

Sobre a construção de um léxico da afetividade para o processamento computacional do português

A Sentiment Lexicon for Portuguese Natural Language Processing

Cláudia Freitas*

Pontifícia Universidade Católica

Rio de Janeiro / Rio de Janeiro – Brasil

RESUMO: O objetivo principal deste artigo é descrever o processo de elaboração de um léxico de elementos afetivos da língua portuguesa e polaridades associadas, construído principalmente para auxiliar a tarefa de análise de sentimento. Busca-se, além disso, mostrar como a exploração da dimensão afetiva da linguagem encontra espaço na descrição da língua e pode, ainda, contribuir para o ensino de português como segunda língua. Assim, o léxico foi criado com um duplo objetivo: fornecer subsídios linguísticos para uma aplicação do processamento de linguagem natural (PLN) e contribuir com a caracterização do vocabulário afetivo da língua portuguesa, sobretudo em contextos informais.

PALAVRAS-CHAVE: léxico de polaridades; léxico de sentimento; linguística com *corpus*; análise de sentimento; prosódia semântica; linguística computacional; análise de opinião.

ABSTRACT: The main goal of this paper is to describe the creation of a sentiment lexicon designed to collaborate in sentiment analysis tasks. We also show how the analysis of the affective language finds room in language description and that it can contribute to foreign language teaching as well. Thus, the lexicon was created with a dual purpose: provide a resource for natural language processing tasks and collaborate to the characterization of Portuguese language affective vocabulary, especially in informal contexts.

KEYWORDS: corpus linguistics; polarity lexicon; sentiment lexicon; sentiment analysis; emotional words; semantic prosody; computational linguistics; opinion mining.

* claudiafreitas@puc-rio.br

1. Introdução

Recursos lexicais são reconhecidamente parte fundamental de sistemas que lidam com o processamento computacional da língua, os quais podem ter como objetivo a realização de tarefas relacionadas à identificação e extração da informação contida em textos. Recentemente, tem havido um crescente interesse não apenas na identificação de informações factuais (como *quem, o quê, quando, onde* etc.), mas também na identificação da informação subjetiva que textos veiculam. A área chamada análise de sentimento ou mineração de opinião¹ busca justamente identificar opiniões, avaliações e atitudes expressas em texto (PANG; LEE, 2008).

No entanto, para processar informação é preciso informação. Essa informação, cada vez mais, é fornecida por léxicos, que, no âmbito do processamento computacional da língua, se referem ao componente de um sistema que contém informação (semântica e/ou gramatical) sobre palavras ou expressões, enquanto o termo *dicionário*, normalmente, remete a objetos (livros impressos ou eletrônicos) destinados a leitores humanos, mas que também podem estar acessíveis a máquinas (GUTHRIE *et al.*, 1996).

Assim, uma das abordagens para a extração automática de sentimento/opinião toma por base léxicos de emoções/sentimentos, que podem ser gerais, como o SentiWordNet (ESULI; SEBASTIANI, 2006), ou específicos para determinados domínios ou tarefas (RILOFF; WIEBE, 2003; POIRIER *et al.*, 2011). Para o português (variante de Portugal²), destacamos o SentiLex (SILVA; CARVALHO; SARMENTO, 2012), léxico construído a partir de um *corpus* composto por comentários (*posts*) em matérias de jornal sobre política. Tais léxicos contêm, além das classes de palavras, a polaridade associada a cada item, que pode ser positiva, negativa e, em alguns casos, neutra.

¹ Como notam Pang e Lee (2008), são variadas as nomenclaturas para os trabalhos que lidam com o tratamento computacional de opinião/sentimento/subjetividade no texto: mineração de opinião (*opinion mining*); análise de sentimento (*sentiment analysis*) e análise de subjetividade (*subjectivity analysis*).

² Ainda que no âmbito de léxicos computacionais não seja relevante aludir a diferenças entre variantes, visto que, para a maior parte do processamento da língua, o que é comum é mais importante do que o específico, a diferença é mencionada devido a variações quanto à atribuição de polaridade de algumas palavras/expressões no Senti-Lex, não sendo possível saber se tais alterações se devem justamente à variante ou a outro fenômeno. O adjetivo *incível*, por exemplo, comentado na seção 2.1, possui polaridade neutra no Senti-Lex, diferentemente do que é proposto neste artigo.

Palavras³ são aglomerados de vários tipos de informação. A informação afetiva, simplificada aqui como polaridade, é mais uma. Considere-se que uma dada palavra tem polaridade quando é sistematicamente utilizada para expressar um sentimento sobre algo. *Perfeito*, *admirável* e *amei* são exemplos de palavras com polaridade positiva.

Quanto à forma de elaboração, léxicos com polaridades podem ser construídos manual ou automaticamente, dando-se preferência à maneira automática. No entanto, a criação automática se sustenta (i) em recursos lexicais criados manualmente, como a WordNet (FELLBAUM, 1998) – para a qual não há, até o momento, equivalente em português em termos qualitativos e quantitativos –, ou (ii) em corpora, que podem, por sua vez, ser manual ou automaticamente anotados com informação referente à polaridade.

O objetivo principal deste artigo é descrever o processo de elaboração de um léxico com elementos afetivos da língua portuguesa e polaridades associadas, o ReLi-Lex, criado a partir de um *corpus* composto por resenhas de livros publicadas na internet e previamente anotado com informação de opinião (FREITAS *et al.*, 2012; 2013). Busca-se, além disso, mostrar como a exploração da dimensão afetiva encontra espaço na descrição da língua e pode, ainda, contribuir para o ensino de línguas estrangeiras. Assim, o léxico foi criado com um duplo objetivo: fornecer subsídios para uma aplicação do processamento de linguagem natural (PLN) e contribuir para a caracterização do vocabulário afetivo da língua portuguesa, sobretudo em contextos informais.

A seção 2 apresenta a perspectiva teórica que norteou a elaboração do léxico, bem como a relação entre afetividade e estudos da linguagem e sua relação com o PLN; a seção 3 descreve o processo de elaboração do léxico e suas características; por fim, a seção 4 traz algumas considerações finais e discute alternativas para a continuação do trabalho.

2. PLN, léxico e significado: enquadramento teórico

A partir sobretudo da década de 1990, com a chamada “revolução do *corpus*”, a linguística computacional/Processamento de Linguagem Natural (PLN) se estabelece como área voltada à resolução de problemas práticos, comprometida principalmente com aplicações. Das tarefas que atualmente preocupam o PLN, a que motiva a realização deste trabalho é a análise de sentimento. Paralelamente, ao tratar da formalização de aspectos do significado para o processamento

³ Utilizo aqui o termo *palavra* consciente da sua dificuldade de delimitação e conceituação, como discutido em Biderman (2001).

computacional, é inevitável o diálogo com a semântica, “domínio de investigação de limites movediços” (ILARI; GERALDI, 1985, p. 6).

De forma bastante simplificada, é possível distinguir três paradigmas que irão problematizar o significado de forma sistemática: realista, mentalista e pragmático. Porém, ainda que didaticamente essa distinção seja útil, teorias realistas e mentalistas têm historicamente compartilhado pressupostos teóricos fundamentais, o que permite, com alguma simplificação, agrupá-las sob o rótulo *representacionistas* ou *essencialistas* (MARTINS, 2004). Para ambos, a linguagem é um sistema de representações de significados fixos e compartilhados; palavras representam algo (entidades mentais para os primeiros e virtuais para os segundos), e essa relação de representação se dá de maneira objetiva e estável.

Já o ponto de vista pragmático diz respeito à linguagem em uso, em diferentes contextos. Há uma mudança de perspectiva, uma vez que a linguagem passa a ser entendida como uma prática intersubjetiva. Dentre as linhas de investigação pragmáticas, contudo, há as que poderiam ser também enquadradas em um paradigma essencialista. Isso porque se, por um lado, mentalistas irão assumir que é pela análise das propriedades dos códigos de linguagem que será possível explicar a prática da comunicação, algumas correntes da pragmática recomendam a análise das propriedades da prática da comunicação como maneira de fornecer uma explicação do que são as línguas e os significados, o que faz com que esta visão pragmática tradicional possa ser compreendida como uma disciplina complementar a uma visão semântica essencialista (MARTINS, 1999; TAYLOR, 1992).

Porém, a crítica que pragmatistas mais radicais farão é que qualquer análise essencialista da linguagem é impossível, por ser impossível um distanciamento do objeto examinado. A relação entre linguagem e realidade seria forjada, na medida em que a própria linguagem constitui a realidade.

No presente trabalho, assumimos, com Ellis (1993), que palavras são (também) resultados de processos de categorização, guiados por interesses compartilhados por uma cultura/comunidade discursiva e não por fatos do mundo real, os quais rotulariam. Ao invés do rótulo, o relacionamento: palavras relacionam “situações”, e grupos de coisas diferentes, ou experiências únicas, formam uma categoria coerente – a categoria “folha”, para usar o exemplo nietzschiano⁴ – porque

⁴ “Assim como é evidente que uma folha não é nunca completamente idêntica à outra, é também bastante evidente que o conceito de folha foi formado a partir do abandono arbitrário destas características particulares e do esquecimento daquilo que diferencia um objeto de outro. O conceito faz nascer a idéia de que haveria na natureza, independentemente das folhas particulares, algo como a folha, algo como

nós, membros de uma dada comunidade, fazemos com que eles formem, por meio de estratégias de simplificação e equivalência. O que de modo algum equivale a afirmar que não existam similaridades entre as coisas, mas não parece que essas sejam a única base para as categorias de uma língua. Assim, situações que não são exatamente as mesmas são categorizadas como sendo as mesmas. E é o recorte – a categorização – dependente e articulado com as intenções que traz, por um lado, a estabilidade para que possamos acordar sobre significados, perceber regularidades e, por outro, a instabilidade quando se busca definir precisamente os próprios significados, dado que este não é entendido como intrínseco às palavras.

Com relação ao significado, assumimos, portanto, uma posição alinhada com a perspectiva antiessencialista, segundo a qual significados (dos quais a polaridade é uma das dimensões) correspondem a usos culturalmente determinados que fazemos das palavras. Assim, sobre as regularidades que inegavelmente observamos quanto ao significado (e que subjazem, por exemplo, à elaboração de dicionários), são consequência de ser, o significado, “construído e atribuído a partir de um tácito acordo comunitário” (ARROJO, 1992, p. 37).

Outro aspecto frequentemente associado à elaboração de um léxico que contém informação relativa às polaridades que podem ser atribuídas às palavras em certos contextos é a aproximação com modelos que lidam com a distinção conotativo/denotativo, sendo a polaridade vinculada à conotação. Assumir uma visão antiessencialista, no entanto, é

reconhecer que nenhuma teoria da linguagem conseguiu, até hoje, a partir de pressupostos logocêntricos, distinções objetivas e indiscutíveis entre o literal e o figurado, entre o irônico e o não irônico, ou entre o literário e o não-literário enquanto propriedades textuais intrínsecas. (ARROJO, 1992, p. 36).

A Análise de Sentimento pode ser vista como uma subárea da Extração de Informação que lida com a identificação de opiniões, avaliações e atitudes com relação a entidades como pessoas, produtos e organizações, expressas textualmente. Estimulada sobretudo pela crescente disponibilidade e diversidade de opiniões

uma forma primordial, segundo a qual todas as folhas teriam sido tecidas, desenhadas, cortadas, coloridas, pregueadas, pintadas, mas por mãos tão inábeis que nenhum exemplar teria saído tão adequado ou fiel, de modo a ser uma cópia em conformidade com o original.” (NIETZSCHE, 1911, p. 12).

em *blogs* e redes sociais, a tarefa pode assumir diferentes aspectos: da caracterização de documentos (por exemplo, resenhas), como favoráveis ou desfavoráveis relativamente a um dado objeto, até à identificação de pontos de vista em debates políticos, trazendo desafios interessantes tanto do ponto de vista do processamento computacional quanto do ponto de vista linguístico.

Quando criados automaticamente a partir de *corpus*, léxicos para a análise de sentimento (também chamados de dicionários de polaridades, conforme Taboada *et al.* [2011]) podem partir de uma lista de “palavras-semente” (*seed words*), – um grupo pequeno de palavras com uma polaridade forte e estável, como *ruim*, *excelente*, *pésimo* –, usada para expandir listas de palavras, gerais ou dependentes de domínio. Outra alternativa consiste em utilizar bases lexicais preexistentes, como tesouros (MOHAMMAD; DUNNE; DORR, 2009) ou a WordNet, para a criação de léxicos gerais, como a SentiWordNet (ESULI; SEBASTIANI, 2006). No entanto, lembramos que, para o português, não há equivalente para a WordNet em termos de quantidade e qualidade da informação disponibilizada, ainda que algumas tentativas estejam em andamento e que a conjugação de diferentes recursos já existentes para o português apareça como uma alternativa interessante (para uma revisão da literatura sobre recursos lexicais disponíveis para o processamento computacional do português, veja-se Santos *et al.* [2010]).

Por fim, embora a forma de elaboração manual seja normalmente evitada, devido exclusivamente ao seu alto custo, é aquela que oferece resultados mais confiáveis. De um ponto de vista linguístico, a criação de recursos como léxicos semânticos, de forma manual ou semiautomática, se oferece como uma oportunidade para a observação da língua em ambiente natural, e se feita a partir de *corpus*, com um imenso potencial para a exploração de fatos da língua – quer do ponto de vista das regularidades, quer das irregularidades. Além disso, como salienta Sampson (2001), o trabalho empírico, de taxonomização, é uma atividade valiosa em si mesma, autorizada pelo compartilhamento de esforços e de recursos, e não uma tarefa desinteressante e trivial, anterior ao trabalho de descobertas sobre a língua.

2.1. Afetividade e dicionários

Na perspectiva linguística, o interesse pela linguagem emotiva ou afetiva tem ocupado posição periférica. Como notam Ochs e Schieffelin (1989), linguistas têm subestimado a atuação de estruturas gramaticais e discursivas para fins afetivos, e a tradição lexicográfica não costuma incluir informação

chamada “conotativa” sistematicamente nos dicionários, dada, principalmente, sua alta dependência contextual.

Por exemplo, é uma informação relevante do português saber que *incrível* (na variante brasileira, pelo menos), é usada para caracterizar positivamente: *um ator incrível, um telefone incrível* ou *um restaurante incrível*. A consulta aos dicionários, no entanto, omite essa informação: tanto no *Dicionário eletrônico Houaiss* quanto na versão *on-line* do *Dicionário Priberam da Língua Portuguesa*, são três as acepções de *incrível*. Reproduzo apenas as acepções deste último:⁵

1. Que não pode ser acreditado.
2. Extraordinário.
3. Que custa a acreditar.

Para falantes nativos, é fácil perceber que *restaurante incrível* equivale a *restaurante extraordinário*, ambos expressando um sentimento positivo, mas não há nada que forneça essa informação – e sabemos que *extraordinário*, assim como *incrível*, não necessariamente é usado para exprimir um ponto de vista positivo. A consulta a *extraordinário*, por sua vez, oferece diversas acepções, apenas uma delas relacionada ao uso de *incrível* que nos interessa:⁶

4. Excessivo; muito grande; descomunal; singular; raro; anormal; assombroso; estupendo.

Novamente, falantes nativos não têm dificuldade para reconhecer que apenas *estupendo* se assemelha ao *incrível* positivo, ainda que *assombroso* e *raro* também possam ser usados. Mas, novamente, não há pista no dicionário que nos leve a isso.

Imaginemos agora um dicionário com informação de polaridade. Nesse dicionário, para a palavra *incrível*, teríamos a marcação [+] na acepção 2, *extraordinário*, indicando que, se alguém me diz que foi a um restaurante incrível, provavelmente é porque o restaurante foi bom e que, também, um restaurante *incrível* é, de certa maneira, equivalente a *restaurante extraordinário*. Supondo que não se saiba o que é *extraordinário*, a indicação de polaridade em suas diferentes acepções ajudaria a perceber quais termos seriam mais

⁵ <http://www.priberam.pt/dlpo/default.aspx?pal=incr%C3%ADvel>

⁶ <http://www.priberam.pt/dlpo/default.aspx?pal=extraordin%C3%A1rio>

apropriados como sinônimos. A FIG. 1 ilustra o caminho percorrido, e ao final poderíamos inferir a relativa equivalência entre um restaurante *incrível*, *extraordinário*, *estupendo* e *admirável* – e a inadequação entre a sinonímia de um restaurante *incrível* e um restaurante *anormal*, *excessivo* ou *espantoso*.⁷

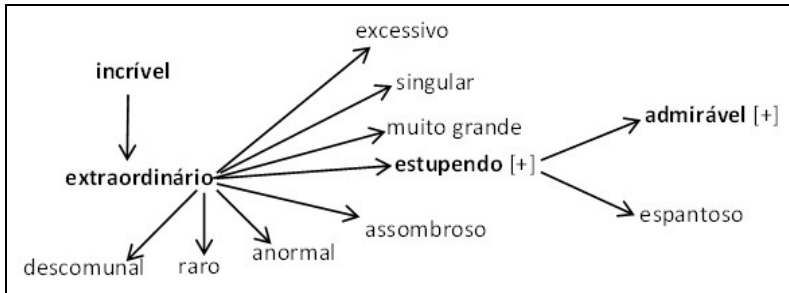


FIGURA 1 – Rede de palavras associadas a *incrível*
 Fonte: Elaboração própria.

2.2. Polaridade e prosódia semântica

A atribuição de polaridade a palavras ou expressões também é compatível com a noção de prosódia semântica (PS). Surgida no contexto da linguística com *corpus*, especialmente na fraseologia, a PS está relacionada diretamente à ideia de naturalidade na linguagem – algumas combinações, ainda que bem formadas sintaticamente, podem não parecer “naturais”, como seria o caso da produção de falantes não nativos de uma língua. Ou seja, a PS indicaria apenas uma preferência, e não uma obrigatoriedade. Além da naturalidade, a PS revela, por meio dos padrões de associação, atitudes favoráveis ou desfavoráveis que impregnariam combinações de palavras que normalmente ocorrem juntas. Uma combinação como *ampliar o número de mortos/feridos* é muito pouco provável,⁸

⁷ Idealmente, *extraordinário* também deveria remeter para *incrível*. Nesse caso, *incrível* – associado a *extraordinário* – teria a marca [+].

⁸ Vale notar que, em uma busca nos *corpora* do projeto AC/DC (Acesso a Corpus/ Disponibilização de Corpus, cf. seção 2.3), há diversas ocorrências de *elegar|aumentar|crescer|subir* o número de mortos/feridos, em que os verbos usados na busca podem ser considerados sinônimos de *ampliar*. Curiosamente, há uma única ocorrência de *ampliar*, mas com viés positivo, já que se fala de “mortos *beneficiados*”: “par=FSP950830-374: Uma delas tem o objetivo de **ampliar** o número de mortos *beneficiados* com a indenização.” Para efetuar a concordância, a expressão de procura foi [pos=”V”] “o” “número” “de” [word=”mortos|feridos”]

assim como *ampliar as dificuldades*; *ampliar as críticas ou ampliar a tristeza*, porque *ampliar* tende a se associar a palavras “positivas” – diz-se que *ampliar* tem prosódia semântica positiva⁹ – ainda que, nos dicionários consultados, não haja qualquer menção a essa característica.

Como usual na metalinguagem relacionada ao significado, o conceito de PS tem sido usado de diferentes maneiras, com diferentes nuances, algumas inclusive discordantes entre si (HUNSTON, 2007). Não é o objetivo deste trabalho, no entanto, problematizar a noção de PS, e por isso o foco está na atribuição de atitudes/pontos de vista, positivos ou negativos, em contexto, a partir de *corpus*.

Na elaboração do léxico, assumimos que a atribuição de polaridade às palavras não pode ser isolada do ponto de vista. Admitir que algo pode ser intrinsecamente bom ou ruim é incompatível com as posições defendidas quanto ao significado. Assim, é razoável supor que frequentemente não existe uma interpretação indisputável. O verbo *reconhecer*, por exemplo, evidencia a concordância com uma proposição que vai contra um ponto de vista mais geral. No exemplo adaptado de Hunston (2007), “X é talvez o primeiro historiador branco a *reconhecer* que sem as armas de fogo, a África poderia não ter sido escravizada”, infere-se, a partir da escolha pelo verbo *reconhecer*, que quem narra o fato acha que as armas de fogo contribuíram para a escravidão da África,

⁹ Outros exemplos retirados do *corpus* Floresta, por meio da interface AC/DC:

(1) Além disso, o ANC vai reter e até **ampliar** as suas infra-estruturas militares e prosseguirá também com a sua campanha contra a violência.

(2) Agora vamos **ampliar** a nossa oferta para os ambientes empresariais com o Digital Unix, que é um dos melhores Unix do mercado.

(3) **Ampliando** suas atividades ao público, o Museu da Imagem e do Som de São Paulo promove oficinas nas áreas de cinema, vídeo, fotografia, novas tecnologias, marketing cultural, som e grafite.

(4) Nos últimos dias, o Banco Nacional de Desenvolvimento Econômico e Social (BNDES) passou a **ampliar** o foco de sua atuação, ajudando a organizar fundos de investimento em empresas emergentes com potencial tecnológico.

(5) Dois ótimos novos procuradores da Internet que **ampliam** as opções do Google:

(6) A IBM, maior empresa de informática do mundo, está **ampliando** sua atuação no mercado financeiro brasileiro.

(7) Lufthansa **amplia** oferta de vôos para Munique.

(8) E o FC Porto acabou mesmo por **ampliar** a vantagem, muito por culpa de Eduardo Filipe, Ricardo Costa e Rui Rocha, um trio que só à sua conta somou 12 golos.

diferentemente do autor do discurso relatado na frase, “X”. No entanto, embora o exemplo ilustre a presença de opinião e de posicionamento, este é difícil de ser enquadrado em termos de polaridade positiva ou negativa. E, por isso, esse tipo de construção não integra, até o momento, o léxico criado.

2.3. Polaridade, emoção e sentimento

Na tradição ocidental, encontramos formulações a respeito da afetividade desde Platão, que conceitua emoção, cognição e conação. Atualmente, o tema interessa a áreas distintas, da antropologia à inteligência artificial, cada uma com seus desafios e suas respostas. A multiplicidade de olhares e interesses leva, naturalmente, à diversidade de nomenclaturas – somando-se à dificuldade intrínseca de tratamento de um tema com dimensões e desdobramentos culturais, cognitivos, filosóficos e linguísticos (veja-se Longhi [2011] para uma descrição detalhada das diferentes abordagens para o estudo das emoções).

Utilizo, neste trabalho, a palavra afetividade para caracterizar o léxico, termo usado para fazer referência à vivência de sentimentos positivos ou negativos. Dado o amplo escopo da afetividade, e o caráter inicial da exploração, a observação foi reduzida a uma das dimensões afetivas da linguagem: a polaridade. A polaridade, por sua vez, perpassa as emoções: considerando, por exemplo, as chamadas “emoções básicas” propostas em Ekman (1999) – medo, raiva, tristeza, alegria, surpresa, aversão – percebemos que, se alegria é uma emoção cuja polaridade é positiva, e raiva e tristeza são regularmente associadas a uma polaridade negativa, a emoção *surpresa* pode ser positiva (frases [1]-[2]), negativa (frases [3]-[4]), ou pode ter uma polaridade muito fraca, se alguma (frases [5]-[6]), como ilustram os exemplos retirados de *corpora* que integram o projeto AC/DC - Acesso a corpos/Disponibilização de corpos (COSTA *et al.*, 2009):¹⁰

- (1) É impressionante, mas a tecnologia continua *surpreendendo* com uma velocidade como nunca se viu na história do homem.
- (2) Me *surpreendi* com a qualidade de trabalhos que eu desconhecia.
- (3) E Carlos Narciso junta a sua voz à aqueles que se *surpreendem* com do facto de Domingos Pereira, condenado a 15 anos pela morte da mulher, ter cumprido apenas seis.

¹⁰ Os *corpora* estão disponíveis em: <<http://www.linguateca.pt/ACDC>>.

- (4) Já pensou, ser *surpreendido* com a morte de algum amigo?
- (5) Não sei por que as pessoas ficam *surpreendidas* com o sucesso da nossa seleção.
- (6) E o pessoal da terceira idade se *surpreende* com tudo.

Além da polaridade, e para dar conta também da aplicação proposta, consideramos, na compilação das entradas do léxico, a presença de opinião. Não é novidade, no entanto, a dificuldade em estabelecer as fronteiras entre o que é opinião e o que não é. Compatível com a visão não essencialista aqui assumida, que não opõe literal/denotativo ao figurado/conotativo, está a não oposição entre fato e julgamento, ou entre “verdade” e retórica (ARROJO, 1992, p. 67). Isso porque a distinção pressuporia uma primeira etapa, um primeiro nível de relação entre sujeito/leitor e realidade/texto, “que pudesse ser objetivamente previsível e determinável, independentemente de uma perspectiva ou de um contexto.” (p. 67). Não é possível negar, no entanto, a existência de enunciados muito mais fortemente – ou explicitamente – carregados de opinião, como em “odiei aquele filme”, que certamente interessa identificar, considerando a aplicação pretendida. Para dar conta dessa frágil distinção, nos apoiamos na solução adotada no processo de anotação do próprio *corpus* de resenhas de livros do qual o léxico deriva: para que uma frase fosse considerada alvo da anotação de opinião, deveria, de alguma maneira, responder à pergunta hipotética “gostou do livro/dessa parte do livro”? Assim, por exemplo, se a resposta dada à pergunta é “achei triste/é triste/a leitura é complexa”, não é possível saber se a pessoa gostou – e portanto, não é possível atribuir polaridade. Já se a resposta é “achei horrível/é horrível”, temos claramente uma opinião negativa. O mesmo procedimento foi adotado na seleção das entradas que comporiam o léxico. Para a pergunta “Gostou de X?”, em que X pode ser um objeto, uma pessoa ou um evento, consideramos que as palavras/expressões do QUADRO 1 não respondem adequadamente à pergunta – e portanto não constam do léxico, ainda que possam apresentar alguma polaridade. Já as palavras/expressões do QUADRO 2 respondem à pergunta, ainda que indiretamente.

QUADRO 1

Palavras eliminadas do léxico de polaridades

ampliar[+]; confiar[+]; cumprir[+]; ganhar[+]; permitir[+]; responder[+]; superar[+];

faltar[-]; perturbar[-]; forçar[-].

Fonte: Elaboração própria.

QUADRO 2

Palavras que integram o léxico de polaridades

adorar[+]; agradar[+]; apreciar[+]; atrair[+]; conquistar[+]; curtir[+]; deliciar[+]; divertir[+]; emocionar[+]; empolgar[+]; fascinar[+]; fluir[+]; inovar[+]; recomendar[+]; sensibilizar[+]; simpatizar[+]; valorizar[+]; vibrar[+]; viciar[+];

aborrecer[-]; arrepender[-]; assustar[-]; atrapalhar[-]; cansar[-]; complicar[-]; decepcionar[-]; desistir[-]; detestar[-]; dificultar[-]; distorcer[-]; empacar[-]; enganar[-]; estragar[-]; estressar[-]; exagerar[-]; faltar[-]; frustrar[-]; incomodar[-]; irritar[-]; largar[-]; odiar[-]; pecar[-]; revoltar[-].

Fonte: Elaboração própria.

3. A elaboração do léxico

O léxico aqui apresentado foi derivado do *corpus* ReLi (FREITAS *et al.*, 2012; 2013), composto por resenhas de livros publicadas na internet. O ReLi caracteriza-se por uma grande ocorrência de material expressivo e um uso variado de registros – tanto informal, típico da internet, quanto um pouco mais formal. Contém 1600 resenhas de treze livros (sete autores), totalizando cerca de 260 mil palavras e 12 mil frases. Para cada livro foram coletadas cerca de 200 resenhas e, quando esse número não pôde ser atingido, completamos com outras obras do mesmo autor até chegarmos a um número próximo a duzentos. O *corpus* foi automaticamente anotado com informação de classes de palavras e manualmente anotado quanto à expressão de opinião. No processo de anotação da opinião, foram marcadas (i) a polaridade da frase que expressa opinião; (ii) segmento(s) que expressa(m) a opinião; (iii) polaridade desses segmentos.¹¹ Uma descrição detalhada do *corpus*, processo de anotação e teste interanotadores encontra-se em (FREITAS *et al.*, 2012; 2013).

¹¹ Nem sempre a polaridade dos segmentos corresponde à polaridade geral da frase, como em “ainda que em alguns momentos a narrativa derrape, o livro é ótimo”.

A partir do *corpus* anotado, o primeiro passo foi extrair as sequências marcadas como núcleos de opinião, e agrupá-las segundo a classe gramatical. Em seguida, como nem todas as palavras núcleo expressam opinião (*repleto, total, necessário* etc.), foi feita, manualmente, uma seleção das candidatas a entradas. Além das palavras e expressões, a análise do *corpus* permitiu também a elaboração de listas relacionadas à inversão de polaridade (estruturas negativas). As classes consideradas foram ADJ, V, N. Para cada classe, há uma lista de entradas com os lemas, obtidos automaticamente com o analisador morfossintático PALAVRAS (BICK, 2000), e polaridades. Expressões multivocabulares também integram o léxico.

O processo de seleção das entradas, além de eliminar as palavras que não indicavam a presença de opinião e polaridade, se justifica principalmente pela preocupação em identificar palavras ou expressões que, embora em um dado contexto sistematicamente apresentassem polaridade/opinião, faziam-no apenas no contexto específico de resenhas de livros. Por exemplo, *imprevisível, perturbador* são características/ações consideradas positivas no *corpus* (exemplos [7]-[8]), mas dificilmente o são em outros contextos. Por outro lado, pareceu interessante não dispensar as particularidades da qualificação nesse contexto. Assim, a opção foi por criar entradas separadas com elementos específicos do domínio cultural.

(7) O mais interessante é que o final foi *imprevisível* e surpreendente.

(8) Sim, o final é ainda mais genial e *perturbador*!

A fim de obter mais segurança na análise das palavras e expressões, isto é, (i) confirmar que eram elementos que regularmente expressam afetividade e opinião; (ii) verificar se a polaridade expressa era geral ou específica de domínio, confrontamos as palavras das listas com as ocorrências nos diversos *corpora* do projeto AC/DC, especificamente no *corpus* Floresta, que contém textos jornalísticos e textos de *blogs* (FREITAS; SANTOS, 2013; AFONSO *et al.*, 2001). A TAB. 1 apresenta a distribuição da entradas do léxico, por classe de palavra e por polaridade. Para cada entrada, estão disponíveis as informações de lema, polaridade e classe de palavra.¹²

Dois critérios nortearam a incorporação de uma entrada no léxico. De acordo com o primeiro deles, mencionado na seção 2.3, as palavras ou

¹² O léxico está disponível em: <<http://www.linguateca.pt/Repositorio/ReLi>>.

expressões candidatas deveriam conjugar a presença de polaridade e de opinião. De acordo com o segundo, deveriam, ainda, apresentar (relativa) estabilidade quanto ao tipo de polaridade (e, por isso, a importância da verificação em outros *corpora*). Além disso, quando um dado uso se restringia a um tipo de texto (especificamente, textos opinativos), optou-se por manter a entrada. Se, por outro lado, a polaridade/opinião flutuava sem permitir a identificação de qualquer regularidade, não.

O verbo *sofrer* é um bom exemplo. Em uma busca nos *corpora* do AC/DC, embora apareça a tendência a ser acompanhado por palavras de polaridade negativa (*ameaças, problemas, baixa, discriminação*), o que permite considerá-lo um verbo de prosódia semântica negativa, há também uma quantidade considerável de ocorrências nas quais é difícil atribuir alguma polaridade, como em *sofrer modificações/mudanças/alterações*. No entanto, a construção *sofrer para* é usada regularmente para remeter algum tipo de dificuldade, de forma negativa do ponto de vista de quem relata. Assim, no léxico, consta apenas a entrada *sofrer para* (*sofri para terminar o livro, sofri para cancelar a compra, sofri para pedir um sanduíche* etc.).

A próxima seção comenta alguns resultados da exploração das listas de polaridade e do *corpus*, conforme as classes de palavras. O APÊNDICE A apresenta exemplos das palavras que compõem o léxico.

TABELA 1
Distribuição das entradas do léxico, por classe e por polaridade

	Polaridade [+]		Polaridade [-]		TOTAL
	Léxico geral	Léxico específico	Léxico geral	Léxico específico	
ADJ	245	11	116	5	377
V	67	2	52	3	124
N		52		47	99
Intensificadores			14		14
Expressões multivocabulares		27		32	59
Elementos inversores			13		13
Total					686

Fonte: Elaboração própria.

3.1. Adjetivos

A maioria das abordagens para a identificação de opinião baseadas em léxico privilegia adjetivos como indicadores da orientação semântica de um texto. No entanto, embora adjetivos sejam qualificadores por excelência, nem todos interessam a um léxico de polaridades – ou por terem um comportamento demasiado flutuante (por exemplo, *comum*), dependente do domínio ou do objeto qualificado (*celular pequeno* [+] vs. *memória pequena* [-]), ou pela dificuldade na própria atribuição da polaridade (*triste* em *O final da história é triste*).

De um ponto de vista semântico, adjetivos podem ser descritores, isso é, atribuir propriedades, ou classificadores, cuja função principal é delimitar a expressão/entidade referida. Para a análise de opinião são relevantes os adjetivos descritores, que podem ainda se subdividir em diferentes classes semânticas, das quais interessa a classe avaliativa/emotiva, que inclui adjetivos como *ruim*, *lindo*, *bom* (BIBER *et al.*, 1999, p. 509). No entanto, como nota Mória (1992), o comportamento semântico dos adjetivos está longe de ser homogêneo, e parece não haver propriedades distribucionais distintivas para as diferentes subclasses semânticas de adjetivos, mesma conclusão a que chega Rio-Torto (2006), quando afirma que as propriedades morfológico-sintático-semânticas do adjetivo estão a serviço das diferentes modalidades de funcionamento e de interpretação do adjetivo. Por isso a relevância e necessidade de uma análise manual cuidadosa dos adjetivos.

O adjetivo *arrasador*, por exemplo, aparece como candidato à entrada positiva, pois suas ocorrências no *corpus* de resenhas indicam sistematicamente uma opinião favorável. A busca no AC/DC revela, no entanto, que embora *arrasador* seja usado com frequência para expressar pontos de vista favoráveis (*um arrasador início de campeonato; a mais arrasadora personalidade feminina da Nova York; uma exibição arrasadora; um novo produto verdadeiramente arrasador*), esse nem sempre é o caso – o que, aliás, está de acordo com as definições lexicográficas de *arrasar*, que referem-se a destruição e devastação. No entanto, observamos que, no contexto dos objetos culturais, *arrasador* é sistematicamente usado de maneira positiva, como ilustram as ocorrências 9-12, retiradas do *corpus* Floresta do projeto AC/DC:

- (9) O livro é uma análise sofisticada e *arrasadora* do sistema político dos EUA no fim do século 19.
- (10) O resultado: uma comédia *arrasadora*, onde a trama perde a importância frente à interpretação e a graça das falas.

- (11) Quem fecha o disco é o U2, com um *arrasador* remix de «Lemon».
- (12) Em 90 minutos de um concerto *arrasador*, a musa do Carnaval da Baía mexeu, remexeu, rebolou, embalou, contagiando a multidão que encheu por completo o Coliseu da cidade.

Por isso, *arrasador* integra o léxico, mas em uma seção separada, com a indicação de que é um adjetivo com polaridade positiva em textos do domínio cultural.¹³

Alguns adjetivos, como *infantil* ou *descritivo*, podem ser descritores ou classificadores. No entanto, quando acompanhados de intensificadores, assumem uma polaridade negativa: *muito infantil/muito descritivo*; *infantil demais/descritivo demais*; *totalmente infantil/totalmente descritivo*. Essa informação está também no léxico, associada a regras, como ilustra a FIG. 2:

(a) [intensificador] descritivo/intelectual/juvenil/infantil/informal
(b) descritivo/intelectual/juvenil/infantil/informal [intensificador]

FIGURA 2 – Exemplo de entradas associadas a regras
Fonte: Elaboração própria

3.2. Verbos

No léxico, de modo semelhante aos adjetivos, os verbos foram separados em dois grupos: aqueles cuja polaridade é altamente estável, e aqueles cuja polaridade atribuída parece ter forte dependência do domínio/tipo de texto.

Além da propriedade predicadora dos verbos de relacionarem, pelo menos, dois argumentos, outro aspecto relevante no tratamento dos verbos diz respeito à diversidade de sentidos, que pode se refletir em diferentes polaridades.

O verbo *arrebentar*, embora sintaticamente dispense complemento nos exemplos (13) e (14), apenas em (14), utilizado em um registro informal, tem polaridade positiva:

¹³ Uma outra solução seria indicar que *arrasador* tem polaridade positiva quando o argumento não é agente. No entanto, não temos, ainda, como fornecer descrição de papéis semânticos.

- (13) A corda sempre *arrebenta* do lado mais fraco.
- (14) Aplauso para os caras que vieram do Acre e *arrebentaram* e a banda deles tá virando.

O verbo *arrebentar*, portanto, faz parte do léxico, mas integra a seção específica de domínio/registo. Dado o amplo uso em um registo informal com uma polaridade positiva regular, não há motivos para dispensar essa informação.

Outro exemplo interessante é o do verbo *amadurecer*. Embora, em termos gerais, seja difícil considerá-lo um verbo com polaridade, como ilustram os exemplos (15)-(18), retirados do *corpus* Floresta (AFONSO *et al.*, 2001, [s. p.]):

- (15) «Foi uma decisão muito ponderada, muito difícil e muito *amadurecida*», disse ao PÚBLICO Anabela Moutinho, que foi lacônica em o que se refere aos motivos.
- (16) Ele se nega a sugerir em público uma data ideal, mas entende que a adesão à URV ainda tem o que *amadurecer*.
- (17) Ele virá quando o país estiver *amadurecido* e sentir que entendeu as regras, que já não vamos fazer nenhuma injustiça aqui e ali e que é o momento de passar para o real.
- (18) “Estamos *amadurecendo* hipóteses sobre as razões do maior sucesso no Pará”, disse o pesquisador.

Quando tem como complemento entidades do domínio cultural, o verbo parece ser sistematicamente utilizado para expressar sentimentos positivos, indicando algum aspecto de melhoria e, conseqüentemente, a presença de uma opinião positiva, como ilustram (19)-(23), também retirados do *corpus* Floresta:

- (19) A fórmula pop da Calypso *amadureceu*.
- (20) É mesmo algo totalmente novo no gênero, que *amadureceu* esse tempo todo na cabeça e no sampler do DJ Sany Pitbull, o inventor das 5 faixas do novo EP.
- (21) Além da já conhecida habilidade em criar tramas onde o riso predomina e em saber o que esperar de cada ator que escala, ele mostrou que sua escrita *amadureceu*, e não se restringiu à coletânea de piadas de gosto duvidoso exibidas em doses diárias de 45 minutos.

- (22) Nesse processo, *amadureceram* como artistas e se apropriaram de todos os truques pop que foram pilhando pelo caminho.
- (23) Esse já é o 24º trabalho cinematográfico de Cícero, que visivelmente *amadureceu* no trabalho e teve como reflexo a qualidade superior do filme.

O uso regularmente positivo de *amadurecer* corresponderia, de certa maneira, às acepções 4 e 5 de *amadurecer* segundo o *Dicionário eletrônico Houaiss* (v. 3.0), que expressam uma qualidade positiva, ainda que não esteja explicitado:

4 tornar mais elaborado, desenvolvido, acabado

Ex.: a. um projeto, uma ideia

5 tornar-se equilibrado, ponderado, sensato

Ex.: a relação do casal, a princípio caótica, acabou por a.

(HOUAISS, 2009).

Palavras mais diretamente associadas a sentimento também são um campo interessante para a exploração da afetividade na linguagem. O verbo *apavorar*, por exemplo, indica que aquilo que apavora é capaz de despertar o sentimento de medo. Mas não necessariamente o sentimento será negativo, como exemplifica (24), retirado do *corpus* ReLi, de uma resenha claramente positiva com relação ao livro:

- (24) Este livro, para resumir, me *apavorou*. [...] O pavor vem, é claro, do envolvimento com a história [...]

Além disso, em contextos altamente informais, *apavorar* pode ser equivalente a *ter um bom desempenho*, e pode portanto ser usado de forma positiva, ao menos na variante brasileira do português, como ilustram os exemplos abaixo, retirados da internet e nos quais, conjugam-se polaridade positiva e opinião:

- (25) Ele *apavorou* aqui em Brasília!!! Pra variar um pouquinho o Luiz Lima detonou aqui em Brasília na 4ª etapa do Brasileiro!!!! Invicto e com mais uma vitória!!!
- (26) Depois de 400 fitas e muitos testes, achamos Zachary Weilandt, recomendação do fotógrafo Mark Weiss. Ele *apavorou* no teste.
- (27) [...] e o mais legal foi a sua técnica com Slap, aí ele *apavorou*, foi realmente empolgante a velocidade e o bom senso musical do Tracy [...]

Por fim, quanto aos verbos, vale lembrar que expressões multivocabulares verbais têm um comportamento mais previsível, e por isso compõem boa parte do léxico verbal. Se *invejar*, por exemplo, tem um comportamento de difícil formalização em termos de polaridade, a expressão *de fazer inveja* é usada regularmente em situações positivas. No apêndice listamos expressões que constam do léxico, mesmo as que não contêm verbos.

3.3. Negação e inversores de polaridade

Em trabalhos relacionados à atribuição de polaridade, é fundamental considerar elementos de negação, dada a sua capacidade de inverter a polaridade de determinadas palavras ou expressões: *não gostei* é negativo, pois o *não* inverte a polaridade positiva de *gostei*. Mas nem apenas advérbios exercem a função de inversores. O adjetivo *impossível* também pode funcionar como inversor (*impossíveis de largar* *impossível abandonar o livro por a metade*, em que *abandonar* e *largar* são verbos com polaridade negativa, e o inversor *impossível* faz com que os trechos passem a indicar comentários favoráveis ao livro.).

Os exemplos acima ilustram dependência contextual e a impossibilidade de polaridades/atribuições estanques, assim como de significados estanques. A polaridade indicada nos itens do léxico revela o uso mais convencional, e busca refletir as generalizações feitas pelos falantes do português. Sabemos, no entanto, do papel fundamental da criatividade na linguagem humana, ainda mais fortemente presente em contextos expressivos e opinativos. Por isso a importância do *corpus*, entendido como o espaço que simultaneamente reflete o uso convencional e fornece amostras do potencial criativo, nos confrontando com a impossibilidade de regularidades absolutas e comportamentos totalmente previsíveis.

4. Considerações finais

Apresentamos aqui algumas considerações relativas à construção de um léxico do português com elementos expressivos e polaridades associadas a partir da sua identificação em um *corpus* de resenhas de livros.

O processo de elaboração manual do léxico se justifica principalmente pela preocupação em eliminar as palavras que, embora em um dado contexto aparecessem como portadoras de polaridade/opinião, faziam-no apenas em contextos muito restritos. Adicionalmente, a existência de um léxico com elementos expressivos da língua portuguesa e polaridades associadas pode ser de

grande valor também de um ponto de vista lexicográfico. Por isso, o processo de seleção das palavras não consistiu apenas na leitura e eventual descarte de algumas palavras ou expressões, mas incluiu a consulta a *corpora* diversificados, com o objetivo de observar o comportamento das palavras em diferentes contextos, e não apenas no âmbito de resenhas de livros publicadas na internet. Assim, o léxico foi criado com um duplo objetivo: fornecer subsídios para uma aplicação do PLN, nomeadamente a tarefa de análise de opinião, e contribuir para a caracterização do vocabulário expressivo da língua portuguesa, sobretudo em contextos informais.

A inexistência de dicionários para a linguagem expressiva dificulta a realização da tarefa do ponto de vista linguístico, ao mesmo tempo em que motiva a investigação sobre essa dimensão da linguagem. É sabido que o vocabulário de expressividade é de grande relevância para aqueles que desejam aprender uma língua. Por outro lado, ao menos com relação ao ensino de português, são escassos os recursos didáticos que atentam para esse aspecto tão humano da interação. A elaboração de dicionários para aprendizes, que incorporem esse tipo de informação é, portanto, uma atividade de grande relevância no contexto do ensino de línguas, e o presente trabalho busca também sensibilizar a comunidade de estudos da linguagem para esse aspecto do português, bem como lançar algumas sugestões sobre como abordar a questão.

As palavras e expressões do ReLi-Lex podem contribuir para a investigação, na língua, de outras dimensões semânticas, como as emoções. Palavras com polaridade têm, também, sentimento, embora nem sempre um dado sentimento esteja associado à mesma polaridade: um livro, filme ou enredo *angustiante* ou *aterrorizante* podem ser considerados bons, ainda que a angústia ou o terror não sejam, normalmente, sentimentos positivos. Assim, uma possibilidade futura é o estudo sobre como a polaridade dos sentimentos varia, em complementação ao que vem sendo feito por Maia e Santos (2012). Outro caminho a ser explorado é a incorporação de informação relativa à intensidade: algo que não é excelente é, necessariamente, ruim? Por fim, investigar em que medida a combinação de palavras com polaridades distintas pode indicar a presença de ironia também parece um caminho instigante. Do ponto de vista da descrição do português, espera-se que o trabalho possa contribuir para uma caracterização da classe semântica dos adjetivos avaliativos/emotivos.

É interessante constatar que nosso léxico é bem mais enxuto que os relatados na literatura (SILVA *et al.* [2012], TABOADA *et al.* [2011] e o

SentiWordNet, para citar alguns). Não consideramos, por exemplo, palavras com polaridade neutra, o que explica boa parte da diferença numérica. Por outro lado, um léxico maior e mais abrangente pode ser uma armadilha. Cada domínio ou área tem suas preferências discursivas e jargões, e dar conta de todos os casos talvez seja dar conta de nenhum. *Aterrorizante, sufocante, surpreendente, imprevisível, perturbador* podem ser positivos para objetos culturais como filmes e livros, mas certamente não o são em outros domínios. Por isso, parece interessante a ideia de um léxico geral, enxuto, associado a léxicos-satélites, específicos de domínio, que podem ser construídos por demanda. No ReLi-Lex, as entradas cuja polaridade pareceu específica do domínio “livros” foram separadas.

Do ponto de vista do PLN, o desenvolvimento mais imediato é a investigação sobre formas automáticas ou semiautomáticas de expansão do léxico, fazendo uso de pistas léxico-sintáticas como a coordenação.

Do ponto de vista teórico, assumir uma perspectiva não essencialista com relação ao significado tem como desdobramento o reconhecimento de que qualquer palavra é capaz de encerrar uma carga expressiva que pode, inclusive, ser variável. No entanto, não há como negar regularidades no uso da língua em uma dada comunidade discursiva, e o léxico proposto tenta capturar tais regularidades no que se refere à expressividade. A intenção, assim, é revelar quais significados são selecionados em um dado momento em uma dada comunidade, já que interpretações semelhantes resultam de convenções institucionalizadas. Entender o significado atribuído é, portanto, entender uma das faces das convenções (GRIGOLETTO, 1992, p. 95).

Deste modo, o léxico captura o que pode ser generalizado, as convenções, as estabilidades. Os usos particulares, por sua vez, permanecem no *corpus*, como exemplo da resistência; da criatividade no uso.

Quando se associa ao PLN, uma perspectiva não essencialista sobre a linguagem, que busca “construir generalizações que, se não são definitivas, buscam lançar alguma luz sobre nossas práticas linguísticas” (MARTINS, 1999, p. 144), se mostra produtiva na medida em que assume o auxílio à resolução de tarefas de sistemas que manipulam a língua como um objetivo tão legítimo quanto a compreensão do funcionamento da mente com relação à linguagem ou o ensino de línguas, por exemplo. Uma linguística que assume suas/as limitações na busca por um conhecimento objetivo e verdadeiro parece a mais indicada para colaborar em uma área como o PLN, que precisa tanto do conhecimento linguístico quanto de flexibilidade na utilização desse

conhecimento. Vale lembrar, por fim, que este conhecimento não deve, idealmente, ser limitado demais a ponto de resolver um problema de um único sistema, o que possivelmente acarretaria em um resultado de pouco interesse linguístico, ainda que com impacto do ponto de vista prático. A ideia é que a descrição seja *motivada* por uma aplicação, mas não *subordinada* a ela.

Referências

- AFONSO, S. *et al.* Floresta sintá(c)tica: um treebank para o português. In: ENCONTRO DA ASSOCIAÇÃO PORTUGUESA DE LINGUÍSTICA, 17., out. 2001, Lisboa. *Actas...* Disponível em: <<http://www.linguateca.pt/Diana/download/AfonsoetalAPL2001.pdf>>. Acesso em: 27 jul. 2013. Não paginado.
- ARROJO, R. A desconstrução do logocentrismo e a origem do significado. In: _____ (Org.). *O signo desconstruído*. Campinas: Pontes, 1992. p. 41-46.
- ATKINS, B. T. S. *The Oxford Guide to Practical Lexicography*. Oxford (UK): Oxford University Press, 2008.
- BIBER, D. *et al.* *Longman Grammar of Spoken and Written English*. New York: Pearson Education, 1999.
- BICK, E. *The Parsing System "Palavras": Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. Aarhus: Aarhus University Press, 2000.
- BIDERMAN, M. T. *Teoria lingüística (teoria lexical e lingüística computacional)*. 2. ed. São Paulo: Martins Fontes, 2001.
- CORPOS de português: Projecto AC/DC. Disponível em: <<http://www.linguateca.pt/ACDC>>. Acesso em: 27 jul. 2013.
- CORPUS ReLi. Disponível em: <<http://www.linguateca.pt/Repositorio/ReLi>>. Acesso em: 27 jul. 2013.
- COSTA, L.; SANTOS, D.; ROCHA, P. Estudando o português tal como é usado: o serviço AC/DC. In: Brazilian Symposium in Information and Human Language Technology (STIL 2009), 7., 8-11 set. 2009, São Carlos. *Anais...* Disponível em: <<http://www.lbd.dcc.ufmg.br/colecoes/stil/2009/020.pdf>>. Acesso em: 27 jul. 2013. Não paginado.
- EKMAN, P. Basic Emotions. In: DALGLEISH, T.; POWER, T. (Ed.). *The Handbook of Cognition and Emotion*. Sussex: John Wiley & Sons, 1999. p. 45-60.
- ELLIS, J. M. *Language, Thought and Logic*. Evanston: Northwestern University Press, 1993.

- ESULI, A.; SEBASTIANI, F. SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. In: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION (LREC), 5., May 2006, Genoa. *Proceedings...* Disponível em: <http://www.lrec-conf.org/proceedings/lrec2006/pdf/384_pdf.pdf>. Acesso em: 27 jul. 2013. p. 417-422.
- EXTRAORDINÁRIO. In: *Dicionário Priberam da Língua Portuguesa*. Disponível em: <<http://www.priberam.pt/dlpo/default.aspx?pal=extraordin%C3%A1rio>>. Acesso em: 27 jul. 2013.
- FELLBAUM, C. *WordNet: An Electronic Lexical Database*. Cambridge (MA): MIT Press, 1998.
- FREITAS, C. *et al.* Sparkle Vampire... LoL! Annotating opinions in a book review corpus. In: ALUÍSIO, S.; TAGNIN, S. (Org.). *11th Corpus Linguistics Conference*. Cambridge (UK): Cambridge Scholars Publishing, 2013. No prelo.
- FREITAS, C. *et al.* Vampiro que brilha... rá! Desafios na anotação de opinião em um *corpus* de resenhas de livros. In: ENCONTRO DE LINGUÍSTICA DE CORPUS, 11., 2012, São Carlos. *Anais...* Disponível em: <http://www.linguateca.pt/Repositorio/ReLi/Anais_ELC2012_Freitasetal.pdf>. Acesso em: 27 jul. 2013. Não paginado.
- FREITAS, C.; SANTOS, D. Blogs, Amazônia e a Floresta Sintá(c)tica: um corpus de um novo gênero? In: SARMENTO, S. *et al.* (Org.). *Pesquisas e perspectivas em linguística de corpus*. Campinas: Mercado de Letras, 2013.
- GRIGOLETTO, M. A desconstrução do signo e a ilusão da trama. In: ARROJO, R. (Org.). *O signo desconstruído*. Campinas: Pontes, 1992. p. 31-34.
- GUTHRIE, L. *et al.* The Role of Lexicons in Natural Language Processing. *Communications of the ACM*, v. 39, n. 1, p. 63-72, Jan. 1996.
- HOUAISS, A. *Dicionário eletrônico Houaiss da língua portuguesa*. Versão 3.0. Rio de Janeiro: Objetiva, 2009. 1 CD-ROM.
- HUNSTON, S. Semantic Prosody Revisited. *International Journal of Corpus Linguistics*, v. 12, n. 2, p. 249-268, 2007. Disponível em: <http://www.researchgate.net/publication/228929277_Semantic_prosody_revisited>. Acesso em: 27 jul. 2013.
- ILARI, R.; GERALDI, J. W. *Semântica*. São Paulo: Ática, 1985.
- INCRÍVEL. In: *Dicionário Priberam da Língua Portuguesa*. Disponível em: <<http://www.priberam.pt/dlpo/default.aspx?pal=incr%C3%ADvel>>. Acesso em: 27 jul. 2013.

LONGHI, M. T. *Mapeamento de aspectos afetivos em um ambiente virtual de aprendizagem*. 2011. 273 f. Tese (Doutorado em Informática na Educação) – Centro de Estudos Interdisciplinares em Novas Tecnologias na Educação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2011. Disponível em: <<http://pct.capes.gov.br/teses/2011/42001013075P9/TES.PDF>>. Acesso em: 27 jul. 2013.

MAIA, B.; SANTOS, D. “Who’s afraid of ... what?” – in English and Portuguese. *Varieng* – Studies in Variation, Contacts and Change in English, v. 12, 2012. Disponível em: <http://www.helsinki.fi/varieng/journal/volumes/12/maia_santos>. Acesso em: 27 jul. 2013. Não paginado.

MARTINS, H. *Metáfora e polissemia no estudo das línguas do mundo: uma aproximação não representacionista*. 1999. 190 f. Tese (Doutorado em Linguística e Filologia) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1999.

MARTINS, H. Três caminhos na filosofia da linguagem. In: MUSSALIM, F.; BENTES, A. C. (Org.). *Introdução à Linguística*. São Paulo: Cortez, 2004. v. 3, p. 439-474.

MOHAMMAD, S., DUNNE, C., DORR, B. Generating High-Coverage Semantic Orientation Lexicons From Overtly Marked Words and a Thesaurus. 2009 CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING, 6-7 ago. 2009, Singapura. *Proceedings...* Disponível em: <<http://www.cs.umd.edu/~cdunne/hcil/pubs/Mohammad09Generatinghigh-coveragesemantic.pdf>>. Acesso em: 27 jul. 2013. p. 599-608.

MÓIA, T. Sobre classes semânticas de adjetivos. *Cadernos de Semântica*, Lisboa, v. 7, p. 1-34, 1992.

NIETZSCHE, F. W. Sobre verdade e mentira no sentido extra-moral. In: *Obras incompletas/Friedrich Nietzsche*. Seleção de textos de Gérard Lebrun; tradução e notas de Rubens Rodrigues Torres Filho. 4. ed. São Paulo: Nova Cultura, 1911 [1987]. (Os Pensadores).

OCHS, E.; SCHIEFFELIN, B. Language has a heart. *Text – Interdisciplinary Journal for the Study of Discourse*, v. 9, n. 1, p. 7-26, 1989.

PANG, B.; LEE, L. Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, v. 2, n. 1-2, p. 1-135, 2008.

RIO-TORTO, G. Para uma gramática do adjetivo. *Alfa*, São Paulo, v. 50, n. 2, p. 103-129, 2006.

SAMPSON, G. *Empirical Linguistics*. London: Continuum, 2001.

SANTOS, D. *et al.* Relações semânticas em português: comparando o TeP, o MWN.PT, o Port4NooJ e o PAPEL. In: BRITO, A. M. *et al.* (Ed.). *XXV Encontro Nacional da Associação Portuguesa de Linguística*: Textos seleccionados. Associação Portuguesa de Linguística: Lisboa, 2010. p. 681-700. Disponível em: <<http://www.apl.org.pt/apl-actas/xv-encontro-nacional-da-apl.html>>. Acesso em: 27 jul. 2013.

SILVA, M.; CARVALHO, C.; SARMENTO, L. Building a Sentiment Lexicon for Social Judgement Mining. In: *Computational Processing of the Portuguese Language*. 10th International Conference, PROPOR 2012, Coimbra, Portugal, April 17-20, 2012. Berlin: Springer Berlin Heidelberg, 2012. p. 218-228. (Lecture Notes in Computer Science, 7243).

TABOADA, M. *et al.* Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, v. 37, n. 2, p. 267-307, Jun. 2011. Disponível em: <http://cgi.sfu.ca/~mtabuada/docs/Taboada_etal_SO-CAL.pdf>. Acesso em: 27 jul. 2013.

TAYLOR, T. *Mutual Misunderstanding: Scepticism and the Theorizing of Language and Interpretation (Post-Contemporary Interventions)*. Durham; London: Duke University Press, 1992.

WILSON, T.; WIEBE, J.; HOFFMANN, P. Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In: 2005 HUMAN LANGUAGE TECHNOLOGY CONFERENCE AND THE CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING, 2005, Vancouver. *Proceedings...* Disponível em: <<http://www.cs.utexas.edu/~ml/HLT-EMNLP05>>. Acesso em: 27 jul. 2013. p. 347-354.

WITTGENSTEIN, L. *Investigações filosóficas*. São Paulo: Abril Cultural, 1979. (Os Pensadores).

APÊNDICE A – Entradas do léxico

	SUBSTANTIVOS [+]	SUBSTANTIVOS [-]
	[alegria]; [astúcia]; [atemporalidade]; [beleza]; [bem-estar]; [best-seller]; [brilho]; [carisma]; [clímax]; [companheirismo]; [compreensão]; [deleite]; [densidade]; [diferencial]; [diversão]; [doçura]; [emoção]; [empatia]; [entusiasmo]; [envolvente]; [êxtase]; [fã]; [facilidade]; [fascinação]; [favorito]; [fluência]; [fluidez]; [genialidade]; [gostinho]; [grandeza]; [inteligência]; [ironia]; [jóia]; [magnetismo]; [mérito]; [obra-prima]; [originalidade]; [paixão]; [perfeição]; [prazer]; [predileto]; [preferido]; [primor]; [sensibilidade]; [sufocamento]; [talento]; [taquicardia]; [ternura]; [top]; [virtude].	[anticlímax]; [besteira]; [bobagem]; [boçal]; [bosta]; [chatice]; [clichê]; [decepção]; [defeito]; [demorar]; [desastre]; [desconforto]; [desesperança]; [desgraça]; [desprazer]; [droga]; [engano]; [erro]; [exagero]; [exaustão]; [excesso]; [falhar]; [falho]; [falta]; [furo]; [imaturidade]; [incômodo]; [inexperiente]; [lixar]; [meleca]; [melodrama]; [melodramático]; [meloso]; [monótono]; [mundo]; [nojo]; [passividade]; [pena]; [pieguice]; [porca]; [porcaria]; [problema]; [sacrifício]; [sentimentalismo]; [superficialidade]; [vergonha].
	VERBOS [+]	VERBOS [-]
Léxico geral	[adorar]; [agradar]; [amadurecer]; [amar]; [animar]; [apaixonar]; [apegar]; [apreciar]; [aprender]; [arrebatar]; [arrebentar]; [atrair]; [cativar]; [conquistar]; [curtir]; [deliciar]; [destacar-se]; [divertir]; [emocionar]; [emplacar]; [empolgar]; [encantar]; [enternecer]; [entreter]; [facilitar]; [fascinar]; [fluir]; [gostar]; [identificar]; [inovar]; [recomendar]; [sensibilizar]; [simpatizar]; [valorizar]; [vibrar]; [viciar].	[abandonar]; [aborrecer]; [arrepender]; [assustar]; [aterrorizar]; [atormentar]; [atrapalhar]; [cansar]; [chocar]; [complicar]; [decepcionar]; [deprimir]; [desanimar]; [desgastar]; [desistir]; [desmerecer]; [destorcer]; [detestar]; [dificultar]; [distorcer]; [empacar]; [enfraquecer]; [enganar]; [engolir]; [errar]; [estragar]; [estressar]; [exagerar]; [faltar]; [frustrar]; [incomodar]; [irritar]; [largar]; [limitar]; [odiar]; [pecar]; [perder]; [prolongar]; [revoltar].
Verbos com polaridade no âmbito de objetos culturais	[apavorar]; [brilhar]; [comover]; [devorar] [prender]; [rir]; [saborear] [surpreender]; [tocar].	

	ADJETIVOS [+]	ADJETIVOS [-]
Léxico geral	<p>[admirável]; [adorável]; [agradável]; [alegre]; [alucinante]; [apaixonante]; [apoteótico]; [ardente]; [arrasador]; [arreatador]; [astuto]; [atemporal]; [aterrador]; [atraente]; [atual]; [atuar]; [autêntico]; [avassalador]; [bacana]; [bárbaro]; [batuta]; [belíssimo]; [belo]; [bem]; [bem-humorado]; [bom]; [bonito]; [brilhante]; [carinhoso]; [carismático]; [cativante]; [cativar]; [certo]; [claro]; [classe]; [clássico]; [coerente]; [comédia]; [cômico]; [comovente]; [compreensível]; [comum]; [consistente]; [contagante]; [contemporâneo]; [contente]; [convicente]; [corajoso]; [criativo]; [crivar]; [curto]; [decente]; [delicado]; [delicioso]; [demais]; [denso]; [desafiador]; [despretensioso]; [devanear]; [digno]; [dinâmico]; [direto]; [distrativo]; [divertido]; [divino]; [doce]; [duro]; [elegante]; [eletrizante]; [emocionante]; [empolgante]; [encantador]; [encantável]; [engenhoso]; [engraçado]; [enigmático]; [enriquecedor]; [envolvente]; [especial]; [esperto]; [espetacular]; [espirituoso]; [esplêndido]; [essa]; [estonteante]; [excelente]; [excepcional]; [excitante]; [exemplar]; [extraordinário]; [fabuloso]; [face]; [fácil]; [fama]; [fantástico]; [fascinante]; [fascinate]; [favorito]; [feliz]; [fluente]; [fluido]; [foder]; [fofo]; [forte]; [frenético]; [fundamental]; [genial]; [gostoso]; [gotoso]; [grande]; [grandioso]; [gratificante]; [grato]; [hilário]; [humano]; [ideal]; [imcompar]; [impactante]; [impagável]; [impar]; [ímpar]; [impecável]; [imperdível];</p>	<p>[aborrecente]; [água-com-açúcar]; [anacrônico]; [besta]; [bestar]; [bizarro]; [bobo]; [burro]; [cansativo]; [chato]; [chocante]; [chulo]; [clichê]; [confuso]; [construir]; [decepcionante]; [defeituoso]; [deplorável]; [depressivo]; [deprimente]; [desagradável]; [desconexo]; [desgastante]; [desinteressante]; [desnecessário]; [desprezível]; [difícil]; [dispensável]; [doentio]; [egoísta]; [enfadonho]; [enjoado]; [enjoativo]; [entediante]; [esdrúxulo]; [estereotipado]; [estranho]; [falsar]; [falso]; [fraco]; [frio]; [frustrante]; [fútil]; [graça]; [horrível]; [horrorosa]; [idiota]; [idiotece]; [imaturado]; [impaciente]; [impressionar]; [incompreensível]; [inconsistente]; [ingênuo]; [injustificável]; [insuportável]; [interminável]; [inútil]; [irritante]; [lamentável]; [maçante]; [machista]; [mal]; [manjado]; [massa]; [mau]; [mediano]; [médio]; [meloso]; [mero]; [modinha]; [monótono]; [morno]; [negativo]; [normal]; [obsessivo]; [óbvio]; [oco]; [opressivo]; [paranóico]; [patético]; [pesaroso]; [péssimo]; [piegas]; [piorar]; [plana]; [pobre]; [pomposo]; [preguiçoso]; [previsível]; [puritano]; [rasa]; [raso]; [razoável]; [reacionário]; [reclamo]; [repa]; [repetitivo]; [repulsivo]; [revoltante]; [ridículo]; [ruim]; [sem]; [simplista]; [sofrível]; [superficial]; [surreal]; [tédio]; [tedioso]; [tenda]; [tongo]; [tosco]; [triste]; [vazio]; [violento]; [volúvel];</p>

ADJETIVOS [+]

ADJETIVOS [-]

[importante]; [impressionante];
[inabalável]; [inacreditável]; [incrível];
[indescritível]; [indispensável];
[inedito]; [inesquecível]; [infalível];
[inovador]; [inquietante]; [inquieto];
[inspirador]; [instigante]; [instingar];
[inteligente]; [intenso]; [interessante];
[íntimo]; [intrigante]; [irado]; [irônico];
[irresistível]; [jovial]; [legal]; [legítimo];
[leve]; [liça]; [light]; [linda]; [lindo];
[lírico]; [mágico]; [magistral];
[magnético]; [magnificar]; [magnífico];
[majestoso]; [maravilha]; [maravilhoso];
[marcante]; [massa]; [máximo];
[melífluo]; [memorável]; [minucioso];
[misterioso]; [moderno]; [notável];
[novo]; [obra]; [ode]; [original];
[ótimo]; [palpável]; [passável];
[perfeito]; [pertinente]; [plausível];
[poderoso]; [poético]; [pontual];
[positivo]; [precioso]; [predileto];
[primo]; [primoroso]; [profético];
[profundo]; [pungente]; [rápido];
[raro]; [razoável]; [real]; [realista];
[realístico]; [recente]; [recomendar];
[recomendável]; [reflexivo]; [relaxante];
[relevante]; [rico]; [sagaz];
[satisfatório]; [seco]; [sedutor];
[seguro]; [sensacional]; [sensível];
[sensual]; [sentimental]; [sério]; [sexy];
[significativo]; [simples]; [sincero];
[singelo]; [singular]; [sólido]; [sublime];
[subversivo]; [surpreso]; [surreal];
[teres]; [tocante]; [tranquilo];
[translúcido]; [último]; [válido];
[verdadeiro]; [verossímil]; [vertiginoso];
[viciante]; [vicioso]; [vigoroso];
[visceral]; [vital]; [voraz].

Adjetivos com polaridade no âmbito de objetos culturais	[arrasador]; [aterrorizante]; [assustador]; [angustiante]; [compulsivo]; [frenético]; [imprevisível]; [surpreendente]; [sufocante]; [perturbador]; [sombrio].	[estático]; [idêntico]; [lento]; [linear] [óbvio].
--	---	--

	EXPRESSÕES [+]	EXPRESSÕES [-]
Léxico geral	[abrir (a) mente]; [amor à primeira vista]; [arrancar lágrimas/ risadas/ gargalhadas]; [capturar a atenção]; [chamar a atenção]; [com chave de ouro]; [dar certo]; [de corpo e alma]; [de encher os olhos]; [de fazer inveja]; [de tirar o fôlego]; [deixar de boca aberta]; [[estar louco para]; faltar adjetivos]; [faltar palavras]; [fazer pensar]; [fora de série]; [gostinho de quero mais]; [grata surpresa]; [ir a nocaute]; [lindo de morrer]; [morrer de rir]; [não ter adjetivos para]; [pra ninguém botar defeito]; [ser (um) show]; [ser um achado]; [ser um tapa na cara]; [soco na boca de o estômago]; [ter fôlego]; [tudo de bom]; [valer cada N]; [valer a pena].	[água com açúcar]; [cabeça-dura]; [carregar nas tintas]; [dar náuseas]; [dar raiva]; [dar sono]; [deixar a desejar]; [embrulhar o estômago]; [esperar mais]; [forçar a barra]; [mala sem alça]; [melhor esquecer]; [não aturar]; [não engolir]; [não prestar]; [não se enxergar]; [perder o fôlego]; [picolé de chuchu]; [pra criancinha]; [quebrou o encanto]; [sem graça]; [sem sal]; [sem-noção]; [ser um balde de água fria]; [ser um saco]; [tempo perdido / perder tempo/perda de tempo].

Recebido em 28/02/2013. Aprovado em 13/05/2013.