática de estudos desenvolvidos para AS exclusivamente em português; boa referência de recursos disponíveis para AS.
- **Aspectos negativos:** Não realizou avaliações experimentais dos LAs.

5. Comparação sistemática de ferramentas automatizadas de análise de conteúdo

- **Aspectos positivos:** Estudos comparativos entre LAs; Utilização de diferentes conjunto de dados para análise.
- **Aspectos negativos:** Utilização de LAs somente em inglês.

4- Metodologia

Este capítulo apresenta detalhes do processo utilizado para a execução dos experimentos desta pesquisa. Na Seção 4.1 será detalhada a linguagem de programação utilizada. Na Seção 4.2 serão abordados os conjuntos de dados utilizados, bem como os critérios de seleção empregados. Na Seção 4.3 será detalhado o *framework* construído para a execução dos experimentos. Na Seção 4.4 é apresentado os detalhes do método utilizado para AS e a Seção 4.5 detalha o processo de normalização de dados.

4.1- Linguagem de programação

Nesta pesquisa, os experimentos foram executados utilizando a linguagem de programação Python¹. Essa escolha se deu devido as suas características como linguagem *Open Source*, suporte para programação orientada à objeto, facilidade em operações aritméticas entre matrizes (álgebra linear). Também considerou-se características oferecidas pela linguagem como suporte a programação paralela (*multi threads*), ampla comunidade para suporte, utilização em diversas pesquisas envolvendo Mineração de Textos e ampla disponibilidade de bibliotecas para apoio em operações de AS.

4.2- Conjunto de dados

Pretendendo produzir resultados de maior generalização, esta pesquisa adotou o critério em utilizar conjunto de dados diferentes nos experimentos aplicados. Foram realizadas pesquisas na literatura e em sites de artigos científicos indexados (e.g. Scopus²), selecionando trabalhos relacionados que disponibilizassem publicamente os dados rotula-

¹<https://www.python.org/>

²<https://www.scopus.com/>

dos em português. Nesta vertente, quatro conjunto de dados foram selecionados: Meu Querido Diário - Dataset (MQD), Sentiment Analysis Dataset Twitter-pt-br (SADT), TOPIE e TweetSentBR. Todos estes conjuntos de dados estão rotulados em três classes: positivo, negativo e neutro. Vale também ressaltar que eles estão estruturados em sentenças. Os detalhes de cada conjunto de dados apresentam-se nas próximas subseções.

4.2.1 MQD

O MQD é um conjunto de dados produzido por alunos do curso do Programa de Pós-graduação em Ciência da Computação (PPCIC), do CEFET/RJ, inscritos na disciplina de Mineração de Textos, no período 3/2020. Como atividade de aula, os alunos rotularam os dados, de maneira que pudessem ser utilizados posteriormente como conteúdo didático da disciplina. Os dados iniciais foram extraídos do site Meu Querido Diário³, uma rede social que funciona como um “diário online” em que os usuários expõem seus pensamentos abordando diversos assuntos.

A estrutura final dos dados é composta por 1.457 frases distribuídas em três classes: positivo (508), negativo (539) e neutro (410). Tais dados atenderam a um critério de concordância mínima de 2, entre os 3 juízes (alunos) que realizaram a atividade de rotular as frases de forma manual. A tabela 20 apresenta um exemplo dos dados que compõem o MQD.

Tabela 20 – Exemplo de dados do MQD.

Comentário	Classe
<i>“Estou muito feliz e espero que você também”</i>	positivo
<i>“Vou morrer de vergonha quando o resultado sair”</i>	negativo
<i>“Hoje acaba uma guerra”</i>	neutro

³<http://www.meuqueridodiario.com.br/>

4.2.2 SADT

O SADT é um conjunto de dados elaborado por Silva et al. [2019]. É composto por comentários extraídos do Twitter que utilizou como palavras-chave para busca, adjetivos da língua portuguesa extraídos do thesaurus TeP 2.0 [Dias-Da-Silva and Moraes, 2003], considerando o período das postagens entre 24/09/2018 e 06/12/2018. É rotulado em três classes: positivo, negativo e neutro. A atividade de rotular os dados foi desempenhada por alunos do curso de Bacharelado em Sistemas de Informação da Universidade Federal da Paraíba utilizando uma ferramenta *web* como apoio.

A ferramenta disponibilizava ao usuário uma tela contendo um *tweet* escolhido aleatoriamente na base de dados. Em seguida, o aluno deveria escolher uma das três opções disponíveis na tela para rotular o sentimento evocado no texto (i.e. positivo, negativo, neutro). Feito isto, um novo *tweet* aleatório era disponibilizado, repetindo o processo até que não houvesse mais *tweets* sem rótulo.

A estrutura de dados final é composta por 1.867 frases distribuídas em três classes: positivo (559), negativo (597) e neutro (711). A Tabela 21 apresenta um exemplo dos dados que compõem o SADT.

Tabela 21 – Exemplo de dados do SADT.

Comentário	Classe
<i>“Monise está muito linda, to apaixonado”</i>	positivo
<i>“Ah da fazenda. Tem o agressor, Douglas”</i>	negativo
<i>“Não posso sentir um calorzinho que a pressão vai lá em baixo”</i>	neutro

4.2.3 TOPIE

O TOPIE [Souza et al., 2016] é um conjunto de dados elaborado por pesquisadores do grupo de estudo MiningBR Research Group, da Universidade Federal Rural de Pernambuco (UFRPE). É composto por comentários extraídos do Twitter coletados por

meio de um *webcrawler*⁴.

Inicialmente, foram selecionados mais de 30.000 *tweets* contendo comentários referentes às empresas que mais receberam reclamações em 2013, de acordo com a Agência do Programa de Defesa e Proteção ao Consumidor (PROCON). Segundo os autores, a motivação para a escolha das empresas mais citadas objetivou comparar comentários expostos no Twitter relacionados à estas empresas.

Para o conjunto de dados final, foram selecionados 543 *tweets*, que foram rotulados manualmente pela equipe, e distribuídos de forma equivalente em três classes: positivo (181), negativo (181) e neutro (181). A Tabela 22 apresenta um exemplo dos dados que compõem o TOPIE.

Tabela 22 – Exemplo de dados do TOPIE.

Comentário	Classe
<i>“Sendo muito bem atendido pelo pessoal do @alobradesco”</i>	positivo
<i>“Serviço horrivelllllll; @Vivoemrede”</i>	negativo
<i>“Finalmente comprei meu chip da @Claro”</i>	neutro

4.2.4 TweetSentBR

O TweetSentBR [Brum and das Graças Volpe Nunes, 2018] é um conjunto de dados composto por comentários do Twitter extraídos no período de janeiro a julho de 2017. Segundo os autores, o critério para seleção dos dados foi considerar comentários específicos de programas de televisão durante as suas respectivas exibições.

A escolha de um público-alvo, o “espectador de TV”, deu-se pelo motivo da grande quantidade de conteúdo que é produzido pelos telespectadores durante a exibição dos programas de televisão. Os apresentadores de tais programas induzem a participação de seus telespectadores nas RS por meio da utilização de *hashtags* (e.g. #thenoite, #encontro). Conseqüentemente, os telespectadores acabam classificando dados para um domínio específico (e.g. comentários sobre o programa The Noite, comentários sobre o programa Encontro).

⁴Um sistema utilizado para encontrar, ler e indexar páginas de um site.

Sua estrutura é composta por 15.000 comentários, distribuídos em três classes: positivo (6.648), negativo (4.426) e neutro (3.926). A Tabela 23 apresenta um exemplo dos dados que compõem o TweetSentBR.

Tabela 23 – Exemplo de dados do TweetSentBR.

Comentário	Classe
“#VídeoShowAoVivo a música ficou mais bonita na voz do Chico”	positivo
“Cada pergunta idiota q esse povo da plateia faz #altashoras”	negativo
“Lembrei q hj tem #MasterChefBR e já me deu fome”	neutro

4.3- Framework para Análise de Sentimento

Para proceder com a tarefa de AS utilizando os LA mencionados no Capítulo 2, algumas etapas tornaram-se necessárias para a execução dos experimentos aplicados nesta pesquisa. Na ausência de ferramentas que viabilizassem a análise e comparação de múltiplos LAs, foi desenvolvido um *framework* que possibilitou a execução de atividades importantes e conseqüentemente a realização dos experimentos.

Este *framework* é composto por 4 etapas: 1) Entrada de dados: consiste na leitura dos arquivos em CSV contendo os conjuntos de dados utilizados (Seção 4.2); 2) Pré-processamento: consiste na eliminação de ruídos dos dados, como remoção de links, remoção das *stopwords*⁵, remoção de pontuação, transformação dos caracteres para minúsculo (*case folding*⁶); 3) Análise de Sentimento: consiste na AS baseada em LAs utilizando abordagem adotada neste estudo (Seção 4.4); 4) Normalização: consiste na normalização das valências associadas às palavras do LA (Seção 4.5); 5) Avaliação: consiste na aplicação das medidas que avaliam o desempenho dos LA, analisando os resultados. A Figura 7 fornece uma visão ampla do *framework* desenvolvido.

⁵Palavras sem relevância no contexto em que se apresentam (e.g. a, as, o, que, um, com, por)

⁶Técnica que consiste em converter os caracteres para letras minúsculas

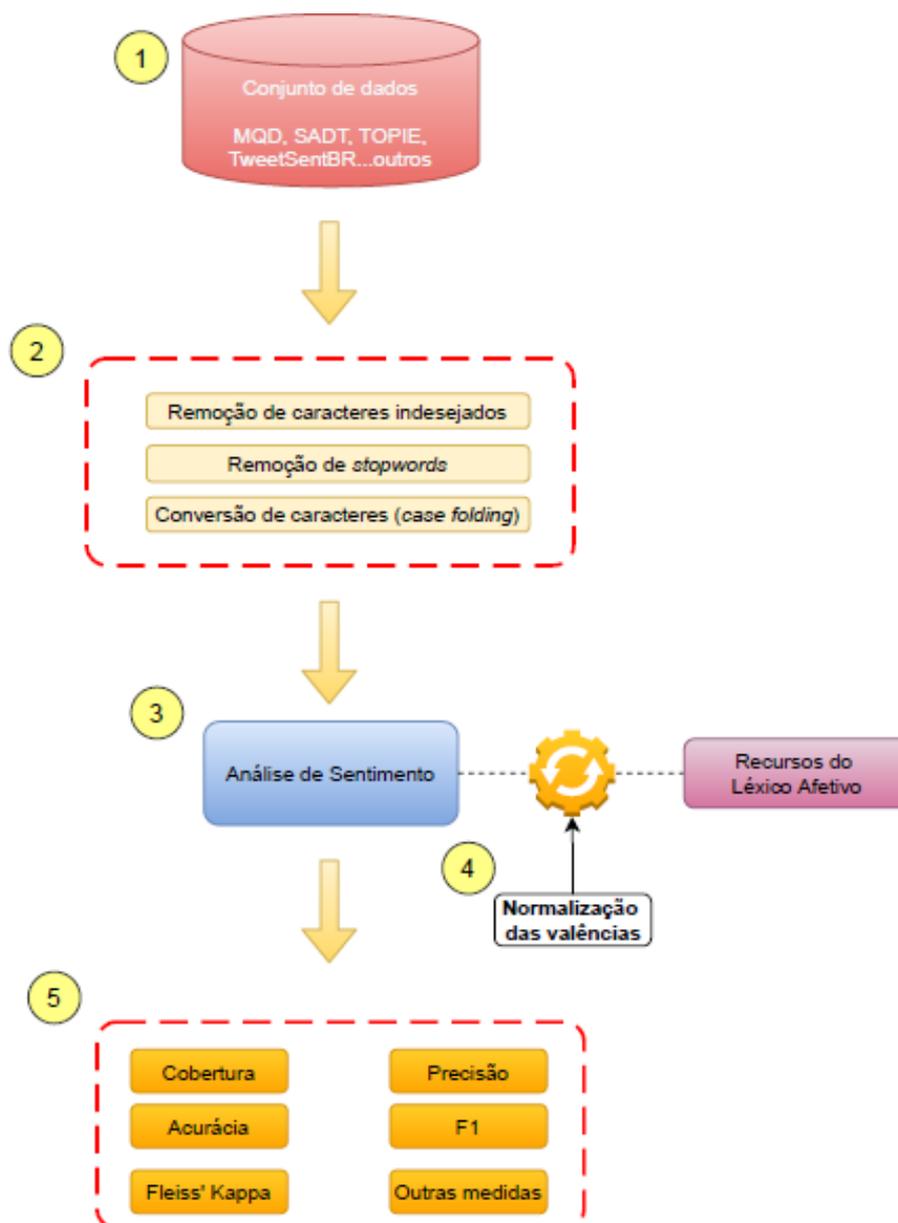


Figura 7 – Visão geral do *framework* elaborado para a execução dos experimentos. Na etapa 1 ocorre a coleta dos dados. Na etapa 2 ocorre as atividades de pré-processamento. Na etapa 3 ocorre a análise de sentimento com apoio da etapa 4 (normalização das valências). Na etapa 5 ocorre a aplicação das medidas de avaliação. Fonte: Elaborada pelo autor.

4.4- Método Utilizado para Análise de Sentimento

De forma análoga a Kim and Hovy [2004] e Musto et al. [2014], esta pesquisa adotou para AS a técnica de Abordagem Baseada em Dicionário que consiste na utilização

de recursos do LA para detectar sentimentos. O Algoritmo 2 descreve os detalhes desta técnica aplicada à esta pesquisa.

Algoritmo 2 – Pseudocódigo para detectar sentimento utilizando a técnica de Abordagem Baseada em Dicionário

```

1  Entrada
2      sentença  $S$ 
3      léxico  $L$ 
4  Saída
5      sentimento  $R$ 
6
7  Procedimento DetectarSentimento ( $S, L$ )
8      pontuação  $P \leftarrow 0$ 
9      valência afetiva  $v \leftarrow 0$ 
10
11      $S \leftarrow pre\_processamento(S)$ 
12
13     para cada palavra  $p$  em  $S$  faça:
14         para cada palavra  $q$  em  $L$  faça:
15             se  $p = q$  então:
16                  $v \leftarrow valenc(q, L)$ 
17                  $v \leftarrow normalizar\_valencia(v)$ 
18                  $P \leftarrow P + v$ 
19             fim-se
20         fim-para
21     fim-para
22
23     se  $P \geq 0,003$  então:
24          $R \leftarrow positivo$ 
25     senão se  $P \leq -0,003$  então:
26          $R \leftarrow negativo$ 
27     senão:
28          $R \leftarrow neutro$ 
29     fim-se
30
31     retorna  $R$ 

```

O algoritmo recebe dois argumentos de entrada (sentença S , léxico L) retornando como saída o sentimento R . Na linha 8 inicializa-se com 0 (zero) a variável de **pontuação** P aplicando o mesmo valor (0) à variável **valência afetiva** (linha 9). Na linha 11 ocorre a tarefa de pré-processamento, removendo ruídos da sentença S atribuindo-lhe um novo valor (sem ruídos). Na linha 13 ocorre a iteração pelas palavras p contidas na sentença S . Na linha 14 ocorre a iteração das palavras q contidas no léxico L . Na linha 15, de maneira literal (e.g. a=a, b=b), ocorre a checagem da equivalência entre a palavra da sentença

(p) e a palavra do léxico (q). Na linha 16, se as palavras p e q forem equivalentes ocorre a extração no léxico L da VA associada à palavra q . Na linha 17 ocorre o processo de normalização da valência (detalhes na Seção 4.5) na variável v atribuindo-lhe um novo valor. Na linha 18, computa-se o valor normalizado v acumulando-o na variável P .

A polaridade é caracterizada por uma escala numérica que varia entre $[-1 .. 1]$. Valores próximos a extremidade mais baixa (-1) indicam que o sentimento evocado é mais negativo. Valores próximos a extremidade mais alta (1) indicam que o sentimento evocado é mais positivo. Valores próximo a zero (0) indicam um sentimento neutro. Após o término da iteração das palavras da sentença (linha 21), de maneira semelhante ao estudo de Silva et al. [2019], a linha 23 adota um limiar na pontuação acumulada à variável P para definir o sentimento R (Figura 8). Por fim, retorna-se o sentimento detectado na sentença S definido na variável R (linha 31).

$$R = \begin{cases} P \geq 0,003 & \rightarrow \textit{positivo} \\ P \leq -0,003 & \rightarrow \textit{negativo} \\ P > -0,003; P < 0,003 & \rightarrow \textit{neutro} \end{cases}$$

Figura 8 – Limiar da pontuação para detecção de sentimento. Se o valor da pontuação P for maior ou igual a $0,003$, define-se o sentimento como **positivo**. Se o valor da pontuação P for menor ou igual a $-0,003$, define-se o sentimento como **negativo**. Se o valor da pontuação P estiver entre $-0,003$ e $0,003$, define-se o sentimento como **neutro**. Fonte: Elaborada pelo autor.

4.5- Normalização das valências dos Léxicos Afetivos

Apesar de almejam o mesmo propósito, nem sempre os LAs descritos no Capítulo 2 implementam as VAs nas palavras de modo semelhante. Existem LAs que representam estas valências de forma numérica categórica (e.g. 1 e -1) ou por meio de uma escala numérica (e.g. $[-1 .. 1]$). Há também aqueles que representam as valências em textos (e.g. *positive*, *negative*, *neutral*) ou em símbolos (e.g. $+$ e $-$).

Neste sentido, há LAs que complementam suas VAs por meio da utilização de **intensificadores** cujo o objetivo é dar um “peso” maior/menor à palavra referente ao sentimento evocado. Os intensificadores também são representados de diversas maneiras pelos LAs (e.g. *strongsubj*, *weaksbj*). Como se percebe, há diversas maneiras de

representar VA entre os LAs. A Tabela 24 apresenta um exemplo das variedades de representações das VAs pelos LAs.

Tabela 24 – Exemplo de múltiplos tipos de representações dos VA nas palavras.

Léxico	V-Pos	V-Neg	V-Neu	Intensificador
ANEW-BR	valor > 6	valor ≥ 1 e < 4	valor ≥ 4 e < 6	Aplicado à escala
BabelSenticnet	valor > 0	valor < 0	valor = 0	Aplicado à escala
MPQA	positive	negative	neutral	strongsubj
Polyglot	+1	-1	N/A	N/A
WordNetAffectBR	+	-	N/A	N/A

Diante da multiplicidade de representação das valências pelos LAs, tornou-se necessário aplicar um processo para **normalização** destas VAs, possibilitando a detecção de sentimento na abordagem utilizada nesta pesquisa (Seção 4.4). O objetivo da normalização é alterar as VAs dos léxicos (quando necessário) para uma escala comum ($[-1..1]$), sem distorcer as diferenças nos intervalos de valores. Esse método é necessário apenas quando os parâmetros tiverem intervalos muito diferentes. A Tabela 25 sintetiza os léxicos que são normalizados e em seguida, os detalhes do processo de normalização para cada LA.

1. **AffectPT-br** (seção 2.1)

- (a) **Valência positiva:** As palavras no léxico associadas à categoria 31-*posemo* (*Positive emotions*) tem suas VAs normalizadas para o valor 1. Exemplo: Palavra “feliz” $v = 1$.
- (b) **Valência negativa:** As palavras no LA associadas à categoria 32-*negemo* (*Negative emotions*) tem suas valências normalizadas para o valor -1 . Exemplo: Palavra “tristeza” $v = -1$.
- (c) **Valência neutra:** Não se aplica nas palavras deste léxico.

2. **ANEW-Br** (seção 2.2)

- (a) **Valência positiva:** As palavras no LA com valor $v \geq 6$ tem suas VAs normalizadas pela equação $v = p - 6$, sendo p o valor associado à palavra, 6 o limiar mínimo para valência positiva e v o valor utilizado para cálculo, gerando uma

Tabela 25 – Resumo da normalização efetuada nos Lexicos Afetivos. Pode-se notar que alguns lexicos não são normalizados (e.g., Polyglot).

Léxico	Normaliza V-Pos	Normaliza V-Neg	Normaliza V-Neu
AffectPT-br	x	x	
ANEW-Br	x	x	x
BabelSenticnet	x	x	
LIWC2007-PT-br	x	x	
LIWC2015-PT-br	x	x	
MPQA-PT	x	x	x
OpLexicon			x
Polyglot			
ReLi-Lex			
Sentilex-PT	x	x	x
SentiStrength			
WordNetAffectBR	x	x	
WordNetAffectBR(adapt)	x	x	

escala de intensidade de $v = [0..3]$. Exemplo: Palavra “admirado”, $p = 7$, então $v = (7 - 6)$, resultando $v = 1$.

- (b) **Valência negativa:** As palavras no LA com valor $v \geq 1$ e $v < 4$ tem suas valências normalizadas pela equação $v = (4 - p) \cdot -1$, sendo p o valor associado à palavra, 4 o limiar mínimo para valência negativa e v o valor utilizado para cálculo, gerando uma escala de intensidade de $v = [-3..0]$. Exemplo: Palavra “agonia”, $p = 2$, então $v = (4 - 2) \cdot -1$, resultando $v = -2$.
- (c) **Valência neutra:** As palavras no LA com valor $v \geq 4$ e $v < 6$ tem suas VAs normalizadas para o valor 0,0001. Exemplo: Palavra “agulha” $v = 0,0001$.

3. BabelSenticnet (seção 2.3)

Retorna os valores [$v > 0$, $v \leq 1$] para *positivo*, [$v \geq -1$, $v < 0$] para *negativo*. Não se aplica *neutro* para este léxico.

4. LIWC2007-PT-br (seção 2.4)

- (a) **Valência positiva:** As palavras no LA associadas à categoria 126-*posemo* (*Positive emotions*) tem suas VAs normalizadas para o valor 1. Exemplo: Palavra “legal” $v = 1$.
- (b) **Valência negativa:** As palavras no LA associadas à categoria 127-*negemo* (*Negative emotions*) tem suas VAs normalizadas para o valor -1 . Exemplo: Palavra “péssimo” $v = -1$.
- (c) **Valência neutra:** Não se aplica nas palavras deste léxico.

5. LIWC2015-PT-br (seção 2.5)

- (a) **Valência positiva:** As palavras no LA associadas à categoria 31-*posemo* (*Positive emotions*) tem suas VAs normalizadas para o valor 1. Exemplo: Palavra “contente” $v = 1$.
- (b) **Valência negativa:** As palavras no LA associadas à categoria 32-*negemo* (*Negative emotions*) tem suas VAs normalizadas para o valor -1 . Exemplo: Palavra “raiva” $v = -1$.
- (c) **Valência neutra:** Não se aplica nas palavras deste léxico.

6. MPQA-PT (seção 2.6)

- (a) **Valência positiva:** As palavras ou expressões no LA com valor categórico *positive*, tem suas VAs normalizadas com o valor $v = 1$. Se a palavra faz uso do intensificador *strongsubj*, o valor é acrescido de $+1$. Exemplo: Palavra “aclamar”, valor “*positive*”, então $v = 1$. Intensificador “*strongsubj*”, então $v = 2$.
- (b) **Valência negativa:** As palavras ou expressões no LA com valor categórico *negative*, tem suas VAs normalizadas com o valor $v = -1$. Se a palavra faz uso do intensificador *strongsubj*, o valor é acrescido de -1 . Exemplo: Palavra “adoentado”, valor “*negative*”, então $v = -1$. Intensificador “*strongsubj*”, então $v = -2$.
- (c) **Valência neutra:** As palavras ou expressões no LA com valor categórico *neutral*, tem suas VAs normalizadas com o valor $0,0001$. Exemplo: Palavra “alarme” $v = 0,0001$.

7. OpLexicon (seção 2.7)

- (a) **Valência positiva:** Não se aplica nas palavras deste léxico.

(b) **Valência negativa:** Não se aplica nas palavras deste léxico.

(c) **Valência neutra:** As palavras no léxico com valor $v = 0$, tem suas VAs normalizadas para o valor $v = 0,0001$. Exemplo: Palavra “abaixar”, $v = 0,0001$.

8. **Polyglot** (seção 2.8)

Retorna os valores $[v > 0, v \leq 1]$ para *positivo*, $[v \geq -1, v < 0]$ para *negativo*. Não se aplica *neutro* para este léxico.

9. **ReLi-Lex** (seção 2.9)

Retorna os valores $[v > 0, v \leq 1]$ para *positivo*, $[v \geq -1, v < 0]$ para *negativo*. Não se aplica *neutro* para este léxico.

10. **Sentilex-PT** (seção 2.10)

(a) **Valência positiva:** As palavras ou expressões no LA com valor categórico de POL:N0=1, tem suas VAs normalizadas para o valor $v = 1$. Exemplo: Palavra “elogio”, $v = 1$.

(b) **Valência negativa:** As palavras ou expressões no LA com valor categórico de POL:N0=-1, tem suas VAs normalizadas para o valor $v = -1$. Exemplo: Palavra “estúpido”, $v = -1$.

(c) **Valência neutra:** As palavras no LA com valor categórico de POL:N0=0, tem suas VAs normalizadas para o valor $v = 0,0001$. Exemplo: Palavra “garrido”, $v = 0,0001$.

11. **SentiStrength** (seção 2.11)

Retorna os valores $[v > 0, v \leq 1]$ para *positivo*, $[v \geq -1, v < 0]$ para *negativo*. Não se aplica *neutro* para este léxico.

12. **WordNetAffectBR** (seção 2.12)

(a) **Valência positiva:** As palavras no LA com valor categórico de “+”, tem suas valências normalizadas para o valor $v = 1$. Exemplo: Palavra “animado”, $v = 1$.

(b) **Valência negativa:** As palavras no LA com valor categórico de “-”, tem suas VAs normalizadas para o valor $v = -1$. Exemplo: Palavra “conflito”, $v = -1$.

(c) **Valência neutra:** Não se aplica nas palavras deste léxico.

13. **WordNetAffectBR(adapt)** (seção 2.13)

- (a) **Valência positiva:** As palavras no LA com valor categórico de “+”, tem suas VAs normalizadas para o valor $v = 1$. Exemplo: Palavra “afeição”, $v = 1$.
- (b) **Valência negativa:** As palavras no LA com valor categórico de “-”, tem suas VAs normalizadas para o valor $v = -1$. Exemplo: Palavra “antipático”, $v = -1$.
- (c) **Valência neutra:** Não se aplica nas palavras deste léxico.

5- Resultados

Este capítulo apresenta os resultados dos experimentos aplicados nesta pesquisa. Inicialmente na Seção 5.1 se fez uma avaliação dos LAs em tarefas de AS comparando os resultados por meio das medidas de precisão, revocação, acurácia e F1. Em seguida na Seção 5.2, devido a grande quantidade de LAs abordados nesta pesquisa, foram selecionados os três léxicos de melhor desempenho na AS para proporcionar melhor visualização dos resultados. Em sequência na Seção 5.3, é exibido a MC dos três LAs aplicados ao conjunto de dados MQD. Por fim nas Seções 5.4 e 5.5, buscou-se compreender a concordância afetiva nas sentenças e palavras entre os léxicos.

5.1- Avaliação dos Léxicos Afetivos

O objetivo principal desta pesquisa é evidenciar a contribuição dos LAs em português para a tarefa de AS. São utilizados quatro conjuntos de dados (descritos na Seção 4.2). As medidas utilizadas para avaliar o quão bom são os LAs em realizarem tarefas de AS utilizando ABD, incluem precisão, revocação, acurácia e F1. Dentre estas, destacamos a medida F1, que é a média harmônica entre precisão e revocação.

A Figura 9 ilustra os resultados da medida F1 referente a AS executada pelos LAs. O conjunto de dados utilizado foi o MQD. Nota-se que o léxico AffectPT-br apresenta o melhor desempenho, seguido dos léxicos LIWC2015-PT-br e SentiStrength. Os léxicos Reli-Lex, WordNetAffectBR e BabelSenticNet ocupam as últimas posições.

Na Figura 10 observa-se os resultados da medida F1 aplicados ao conjunto de dados SADT. Percebe-se que o léxico AffectPT-br também apresenta o melhor desempenho, seguido dos léxicos Sentilex-PT e SentiStrength. Os léxicos MPQA-PT, WordNetAffectBR e BabelSenticNet ocupam as últimas posições.

Na Figura 11 observa-se os resultados da medida F1 aplicados ao conjunto de dados TOPIE. Constata-se que o léxico SentiStrength é o que apresenta o melhor desempenho, seguido dos léxicos LIWC2015-PT-br e AffectPT-br. Os léxicos WordNetAf-

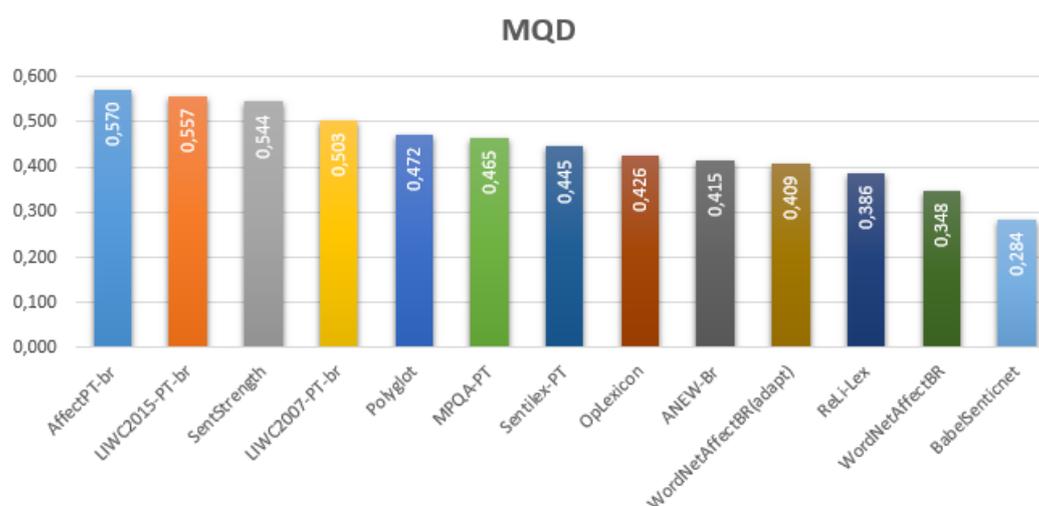


Figura 9 – Medida F1 no conjunto de dados MQD. O léxico AffectPT-br apresenta o melhor resultado entre os demais na análise de sentimento. Fonte: Elaborada pelo autor.

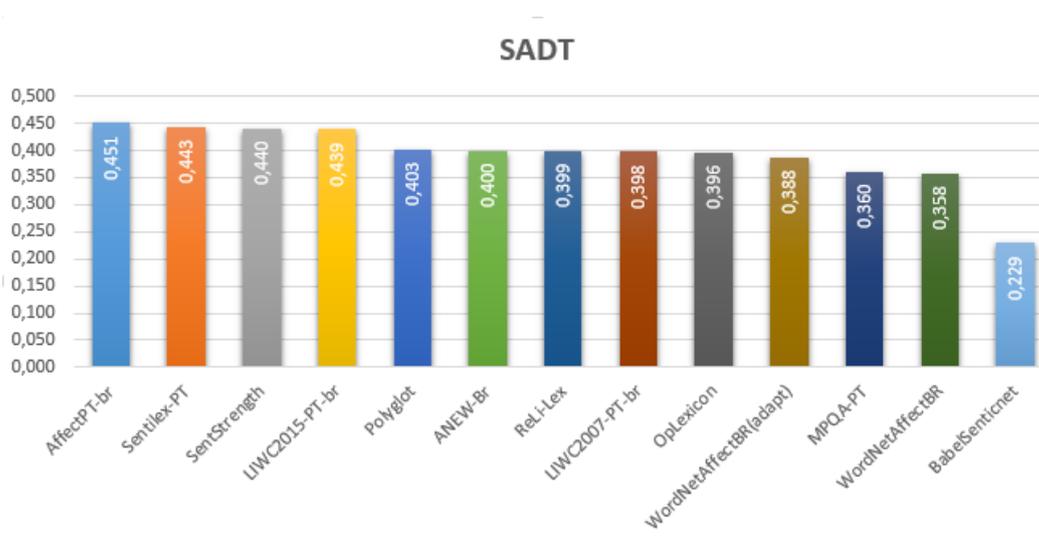


Figura 10 – Medida F1 no conjunto de dados SADT. O léxico AffectPT-br apresenta o melhor resultado entre os demais na análise de sentimento. Fonte: Elaborada pelo autor.

fectBR(adapt), BabelSenticNet e WordNetAffectBR ocupam as últimas posições.

Na Figura 12 observa-se os resultados da medida F1 aplicados ao conjunto de dados TweetSentBR. Verifica-se que o léxico LIWC2015-PT-br apresenta o melhor desempenho, seguido dos léxicos LIWC2007-PT-br e AffectPT-br. Os léxicos BabelSenticNet, WordNetAffectBR(adapt) e WordNetAffectBR ocupam as últimas posições.

Objetivando melhor legibilidade, os resultados completos dos experimentos aplicados incluindo todas as medidas de avaliação (precisão, revocação e acurácia) podem ser visualizados no Apêndice A deste documento.

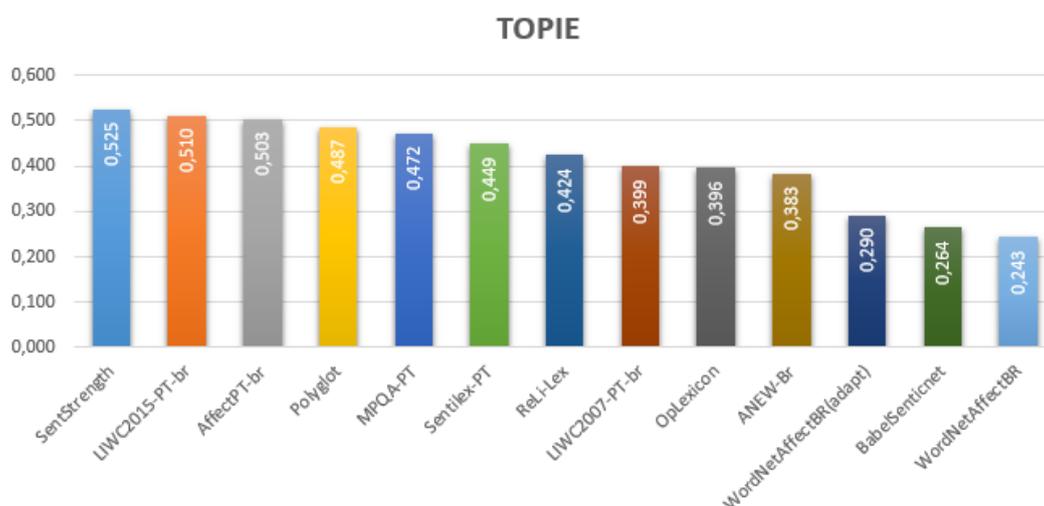


Figura 11 – Medida F1 no conjunto de dados TOPIE. O léxico SentiStrength apresenta o melhor resultado entre os demais na análise de sentimento. Fonte: Elaborada pelo autor.

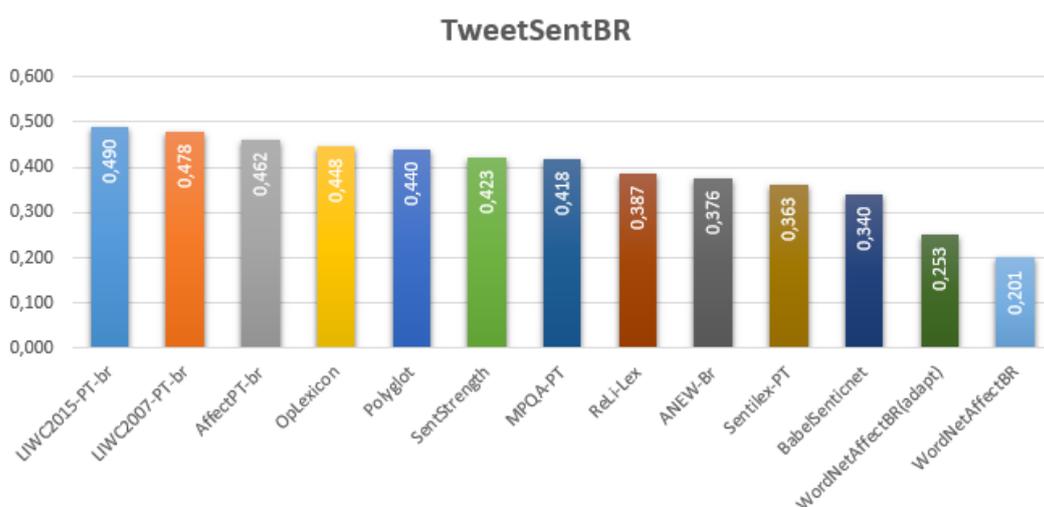


Figura 12 – Medida F1 no conjunto de dados TweetSentBR. O léxico LIWC2015-PT-br apresenta o melhor resultado entre os demais na análise de sentimento. Fonte: Elaborada pelo autor.

5.2- Seleção dos melhores léxicos para apresentação dos resultados

Em consequência da grande quantidade de LAs abordados neste estudo (Capítulo 2), destacaremos apenas os três léxicos que apresentaram os melhores resultados da medida F1 na tarefa de AS. O objetivo é proporcionar melhor legibilidade e facilitar a compreensão de outros experimentos aplicados.

Para a escolha dos melhores léxicos, aplicou-se como critério o teste de significância estatística de Friedman (Seção 1.6) que envolve a classificação dos dados para detectar diferenças entre os grupos. O teste resultou em uma estatística de 11.95890, um valor de $p = 0.0001$, sendo $\alpha = 0.05$ e $p < \alpha$ portanto, rejeita-se a hipótese nula (H_0) ou seja, variáveis numéricas não procedem da mesma população. Os três melhores léxicos com significância estatística são: AffectPT-br, LIWC2015-PT-br e SentiStrength. O teste foi aplicado utilizando uma ferramenta *online*¹ [Rodríguez-Fdez et al., 2015]. A Tabela 26 apresenta o resultado ordenado pelo teste de Friedman e a Tabela 27 apresenta a medida F1 dos léxicos selecionados.

Tabela 26 – Tabela dos melhores léxicos classificados pelo teste estatístico de Friedman.

Posição	Léxico
1	AffectPT-br
2	LIWC2015-PT-br
3	SentiStrength
4	Polyglot
5	LIWC2007-PT-br
6	Sentilex-PT
7	MPQA-PT
8	OpLexicon
9	ReLi-Lex
10	ANEW-Br
11	WordNetAffectBR(adapt)
12	BabelSenticnet
13	WordNetAffectBR

¹<http://tec.citius.usc.es/stac/index.html>

Tabela 27 – Medida F1 dos três melhores léxicos. Percebe-se que o léxico AffectPT-br ocupa o 1º lugar nos conjuntos de dados MQD e SADT, ficando em 2º lugar no TweetSentBR e 3º no TOPIE. O léxico LIWC2015-PT-br ocupa o 1º lugar no conjunto de dados TweetSentBR, ficando em 2º no MQD e TOPIE e na 3ª posição no SADT. O léxico SentiStrength ocupa o 1ª lugar no conjunto de dados TOPIE, 2ª lugar no SADT e 3º lugar nos conjuntos de dados MQD e TweetSentBR.

Léxico	MQD	SADT	TOPIE	TweetSentBR
AffectPT-br	0,570	0,451	0,503	0,462
LIWC2015-PT-br	0,557	0,439	0,510	0,490
SentiStrength	0,544	0,440	0,525	0,423

5.3- Matriz de Confusão

Para melhor compreender como os LAs se comportam com a abordagem proposta nesta pesquisa, foram criadas matrizes de confusão para observar detalhadamente os resultados da AS executada pelos léxicos. Objetivando proporcionar melhor legibilidade e compreensão dos resultados, nesta seção apenas o conjunto de dados MQD é abordado (escolhido de maneira aleatória). A Figura 13 apresenta a MC abordando os três melhores léxicos no conjunto de dados MQD.

Observa-se que o AffectPT-br apresenta melhor desempenho em detectar sentimentos *positivos*, detectando corretamente 310 de 508 exemplos (61%). Entretanto, ele apresenta baixo desempenho ao lidar com sentimentos *negativos*, detectando apenas 251 de 539 exemplos (46,6%). O LIWC2015-PT-br também apresenta melhor desempenho em detectar sentimentos *positivos* detectando corretamente 314 de 508 exemplos (61,8%). Também apresenta baixo desempenho ao lidar com sentimentos *negativos*, detectando apenas 241 de 539 exemplos (44,7%). Já o SentiStrength apresenta melhor desempenho em detectar sentimentos *negativos*, detectando corretamente 334 de 539 exemplos (62%). Seu baixo desempenho está em detectar sentimentos *neutros*, detectando apenas 197 de 410 exemplos (48%). As MCs para os conjuntos de dados e LAs podem ser visualizados no Apêndice B deste documento.

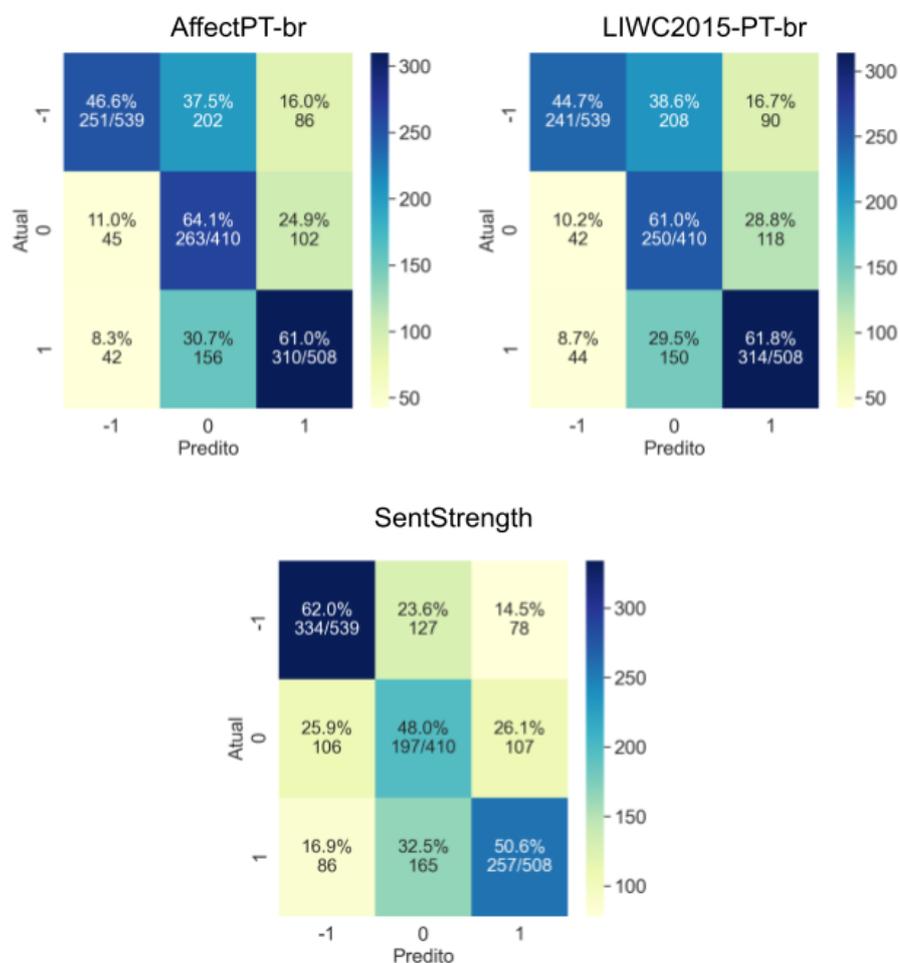


Figura 13 – Matriz de confusão - MQD. O eixo x (Predito) representa os valores preditos pelo léxico e o eixo y (Atual) representa os valores rotulados. O sentimento *positivo* é caracterizado nos eixos pelo número 1, seguidamente de -1 para *negativo*, e 0 para *neutro*. Fonte: Elaborada pelo autor.

5.4- Concordância Afetiva nas sentenças

Com muitos LAs executando tarefas de AS, é importante saber como estes léxicos concordam afetivamente entre si ao analisarem as sentenças (e.g. dada uma sentença, quantos concordaram com o sentimento *positivo*?). Para melhor entender esta concordância, foi utilizado o método Fleiss' Kappa (Seção 1.5) para medir o grau de concordância entre eles. A Figura 14 apresenta os níveis de concordância dos LAs aplicados em cada conjunto de dados e a Tabela 28 apresenta a escala para interpretação dos resultados.

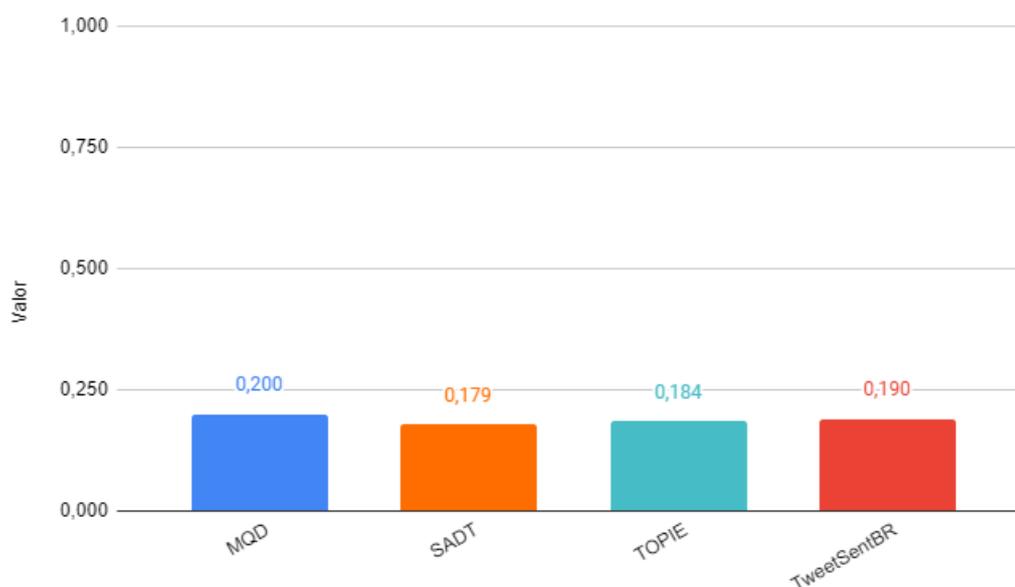


Figura 14 – Níveis de concordância afetiva entre os léxicos. Nota-se que os léxicos apresentam baixa concordância afetiva, indicando grande divergência entre si ao detectar sentimentos. Fonte: Elaborada pelo autor.

Tabela 28 – Escala para interpretação de valores do método Fleiss' Kappa.

Valor	Interpretação
< 0,01	Nenhuma ou quase nenhuma concordância
0,01–0,20	Breve concordância
0,21–0,40	Concordância razoável
0,41–0,60	Concordância moderada
0,61–0,80	Concordância considerável
0,81–1,00	Concordância plena

Observa-se que em todos os conjuntos de dados (MQD, SADT, TOPIE e TweetSentBR), a concordância afetiva entre os léxicos apresentaram-se na escala “breve concordância” ou seja, concordância extremamente baixa conforme indicado na Tabela 28. Isto evidencia que há grande divergência afetiva entre os léxicos para detectar o mesmo sentimento. Neste contexto, o mesmo experimento foi aplicado somente aos três melhores léxicos selecionados na Seção 5.2: AffectPT-br, LIWC2015-PT-br e SentiStrength (Figura 15). Ao executar o experimento, observou-se que os resultados apresentaram-se na escala “Concordância moderada” o que evidencia maior correspondência entre estes três

léxicos para detectar o mesmo sentimento.

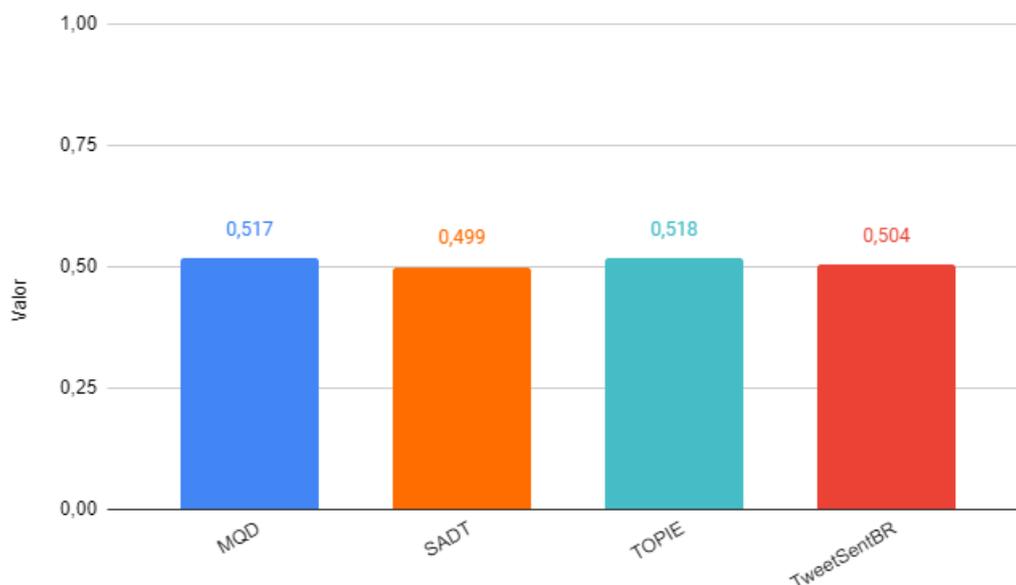


Figura 15 – Fleiss' Kappa dos Léxicos Afetivos mais influentes. Fonte: Elaborada pelo autor.

5.5- Concordância nas palavras entre os Léxicos Afetivos

Visto que já se conhece a concordância afetiva entre os LAs em nível de sentença, também é importante saber como as palavras que compõem estes léxicos concordaram entre si. Isto significa que se um LA inclui uma palavra considerando sua VA (e.g. “triste=*negativa*”), dentre os outros léxicos abordados nesta pesquisa, quantos abrangeram a mesma palavra e quantos consideraram a mesma afetividade?

Para executar este experimento, dois conjuntos de palavras foram produzidos: Conjunto X_1 e conjunto X_2 . O conjunto X_1 representa a coleção de palavras equivalentes entre um léxico observado e os demais léxicos. O conjunto X_2 representa o subconjunto de X_1 contendo palavras de VA equivalentes entre o léxico observado e demais. A Figura 16 apresenta um gráfico de concordância entre palavras aplicadas do léxico AffectPT-br com os demais. As Figuras 17 e 18 apresentam o mesmo gráfico aplicado aos léxicos LIWC2007-PT-br e SentiStrength. Os gráficos contendo a concordância entre palavras dos demais LAs podem ser visualizados no Apêndice C deste documento.

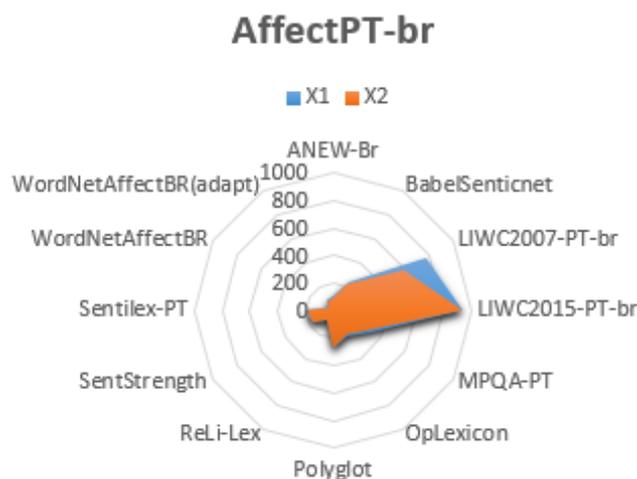


Figura 16 – Concordância entre palavras do léxico AffectPT-br. X_1 representa palavras equivalentes entre os léxicos. X_2 representa palavras de valência afetiva equivalentes. Observa-se que para o léxico AffectPT-br há maior concordância com o léxico LIWC2015-PT-br. Fonte: Elaborada pelo autor.

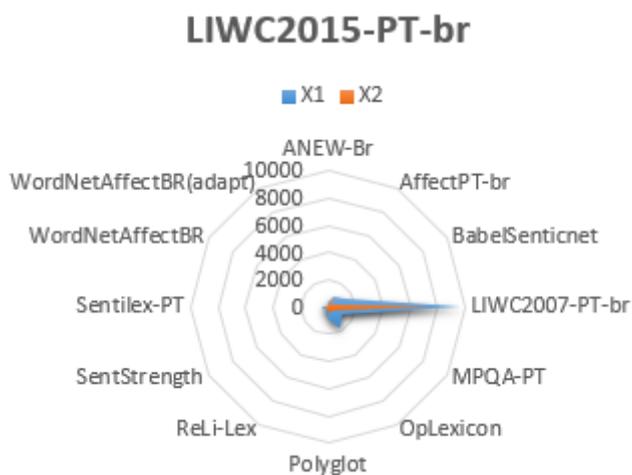


Figura 17 – Concordância entre palavras do léxico LIWC2015-PT-br. X_1 representa palavras equivalentes entre os léxicos. X_2 representa palavras de valência afetiva equivalentes. Observa-se que para o léxico LIWC2015-PT-br há maior concordância com o léxico LIWC2007-PT-br. Fonte: Elaborada pelo autor.

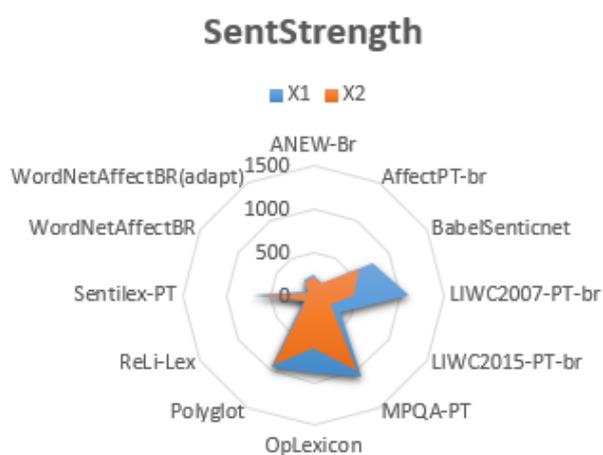


Figura 18 – Concordância entre palavras do léxico SentiStrength. X_1 representa palavras equivalentes entre os léxicos. X_2 representa palavras de valência afetiva equivalentes. Observa-se que para o léxico SentiStrength há maior concordância com os léxicos MPQA-PT e Polyglot. Fonte: Elaborada pelo autor.

6- Discussão dos Resultados

Neste capítulo, são abordados os resultados dos experimentos aplicados nesta pesquisa. Inicialmente são discutidos os resultados apresentados nas medidas de avaliação relativos à tarefa de AS entre os LAs. Em seguida, discute-se sobre a abordagem da utilização de diferentes conjuntos de dados utilizados nesta pesquisa. Debate-se também observações aplicadas nos experimentos referentes à anotações nos conjuntos de dados utilizados. Seguidamente, discute-se as divergências apresentadas entre os LAs. Finalizando, ocorre a discussão sobre os melhores LAs identificados nesta pesquisa e as conclusões finais.

6.1- Precisão, revocação, Acurácia e F1

Nos experimentos aplicados, os registros dos melhores resultados das medidas precisão, revocação, acurácia e F1 apresentaram-se no léxico AffectPT-br, no conjunto de dados MQD, sendo os valores respectivamente: *precisão* = 0,611; *revocação* = 0,566; *acurácia* = 0,566; *F1* = 0,570 (Apêndice A). As medidas de avaliação mais baixas apresentaram-se no léxico WordNetAffectBR no conjunto de dados TweetSentBR, sendo os valores respectivamente: *precisão* = 0,534; *revocação* = 0,298; *acurácia* = 0,298; *F1* = 0,201.

É interessante destacar que nestes resultados o léxico WordNetAffectBR, *a priori*, identifica corretamente (precisão de 0,534) um número de exemplos que proporcionalmente aproxima-se do léxico AffectPT-br (precisão de 0,611). Entretanto, sua baixa revocação no TweetSentBR (0,298) indica que o léxico não consegue abranger corretamente grande quantidade de exemplos. Este mesmo cenário também ocorre em outro conjunto de dados, o MQD, com uma precisão de 0,621 e revocação de 0,396.

Ao observar a quantidade de palavras que compõem o léxico WordNetAffectBR em relação aos outros LAs, percebe-se que ele é o LA de menor tamanho em estrutura (289 palavras) o que desperta uma possível correlação com seu desempenho. No entanto, uma

grande quantidade de palavras não é fator determinante para melhorias no desempenho dos LA. Isto pode ser observado ao analisar o léxico LIWC2007-PT-br. Ao considerar a quantidade de palavras deste léxico (127.149) percebe-se que é o LA de maior estrutura. Todavia, seu desempenho não se apresenta significativo de modo geral, ocupando a 4ª posição no conjunto de dados MQD, 8ª posição no conjunto de dados SADT e TOPIE e a 2ª posição no conjunto de dados TweetSentBR. Este mesmo cenário também ocorre em outro léxico, o OpLexicon (32.113 palavras), ao ocupar a 8ª posição no conjunto de dados MQD, 9ª posição no conjunto de dados SADT e TOPIE, e a 4ª posição no conjunto de dados TweetSentBR.

6.2- Utilização de conjunto de dados de diferentes domínios

A aplicação dos experimentos utilizando conjunto de dados de diferentes domínios ou seja, elaborados em contextos diferentes (e.g. TweetSentBR, criado a partir de comentários de programas de televisão; TOPIE, criado a partir de reclamações de empresas), transparece resultados mais consistentes quando se objetiva a generalização da AS baseada em LAs. Por exemplo, observa-se que o léxico Sentilex-PT ocupa a 2ª posição de melhor desempenho no conjunto de dados SADT. Entretanto, no conjunto de dados TOPIE o léxico ocupa a 6ª posição, no MQD ocupa a 7ª posição e no TweetSentBR ocupa a 10ª posição. Tais números indicam que deve-se reconsiderar o desempenho deste léxico no conjunto de dados SADT, pois ao ser aplicado em domínios diferentes, resultados não apresentaram o mesmo comportamento.

Todavia, o léxico AffectPT-br apresenta maior consistência em seu desempenho ao ser aplicado em dados de domínios diferentes. Nos conjunto de dados MQD e SADT, o léxico ocupa a 1ª posição. No TweetSentBR e no TOPIE, o léxico ocupa a 3ª posição. Tais números indicam a relevância deste léxico em tarefas de AS, pois o mesmo consegue se manter entre os 3 melhores resultados em todos os conjuntos de dados (MQD, SADT, TOPIE e TweetSentBR), ocupando a 1ª posição em ao menos dois (MQD e SADT). Por esse motivo, esse léxico foi considerado o melhor léxico para análise de sentimentos com a Abordagem Baseada em Dicionário em Português.

6.3- Anotações incorretas no corpus

As investigações para compreensão do desempenho dos LAs revelaram, em todos os conjunto de dados utilizados nesta pesquisa, que algumas sentenças apresentaram-se rotuladas de maneira não condizente com o sentimento evocado. Tomemos como exemplo uma sentença presente no conjunto de dados SADT: *“Uma vez o i do teclado de uma amiga quebrou e ela usou o l minúsculo como substituto pro i maiúsculo”*. Esta sentença está anotada como sentimento **positivo** entretanto, é evidente que o sentimento evocado na sentença é um sentimento **neutro**. Um outro exemplo manifesta-se no conjunto de dados TOPIE ao analisar a sentença: *“Boa tarde! Alguém sabe o que está acontecendo com os celulares da @Vivo que não completam ligação?”*. Tal sentença está anotada como sentimento **negativo** entretanto, a sentença refere-se à uma indagação não expressando nenhum tipo de palavra que evoque um sentimento relacionado a sua anotação. A Tabela 29 expressa mais exemplos semelhantes.

Tabela 29 – Exemplo de anotações divergentes no corpus.

Conjunto de dados	Sentença	Sentimento anotado	Sentimento evocado
MQD	<i>“Declarações de amizade e dos desenhos doidos no caderno de escola”</i>	negativo	neutro
SADT	<i>“Até achei estranho pq ele nunca me da nada”</i>	positivo	neutro
TOPIE	<i>“Querida @Oi sua internet é boa, mas quando cai, parece que morre, porque não volta mais”</i>	neutro	negativo
TweetSentBR	<i>“Meu marido é mas novo q eu e somos felizes. Idade nao muda nada #encontro”</i>	neutro	positivo

Anotações incorretas no corpus contribuem para dificultar o desempenho das tarefas de AS, bem como a consistência das avaliações dos LA. Pode-se destacar algumas dificuldades para anotação: i) uso de palavras de sentimento pode ser enganoso; ii) uso de ironia/sarcasmo em muitos domínios; iii) a opinião pode depender do observador; e iv) polaridade do comentário nem sempre é objeto de consenso [Ramos and Freitas, 2019].

Esta pesquisa não quantificou a recorrência de tais casos e tampouco os corrigiu,

pois compreende que tal atividade restringe-se às extensões dos trabalhos que produziram os conjuntos de dados. Entretanto, considera-se importante destacar tal tipo de observação.

6.4- Divergências entre os Léxicos Afetivos

Esta pesquisa buscou avaliar o quão harmônico os LAs se mostram em executar tarefas de AS. O resultado apresentado pelo método Fleiss' Kappa, indica baixa harmonia entre os LA ao detectar o sentimento em nível de sentença, ou seja, se para uma sentença um léxico atribui um sentimento **negativo**, outros léxicos predizem sentimentos diferentes para a mesma sentença (*positivo* ou *neutro*). É importante ressaltar que tais divergências ocorrem em todos os conjunto de dados utilizados nesta pesquisa. Sendo assim, observa-se que as estruturas dos LA rumam em direções diferentes do interesse coletivo ao tentar detectar sentimentos. Isto enfatiza a necessidade de pesquisas que lidem com um problema já previsto pela literatura, referente a palavras que expressam orientações distintas entre domínios diferentes (e.g. “obrigado”, no sentido de **gratidão**, e “obrigado”, no sentido de **imposição**), distanciando-se de sua semântica padrão em um LA para uso geral podendo potencialmente ocasionar baixo desempenho em tarefas de AS [Liu, 2017].

As divergências entre os LA podem ser visualizadas em maior profundidade observando a concordância de palavras e as afetividade entre si. O léxico SentiStrength apresenta concordância moderada de palavras (palavras equivalentes) com o léxico LIWC2007-PT-br (1.090 palavras de 2.112) entretanto, há uma disparidade considerável na concordância de afetividade (460 palavras de 1.090). Isto indica um contraste entre os desenvolvedores dos léxicos no que se refere à afetividade aplicada na palavra. Tal contraste também se apresenta nos demais léxicos, evidenciando o problema já detectado na literatura, referente a disponibilização de muitos LA que proporcionam inconsistências e erros [Liu, 2017].

6.5- Melhores Léxicos Afetivos

Os LA mais influentes selecionados pelo teste estatístico de Friedman (AffectPT-br, LIWC2015-PT-br, SentiStrength), indicam que as estruturas de tais léxicos caminham para um conjunto de palavras de melhor harmonização nas tarefas de AS. Isto pode ser observado pelo método Fleiss' Kappa, ao apresentar uma concordância moderada entre tais léxicos. Este resultado proporciona estudos que avaliem as estruturas destes léxicos e consequentemente, a elaboração de novos LA a partir de um *baseline*.

Observando a Matriz de Confusão do léxico AffectPT-br no conjunto de dados MQD, notou-se que este léxico apresentou melhor desempenho em detectar sentimentos *positivos*, acertando 310 de 508 casos (61%). Contudo, o léxico SentiStrength apresentou melhor desempenho em detectar sentimentos *negativos*, acertando 334 de 539 casos (62%). Nesta vertente, avaliar as palavras que proporcionaram tais resultados motiva o desenvolvimento de novos estudos que objetivem melhor compreensão da distribuição de emoções evocadas por termos em diferentes partes de discursos [Mohammad and Turney, 2010], potencializando novos LA mais eficazes em tarefas de AS na língua portuguesa.

7- Conclusão

Os resultados apresentados nesta pesquisa revelam o potencial que os LAs em Português expressam em tarefas de AS. Destaca-se os três melhores LAs identificados por meio da medida F1, sendo eles AffectPT-br, LIWC2015-PT-br e SentiStrength (Seção 5.2).

A realização de experimentos utilizando quatro conjunto de dados de diferentes domínios (MQD, SADT, TOPIE e TweetSentBR), produziram resultados que contribuíram para a detecção de pontos fracos que dificilmente seriam identificados caso os experimentos fosse realizados em um conjunto de dados específico. Como por exemplo, o desempenho LA Sentilex-PT que ocupa a 2ª posição no conjunto de dados SADT porém, seu desempenho degrada ao ser aplicado com a mesma abordagem ao conjunto de dados TOPIE, sendo rebaixado à 7ª posição. E o mesmo cenário ocorre no conjunto de dados TweetSentBR, sendo rebaixado à 10ª posição. Isto revela o viés identificado no conjunto de dados SADT, direcionando esforços para uma reconsideração do resultado apresentado.

Observando como os LAs concordavam afetivamente entre si ao detectarem o sentimento em uma sentença, os resultados indicaram que houve grande divergência entre si ao tentarem detectar sentimentos em textos. O método Fleiss' Kappa revelou que em todos os conjunto de dados (MQD, SADT, TOPIE e TweetSentBR) os LAs possuem uma “breve concordância” ao detectarem sentimentos. Isto significa que dada uma sentença, um LA detecta um sentimento **positivo**, um segundo LA detecta a mesma sentença como **negativo** e um terceiro LA detecta como **neutro**. Ao observar a concordância afetiva considerando apenas os três melhores léxicos (AffectPT-br, LIWC2015-PT-br e SentiStrength) aplicados em todos os conjunto de dados, o método Fleiss' Kappa revelou uma “concordância moderada” entre si, indicando maior consenso entre este LAs ao detectarem sentimentos em textos.

De maneira semelhante, discordâncias afetivas também se mostraram presentes nos experimentos ao observar a orientação da polaridade em palavras equivalentes entre os LAs. Isto significa que se uma está palavra presente na estrutura de dois léxicos distintos, a polaridade desta palavra é diferente entre eles. Tomemos com exemplo,

a palavra “obrigado” que no sentido de **gratidão**, no LA SentiStrength esta palavra tem a polaridade definida como **positivo**. Entretanto, a mesma palavra no sentido de **imposição** aparece com a polaridade definida como **negativa** nos léxicos OpLexicon e Sentilex-PT. Sendo assim, percebe-se uma contradição entre os LAs. Este resultado pode auxiliar trabalhos futuros que objetivem resolver problemas já identificados pela literatura, como por exemplo o desafio em identificar e lidar com palavras e frases que têm o sentimento dependente de domínio ou de contexto [Liu, 2017].

Os resultados apresentados na Matriz de Confusão (Seção 5.3) indicam em qual classe (positivo, negativo e neutro) os melhores léxicos (AffectPT-br, LIWC2015-PT-br e SentiStrength) apresentam melhor desempenho. Por exemplo, os léxicos AffectPT-br e LIWC2015-PT-br são melhores em predizer sentimentos das classes **positivo** e **neutro**. Entretanto, o léxico SentiStrength é melhor em predizer sentimentos da classe **negativo**. Isto motiva a elaboração de trabalhos futuros que combinem palavras dos LAs que contribuiriam para que o léxico apresentasse melhor desempenho nas classes, formando então um novo LA e conseqüentemente, com potencial para apresentar melhor desempenho em tarefas de AS.

Não há trabalhos na literatura que se assemelhem à investigação desenvolvida por esta pesquisa. Portanto, de um modo geral, esta pesquisa contribui de três maneiras: 1) Ser um referencial para identificação de LA em Português, auxiliando trabalhos relacionados na escolha do melhor LA; 2) Servir de *baseline* fornecendo recursos para que trabalhos futuros sejam capazes de elaborar novos LA em Português ou aprimorar os que foram apresentados nesta pesquisa, objetivando melhores resultados em tarefas de AS; 3) Colaborar na área da CA, de maneira a fornecer suporte a futuros trabalhos que utilizem LAs para efetuar AS. Isso, atuando como componente principal ou como instrumento de apoio e auxiliando modelos que façam utilização de algoritmos de Aprendizado de Máquina (e.g. SVM, Naive Bayes, Random Forest), ou em arquitetura de redes neurais profundas (*deep learning*). O propósito é ampliar a generalidade para predição de sentimentos em textos na língua portuguesa.

Referências Bibliográficas

- Agarwal, A., Sharma, V., Sikka, G., and Dhir, R. (2016). Opinion mining of news headlines using sentiwordnet. In *2016 Symposium on Colossal Data Analysis and Networking (CDAN)*, pages 1–5.
- Ahire, S. (2014). A survey of sentiment lexicons. *Computer Science and Engineering IIT Bombay, Bombay*.
- Al-Rfou', R., Perozzi, B., and Skiena, S. (2013). Polyglot: Distributed word representations for multilingual NLP. In *Proceedings of the Seventeenth Conference on Computational Natural Language Learning*, pages 183–192, Sofia, Bulgaria. Association for Computational Linguistics.
- Al-Thubaity, A., Alharbi, M., Alqahtani, S., and Aljandal, A. (2018). A saudi dialect twitter corpus for sentiment and emotion analysis. In *2018 21st Saudi Computer Society National Computer Conference (NCC)*, pages 1–6.
- Almeida, M. S. C., Pinto, C., Figueira, H., Mendes, P., and Martins, A. F. T. (2015). Aligning opinions: Cross-lingual opinion mining with dependencies. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 408–418, Beijing, China. Association for Computational Linguistics.
- Althoff, T., Jindal, P., and Leskovec, J. (2017). Online actions with offline impact: How online social networks influence online and offline user behavior. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pages 537–546. ACM.
- Aly, A. and Tapus, A. (2013). A model for synthesizing a combined verbal and nonverbal behavior based on personality traits in human-robot interaction. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 325–332. IEEE.
- Araujo, M., Diniz, J. P., Bastos, L., Soares, E., Ferreira, M., Ribeiro, F., and Benevenuto,

- F. (2016). ifeel 2.0: A multilingual benchmarking system for sentence-level sentiment analysis. In *Tenth International AAI Conference on Web and Social Media*.
- Araújo, M., Gonçalves, P., Cha, M., and Benevenuto, F. (2014). ifeel: a system that compares and combines sentiment analysis methods. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 75–78. ACM.
- Araya, N., Leger, P., and López, M. (2019). Whom do I choose to diffuse information on Twitter? An agent-based model approach. *IEEE Latin America Transactions*, 17(04):677–683.
- Arendt, D., Huang, Z., Shrestha, P., Ayton, E., Glenski, M., and Volkova, S. (2020). Crosscheck: Rapid, reproducible, and interpretable model evaluation. *arXiv preprint arXiv:2004.07993*.
- Asghar, M. Z., Sattar, A., Khan, A., Ali, A., Masud Kundi, F., and Ahmad, S. (2019). Creating sentiment lexicon for sentiment analysis in urdu: The case of a resource-poor language. *Expert Systems*, 36(3).
- Avanco, L. V. and Nunes, M. d. G. V. (2014). Lexicon-based sentiment analysis for reviews of products in brazilian portuguese. In *2014 Brazilian Conference on Intelligent Systems*, pages 277–281. IEEE.
- Baccianella, S., Esuli, A., and Sebastiani, F. (2010). Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In *Lrec*, volume 10, pages 2200–2204.
- Balage Filho, P., Pardo, T. A. S., and Aluísio, S. (2013). An evaluation of the brazilian portuguese liwc dictionary for sentiment analysis. In *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology*.
- Basiri, M. E. and Kabiri, A. (2017). Translation is not enough: Comparing lexicon-based methods for sentiment analysis in persian. In *2017 International Symposium on Computer Science and Software Engineering Conference (CSSE)*, pages 36–41.
- Basiri, M. E. and Kabiri, A. (2018). Words are important: Improving sentiment analysis in the persian language by lexicon refining. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, 17(4).

- Bestgen, Y. et al. (2008). Building affective lexicons from specific corpora for automatic sentiment analysis. In *LREC*. Citeseer.
- Bian, C., Zhang, Y., Wang, D., Liang, Y., Wu, B., and Lu, W. (2018). An academic emotion database and the baseline evaluation. In *2018 13th International Conference on Computer Science Education (ICCSE)*, pages 1–6.
- Bradley, M., L. P. and Cuthbert, B. (1999). Affective norms for english words (anew). *Gainesville, Fla, NIMH Centre for the Study of Emotion and Attention, University of Florida*.
- Bradley, M. M. and Lang, P. J. (1999). Affective norms for english words (anew): Instruction manual and affective ratings. Technical report, Technical report C-1, the center for research in psychophysiology.
- Bruckschen, M., Muniz, F., Souza, J., Fuchs, J., Infante, K., Muniz, M., Gonçalves, P., Vieira, R., and Aluisio, S. (2008). Anotação lingüística em xml do corpus pln-br. *Série de relatórios do NILC, ICMC-USP*.
- Brum, H. and das Graças Volpe Nunes, M. (2018). Building a Sentiment Corpus of Tweets in Brazilian Portuguese. In chair), N. C. C., Choukri, K., Cieri, C., Declerck, T., Goggi, S., Hasida, K., Isahara, H., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J., Piperidis, S., and Tokunaga, T., editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).
- Cambria, E., Li, Y., Xing, F. Z., Poria, S., and Kwok, K. (2020). Senticnet 6: Ensemble application of symbolic and subsymbolic ai for sentiment analysis. In *Proceedings of the 29th ACM International Conference on Information amp; Knowledge Management, CIKM '20*, page 105–114, New York, NY, USA. Association for Computing Machinery.
- Cambria, E., Olsher, D., and Rajagopal, D. (2014). Senticnet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis. In *Proceedings of the twenty-eighth AAAI conference on artificial intelligence*, pages 1515–1521.
- Carvalho, F., Rodrigues, R., Santos, G., Cruz, P., Ferrari, L., and Guedes, G. (2019). Avaliação da versão em português do liwc lexicon 2015 com análise de sentimentos

- em redes sociais. In *Anais do VIII Brazilian Workshop on Social Network Analysis and Mining*, pages 24–34, Porto Alegre, RS, Brasil. SBC.
- Carvalho, F., Santos, G., and Guedes, G. P. (2018). Affectpt-br: an affective lexicon based on liwc 2015. In *2018 37th International Conference of the Chilean Computer Science Society (SCCC)*, pages 1–5.
- Carvalho, P. and Silva, M. (2015). Sentilex-pt: Principais características e potencialidades. *Oslo Studies in Language*, 7.
- Chang, C., Wu, M., and Hwang, S. (2019). An approach to cross-lingual sentiment lexicon construction. In *2019 IEEE International Congress on Big Data (BigDataCongress)*, pages 129–131.
- Chen, Y. and Skiena, S. (2014). Building sentiment lexicons for all major languages. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*.
- Cheng, L. and Tsai, S. (2019). Deep learning for automated sentiment analysis of social media. In *2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 1001–1004.
- Cirqueira, D., Fontes Pinheiro, M., Jacob, A., Lobato, F., and Santana, (2018). A literature review in preprocessing for sentiment analysis for brazilian portuguese social media. In *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pages 746–749.
- Clore, G. L., Ortony, A., and Foss, M. A. (1987). The psychological foundations of the affective lexicon. *Journal of personality and social psychology*, 53(4):751.
- Cruz, P. P., Rodrigues, R. G., Belloze, K. T., and Guedes, G. P. (2017). Uma revisão sistemática sobre léxicos afetivos para o português do brasil. In *Congresso Internacional de Informatica ´ Educativa, 2017*.
- da Silva Nascimento, R., Carvalho, F., and Guedes, G. P. (2019). Identificando sintomas depressivos: um estudo de caso no youtube. In *Anais do VIII Brazilian Workshop on Social Network Analysis and Mining*, pages 119–130. SBC.

- Daily, S. B., James, M. T., Cherry, D., Porter, J. J., Darnell, S. S., Isaac, J., and Roy, T. (2017). Affective computing: Historical foundations, current applications, and future trends. *Emotions and Affect in Human Factors and Human-Computer Interaction*, page 213–231.
- De Choudhury, M., Gamon, M., Counts, S., and Horvitz, E. (2013). Predicting depression via social media. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 7.
- Dias-Da-Silva, B. C. and Moraes, H. R. d. (2003). A construção de um thesaurus eletrônico para o português do brasil. *ALFA: Revista de Linguística*.
- Dodds, P. S. and Danforth, C. M. (2010). Measuring the happiness of large-scale written expression: Songs, blogs, and presidents. *Journal of happiness studies*, 11(4):441–456.
- Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A., and Danforth, C. M. (2011). Temporal patterns of happiness and information in a global social network: Hedonometrics and twitter. *PloS one*, 6(12):e26752.
- Eskander, R. and Rambow, O. (2015). Slsa: A sentiment lexicon for standard arabic. In *Proceedings of the 2015 conference on empirical methods in natural language processing*, pages 2545–2550.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5):378.
- Freitas, C. (2013). Sobre a construção de um léxico da afetividade para o processamento computacional do português. *Revista Brasileira de Linguística Aplicada*, 13:1031 – 1059.
- Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the american statistical association*, 32(200):675–701.
- Gelbukh, A., Sidorov, G., Han, S.-Y., and Hernández-Rubio, E. (2004). Automatic enrichment of very large dictionary of word combinations on the basis of dependency formalism. In *Mexican International Conference on Artificial Intelligence*, pages 430–437. Springer.

- Gilbert, C. and Hutto, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth International Conference on Weblogs and Social Media (ICWSM-14)*. Available at (20/04/16) <http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf>, volume 81, page 82.
- Go, A., Bhayani, R., and Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N project report, Stanford*, 1(12):2009.
- Golbeck, J., Robles, C., and Turner, K. (2011). Predicting personality with social media. In *CHI'11 extended abstracts on human factors in computing systems*, pages 253–262.
- Gonçalves, P., Araújo, M., Benevenuto, F., and Cha, M. (2013a). Comparing and combining sentiment analysis methods. In *Proceedings of the first ACM conference on Online social networks*, pages 27–38.
- Gonçalves, P., Benevenuto, F., and Cha, M. (2013b). Panas-t: A psychometric scale for measuring sentiments on twitter. *arXiv preprint arXiv:1308.1857*.
- González-Bailón, S. and Paltoglou, G. (2015). Signals of public opinion in online communication: A comparison of methods and data sources. *The ANNALS of the American Academy of Political and Social Science*, 659(1):95–107.
- Gonçalo Oliveira, H. (2018). A survey on portuguese lexical knowledge bases: Contents, comparison and combination. *Information*, 9(2).
- Goularas, D. and Kamis, S. (2019). Evaluation of deep learning techniques in sentiment analysis from twitter data. In *2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML)*, pages 12–17.
- Guellil, I., Adeel, A., Azouaou, F., and Hussain, A. (2018). Sentialg: Automated corpus annotation for algerian sentiment analysis. In Ren, J., Hussain, A., Zheng, J., Liu, C.-L., Luo, B., Zhao, H., and Zhao, X., editors, *Advances in Brain Inspired Cognitive Systems*, pages 557–567, Cham. Springer International Publishing.
- Gupta, S., Singh, R., and Singh, J. (2020). A hybrid approach for enhancing accuracy and detecting sarcasm in sentiment analysis. In *2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON)*, pages 130–134.

- Han, J., Kamber, M., and Pei, J. (2011). Data mining concepts and techniques third edition. *The Morgan Kaufmann Series in Data Management Systems*, 5(4):83–124.
- Hannak, A., Anderson, E., Barrett, L. F., Lehmann, S., Mislove, A., and Riedewald, M. (2012). Tweetin' in the rain: Exploring societal-scale effects of weather on mood. In *ICWSM*.
- Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177.
- Huang, S., Niu, Z., and Shi, C. (2014). Automatic construction of domain-specific sentiment lexicon based on constrained label propagation. *Knowledge-Based Systems*, 56:191–200.
- Imran, A. S., Daudpota, S. M., Kastrati, Z., and Batra, R. (2020). Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on covid-19 related tweets. *IEEE Access*, 8:181074–181090.
- Kim, S.-M. and Hovy, E. (2004). Determining the sentiment of opinions. In *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics*, pages 1367–1373.
- Kristensen, C. H., Gomes, C. F. A. d. A., Justo, A. R., and Vieira, K. (2011). Normas brasileiras para o Affective Norms for English Words. *Trends in Psychiatry and Psychotherapy*, 33:135 – 146.
- Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, pages 159–174.
- Levallois, C. (2013). Umigon: sentiment analysis for tweets based on lexicons and heuristics. In *7th International Workshop on Semantic Evaluation (SemEval 2013)At: Atlanta, Georgia*.
- Liu, B. (2017). *Sentiment analysis: mining opinions, sentiments, and emotions*. Cambridge University Press.
- Liu, B. and Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In *Mining text data*, pages 415–463. Springer.

- Longhi, M. T. (2011). *Mapeamento de aspectos afetivos em um ambiente virtual de aprendizagem. 2011. 273 f.* PhD thesis, Tese (Doutorado em Informática na Educação), UFRGS.
- Lu, Y., Castellanos, M., Dayal, U., and Zhai, C. (2011). Automatic construction of a context-aware sentiment lexicon: an optimization approach. In *Proceedings of the 20th international conference on World wide web*, pages 347–356.
- Machado, A., Longhi, M., Nunes, M. A. S. N., and Pardo, T. (2015). Personalitatem lexicon: Um léxico em português brasileiro para mineração de traços de personalidade em textos. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 26, page 1122.
- Maziero, E. G., Pardo, T. A. S., Di Felippo, A., and Dias-da Silva, B. C. (2008). A base de dados lexical e a interface web do tep 2.0: Thesaurus eletrônico para o português do brasil. In *Companion Proceedings of the XIV Brazilian Symposium on Multimedia and the Web, WebMedia '08*, page 390–392, New York, NY, USA. Association for Computing Machinery.
- Medjden, S., Ahmed, N., and Lataifeh, M. (2020). Adaptive user interface design and analysis using emotion recognition through facial expressions and body posture from an rgb-d sensor. *PLOS ONE*, 15(7):1–37.
- Miller, G. A. (1995). Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.
- Minsky, M. (2007). *The emotion machine: Commonsense thinking, artificial intelligence, and the future of the human mind*. Simon and Schuster.
- Mohamad Beigi, O. and Moattar, M. H. (2021). Automatic construction of domain-specific sentiment lexicon for unsupervised domain adaptation and sentiment classification. *Knowledge-Based Systems*, 213:106423.
- Mohammad, S. (2012). # 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)*, pages 246–255.

- Mohammad, S. and Turney, P. (2010). Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text*, pages 26–34.
- Mohammad, S. M., Kiritchenko, S., and Zhu, X. (2013). Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. *arXiv preprint arXiv:1308.6242*.
- Mohammad, S. M. and Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 29(3):436–465.
- Musto, C., Semeraro, G., and Polignano, M. (2014). A comparison of lexicon-based approaches for sentiment analysis of microblog posts. In *DART@ AI* IA*, pages 59–68. Citeseer.
- Nascimento, R., Nascimento, G., Carvalho, F., and Guedes, G. (2020). Mineração de opiniões com liwc: abordagem prática sobre sistemas judiciais eletrônicos brasileiros. In *Anais do IX Brazilian Workshop on Social Network Analysis and Mining*, pages 132–141, Porto Alegre, RS, Brasil. SBC.
- Nielsen, F. Å. (2011). A new anew: Evaluation of a word list for sentiment analysis in microblogs. *arXiv preprint arXiv:1103.2903*.
- Olson, D. L. and Delen, D. (2008). *Advanced data mining techniques*. Springer Science & Business Media.
- Paltoglou, G. and Thelwall, M. (2012). Twitter, myspace, digg: Unsupervised sentiment analysis in social media. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 3(4):1–19.
- Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, 2(1–2):1–135.
- Pappas, N. and Popescu-Belis, A. (2013). Sentiment analysis of user comments for one-class collaborative filtering over ted talks. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 773–776.

- Park, S. and Kim, Y. (2016). Building thesaurus lexicon using dictionary-based approach for sentiment classification. In *2016 IEEE 14th International Conference on Software Engineering Research, Management and Applications (SERA)*, pages 39–44. IEEE.
- Pasqualotti, P. R. and Vieira, R. (2008). Wordnet affect br: uma base lexical de palavras de emoções para a língua portuguesa. *RENOTE-Revista Novas Tecnologias na Educação*, 6(1).
- Patel, P., Patel, D., and Naik, C. (2021). Sentiment analysis on movie review using deep learning rnn method. In Satapathy, S. C., Zhang, Y.-D., Bhateja, V., and Majhi, R., editors, *Intelligent Data Engineering and Analytics*, pages 155–163, Singapore. Springer Singapore.
- Pennebaker, J. W., Boyd, R. L., Jordan, K., and Blackburn, K. (2015). The development and psychometric properties of liwc2015. In *The University of Texas at Austin*.
- Pennebaker, J. W., Francis, M. E., and Booth, R. J. (2001). Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, 71(2001):2001.
- Pereira, D. A. (2021). A survey of sentiment analysis in the portuguese language. *Artificial Intelligence Review*, 54(2):1087–1115.
- Picard, R. W. (2000). *Affective computing*. MIT Press.
- PNAD, I. (2019). Pesquisa nacional por amostra de domicílios. *Rio de Janeiro: IBGE*.
- Poornima, A. and Priya, K. S. (2020). A comparative sentiment analysis of sentence embedding using machine learning techniques. In *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, pages 493–496.
- Pozzi, F., Fersini, E., Messina, E., and Liu, B. (2017). Challenges of sentiment analysis in social networks. *Sentiment Analysis in Social Networks*, page 1–11.
- Pustejovsky, J. and Batiukova, O. (2019). *The lexicon*. Cambridge University Press.
- Pustejovsky, J. and Stubbs, A. (2012). *Natural Language Annotation for Machine Learning: A guide to corpus-building for applications*. "O'Reilly Media, Inc."
- Quan, W., Chen, Z., Gao, J., and Hu, X. T. (2018). Comparative study of cnn and lstm based attention neural networks for aspect-level opinion mining. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 2141–2150.

- Rabeya, T., Chakraborty, N. R., Ferdous, S., Dash, M., and Al Marouf, A. (2019). Sentiment analysis of bangla song review- a lexicon based backtracking approach. In *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, pages 1–7.
- Ramos, B. and Freitas, C. (2019). sentimento de quê?": uma lista de sentimentos para a análise de sentimentos. *STIL*, pages 15–18.
- Ribeiro, F. N., Araújo, M., Gonçalves, P., Gonçalves, M. A., and Benevenuto, F. (2016). Sentibench-a benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Science*, 5(1):1–29.
- Rodrigues, R. G., Rodrigues, K. T., Gomes, R. R., Ferrari, L., Ogasawara, E., and Guedes, G. P. (2020). Brapt: A new metric for translation evaluation based on psycholinguistic perspectives. *IEEE Latin America Transactions*, 18(07):1264–1271.
- Rodríguez-Fdez, I., Canosa, A., Mucientes, M., and Bugarín, A. (2015). STAC: a web platform for the comparison of algorithms using statistical tests. In *Proceedings of the 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*.
- Rosenthal, S., Farra, N., and Nakov, P. (2017). Semeval-2017 task 4: Sentiment analysis in twitter. In *Proceedings of the 11th international workshop on semantic evaluation (SemEval-2017)*, pages 502–518.
- Roy, C., Dey, R., Chaudhuri, C., and Das, D. (2019). Emotion predictor using social media text and graphology. In *2019 IEEE 9th International Conference on Advanced Computing (IACC)*, pages 96–102.
- Rumelli, M., Akkuş, D., Kart, , and Isik, Z. (2019). Sentiment analysis in turkish text with machine learning algorithms. In *2019 Innovations in Intelligent Systems and Applications Conference (ASYU)*, pages 1–5.
- S., H. and Ramathmika, R. (2019). Sentiment analysis of yelp reviews by machine learning. In *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, pages 700–704.
- Sabra, K. S., Zantout, R. N., Abed, M. A. E., and Hamandi, L. (2017). Sentiment analysis: Arabic sentiment lexicons. In *2017 Sensors Networks Smart and Emerging Technologies (SENSET)*, pages 1–4.

- Sajadi, S. H., Fazli, M., and Habibi, J. (2018). The affective evolution of social norms in social networks. *IEEE Transactions on Computational Social Systems*, 5(3):727–735.
- Sanguansat, P. (2016). Paragraph2vec-based sentiment analysis on social media for business in thailand. In *2016 8th International Conference on Knowledge and Smart Technology (KST)*, pages 175–178.
- Santos, A., Becker, K., and Moreira, V. (2014). Um estudo de caso de mineração de emoções em textos multilíngues. In *Anais do III Brazilian Workshop on Social Network Analysis and Mining*, pages 140–151, Porto Alegre, RS, Brasil. SBC.
- Shahare, F. F. (2017). Sentiment analysis for the news data based on the social media. In *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*, pages 1365–1370.
- Silva, E. P., Malheiros, Y., Nunes, R. T. A., Antunes, I. L., and Rêgo, T. G. (2019). Um conjunto de dados extraído do twitter para análise de sentimentos na língua portuguesa. In *7^o Symposium in Information and Human Language Technology, 2019*, Salvador, Bahia. STIL.
- Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A. Y., and Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 1631–1642.
- Souza, E., Alves, T., Teles, I., Oliveira, A. L., and Gusmão, C. (2016). Topie: an open-source opinion mining pipeline to analyze consumers' sentiment in brazilian portuguese. In *International Conference on Computational Processing of the Portuguese Language*, pages 95–105. Springer.
- Souza, M., Vieira, R., Buseti, D., Chishman, R., Alves, I. M., and Unisinos, F. D. L. (2011). Construction of a portuguese opinion lexicon from multiple resources. In *In 8th Brazilian Symposium in Information and Human Language Technology - STIL, Mato Grosso*.
- Strapparava, C., Valitutti, A., et al. (2004). Wordnet affect: an affective extension of wordnet. In *Lrec*, volume 4, page 40. Citeseer.
- Taboada, M., Brooke, J., Tofiloski, M., Voll, K., and Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2):267–307.

- Tanantong, T., Sanglerdsinlapachai, N., and Donkhampai, U. (2020). Sentiment classification on thai social media using a domain-specific trained lexicon. In *2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, pages 580–583.
- Thavareesan, S. and Mahesan, S. (2020). Sentiment lexicon expansion using word2vec and fasttext for sentiment prediction in tamil texts. In *2020 Moratuwa Engineering Research Conference (MERCon)*, pages 272–276.
- Thelwall, M. (2017). The heart and soul of the web? sentiment strength detection in the social web with sentistrength. In *Cyberemotions*, pages 119–134. Springer.
- Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., and Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12):2544–2558.
- Vilares, D., Peng, H., Satapathy, R., and Cambria, E. (2018). Babelsentinet: A commonsense reasoning framework for multilingual sentiment analysis. In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1292–1298.
- Wang, H., Can, D., Kazemzadeh, A., Bar, F., and Narayanan, S. (2012). A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In *Proceedings of the ACL 2012 system demonstrations*, pages 115–120.
- Wang, H. and Castanon, J. A. (2015). Sentiment expression via emoticons on social media. In *2015 IEEE International Conference on Big Data (Big Data)*, pages 2404–2408.
- Wang, Q., Sun, L., and Chen, Z. (2019). Sentiment analysis of reviews based on deep learning model. In *2019 IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS)*, pages 258–261.
- Wang, Z., Chong, C. S., Lan, L., Yang, Y., Beng Ho, S., and Tong, J. C. (2016). Fine-grained sentiment analysis of social media with emotion sensing. In *2016 Future Technologies Conference (FTC)*, pages 1361–1364.
- Wiebe, J., Wilson, T., and Cardie, C. (2005a). Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39(2-3):165–210.

- Wiebe, J., Wilson, T., and Cardie, C. (2005b). Annotating expressions of opinions and emotions in language. *Language resources and evaluation*, 39(2):165–210.
- Wilson, T., Wiebe, J., and Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of human language technology conference and conference on empirical methods in natural language processing*, pages 347–354.
- Yamauchi, T., Hayashi, Y., and Nakano, Y. I. (2013). Searching emotional scenes in tv programs based on twitter emotion analysis. In Ozok, A. A. and Zaphiris, P., editors, *Online Communities and Social Computing*, pages 432–441, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Young, L. and Soroka, S. (2012). Affective news: The automated coding of sentiment in political texts. *Political Communication*, 29(2):205–231.
- Zhang, X. and Zheng, X. (2016). Comparison of text sentiment analysis based on machine learning. In *2016 15th International Symposium on Parallel and Distributed Computing (ISPDC)*, pages 230–233.
- Zhang, Z. and Singh, M. P. (2014). ReNew: A semi-supervised framework for generating domain-specific lexicons and sentiment analysis. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 542–551, Baltimore, Maryland. Association for Computational Linguistics.
- Zhao, S., Wang, S., Soleymani, M., Joshi, D., and Ji, Q. (2019). Affective computing for large-scale heterogeneous multimedia data: A survey. *ACM Trans. Multimedia Comput. Commun. Appl.*, 15(3s).

**A- Apêndice: Resultados de acurácia, revocação, precisão e
F1**

Tabela 30 – Resultados da Análise de Sentimento - MQD.

Posição	Léxico	Precisão	Revocação	Acurácia	F1
1	AffectPT-br	0,611	0,566	0,566	0,570
2	LIWC2015-PT-br	0,598	0,553	0,553	0,557
3	SentiStrength	0,551	0,541	0,541	0,544
4	LIWC2007-PT-br	0,547	0,515	0,515	0,503
5	Polyglot	0,476	0,469	0,469	0,472
6	MPQA-PT	0,466	0,466	0,466	0,465
7	Sentilex-PT	0,539	0,447	0,447	0,445
8	OpLexicon	0,447	0,428	0,428	0,426
9	ANEW-Br	0,443	0,417	0,417	0,415
10	WordNetAffectBR(adapt)	0,589	0,431	0,431	0,409
11	ReLi-Lex	0,528	0,410	0,410	0,386
12	WordNetAffectBR	0,621	0,396	0,396	0,348
13	BabelSenticnet	0,369	0,382	0,382	0,284

Tabela 31 – Resultados da Análise de Sentimento - SADT.

Posição	Léxico	Precisão	Revocação	Acurácia	F1
1	AffectPT-br	0,452	0,451	0,451	0,451
2	Sentilex-PT	0,443	0,445	0,445	0,443
3	SentiStrength	0,447	0,445	0,445	0,440
4	LIWC2015-PT-br	0,441	0,438	0,438	0,439
5	Polyglot	0,419	0,415	0,415	0,403
6	ANEW-Br	0,402	0,401	0,401	0,400
7	ReLi-Lex	0,439	0,422	0,422	0,399
8	LIWC2007-PT-br	0,423	0,405	0,405	0,398
9	OpLexicon	0,404	0,400	0,400	0,396
10	WordNetAffectBR(adapt)	0,438	0,419	0,419	0,388
11	MPQA-PT	0,384	0,384	0,384	0,360
12	WordNetAffectBR	0,486	0,428	0,428	0,358
13	BabelSenticnet	0,300	0,327	0,327	0,229

Tabela 32 – Resultados da Análise de Sentimento - TOPIE.

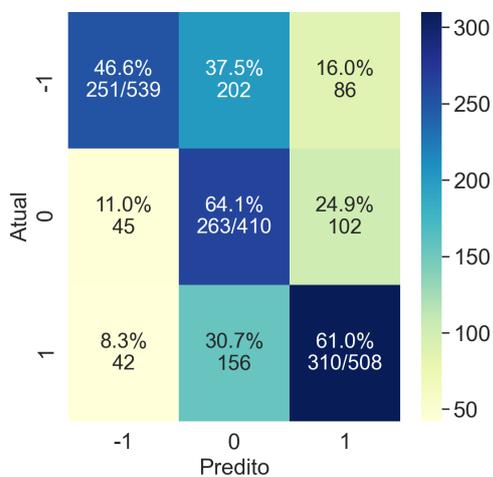
Posição	Léxico	Precisão	Revocação	Acurácia	F1
1	SentiStrength	0,545	0,524	0,524	0,525
2	LIWC2015-PT-br	0,562	0,508	0,508	0,510
3	AffectPT-br	0,555	0,502	0,502	0,503
4	Polyglot	0,489	0,488	0,488	0,487
5	MPQA-PT	0,472	0,476	0,476	0,472
6	Sentilex-PT	0,496	0,470	0,470	0,449
7	ReLi-Lex	0,546	0,448	0,448	0,424
8	LIWC2007-PT-br	0,454	0,418	0,418	0,399
9	OpLexicon	0,421	0,408	0,408	0,396
10	ANEW-Br	0,426	0,398	0,398	0,383
11	WordNetAffectBR(adapt)	0,381	0,363	0,363	0,290
12	BabelSenticnet	0,567	0,378	0,378	0,264
13	WordNetAffectBR	0,419	0,357	0,357	0,243

Tabela 33 – Resultados da Análise de Sentimento - TweetSentBR.

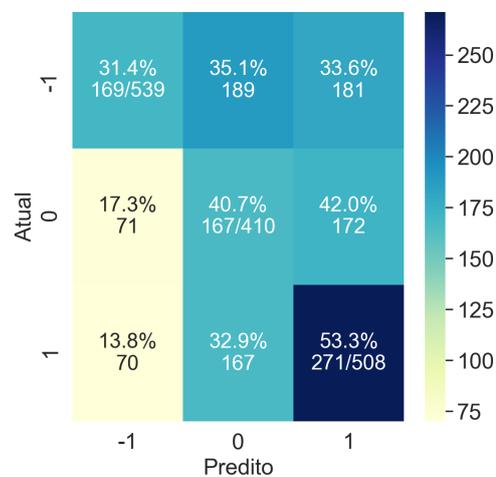
Posição	Léxico	Precisão	Revocação	Acurácia	F1
1	LIWC2015-PT-br	0,574	0,485	0,485	0,490
2	LIWC2007-PT-br	0,512	0,478	0,478	0,478
3	AffectPT-br	0,544	0,456	0,456	0,462
4	OpLexicon	0,504	0,443	0,443	0,448
5	Polyglot	0,479	0,431	0,431	0,440
6	SentiStrength	0,506	0,423	0,423	0,423
7	MPQA-PT	0,465	0,414	0,414	0,418
8	ReLi-Lex	0,584	0,414	0,414	0,387
9	ANEW-Br	0,429	0,381	0,381	0,376
10	Sentilex-PT	0,553	0,382	0,382	0,363
11	BabelSenticnet	0,393	0,457	0,457	0,340
12	WordNetAffectBR(adapt)	0,535	0,324	0,324	0,253
13	WordNetAffectBR	0,534	0,298	0,298	0,201

B- Apêndice: Matriz de Confusão

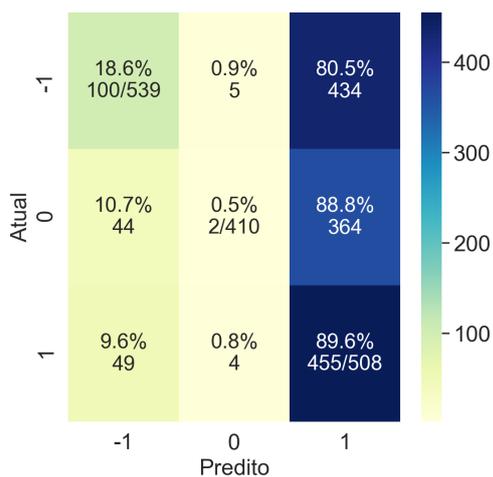
Figura 19 – Conjunto de dados - MQD. Fonte: Elaborada pelo autor.



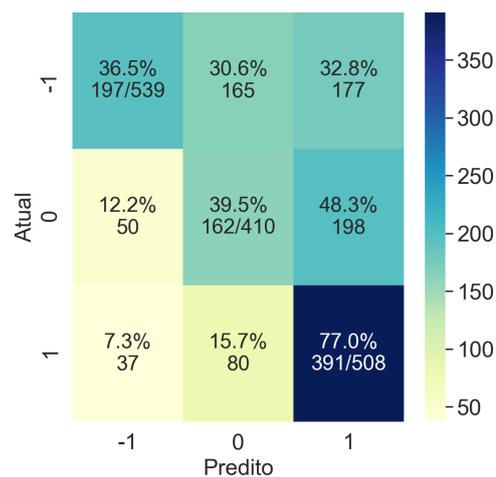
(a) AffectPT-br



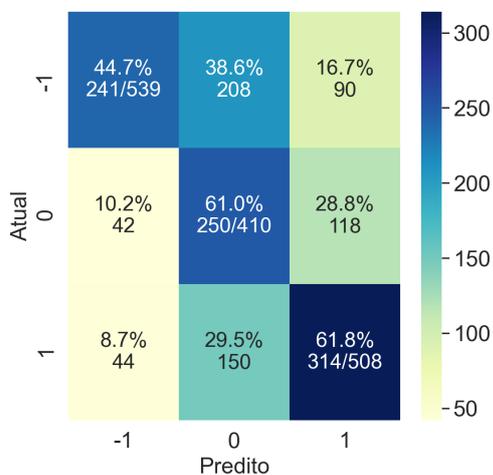
(b) ANEW-Br



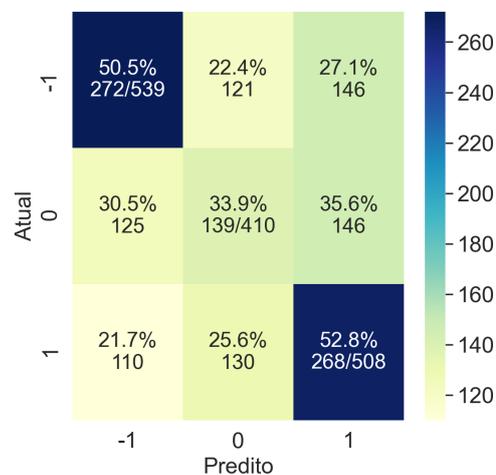
(c) BabelSenticnet



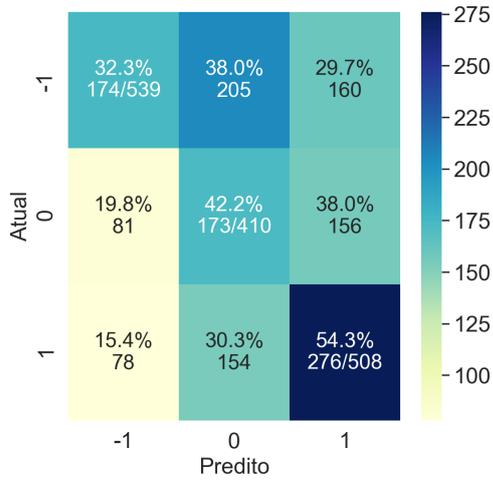
(d) LIWC2007-PT-br



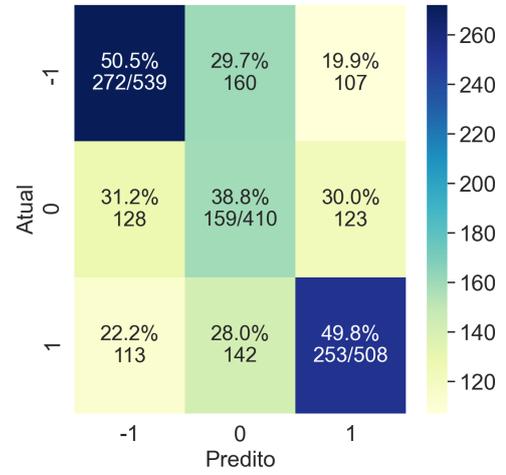
(e) LIWC2015-PT-br



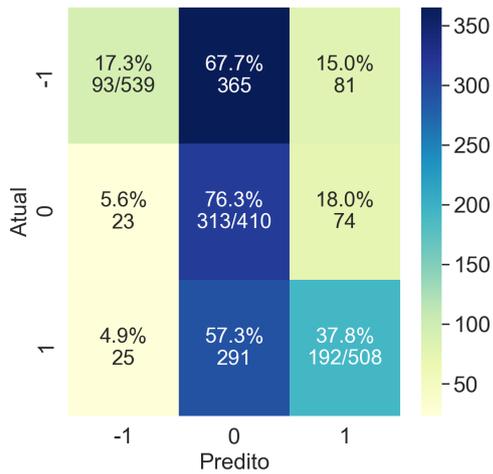
(f) MPQA-PT



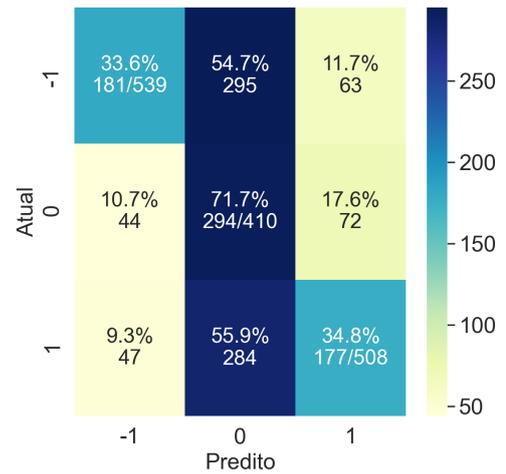
(g) OpLexicon



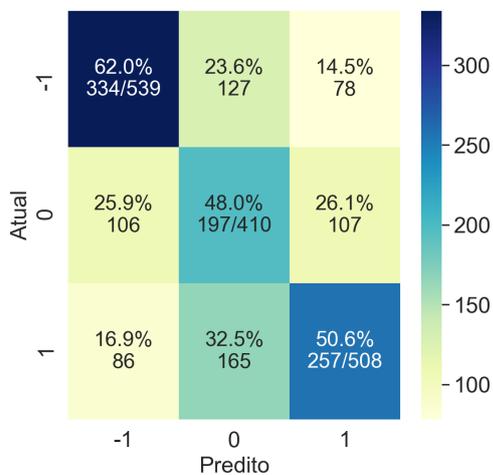
(h) Polyglot



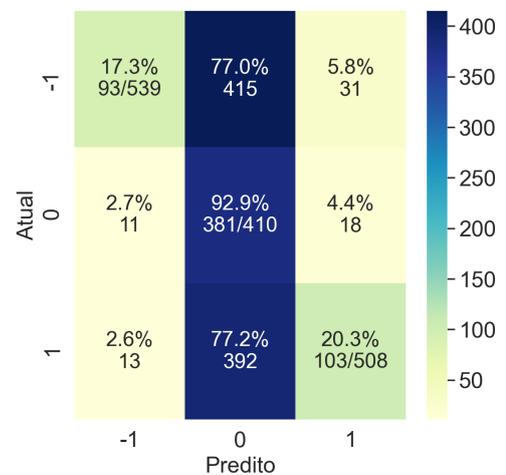
(i) ReLi-Lex



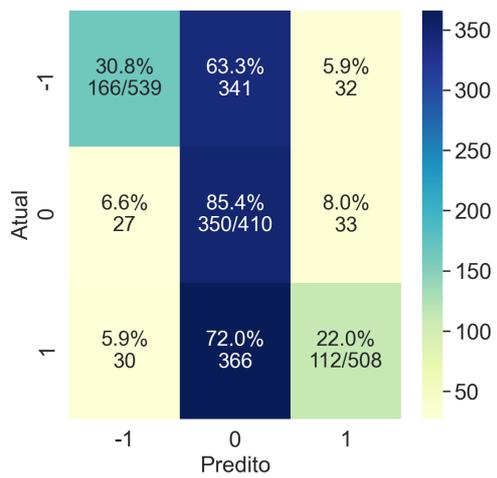
(j) Sentilex-PT



(k) SentiStrength

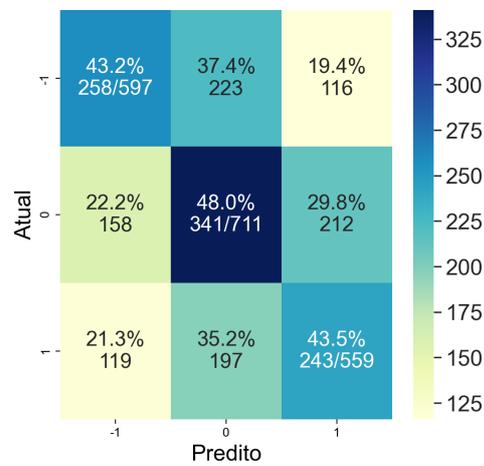


(l) WordNetAffectBR

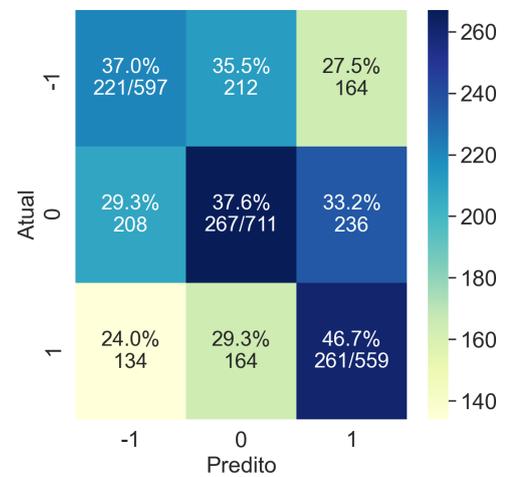


(m) WordNetAffectBR(adapt)

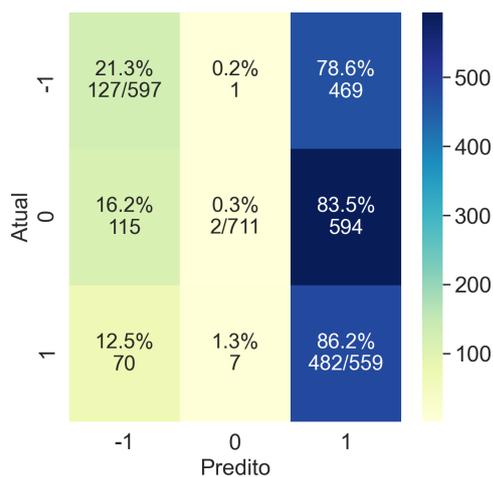
Figura 20 – Conjunto de dados - SADT. Fonte: Elaborada pelo autor.



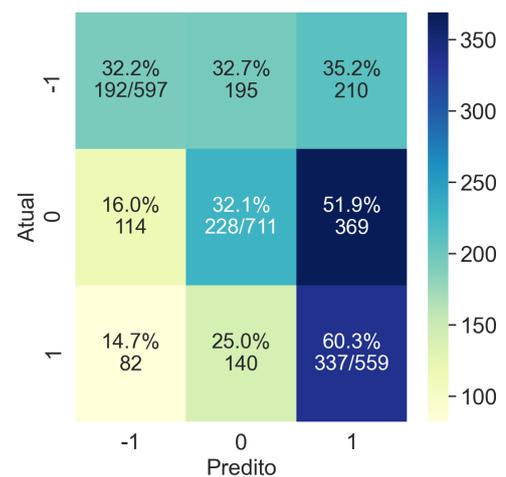
(a) AffectPT-br



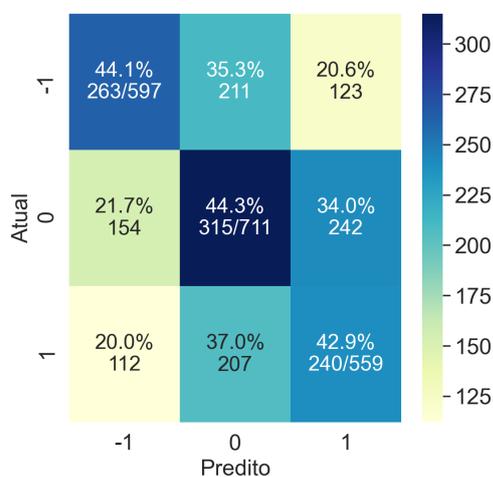
(b) ANEW-Br



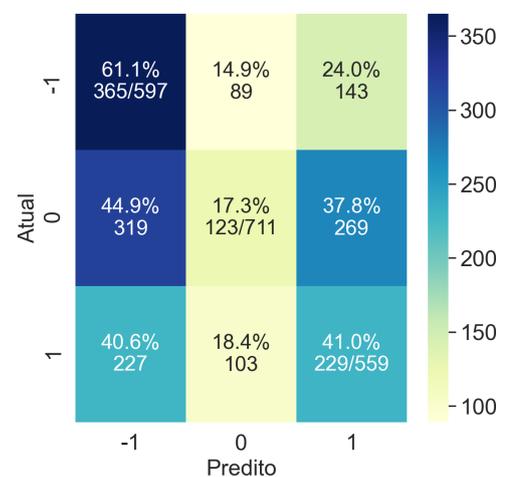
(c) BabelSenticnet



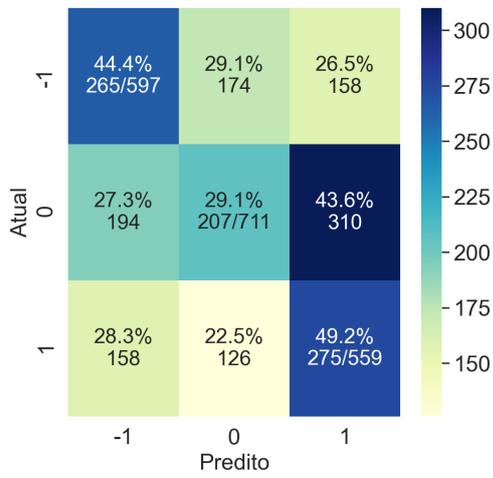
(d) LIWC2007-PT-br



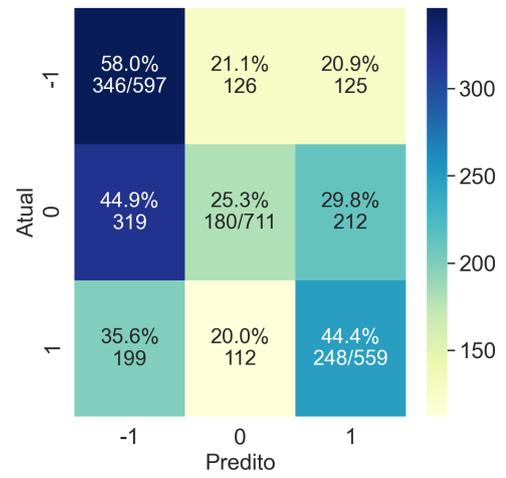
(e) LIWC2015-PT-br



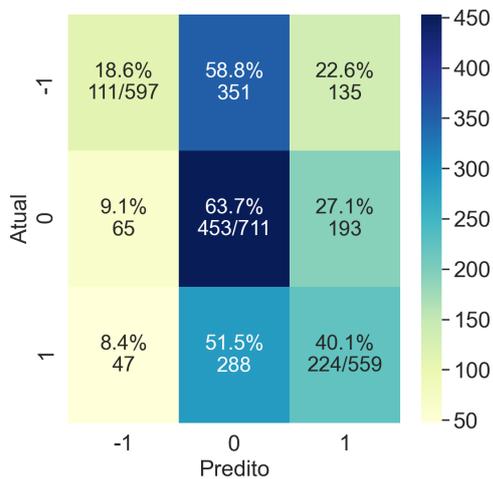
(f) MPQA-PT



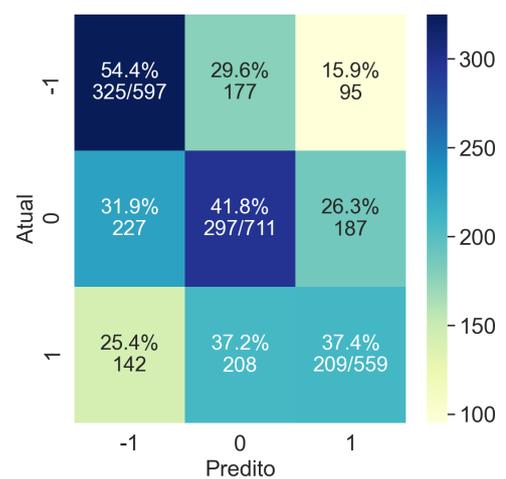
(g) OpLexicon



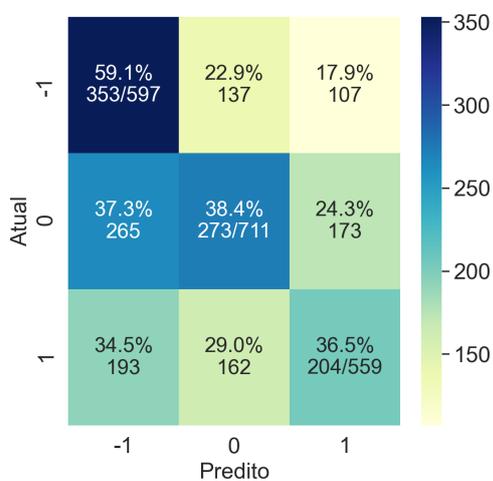
(h) Polyglot



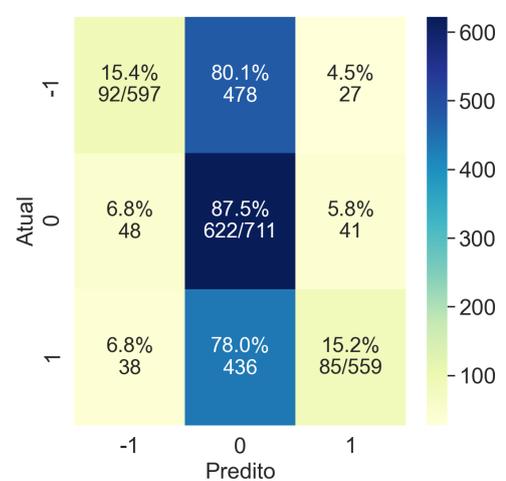
(i) ReLi-Lex



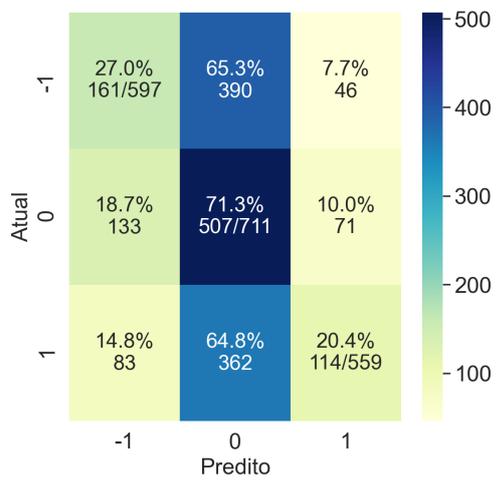
(j) Sentilex-PT



(k) SentiStrength

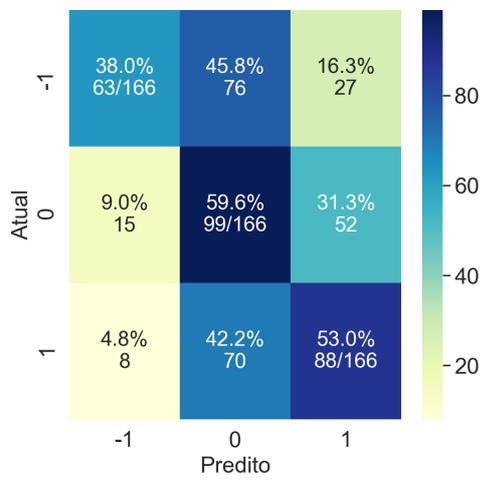


(l) WordNetAffectBR

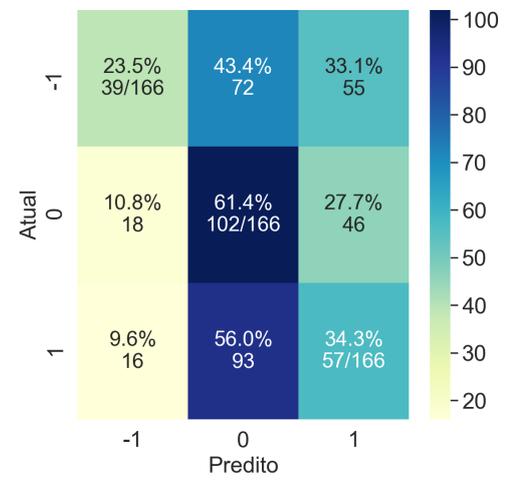


(m) WordNetAffectBR(adapt)

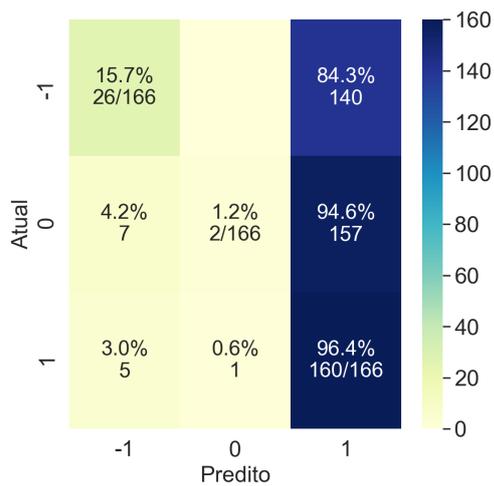
Figura 21 – Conjunto de dados - TOPIE. Fonte: Elaborada pelo autor.



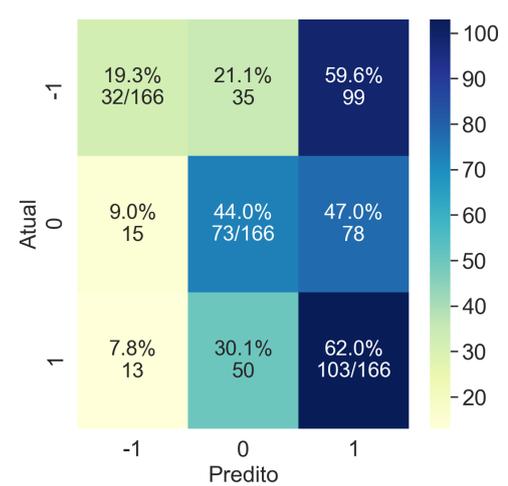
(a) AffectPT-br



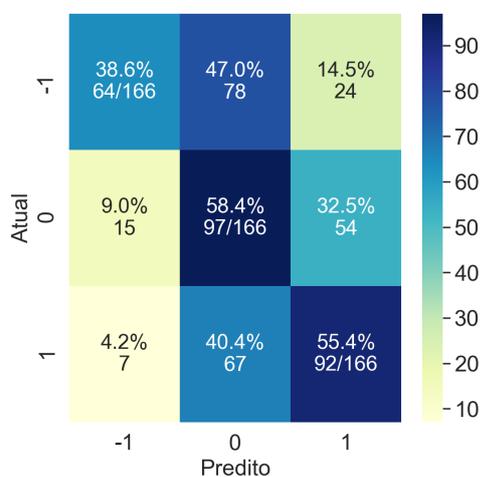
(b) ANEW-Br



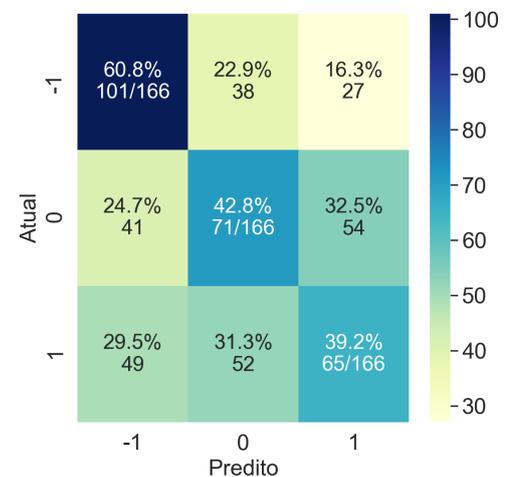
(c) BabelSenticnet



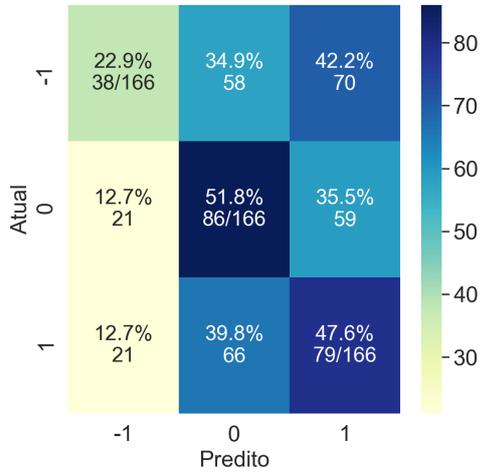
(d) LIWC2007-PT-br



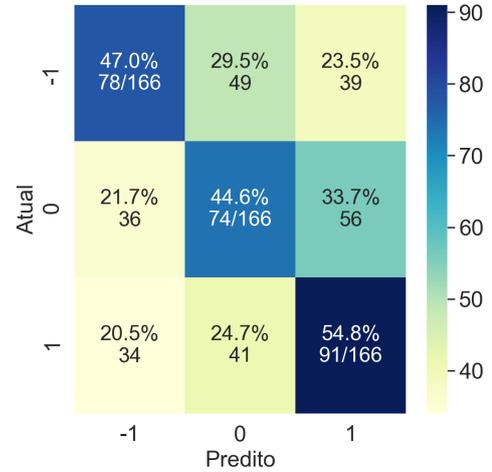
(e) LIWC2015-PT-br



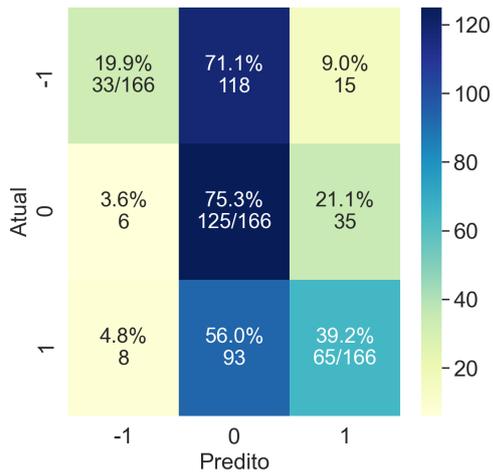
(f) MPQA-PT



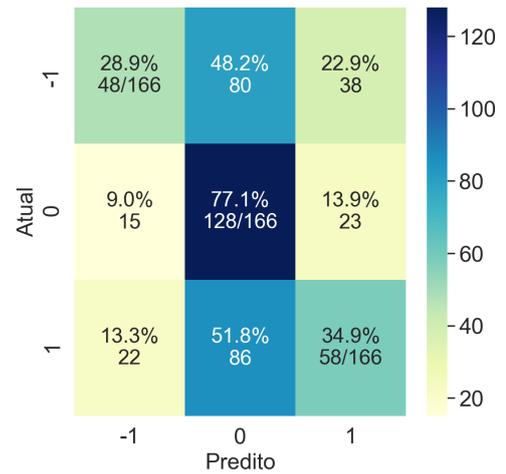
(g) OpLexicon



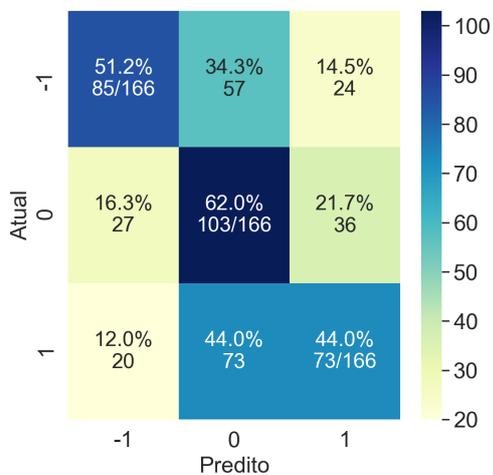
(h) Polyglot



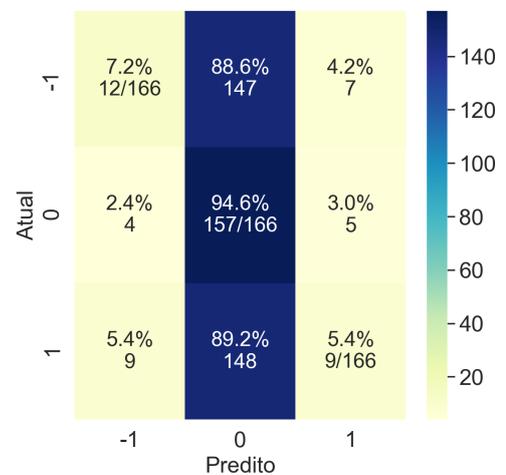
(i) ReLi-Lex



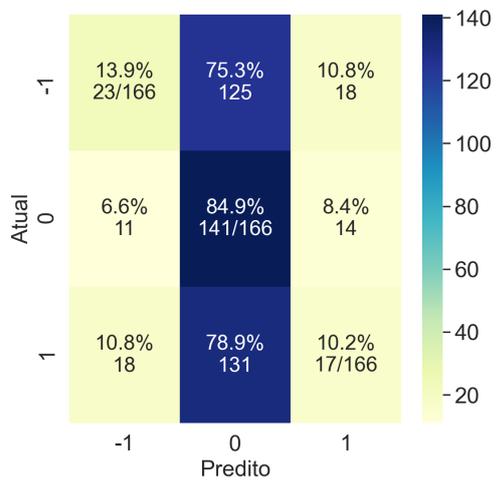
(j) Sentilex-PT



(k) SentiStrength

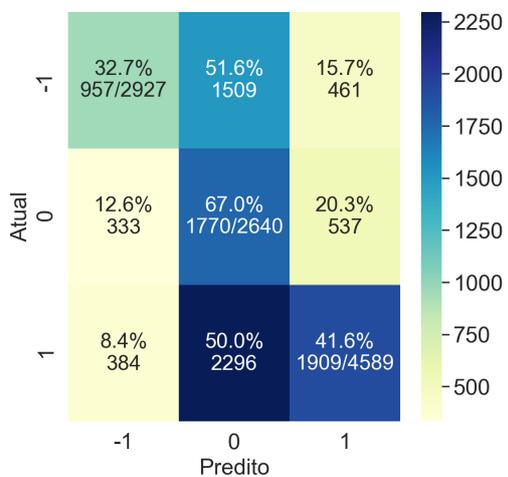


(l) WordNetAffectBR

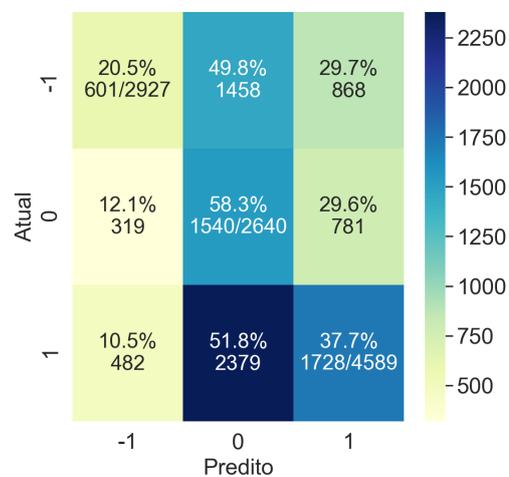


(m) WordNetAffectBR(adapt)

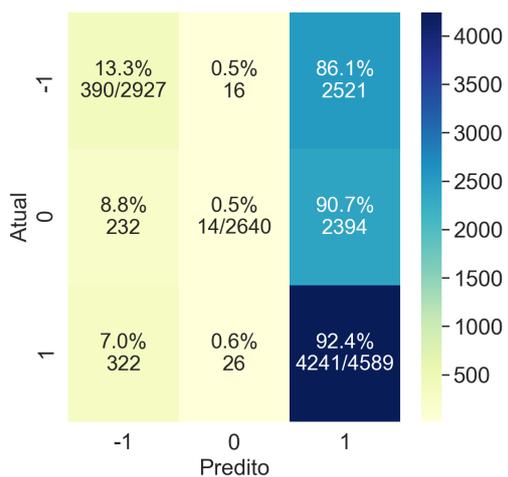
Figura 22 – Conjunto de dados - TweetSentBR. Fonte: Elaborada pelo autor



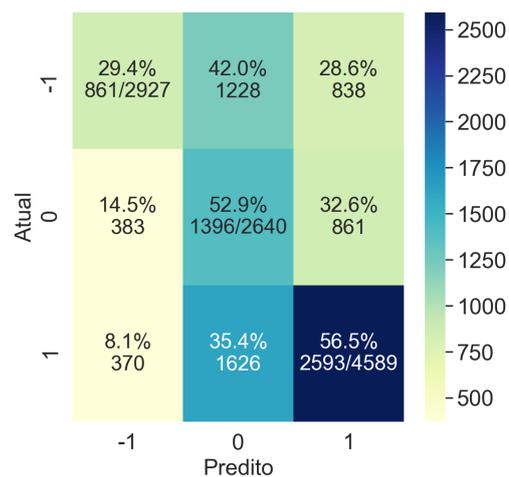
(a) AffectPT-br



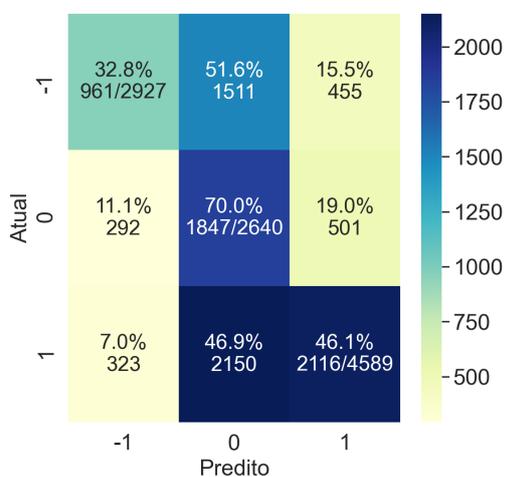
(b) ANEW-Br



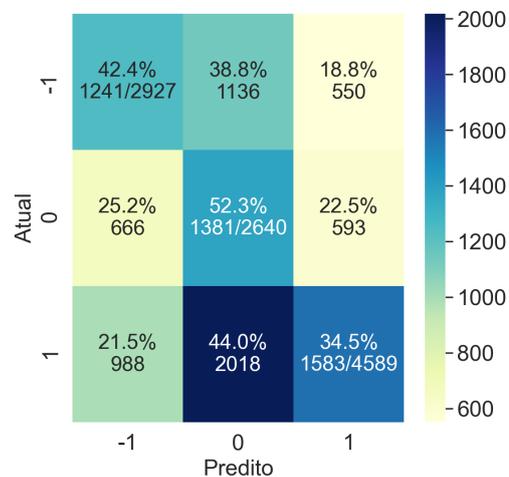
(c) BabelSenticnet



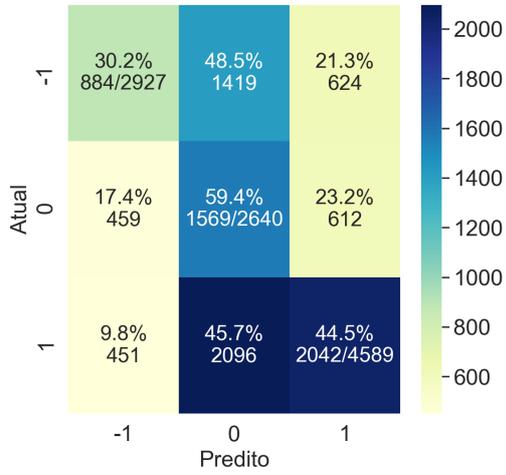
(d) LIWC2007-PT-br



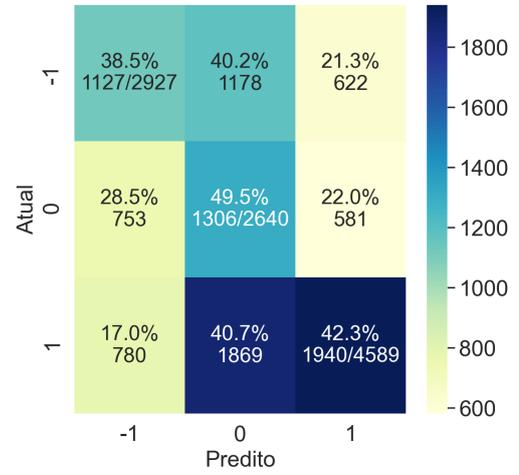
(e) LIWC2015-PT-br



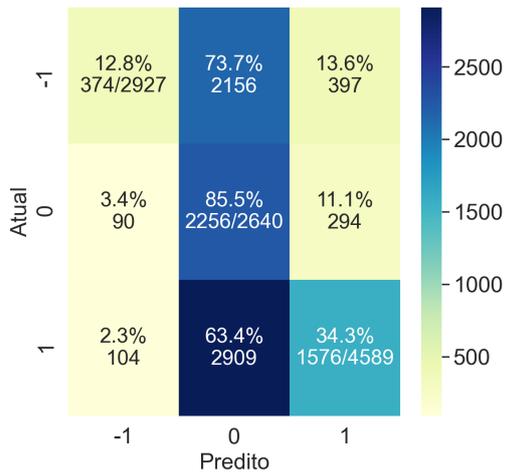
(f) MPQA-PT



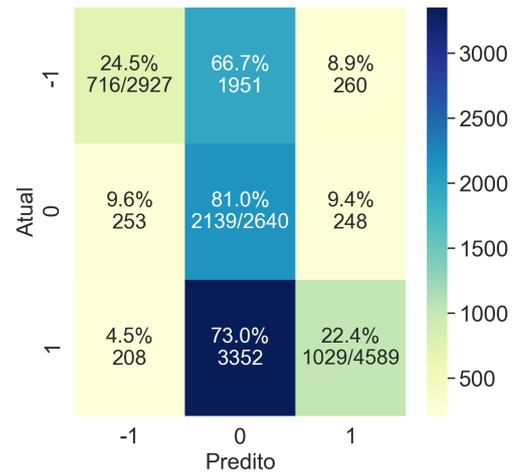
(g) OpLexicon



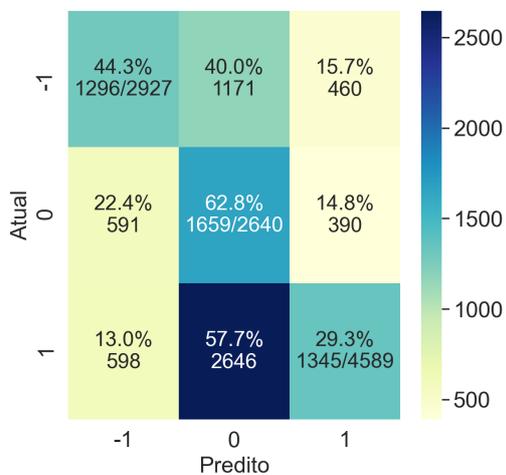
(h) Polyglot



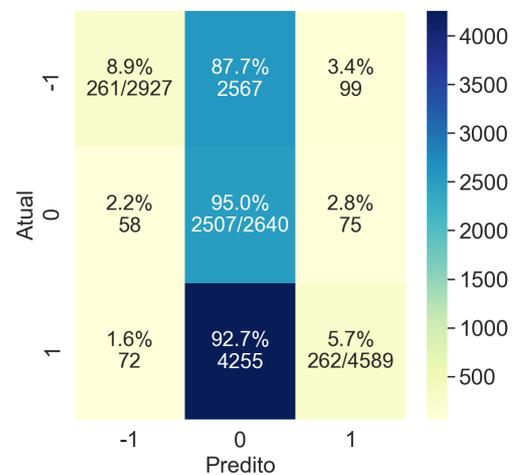
(i) ReLi-Lex



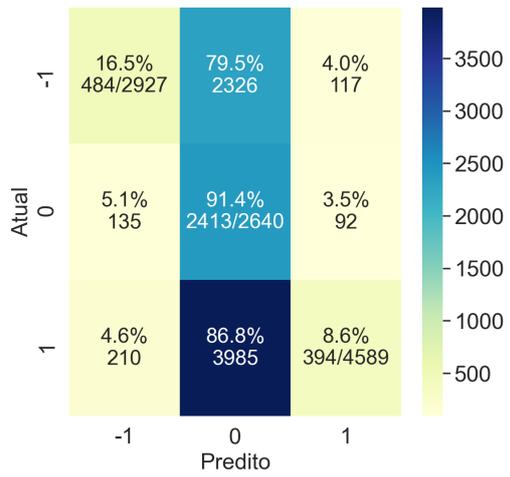
(j) Sentilex-PT



(k) SentiStrength



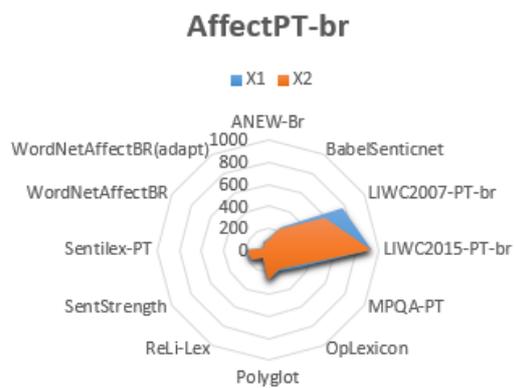
(l) WordNetAffectBR



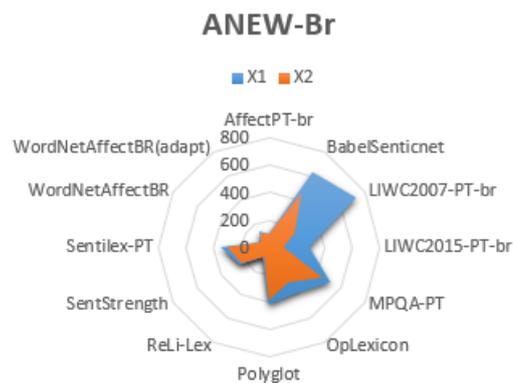
(m) WordNetAffectBR(adapt)

**C- Apêndice: Concordância nas palavras entre os Léxicos
Afetivos**

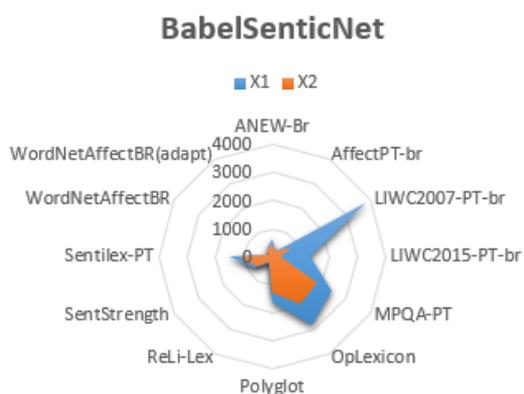
Figura 23 – Gráfico de concordância entre palavras. X_1 representa palavras equivalentes entre os léxicos. X_2 representa palavras de valência afetiva equivalentes. Fonte: Elaborada pelo autor.



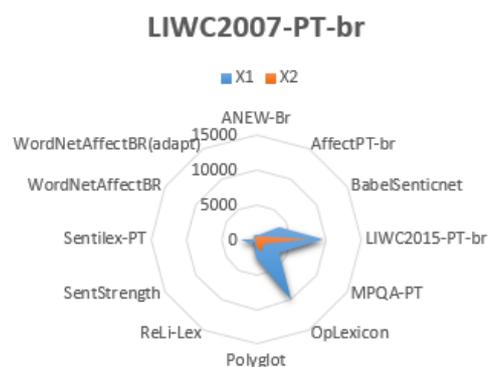
(a) AffectPT-br



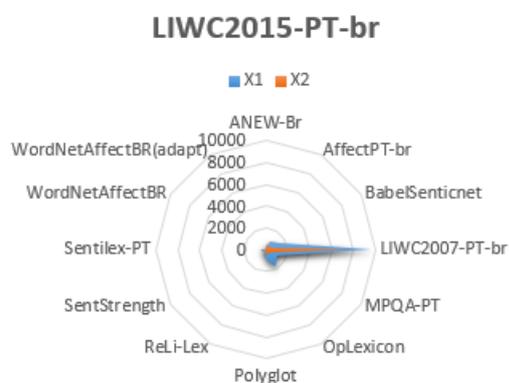
(b) ANEW-Br



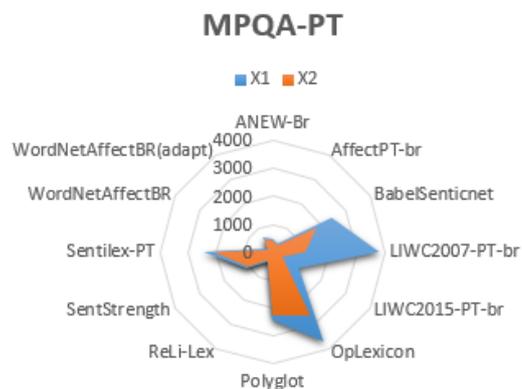
(c) BabelSenticnet



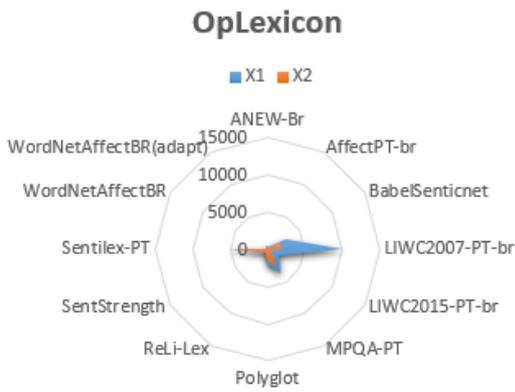
(d) LIWC2007-PT-br



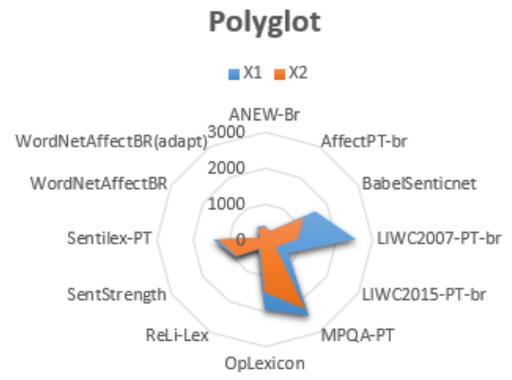
(e) LIWC2015-PT-br



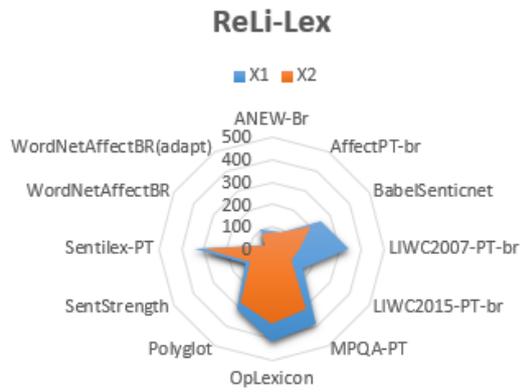
(f) MPQA-PT



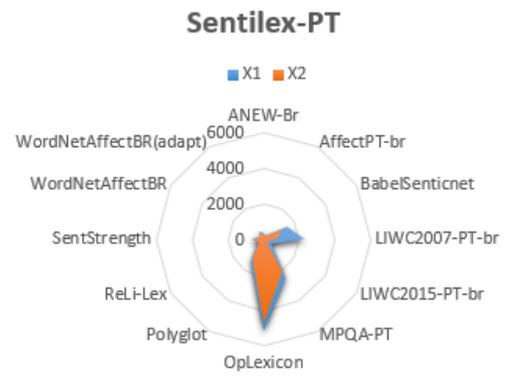
(g) OpLexicon



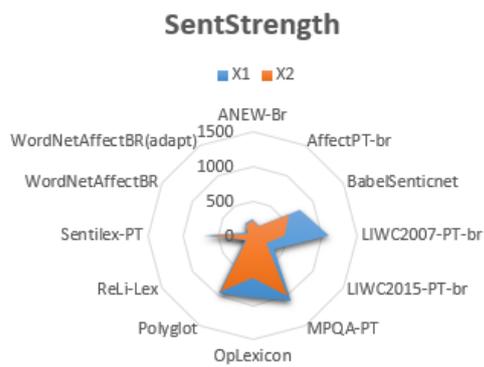
(h) Polyglot



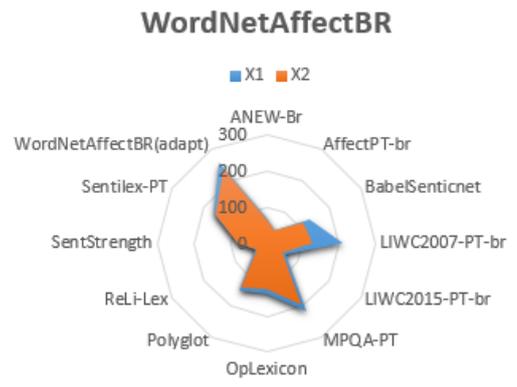
(i) ReLi-Lex



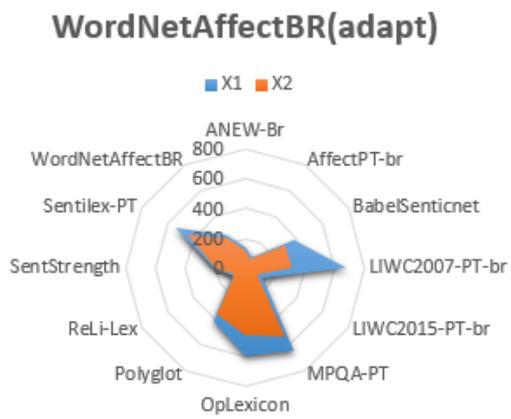
(j) Sentilex-PT



(k) SentiStrength



(l) WordNetAffectBR



(m) WordNetAffectBR(adapt)