

Uma Revisão Sistemática sobre Léxicos Afetivos para o Português do Brasil

Pedro P. Cruz, Rafael G. Rodrigues
CEFET/RJ
Av. Maracanã, 229 - Maracanã,
Rio de Janeiro, RJ CEP: 20271-110
pedroparreiraacruz@gmail.com,
rafael.rodrigues@cefet-rj.br

Kele T. Belloze, Gustavo P. Guedes
CEFET/RJ
Av. Maracanã, 229 - Maracanã,
Rio de Janeiro, RJ CEP: 20271-110
kele.belloze@cefet-rj.br,
gustavo.guedes@cefet-rj.br

ABSTRACT

The literature in the informatics and education area contains many studies that use affective lexicons with the purpose of providing improvements in the teaching/learning process. However, few studies are concentrated on the development of affective lexicons for Brazilian Portuguese. In this scenario, the present work offers a Systematic Review of Literature on affective lexicons for Brazilian Portuguese. In this panorama, the objective is to contribute to the development of new research papers and affective lexicons in Brazilian Portuguese.

RESUMO

A literatura na área de informática e educação contém diversos estudos que utilizam léxicos afetivos com a finalidade de proporcionar melhorias no processo de ensino/aprendizagem. No entanto, poucos estudos são concentrados no desenvolvimento desses léxicos para o português do Brasil. Nesse cenário, o presente artigo oferece uma Revisão Sistemática da Literatura sobre léxicos afetivos para o português do Brasil. Nesse panorama, o objetivo é que o presente estudo contribua para o desenvolvimento de novos artigos de pesquisa e léxicos afetivos para o português do Brasil.

Categories and Subject Descriptors

J.4. [Computer Applications]: Social and behavioral sciences.

Termos Gerais

Documentation, Human Factors, Languages.

Keywords

Affective lexicons; Systematic review; Affective Computing.

1. INTRODUÇÃO

A Computação Afetiva (CA) é uma área de pesquisa voltada ao reconhecimento, interpretação e expressão de emoções por computadores [31]. Rosalind Picard empregou o termo Computação Afetiva, do inglês *Affective Computing*, pela primeira vez em 1995, destacando que computadores genuinamente inteligentes dependem da habilidade de reconhecer, interpretar e expressar emoções para, dessa maneira, interagirem naturalmente com humanos [31]. Seguindo a mesma linha, há outros estudos que enfatizam a importância de se adicionar características emocionais à interação humano-computador, dentre os quais é possível realçar os estudos descritos em [32, 34].

O estudo das emoções também pode ser efetuado com base em textos em linguagem natural, o que tem se tornado crucial na área da interação humano-computador [38]. Inúmeras aplicações

podem emergir deste cenário, desde a inferência de fatores da personalidade [17] até o desenvolvimento de robôs emocionais [3]. Também é possível destacar estudos capazes de identificar diversas patologias, dentre elas, a depressão [10].

Os estudos realizados nos trabalhos supracitados envolveram a utilização de léxicos afetivos. Os léxicos afetivos são úteis para o estudo de emoções em textos e são compostos de listas de palavras anotadas com suas avaliações emocionais. Essas avaliações podem ser representadas por um conjunto de pontuações numéricas ou rótulos de emoções, como, por exemplo, raiva e medo [28]. Assim, os léxicos afetivos são utilizados para a identificação de características emocionais em palavras escritas por indivíduos em seus textos.

Também é possível destacar diversos estudos na área da educação que apontam para a necessidade de melhorias nos ambientes de aprendizagem [2, 11]. Alguns desses estudos utilizam léxicos afetivos para a detecção do estado afetivo dos estudantes durante o processo de aprendizagem [13]. Nesse aspecto, o propósito consiste no aperfeiçoamento do ambiente de aprendizagem com o ajuste do material, de maneira que este possa ser compatível com o estado emocional dos estudantes. Desse modo, o conteúdo apresentado nesses ambientes pode, inclusive, ser adaptado em razão da frustração, por exemplo. Vale ressaltar que a experiência de estados afetivos como frustração e tédio possuem relação negativa com o aprendizado [8].

A literatura possui estudos comprovando que ambientes emocionalmente inteligentes podem adaptar o conteúdo do aprendizado de acordo com o estado emocional/afetivo do aluno [13]. É sabido que esse estado pode comprometer positivamente ou negativamente o aprendizado, visto que, dentre outros aspectos, a atenção e a memória, por exemplo, são diretamente associados às emoções [18].

Com base no cenário apresentado até agora, esse estudo tem o objetivo de contribuir com uma revisão sistemática sobre os léxicos afetivos existentes para o Português do Brasil (PB). Embora poucos trabalhos na literatura produzam léxicos afetivos em PB, não existe uma revisão sistemática sobre o assunto. Inicialmente, como cenário motivador, são apresentados estudos na área de educação que utilizam léxicos afetivos em outras linguagens. Posteriormente, recomenda-se a revisão sistemática da literatura que embasa esse trabalho e, por fim, a descrição dos léxicos afetivos em PB. Dessa maneira, com a descrição dos estudos da área e a revisão sistemática proposta, objetiva-se motivar e contribuir com o desenvolvimento de novos estudos capazes de detectar emoção em textos em PB.

O restante deste artigo encontra-se organizado da seguinte maneira: A Seção 2 aborda a utilização de léxicos afetivos na

educação. A Seção 3 detalha a condução desta revisão sistemática da literatura. A Seção 4 mostra os léxicos afetivos encontrados. Por fim, a Seção 5 apresenta as conclusões acerca desse estudo.

2. UTILIZAÇÃO DE LÉXICOS AFETIVOS NA EDUCAÇÃO

A computação afetiva está se tornando uma promessa na área da educação, onde alguns resultados interessantes estão sendo alcançados, especialmente no que diz respeito a melhor adaptação dos estudantes em relação a ambientes de aprendizagem e aprimoramento dos mesmos. Dentre os estudos que utilizam léxicos afetivos para aprimorar o desempenho e a motivação dos estudantes em ambientes inteligentes de aprendizagem, destaca-se o estudo proposto em [33]. Os autores descrevem uma série de sistemas e estudos que exploram a importância de se extrair emoções de textos em ambientes de aprendizagem e apresentam um caso de uso em que os textos de um estudante de primeiro semestre são analisados por um sistema capaz de detectar quatro emoções em textos: alegria, raiva, tristeza e medo. Em seguida, esse sistema recomenda as atividades mais adequadas para o estudante, levando em conta seus estados emocionais.

De acordo com o referido estudo, aspectos afetivos e emocionais do estudante desempenham um papel fundamental nos seus processos de aprendizagem, especialmente no que diz respeito ao engajamento para atuar em ambientes virtuais de aprendizagem. Ao detectar e gerenciar esses aspectos, seria possível contribuir para melhorar a motivação do aluno. Nesse sentido, os pesquisadores realizaram experiências sobre a avaliação das emoções e estados afetivos dos estudantes tanto para melhorar sua interação com os ambientes de aprendizagem quanto para avaliar as emoções e estados afetivos refletidos nos textos escritos por eles.

Dentre outros trabalhos, os autores relatam um estudo que envolve o Chat-SEE, um avaliador emocional semântico para textos [6]. O sistema, baseado em um léxico afetivo, é capaz de fornecer a evolução emocional do estudante ao longo de um período de tempo e tem sido usado para avaliar estados afetivos obtidos nas conversas de bate-papo. Os resultados devem ser analisados e utilizados para enriquecer e ajustar o perfil do aluno, inclusive em relação à melhor interação dos mesmos com os referidos ambientes.

Os resultados foram considerados satisfatórios e as experiências obtidas junto ao Chat-SEE ofereceram resultados interessantes que apontam para um maior aprofundamento das possibilidades de se fazer uso da avaliação de aspectos afetivos para fins de aprendizagem em ambientes virtuais.

3. REVISÃO SISTEMÁTICA DA LITERATURA

A Revisão Sistemática da Literatura (RSL) empregada neste trabalho objetivou, de maneira bem específica, sintetizar os estudos que desenvolvem léxicos afetivos em Português do Brasil (PB). Desta maneira, o presente trabalho possui foco em Artigos, Dissertações e Teses concentrados nesses léxicos. Para isso foi definida a metodologia descrita na Subseção 3.1.

3.1 Metodologia

A metodologia adotada para a realização desse trabalho consistiu em seguir as diretrizes propostas por [20] para esse tipo de estudo.

A RSL está dividida em Etapas (E), algumas delas cíclicas, conforme descrição que se segue:

[E1]: Definir as questões de pesquisa, identificar as palavras-chave e definir *strings de busca* (SB), que, segundo o estudo de [20], são importantes ferramentas para a execução de uma RSL. Em seguida, espera-se obter uma lista de estudos relevantes junto aos mecanismos de pesquisa de trabalhos acadêmicos; [E2]: Definir os critérios de inclusão (CI); [E3]: Definir os critérios de exclusão (CE); [E4]: Realizar uma análise minuciosa a partir da lista de estudos relevantes obtida. Essa análise consiste em ler o resumo, a introdução e a conclusão de cada estudo, aplicar os CI e, conseqüentemente, obter uma nova lista de estudos relevantes e consonantes com o objetivo deste estudo; [E5]: Ler, na íntegra, os trabalhos selecionados na E4, aplicar os CE e obter uma nova lista de estudos relevantes; [E6]: A partir dos estudos selecionados na E5, aplicar as técnicas de *Forward Snowballing* e *Backward Snowballing*, com o intuito de obter novas listas de estudos relevantes. As técnicas de *Snowballing* consistem em, a partir de determinado estudo, encontrar os estudos que o citaram (*Forward Snowballing*) e os estudos que o autor (ou autores) citou (*Backward Snowballing*) [36]; [E7]: Caso sejam encontrados novos estudos a partir da etapa E6, repetir E4, E5 e E6 até que não sejam encontrados novos estudos relevantes.

3.2 Questões de Pesquisa

Esta RSL tem o objetivo muito específico e bem definido. Desta forma, foram definidas as seguintes questões de pesquisa: Quais léxicos afetivos foram criados originalmente para o PB? Quais léxicos afetivos surgiram a partir de adaptações de léxicos já conhecidos em outras línguas? Quais léxicos afetivos foram resultantes da fusão de outros já existentes?

3.3 Identificação das Palavras-chave e Definição das *Strings de Busca*

A questão dessa RSL é bem simples e objetiva: encontrar léxicos afetivos em PB. O termo “léxico computacional” também é bastante comum. Nesse sentido, não foi difícil definir as palavras-chave e seus sinônimos, como destaca-se a seguir: léxico, afetivo, computacional, português, dicionário e língua portuguesa. Assim, a *string de busca* principal foi definida da seguinte forma, com a utilização dos operadores lógicos AND e OR: (“português” OR “língua portuguesa”) AND (“léxico computacional” OR “léxico afetivo” OR “dicionário afetivo”)

Outras *strings de busca* menos restritivas, inclusive em língua inglesa, foram empregadas posteriormente. No entanto, não trouxeram nenhum estudo diferente dos encontrados na aplicação das etapas citadas na subseção 3.1. As pesquisas foram realizadas na ferramenta *Google Scholar* (<https://scholar.google.com.br>) e, posteriormente, em bases de dados especializadas em artigos e documentos indexados (*i.e.*, ACM, IEE- Explorer, Scielo, Science Direct, Scopus, Springer e Web of Science).

3.4 Critérios de Inclusão e Critérios de Exclusão

Para ajudar a nortear a RSL foram definidos Critérios de Inclusão (CI) e Critérios de Exclusão (CE), que permitiram reduzir e filtrar os resultados de uma busca para obter um número razoável de estudos a serem analisados. Os critérios foram definidos da seguinte forma:

Crítérios de Inclusão (CI): [CI 1]: Artigos que resultem na construção de léxicos afetivos em PB; [CI 2]: Dissertações e teses que abordem estudos sobre léxicos afetivos em PB; [CI 3]: Estudos comparativos entre léxicos afetivos em PB e em outras línguas.

Crítérios de Exclusão (CE): [CE 1]: Estudos em outros idiomas que não sejam inglês e PB; [CE 2]: Estudos redundantes tanto em relação à autoria quanto ao léxico abordado.

3.5 Estudos Encontrados Preliminarmente

Após a definição da *string de busca* principal, os resultados compreenderam 152 estudos. A etapa posterior consistiu na leitura das seções resumo, introdução e conclusão de cada trabalho. Por fim, foram aplicados os CI e CE que resultaram em uma parcial de apenas três estudos (EST). A Tabela 1 apresenta tais estudos e o ano de publicação.

Tabela 1: Estudos encontrados preliminarmente

Título do estudo	Ano
EST01: WordnetAffectBR: uma base lexical de palavras de emoções para a língua portuguesa [29].	2008
EST02: Inferência de personalidade a partir de textos de rede social utilizando um léxico afetivo em português brasileiro [24].	2016
EST03: Inferência de personalidade a partir de textos em português brasileiro utilizando léxicos [27].	2016

3.6 Estudos Encontrados a partir da técnica de Snowballing

Para cada estudo mostrado na Tabela 1 foram aplicadas as técnicas de *Forward Snowballing* (FS) e *Backward Snowballing* (BS), o que resultou em pouco mais de 300 outros estudos. Para cada um desses novos estudos, foram aplicadas as etapas 4, 5, 6 e 7 da metodologia, gerando um conjunto menor de estudos. Novamente, ao aplicar a técnica de FS a estes últimos, não foram encontrados outros estudos relevantes. No entanto, a aplicação da técnica de BS resultou em mais cinco estudos, todos obtidos a partir do EST02. Os novos resultados são mostrados na Tabela 2.

As técnicas de Snowballing foram também aplicadas aos estudos da Tabela 2, porém, sem obtenção de nenhum novo estudo. Desta maneira, passamos a definir os léxicos afetivos encontrados e seus respectivos estudos.

Tabela 2: Estudos encontrados a partir de Backward Snowballing

Título do Estudo	Ano
EST04: Sobre a construção de um léxico da afetividade para o processamento computacional do português [15].	2013
EST05: Brazilian norms for the affective norms for English words [21].	2011
EST06: An evaluation of the Brazilian Portuguese LIWC dictionary for sentiment analysis [4].	2013
EST07: Construction of a portuguese opinion lexicon from multiple resources [35].	2011
EST08: Mapeamento de aspectos afetivos em um ambiente virtual de aprendizagem [22].	2011

3.7 Definição dos Léxicos Afetivos

Após a Revisão Sistemática da Literatura, foram encontrados sete léxicos afetivos e seus respectivos estudos, a saber: ANEW-Br (EST05), LIWC (EST06), Opinion Lexicon (EST07), Personalitatem Lexicon (EST02), ReLi-Lex (EST04), Wordnet AffectBR (EST01) e WordnetAffectBR_adapt (EST08). Além dos léxicos afetivos supracitados, cabe ressaltar que, a partir da aplicação de BS ao estudo EST03, também foi encontrado o léxico Sentilex-PT resultado do trabalho de [7]. Contudo, ao aplicar os CI e CE, o mesmo foi excluído por se tratar de um léxico em português de Portugal e não do Brasil.

A maioria dos léxicos surgiu entre os anos de 2011 e 2016. Esse dado vai ao encontro do crescimento de cursos a distância e ambientes virtuais de aprendizagem no Brasil, cujos dados do Censo da Educação de 2016 indicam que o número de matrículas a distância vem crescendo ano a ano, atingindo, em 2016, uma participação de 18,6% (quase 1,5 milhão) do total de matrículas da educação superior [9]. Esses números sugerem que novos estudos sobre a utilização de léxicos afetivos podem ser aplicados aos ambientes virtuais de aprendizagem de modo a entender o comportamento dos alunos e, conseqüentemente, tornar esses ambientes cada vez mais personalizados.

4 LÉXICOS AFETIVOS

4.1 ANEW-Br

O léxico ANEW-Br [21] foi construído a partir de uma tradução do *Affective Norms for English Words* (ANEW) para o PB. No estudo de [5], que deu origem ao ANEW, foi construído um glossário a partir de palavras com características emotivas. Nesse léxico, as palavras foram avaliadas em três dimensões: valência, indicando o quão agradável ou desagradável é um estímulo; alerta, indicando quão relaxado ou estimulado um estímulo nos deixa e; dominância, indicando o grau de controle exercido pelo estímulo.

Para avaliar essas diferentes dimensões foi utilizada a escala *Self-Assessment Manikin* (SAM), elaborada pelos mesmos criadores do ANEW em 1980. A representação gráfica da escala SAM é realizada com a utilização de bonecos que apresentam variações nas dimensões citadas anteriormente. A representação de valência, por exemplo, varia entre uma figura de um boneco com expressão agradável (sorridente) a um boneco com expressão desagradável (descontente). O ANEW compreende um total de 1.034 palavras.

A tradução do ANEW para o ANEW-Br envolveu um linguista, pesquisadores com fluência nas duas línguas e juizes independentes no processo de tradução/avaliação das 1.034 palavras (existentes no ANEW) para o PB. Após a tradução, algumas palavras foram removidas por não possuírem equivalência semântica. No entanto, novas palavras sobre emoções foram adicionadas, o que fez com que o ANEW-Br totalizasse 1.046 palavras. Vale ressaltar que, no caso do ANEW-Br, 755 universitários utilizaram as escalas de valência e alerta do SAM para avaliar as 1.046 palavras.

4.2 LIWC

O *Linguistic Inquiry Word Count* (LIWC) [30] é uma ferramenta que possui léxicos com palavras relacionadas a uma ou mais categorias que refletem processos linguísticos, psicológicos e sociais (e.g., *prep* (preposições), *posemo* (emoções positivas),

negemo (emoções negativas), *anger* (raiva), *health* (saúde)). Em seu trabalho inicial, os autores desenvolveram um léxico para a língua inglesa contendo 2.300 palavras. Além do léxico desenvolvido em 2001, foram produzidos mais dois para a língua inglesa: um deles em 2007, contendo 4.500 palavras e; o outro em 2015, com um total de 6.400 palavras.

A construção do léxico em PB foi desenvolvida por Balage Filho e colaboradores [4]. Para construir esse léxico, foi efetuada uma tradução das palavras existentes na versão do léxico do LIWC de 2007. Essa tradução foi realizada por quatro tradutores entre os anos de 2011 e 2012. Dessa maneira, a versão do léxico em PB foi constituída por 127.149 palavras distribuídas em 64 categorias que refletem aspectos linguísticos, psicológicos e sociais. No LIWC, cada palavra pode ser relacionada a uma ou mais categorias. É importante ressaltar que o número de palavras é bastante superior à versão da língua inglesa. Isso ocorre devido ao léxico em PB possuir palavras com flexão de gênero, número e grau para os adjetivos (*e.g.*, bonita, bonito, bonitas, bonitinha), verbos conjugados (*e.g.*, corri, corro, correm), dentre outros.

4.3 Opinion Lexicon

O Opinion Lexicon [35] é um léxico originalmente desenvolvido para o PB. Possui valores de valência para 7.077 palavras e expressões, indicando quão agradável ou desagradável é uma palavra ou expressão.

O processo de criação desse léxico consistiu na aplicação de três diferentes métodos: i) criar um léxico constituído por uma orientação semântica, entre adjetivos e verbos extraídos de 346 resenhas de filmes e 970 textos jornalísticos sobre vários temas (ambos escritos em português do Brasil). No total foram aproximadamente um milhão de palavras distribuídas em 1317 documentos; ii) como recurso léxico, um dicionário de sinônimos e antônimos para o português do Brasil conhecido como TeP [12] contendo 44.077 palavras e anotações de sinonímia e antonímia. Esse método permitiu calcular o menor caminho entre os sinônimos e antônimos das palavras; iii) baseando-se no corpus do estudo de [25], foi realizada a tradução automática das palavras utilizando uma ferramenta online de tradução conhecida como *Google Translator* (<https://translate.google.com.br>). As palavras que o sistema não pôde traduzir foram descartadas em sua revisão manual.

A aplicação dos três métodos resultaram em três diferentes léxicos para o PB. Em seguida, os autores produziram uma interseção entre esses léxicos, removendo palavras e expressões com valência neutra. O resultado final desse processo originou a criação das 7.077 palavras e expressões já citadas. Os autores destacam os bons resultados alcançados com a avaliação experimental proposta para o Opinion Lexicon.

4.4 Personalitatem Lexicon

O Personalitatem Lexicon foi mencionado pela primeira vez em [23], cuja versão apresentou estudos iniciais para construção de um léxico capaz de minerar traços de personalidade em textos em PB. Posteriormente, os autores desenvolveram a ideia inicial com base no léxico do LIWC em PB [24]. Dessa maneira, a versão final do Personalitatem Lexicon consistiu na utilização das palavras afetivas existentes no LIWC (*i.e.*, “*affect*”, “*posemo*”, “*negemo*”, “*anx*”, “*anger*”, “*sad*”, “*feel*”). Ao final, chegou-se a um total de 1.866 palavras.

O objetivo principal do Personalitatem Lexicon foi a criação de um léxico afetivo com base em posts de usuários da rede social Facebook e questionários para detectar os cinco grandes traços da personalidade humana (*Big Five Personality Traits* ou *Big Five*) [19]. O *Big Five* identifica cinco fatores como centrais para a personalidade: Abertura à Experiência, Conscienciosidade, Extroversão, Agradabilidade e Neuroticismo [19]. Para identificar esses fatores nos usuários do Facebook, foram utilizados os questionários IPIP-NEO 120 [14] e TIPI [14]. Nesse panorama, foi proposto um algoritmo para associar pesos às 1.866 palavras com base nos questionários respondidos e nos textos dos usuários.

Os resultados foram avaliados com estudantes e professores de diferentes instituições de ensino. Esses indivíduos também preencheram os questionários mencionados voluntariamente, além de permitirem que seus textos fossem utilizados. Os autores destacam a importância da construção de um léxico baseado em fatores da personalidade para o PB e mencionam bons resultados alcançados.

4.5 ReLi-Lex

O ReLi-Lex [15] é composto de elementos afetivos associados a polaridades (oposição de significado) e foi criado com dois objetivos específicos: contribuir para a caracterização do vocabulário expressivo da língua portuguesa, especialmente em contextos informais, e fornecer subsídios para aplicação do processamento de linguagem natural, especialmente para análise de opinião. Esse léxico deriva do corpus “Reli”[16], composto por resenhas de livros publicadas na internet e foi anotado com informações de opinião expressa de maneira formal e informal (*e.g.*, linguagem utilizada na internet). Ao todo são 1.600 resenhas de 13 livros, totalizando cerca de 260.000 palavras e 12.000 frases.

Com relação à inclusão de entradas no léxico, as palavras e expressões precisaram de expressão de opinião e polaridade bem definida. No tocante a informações relativas à expressão de opinião, a inclusão foi manual, dada a necessidade de identificar se a informação realmente refletia uma opinião, além de considerar sua polaridade. A inclusão de informações de classes de palavras foi realizada de maneira automática a partir de um analisador morfossintático denominado “Palavras” [1]. As referidas classes são verbos, substantivos, adjetivos e expressões.

A polaridade dos segmentos nem sempre condiz com a polaridade do contexto (*i.e.*, a polaridade geral da frase). Como exemplo, o estudo cita a frase “ainda que em alguns momentos a narrativa derrape, o livro é ótimo”. Em casos como esse, a intervenção humana é insuperável no sentido de confirmar os elementos que regularmente expressam opinião e afetividade, bem como verificar se a polaridade expressa era geral ou específica do domínio. Nessa etapa foi importante considerar também os inversores de polaridade (*e.g.*, não, impossível). A palavra *não* é um inversor natural de polaridade (*e.g.*, não gostei). Na frase “impossível abandonar o livro pela metade” é possível ver que a palavra *impossível* inverte a polaridade negativa do verbo abandonar.

Conforme citado anteriormente, as classes de palavras são representadas por verbos, substantivos, adjetivos e expressões, cada qual com listas de entradas com polaridade positiva e negativa (*i.e.*, verbos[+] e verbos[-]). Dessa forma, o ReLi-Lex possui palavras e expressões distribuídas em quatro classes e duas polaridades.

4.6 Wordnet AffectBR

O Wordnet AffectBR [29] é um léxico que foi adaptado para o PB e sua criação se deu a partir de outros léxicos existentes para a língua inglesa: o Wordnet [26], Wordnet Affect [37] e Affect [37]. O Affect é um léxico que compreende 1.903 palavras contendo informações emocionais anotadas manualmente.

O Wordnet e o Wordnet Affect são regidos por palavras agrupadas em conjuntos de sinônimos (*synsets*). Cada *synset* representa o agrupamento de palavras consolidado pelo seu significado semântico. Como exemplo, o *synset* "car.n.01" simboliza os seguintes sinônimos: *car*, *auto*, *automobile*, *machine* e *motorcar*. A última versão do Wordnet é a 3.0, que possui mais de 150.000 palavras e 110.000 *synsets*. As informações afetivas do Wordnet Affect estão relacionadas com o princípio de que as emoções possuem funções de orientação cognitiva, comunicação de significados, de avaliação cognitiva e são fundamentadas em valores de valência. Para produzir o núcleo do Wordnet Affect, os autores extraíram e incorporaram as informações afetivas do Affect. Em seguida, estenderam esse núcleo com informações provenientes do Wordnet. Dessa maneira, o Wordnet Affect compreende 4.787 palavras e 2.874 *synsets*.

A construção do Wordnet AffectBR foi efetuada com base em traduções realizadas por tradutores com ampla experiência e com vivência em países de língua inglesa. Os autores descrevem que, no processo de tradução, diversas duplicidades foram encontradas. Após a remoção dessas duplicidades, o Wordnet AffectBR totalizou 289 palavras.

4.7 Wordnet AffectBR_adapt

O Wordnet AffectBR_adapt [22] foi gerado a partir de uma expansão do Wordnet AffectBR. Possui mais de 3.000 entradas, contendo *emoticons*, interjeições, advérbios (necessários para reger a intensidade de cada termo afetivo), *stopwords* e abreviações. *Stopwords* são palavras de baixo valor semântico, entretanto, aparecem com grande frequência, podendo ser pronomes, artigos, preposições, verbos e advérbios. Além dessa composição, o que pode ser captado como diferencial em relação a esse léxico é a possibilidade do mesmo ser utilizado por alunos em qualquer ambiente virtual de aprendizagem (AVA), concomitantemente, proporcionando dinamicidade e flexibilidade em face das características de cada aluno. Para cada AVA devem ser feitas as devidas adaptações para recuperação dos textos necessários [22].

5 CONCLUSÃO

Os léxicos afetivos são ferramentas utilizadas para a identificação de características emocionais em textos escritos em linguagem natural. Essas ferramentas são compostas de listas de palavras anotadas com suas avaliações emocionais. Com isso, podem ser utilizados com diferentes objetivos, dentre os quais é possível destacar: inferência de fatores da personalidade; desenvolvimento de robôs emocionais; detecção de patologias e adaptação de conteúdo em ambientes inteligentes de aprendizagem, como é o caso do, já citado, léxico Wordnet AffectBR_adapt, por exemplo.

Diversos trabalhos utilizam léxicos afetivos em ambientes inteligentes de aprendizagem objetivando a melhoria no processo de aprendizagem dos estudantes. No entanto, a maioria desses trabalhos utiliza léxicos afetivos que não suportam o PB. Dessa maneira, neste trabalho foi realizada uma revisão sistemática da

literatura sobre os léxicos afetivos existentes para o PB. Foram encontrados sete léxicos criados entre os anos de 2007 e 2016.

Refletindo sobre o trabalho de pesquisa, faz-se necessário observar que a maioria dos estudos foram obtidos por meio de *Snowballing*, o que reforça a importância de uma *string de busca* bem definida associada ao uso dessa técnica na realização de uma boa RSL.

Embora tenham sido encontrados poucos léxicos afetivos em PB, a descrição dos mesmos com detalhes de suas criações e conteúdos é uma forma de elucidá-los para uso na educação. As possibilidades para uso dos léxicos afetivos tanto podem ir na direção do ambiente presencial de aprendizagem quanto do ambiente virtual de aprendizagem, analisando emoções nos diversos textos e respostas produzidos pelos alunos nas tarefas das disciplinas as quais estão envolvidos. Para este último, os léxicos afetivos ainda podem ser explorados nas ferramentas de comunicação como fóruns, mensagens e *chats* de maneira a apoiar o desenvolvimento de objetos de aprendizagem mais motivadores para os alunos.

Por fim, vale ressaltar que os mapeamentos sistemáticos apresentam o cenário da área em análise e auxiliam o desenvolvimento de estudos mais aprofundados. Dessa maneira, com esse trabalho, espera-se que os léxicos afetivos em português do Brasil possam ser mais divulgados e utilizados de uma maneira mais efetiva, principalmente para auxiliar na área da educação, como já acontece com léxicos afetivos em outros idiomas.

Os autores agradecem ao CEFET/RJ - PIBIC 2017 pelo apoio financeiro.

REFERENCIAS

- [1] Susana Afonso, Eckhard Bick, Renato Haber, and Diana Santos. 2002. Floresta sintá(c)tica: a treebank for Portuguese. In Manuel González Rodrigues; Carmen Paz Suarez Araujo (eds) Proceedings of LREC 2002, the Third International Conference on Language Resources and evaluation.
- [2] Fábio Pereira Alves and Cristiano Maciel. 2016. Codesign de atividades gamificadas: O papel das medalhas no planejamento de unidades de ensino. Proceedings of the XXI Congresso Internacional de Informática Educativa, 2016.
- [3] Amir Aly and Adriana Tapus. 2013. A model for synthesizing a combined verbal and nonverbal behavior based on personality traits in human-robot interaction. In Proceedings of the 8th ACM/IEEE International Conference on Human-robot interaction. IEEE Press, 325–332.
- [4] Pedro P Balage Filho, Thiago AS Pardo, and Sandra M Aluisio. 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 (STIL). 215–219.
- [5] Margaret M Bradley and Peter J Lang. 1999. Affective norms for English words (ANEW): Instruction manual and affective ratings. Technical Report. Technical report C-1, The Center for Research in Psychophysiology, University of Florida.
- [6] C Bueno, JA Rojo, and P Rodriguez. 2011. An experiment on semantic emotional evaluation of chats. In The Fifth International Conference on Advances in Semantic Processing.
- [7] Paula Carvalho and Mário J Silva. 2015. SentiLex-PT: Principais características e potencialidades. Oslo Studies in Language 7, 1.
- [8] Scotty Craig, Arthur Graesser, Jeremiah Sullins, and Barry Gholson. 2004. Affect and learning: an exploratory look into the role of affect in

- learning with AutoTutor. *Journal of educational media* 29, 3, 241–250.
- [9] Censo da Educação Superior. 2015. Notas Estatísticas. https://abmes.org.br/arquivos/documentos/notas_sobre_o_censo_da_educacao_superior_2016_versao_29_08_enviado.pdf.
- [10] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting Depression via Social Media. In ICWSM.
- [11] Kadja de Sousa Cavalcante, Marcelo Santana de Sousa, Letícia Ribeiro P de Oliveira, and Ivo Sócrates M de Oliveira. 2016. Análise de Softwares de Gestão Escolar Open Source para uma Escola Privada do Ensino Básico. Proceedings of the XXI Congresso Internacional de Informática Educativa, 2016.
- [12] Bento Carlos Dias-Da-Silva and Helio Roberto de Moraes. 2003. A construção de um thesaurus eletrônico para o português do Brasil. *ALFA: Revista de Linguística*.
- [13] Sidney D’Mello, Rosalind W Picard, and Arthur Graesser. 2007. Toward an affect-sensitive AutoTutor. *IEEE Intelligent Systems* 22, 4.
- [14] M Brent Donnellan, Frederick L Oswald, Brendan M Baird, and Richard E Lucas. 2006. The mini-IPIP scales: tiny-yet-effective measures of the Big Five factors of personality. *Psychological assessment* 18, 2, 192.
- [15] Cláudia Freitas. 2013. Sobre a construção de um léxico da afetividade para o processamento computacional do português. *Revista Brasileira de Linguística Aplicada* 13, 4, 1013–1059.
- [16] Cláudia Freitas, Eduardo Motta, R Milidiú, and Juliana César. 2012. Vampiro que brilha... rá! Desafios na anotação de opinião em um corpus de resenhas de livros. *Encontro de Linguística de Corpus* 11, 1–13.
- [17] Jennifer Golbeck, Cristina Robles, and Karen Turner. 2011. Predicting personality with social media. In CHI’11 extended abstracts on human factors in computing systems. ACM, 253–262.
- [18] Eva Hudlicka. 2003. To feel or not to feel: The role of affect in human–computer interaction. *International Journal of Human-Computer Studies* 59, 1, 1–32.
- [19] Oliver P John and Sanjay Srivastava. 1999. The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and Research* 2, 102–138.
- [20] Barbara Kitchenham. 2004. Procedures for performing systematic reviews. *Keele, UK, Keele University* 33, 1–26.
- [21] Christian Haag Kristensen, Carlos Falcão de Azevedo Gomes, Alice Reuwsaat Justo, and Karin Vieira. 2011. Brazilian norms for the affective norms for English words. *Trends in Psychiatry and Psychotherapy* 33, 3, 135–146.
- [22] Magali Teresinha Longhi. 2011. Mapeamento de aspectos afetivos em um ambiente virtual de aprendizagem. Ph.D. Universidade Federal do Rio Grande do Sul.
- [23] Antonio AA Machado, Magali T Longhi, Maria ASN Nunes, Thiago Alexandre Salgueiro Pardo, et al. 2015. Personalitatem Lexicon: um léxico em português brasileiro para mineração de traços de personalidade em textos. In Conferência Latino-americana de Objetos e Tecnologias de Aprendizagem, X; Congresso Brasileiro de Informática na Educação, IV; Simpósio Brasileiro de Informática na Educação, XXVI. Universidade Federal de Alagoas-UFAL.
- [24] Antonio Aliberte de Andrade Machado et al. 2016. Inferência de personalidade a partir de textos de rede social utilizando um léxico afetivo em português brasileiro. Universidade Federal de Sergipe-UFS.
- [25] Rada Mihalcea, Carmen Banea, and Janyce M Wiebe. 2007. Learning multilingual subjective language via cross-lingual projections. *Annual meeting-association for computational linguistics*. Vol. 45. No. 1. 2007.
- [26] George A Miller. 1995. WordNet: a lexical database for English. *Commun. ACM* 38, 1, 39–41.
- [27] Aldo Marcelo Paim. 2016. Inferência de personalidade a partir de textos em português brasileiro utilizando léxicos. Ph.D. Pontifícia Universidade Católica do Paraná.
- [28] Alexander Pak and Patrick Paroubek. 2011. Twitter for sentiment analysis: When language resources are not available. In Database and Expert Systems Applications (DEXA), 22nd International Workshop on. IEEE, 111–115.
- [29] Paulo Roberto Pasqualotti and Renata Vieira. 2008. WordnetAffectBR: uma base lexical de palavras de emoções para a língua portuguesa. *RENOTE* 6, 1.
- [30] James W Pennebaker, Martha E Francis, and Roger J Booth. 2001. Linguistic inquiry and word count: LIWC 2001. Mahway: Lawrence Erlbaum Associates 71.
- [31] Rosalind Picard. 1997. *Affective computing*. Vol. 252. MIT press Cambridge.
- [32] Sandro J Rigo, Isa M Alves, Oliver Gazola, Francini Belau, Jorge LV Barbosa, and Cristiano Costa. 2013. Abordagem linguística para identificação da dimensão afetiva expressa em textos de Ambientes Virtuais de Aprendizagem—um Léxico da Emoção. In Brazilian Symposium on Computers in Education, Vol. 24. 738.
- [33] Pilar Rodriguez, Alvaro Ortigosa, and Rosa M Carro. 2012. Extracting emotions from texts in e-learning environments. In Complex, Intelligent and Software Intensive Systems (CISIS), Sixth International Conference on. IEEE, 887–892.
- [34] Andrew Sears and Julie A Jacko. 2009. *Human-Computer Interaction Fundamentals*. CRC Press.
- [35] Marlo Souza, Renata Vieira, Débora Busetti, Rove Chishman, Isa Mara Alves, et al. 2011. Construction of a portuguese opinion lexicon from multiple resources. In 8th Brazilian Symposium in Information and Human Language Technology. 59–66.
- [36] Rosemarie Streeton, Mary Cooke, and Jackie Campbell. 2004. Researching the researchers: Using a snowballing technique. *Nurse researcher* 12, 1, 35–46.
- [37] Alessandro Valitutti, Carlo Strapparava, and Oliviero Stock. 2004. Developing affective lexical resources. *PsychNology Journal* 2, 1, 61–83.