DOI: 10.17851/2237-2083.23.3.641-680

O léxico do corpo e anotação de sentidos em grandes *corpora*: o projeto *Esqueleto*

The lexicon of the human body and sense annotation: a corpus based study

Cláudia Freitas
Pontifícia Universidade Católica do Rio de Janeiro, RJ; Linguateca
claudiafreitas@puc-rio.br

Diana Santos
Universidade de Oslo, Noruega; Linguateca
d.s.m.santos@ilos.uio.no

Bruno Carriço
Pontifícia Universidade Católica, Rio de Janeiro, RJ[1]
carrico85@gmail.com

Cristina Mota
Linguateca
cmota21@gmail.com

Heidi Jansen
Universidade de Oslo, Noruega
heidi.lanorvegese@gmail.com

Resumo: Apresentamos aqui os resultados iniciais de um amplo estudo sobre o léxico do corpo humano e os seus sentidos, realizado por meio da anotação e revisão de *corpora* de grandes dimensões. Ao longo do artigo explicitamos as decisões linguísticas subjacentes à anotação, relatamos o resultado de um estudo sobre as classes de anotação e exploramos o vasto material criado: um *corpus* de entrevistas (1,4 milhão de palavras) e um *corpus* literário (1,2 milhão de palavras) anotados e integralmente revistos, e demais *corpora* do projeto,

[1] Bolsista de Iniciação Científica da FAPERJ/ E-26/200.812/2015.

parcialmente revistos. Todo o material está publicamente disponível para a comunidade.

Palavras-chave: corpo humano; léxico; anotação semântica; *corpus*; descrição do português.

Abstract: This paper presents the first results of a broad study regarding the lexicon of the human body. The study was based on the annotation of large corpora of Portuguese language. We explain the linguistic annotation choices, present the results of an agreement study and explore the material made available: a corpus of interviews (1.4 million words) and a literary corpus (1.2 million words) full annotated and revised, and the remained corpora partially revised. The whole material is publicly available.

Keywords: human body; lexicon; sense annotation; corpus linguistics; Portuguese.

Recebido em: 30 de junho de 2015.
Aprovado em: 20 de novembro de 2015.

1 Apresentação: motivação e interesses

Cada vez mais é reconhecida a importância da informação semântica associada aos *corpora*, ao mesmo tempo em que é indiscutível a dificuldade envolvida no processo de anotação, não apenas do ponto de vista da anotação automática mas também da anotação manual ou semiautomática.

A importância – e interesse – nesse tipo de informação semântica deve-se à sua vasta aplicação: tarefas do processamento automático de linguagem (PLN) se beneficiam de *corpora* semanticamente anotados e pesquisas linguísticas (voltadas para tradução, ensino de línguas e estudos contrastivos) também.

Em geral, a motivação para a anotação costuma vir de aplicações em PLN: a análise de sentimento / opinião motivou a elaboração de *corpora* anotados quanto a opinião e sentimentos (WIEBE *et al*., 2005; CARVALHO *et al.*, 2011; FREITAS *et al*., 2014; MOTA; SANTOS,

2015, VILLENA ROMÁN *et al*., 2015); tarefas de reconhecimento de entidades mencionadas motivaram a anotação semântica de nomes próprios e relações entre eles, como ilustram iniciativas, como o ACE (DODDINGTON *et al*. 2004) e o HAREM (MOTA; SANTOS, 2008), e o interesse na identificação de relações semânticas entre elementos do texto motivou a criação do PropBank (KINGSBURY *et al*., 2002; PALMER *et al*., 2005) e sua versão brasileira, o PropBank-BR (DURAN; ALUÍSIO, 2012), para citar apenas alguns.

Quando a motivação para a anotação vem, sobretudo, do estudo da língua, a inclusão de informação semântica em *corpus* tem sido mais escassa. Com relação à língua portuguesa, temos conhecimento apenas do projeto C-ORAL Brasil (RASO; MELLO, 2012) e do material do AC/DC (COSTA *et al*., 2009), que hoje contém informação relativa ao campo semântico das cores (FREITAS *et al*., 2012) e do vestuário, e está em andamento a anotação de sentimento (SANTOS; MOTA, 2015). A anotação é uma atividade linguística que poderia ser mais explorada, uma vez que propicia o contato intenso com a língua em uso ao mesmo tempo em que obriga o pesquisador_linguista_anotador a sistematizar (enquadrar nas categorias de anotação) porções dessa língua. A anotação linguística nunca é neutra e, portanto, também interessa teoricamente, pois, (a) ao anotar conforme um dado modelo teórico os pesquisadores_anotadores têm a chance de pôr o modelo à prova, reforçando-o, enriquecendo-o ou refutando-o; ou (b) ao anotar conforme uma dada perspectiva ou motivação (que não necessariamente precisa estar vinculada a uma teoria linguística), os pesquisadores_anotadores têm a chance de testar empiricamente suas hipóteses de categorização acerca do fenômeno investigado.

Se a anotação semântica já é uma tarefa morosa, no viés (b) a morosidade é ainda mais evidente: o processo envolve não apenas a anotação propriamente e toda a discussão inerente a esse processo; envolve uma etapa anterior, de modelagem do fenômeno que se deseja anotar / investigar, ausente quando se parte de categorias pré-estabelecidas por uma teoria específica. Envolve, portanto, idas e vindas, à medida que os dados do *corpus* estão sempre prontos a confrontar nossas pré-concepções, nos fazendo rearranjar e reajustar as categorias-hipótese iniciais, em um processo empírico valoroso à prática linguística.

Este trabalho apresenta os resultados do projeto *Esqueleto*: um esforço conjunto de anotação semântica das palavras do corpo humano, que teve como objetivo geral investigar a estruturação do léxico do corpo na língua portuguesa a partir de sua ocorrência em *corpora*, que vem sendo executado na Linguateca por meio de uma colaboração entre a PUC-Rio e a Universidade de Oslo.

A motivação inicial para esse projeto surgiu de resultados anteriores relacionados à identificação de opinião em resenhas (FREITAS *et al*. 2012), quando foi constatada a grande presença de expressões de opinião associadas a palavras do corpo humano. Investigar as palavras do corpo humano, portanto, contribui também para a descrição de como expressamos opinião em português, ao mesmo tempo em que fornece subsídios para o levantamento de pistas lexicais que devem ser consideradas por sistemas interessados em detectar opinião em textos. Qualquer pessoa interessada na expressão de sentimentos ou opiniões tem de se debruçar, quer queira, quer não, sobre (algumas das) expressões associadas ao corpo humano.

Surgiu também da consciência de que o corpo humano era muito frequente e especial na anotação semântica em português da cor (SILVA; SANTOS, 2012), e das diferenças contrastivas conhecidas entre o português e as línguas germânicas em relação a partes do corpo. Assim, um estudo sobre o léxico do corpo também contribui para questões vinculadas a especificidades culturais. Shigehisa Kuriyama (1999 *apud* GREINER, 2005), em um estudo comparativo entre as diferentes concepções do corpo na China e no Ocidente, explica que a noção de corpo na China

> nunca foi um substantivo (um corpo com nome), e aparece descrita de forma mais próxima de adjetivos (...) [caracterizados] pela descrição de posturas, de atitudes, de gestos, como [...] corpo sentado, corpo em pé, corpo andando, corpo risonho, corpo que chora(...) (GREINER, 2005, p. 22.)

Ainda quanto a especificidades culturais, a indicação de traços de caráter, como alguém "bundão" ou "linguarudo" e expressão de sentimentos, como "nó na garganta" ou "dor de cotovelo", podem trazer

dificuldades para sistemas de apoio à tradução ou para aprendizes de uma segunda língua.

Outra característica relevante do léxico do corpo é sua alta frequência na língua, o que o torna especialmente interessante para estudos baseados em *corpora*. Adicionalmente à ampla ocorrência, uma busca superficial por palavras do corpo humano (e, como amplamente relatado nas abordagens de viés cognitivista (LAKOFF; JOHNSON, 1980) nos revela que boa parte das ocorrências refere-se a usos "não físicos". O interesse na identificação das palavras do corpo, portanto, reside não apenas na exploração de como, em português, falamos sobre o nosso corpo, mas também na investigação sobre a que tipos de outras coisas nos referimos quando usamos palavras do corpo.

Nos estudos pós-estruturalistas vinculados à desconstrução, reconhecendo-se o amplo papel da linguagem na constituição das identificações, o corpo, e especificamente as maneiras de falar do corpo, de referi-lo, são parte dos processos de identificação dos sujeitos, isto é, de marcação social (BUTLER, 2000; PINTO, 2007). Um levantamento amplo das maneiras de caracterizar o corpo – belo, feio, saudável, dócil, áspero, perfumado, alegre – também oferece insumo para explorações nesta perspectiva.

Especificamente, o *Esqueleto* busca responder às seguintes perguntas: (i) quando usamos palavras do léxico do corpo humano, que outros sentidos – que não os do corpo humano – estão em jogo?; (ii) como falamos do corpo humano em português?

Para tanto, conduzimos um processo de anotação semiautomática, que parte de um léxico de palavras do corpo humano para detectar automaticamente as palavras que serão anotadas. A anotação consiste em, considerando as 15 etiquetas (classes semânticas) propostas, atribuir uma ou mais a uma palavra ou expressão, conforme o contexto:

- Sofreu múltiplas fraturas nas duas pernas[sema="corpo"]
- Aí eu fui trabalhando a cabeça[sema="corpo:faculdade"] dele e também a minha pra gente mudar a situação.

Neste artigo, relatamos as opções linguísticas subjacentes à anotação do *Esqueleto*, bem como os resultados deste processo de

anotação, com base na revisão de dois *corpora*: o *corpus* Museu da Pessoa,[2] com 1,4 milhão de palavras, distribuídos em 200 entrevistas que tematizam a vida pessoal dos entrevistados; o *corpus* Obras,[3] atualmente com 1,2 milhão de palavras, composto por 25 obras da literatura brasileira já no domínio público.

Todo o material é público, e está disponível para consulta por meio da interface do serviço AC/DC.[4] Com isso, não apenas apresentamos os resultados de uma pesquisa sobre o léxico do corpo humano baseada em grandes *corpora*, mas também fornecemos à comunidade interessada um rico material para pesquisas futuras.

2 Outros estudos sobre o léxico do corpo humano e o enquadramento do *Esqueleto*

Muitas das pesquisas recentes que se interessam pelo léxico do corpo humano o fazem segundo o viés cognitivista, já que o corpo ocupa um espaço central nesse modelo, estando na base de processos metafóricos (LAKOFF; JOHNSON, 1980). Maalej e Yu (2011) tratam dos usos do léxico do corpo humano em diferentes línguas, em uma perspectiva também contrastiva.

Na perspetiva cognitivista, com relação a estudos envolvendo o corpo e a língua portuguesa, destacamos Soares da Silva (1992) e Leitão de Almeida *et al*. (2009).

Em outras perspectivas, e tratando também apenas da língua portuguesa, Baptista (2000) apresenta o comportamento de descrições de partes do corpo segundo o formalismo da léxico-gramática. Também de um ponto de vista da léxico-gramática, Vale (2013) trata de expressões verbais do tipo [N V (C de N_{hum})[5]], em que C corresponde a uma parte do

---

2 http://www.linguateca.pt/acesso/corpus.php?corpus=MUSEUDAPESSOA

3 http://www.linguateca.pt/OBras/OBras.html

4 http://www.linguateca.pt/ACDC. Esse serviço não só dá acesso a muitos *corpora* cujos donos autorizaram mas também adiciona muita informação linguística, e neste caso semântica, a esse material.

5 Na expressão, C é um substantivo que denomina parte do corpo ou um "elemento inalienável" do N humano. Por exemplo, "(o filme (N_0) encheu(V) o (saco(C) de

corpo humano, levantando a hipótese, a ser posteriormente confirmada em *corpus*, de que a presença de tais expressões fixas seria indicativa da presença de opinião em textos.

Orsi e Zavaglia (2010), agora de um ponto de vista lexicográfico, tratam de expressões idiomáticas tabu que envolvem unidades lexicais obscenas, especificamente referentes à genitália feminina, nas línguas portuguesa e italiana.

Do nosso ponto de vista, o *Esqueleto* é acima de tudo um projeto de anotação de *corpus*. Um projeto de anotação pode estar associado a uma dada teoria linguística, mas também pode estar comprometido mais diretamente com uma motivação empírica, e esse é o nosso caso. Considerando uma das perguntas norteadoras iniciais – quando usamos o léxico do corpo humano, que outros campos do sentido estão em jogo?–, nossa estratégia foi anotar / indicar a presença ou não de um uso associado ao corpo.

Dessa forma evitamos indicar se estamos diante de usos literais ou metafóricos / metonímicos, categorizando as palavras do corpo segundo sua distribuição pelos diferentes campos em que aparecem. Com essa decisão, evitamos, ao longo do processo de anotação, nos posicionar explicitamente no amplo debate sobre a metáfora e a busca de um sentido literal (veja-se ECO, 1991; ARROJO; RAJAGOPALAN, 1992, MARTINS, 2010; para ampla problematização sobre o literal e o metafórico na linguagem).

Mas não negamos os diferentes sentidos que podem ser atribuídos às palavras do corpo como as opções de anotação indicam. Nossa intenção foi agrupar as palavras do corpo segundo seus usos mais gerais. Aos interessados apenas em usos não corporais (metafóricos, segundo a metalinguagem tradicional), é possível uma busca em que se eliminam os demais usos, como será explorado na seção 3.

3 Mão na massa: o processo de anotação

3.1 A escolha do *corpus*

Ana(N_{hum}))", ou "a notícia(N_0) envenenou(V) a (alma(C) de Rui(N_{hum}))". Exemplos extraídos de Vale (2013).

Toda a anotação e estudo conduzidos com o *Esqueleto* tomaram por base os *corpora* do AC/DC. A opção por este material justifica-se, sobretudo, por 3 motivos:

i. livre acesso para pesquisas linguísticas. Dessa maneira, os resultados, bem com toda a documentação das opções de anotação, encontram-se públicos e disponíveis para aqueles interessados em investigar a distribuição e ocorrências em contexto das palavras do léxico do corpo na língua portuguesa;

ii. todo o material já foi anotado morfossintaticamente pelo *parser* PALAVRAS (BICK, 2000), o que se reflete tanto em facilidades no processo de anotação, uma vez que podemos dispor de informações, como lema e classe gramatical, na criação de regras, como em mais possibilidades de exploração – por exemplo, pode-se procurar por verbos que têm como complemento / argumento palavras do léxico do corpo; e

iii. o tamanho e variedade do AC/DC, que conta hoje com mais de 20 *corpora* diferentes, distribuídos em diferentes tipos de texto, como texto jornalístico, acadêmico, entrevistas e obras literárias, totalizando cerca de 1 bilhão de palavras.

Atualmente, já temos dois *corpora* integralmente revistos: o *corpus* OBras e o *corpus* Museu da Pessoa. No entanto, a ênfase na revisão desse material não significa que os demais estejam intocados em relação ao léxico do corpo humano. Devido à maneira como é realizada a anotação, que parte de regras bastante gerais que vão sendo refinadas conforme as ocorrências, todo o material está parcialmente revisto.

3.2 A anotação

O processo de anotação é semiautomático, utilizando uma ferramenta desenvolvida para esse tipo de atividade (SANTOS; MOTA,

2010). As regras são linguisticamente motivadas, e tiramos proveito da informação semântica e morfossintática previamente existente, já incluída no *corpus* pelo PALAVRAS. Em termos gerais, o processo parte de um léxico inicial (no nosso caso, uma lista com palavras do corpo humano, que pode conter palavras simples como "pé" ou expressões como "batata da perna" e "céu da boca") que é aplicado às palavras do *corpus*, anotando-as como palavras relativas ao corpo humano. Em seguida, por meio da análise das palavras inicialmente anotadas, são criadas regras de especialização ou de eliminação, para corrigir casos como "umbigo do mundo", que receberá uma etiqueta semântica específica, e "coluna social", em que "coluna" será desconsiderada como palavra do corpo.

A definição das categorias de anotação consiste no principal desafio do *Esqueleto*. Como mencionamos, optamos por não partir de categorias determinadas aprioristicamente. A estratégia utilizada consistiu em, considerando a observação das ocorrências em *corpus*:

i. estabilizar a primeira grande classificação relevante para as motivações do *Esqueleto*: palavras do corpo humano que se referem ao corpo humano *vs.* palavras e expressões do corpo humano que se distribuem por outros campos semânticos.

ii. a partir da análise das ocorrências, criar subclasses que organizassem as palavras e expressões do corpo humano por outros campos semânticos.

Quando a palavra do corpo é usada para fazer referência ao corpo humano, recebe a etiqueta corpo (procurável por meio da pesquisa [sema="corpo"]). Nos outros casos, receberá semas específicos, na forma[6] [sema="corpo:xxx"], conforme o seu uso.

Considerando a análise integral de dois *corpora* e a análise parcial dos demais *corpora* do AC/DC, temos hoje 15 classes, ou semas, estáveis. O Quadro 1 apresenta os semas, com exemplos de uso.

[6] Daqui em diante usamos a expressão da pesquisa para referir o nome da categoria.

Nesta etapa, cuidamos apenas dos substantivos e adjetivos do corpo.

Quadro1 – Classes semânticas do corpo humano no *Esqueleto*[7]

Sema	Exemplos
corpo	torceu o *pé* na corrida; ter *olhos* azuis
corpo:animal	*orelha* de porco;
corpo:centralidade	Seu departamento é o *cérebro* da operação; Sem revelar o *coração* do plano, Itamar rebatizou o conjunto de medidas(...)
corpo:doenca	não tenho medo do *pé* de atleta; esta medonha epidemia de *bexiga*
corpo:faculdade	uma provocação plástica para *olhos* e *ouvidos* livres
corpo:grupo	*corpo* docente; *coluna* do exército
corpo:lugar	no *coração* da floresta amazônica
corpo:medida	dois *dedos* de pinga; onda de 3 *pés*
corpo:movimento	ir a *pé*; assim que pôs os *pés* na cidade
corpo:opiniao	ele é um *bundão*; tem sempre um *orelhudo* na conversa
corpo:parte	*boca* do fogão; *membro* do Parlamento; *braço* da máfia
corpo:posicao	suplicou de *joelhos*; dormiu em *pé*
corpo:sentimento	com o *coração* apertado; o meu *sangue* ferve por vocês
corpo:vegetal	*dente* de alho; *pé* de jabuticaba
corpo:outros	*boca* da noite; uma *veia* pop

Uma vez que não partimos de categorias familiares a uma teoria específica, explicitamos a seguir as motivações e princípios que nortearam a estabilização de uma classe semântica.

A partir de uma exploração inicial do *corpus*, iniciamos o processo de anotação com cinco classes candidatas: corpo, opinião, sentimento, lugar e outros. À medida que a anotação e revisão avançaram, fomos criando novas classes.[8] Nesse processo, para que um sema se estabilizasse como categoria de anotação, ele deveria englobar (i) diferentes palavras do corpo que compartilhassem o mesmo tipo de sentido; ou (ii) poucas palavras do corpo, mas com um uso muitíssimo frequente e sistemático.

[7] Todos os quadros, tabelas, gráficos e figuras neste trabalho são de nossa autoria.

[8] Uma nova classe só era criada após uma discussão – e consenso – entre todos os envolvidos na anotação e, a cada alteração, havia revisões retroativas, a fim de garantir uniformidade.

Adicionalmente, tínhamos duas preocupações: (iii) evitar uma classificação muito granular do sentido, o que além de levar a um imenso número de classes, poderia contribuir para uma maior discordância quanto ao conteúdo de cada classe – classes mais gerais têm mais possibilidades de acomodar nuances de sentido; e (iv) não inchar a classe "outros" com usos sistemáticos.

A palavra boca ilustra bem nossa preocupação. Sempre que é usada com o sentido de entrada, atribuímos o sentido de lugar [sema="corpo:lugar"].

1. recebeu livre na **boca** da área
2. terá de se identificar na **boca** do caixa

Observando as demais ocorrências de boca, reparamos que, em alguns casos, o sentido de entrada, espacial, vai se direcionando para o sentido de tempo, vinculado a início: por onde se chega ou por onde se começa, em um deslizamento entre espaço e tempo:

3. Na **boca** da safra, as commmodities estão perdendo o fôlego.
4. A conversão no setor acontecerá na **boca** da entressafra, quando a oferta...
5. Que o faça, no entanto, todos os dias do ano, não apenas quando o país está à **boca** da urna, e nos limites da lei.

Encontramos, até o momento, apenas esses três usos / sentidos nos *corpora*, além do uso / sentido sistemático em boca da noite [6] (88 ocorrências em todo o material do AC/DC). Por isso, não criamos [sema="corpo:tempo"], e as ocorrências [3-5] estão anotadas como [sema="corpo:outros"]. Se encontrarmos ainda mais casos de boca ou de outros lemas associados a tempo, é possível que este novo sema seja criado.

6. Primeiro trabalham, depois vão à escola e depois brincam, no fim do dia, na **boca** da noite .

A documentação do *Esqueleto* (FREITAS, 2013) especifica e exemplifica cada uma das classes. Detalhamos aqui os casos que consideramos menos óbvios ou mais interessantes.

- O [sema="corpo:faculdade"] refere-se sobretudo aos 5 sentidos - visão, olfato, paladar, tato, audição – além da faculdade / capacidade do pensamento, frequentemente associada às palavras do corpo "cabeça", "cérebro" ou "miolo". Aos poucos, o sema faculdade foi se ampliando, e atualmente refere-se também a processos realizados pelo corpo, nomeados pelas partes que os realizam – "pulmão", por exemplo, pode ser usado como sinônimo do processo / capacidade de respiração; e "boca" ou "garganta" podem ser usadas para indicar a faculdade / capacidade de falar [7-10].

7. Mas o Stanley tinha **cabeça** para dinheiro, o que eu nunca tive
8. Cafu e Mazinho constituem um meio-de-campo bom de **pulmão**, dinâmico e criativo
9. Posso ser mau de **boca** mas sou bom de **olho**
10. Existem **ouvidos** atentos para cada acorde, cada nota, cada timbre.

- A presença dos sentidos [sema="corpo:sentimento"] e [sema="corpo:opiniao"] indica que, no *Esqueleto*, consideramos sentimento e opinião duas classes distintas, ainda que frequentemente apareçam juntas. A separação foi uma decisão tomada desde o início do projeto. No entanto, a partir de um dado momento, uma das integrantes do projeto pôs em questão a separação. Em favor da unificação, a constatação de que pesquisadores de várias perspectivas chamaram a atenção para que as emoções e as opiniões (ou seja, a atividade do intelecto e do(s) sistemas associado(s) às emoções) estão indissociavelmente ligados. Por isso, na prática, os trabalhos de "análise de sentimentos" e de "garimpo de opiniões" constituem a mesma (sub)disciplina (veja-se MAIA; SANTOS, 2015). A expressão de atitudes, sentimentos e opiniões está intimamente relacionada com o corpo humano, por (pelo menos) três razões diferentes:

a. porque as emoções (pelo menos algumas, as consideradas por alguns psicólogos como básicas) provocam alterações fisiológicas

b. porque a descrição das emoções ou atitudes passa convencionalmente pela descrição da postura do seu dono (facial ou corporal): ficar de boca aberta, encolher os ombros...

c. porque é uma característica (aparentemente universal) das línguas, visto que estas se originaram em tempos sem conhecimento médico preciso, usar órgãos metonimicamente para sentimentos ou características de personalidade (maus fígados, bom coração...)

De toda forma, no *Esqueleto*, mantemos a distinção, e usamos [sema="corpo:opiniao"] para os casos em que o próprio termo ou expressão do corpo se refere a algo já com uma indicação clara de posicionamento [11]. O sentido [sema="corpo:sentimento"], por sua vez, é atribuído quando não há posicionamento ou julgamento explícitos [12]. Em ambos os casos, localizamos a presença do sentimento e / ou opinião apenas na palavra ou expressão, e não no enunciado completo.

11. Ele é um **pé de valsa**

12. Meu **coração** partiu quando ele se foi

- Distinções e sobreposições entre os sentidos [sema="corpo:lugar"], [sema="corpo:centralidade"] e [sema="corpo:parte"]. O sentido [sema="corpo:lugar"] é atribuído a palavras do corpo que se referem a um lugar, como [13-14]. No entanto, diversas outras ocorrências de “coração”, “seio” e “cérebro” em contextos como [15-16] em que está em jogo a ideia de centralidade, puseram em xeque a utilização do sema corpo:lugar, que comporta apenas uma das dimensões de “centro”, deixando de fora as dimensões de espacialidade e importância.

13. Bem no **coração** da floresta amazônica, a cidade é realmente uma bolha.
14. Ele nasceu em São Pedro Alfa, ao **pé** de Coimbra.
15. Deve, também, chegar ao **coração** da sociedade civil, desmascarando atitudes...
16. plantar a desordem no **seio** da família,

Segundo o dicionário Aulete Digital,[9] a palavra "centro" comporta os seguintes sentidos (entre vários outros):

> (a) Ponto que se situa no meio de uma superfície, de uma área ou de um espaço, tendo exata ou aproximadamente a mesma distância das extremidades ou limites
>
> (b) Localidade, região etc. de grande importância em relação às áreas vizinhas, onde se concentram atividades econômicas, administrativas e/ou políticas etc. (ger. especificadas pelo adj.): *A região sudeste é há anos o centro do país*.
>
> (c) Parte ou aspecto principal, mais importante, mais difícil etc.

A acepção (a) dá conta do sentido de "lugar"; a acepção (b) lida com as dimensões de importância e lugar; e a acepção (c) evidencia apenas a dimensão de importância. Para a ideia de "lugar", já dispomos do sema corpo:lugar. Para lidar com o sentido de "importância", criamos [sema="corpo:centralidade"]:

17. Seu departamento é o **cérebro** da operação
18. Sem revelar o **coração** do plano, Itamar rebatizou o conjunto de medidas...

[9] http://www.aulete.com.br

- O sentido [sema="corpo:parte"], além dos casos clássicos, como "pé da cama" e "costas da cadeira", também engloba partes de um todo que não precisa, necessariamente, ser entendido como um objeto. Inicialmente, chamava-se [sema="corpo:partedeobjeto"], mas a ocorrência de diversos casos de partedeobjeto em contextos em que o todo não se caracterizava como objeto [19-21] nos obrigou a repensar a descrição da classe. Com essa opção, unificamos os casos distintos de parte indiferenciada de algo e parte diferenciada, também em coerência com o princípio de evitar classificações demasiado granulares.

19. o **braço** europeu do Cinema Novo
20. para ser o **braço** de financiamento
21. Os **membros** da expedição reuniram-se

Os demais semas listados no Quadro 1 estão exemplificados na documentação.

3.2.1 Classificações múltiplas e vagueza na anotação

É possível que uma mesma palavra, em um mesmo contexto, admita duas ou mais classificações simultaneamente. Nesses casos, aceitamos todas as possibilidades. A palavra "dorso", por exemplo, corresponde à "parte de cima" em [22]. "Parte de cima", indica, simultaneamente, parte de algo (o que justifica o sentido [sema="corpo:parte"]) e localização espacial (o que justifica [sema="corpo:lugar"]). Esses casos, portanto, recebem ambos os semas. Os exemplos [23-26] também ilustram casos que receberam mais de uma classe semântica.

22. as tochas dos penitentes, e a procissão, estendida na linha de cumeadas, traçou uma estrada luminosa no **dorso** [sema="corpo:lugar_parte"] da montanha
23. uma foto de uma fantástica **boca** [sema="corpo:lugar_parte"] de caverna, que é a Gruta dos Brejões

24. e Yusef foram portanto, na tese de Milroye, os **cérebros** [sema="corpo:faculdade_centralidade"] do atentado
25. A primeira notícia refere uma ilha no **coração** [sema="corpo:lugar_centralidade"] da cidade do Porto.
26. Teve gente que **torceu o nariz** [sema="corpo:sentimento_opiniao"], mas eu gostei

3.2.2 Expressões com várias palavras e a relação entre identificação e classificação

É conhecida a participação de palavras do corpo em expressões, bem como a multiplicidade de nomes para o fenômeno de combinação com muitas palavras e a dificuldade quanto a uma abordagem consensual do que sejam tais expressões – nada muito diferente do construto linguístico "palavra", que pode ser estudado por diferentes ângulos (som, forma e sentido), que nem sempre serão convergentes. Em geral, o termo *locução* corresponde a uma perspectiva que privilegia a unidade em torno do sentido (MATTOSO CÂMARA JR., 1984), o termo "colocação" privilegia a convencionalidade / frequência de coocorrência, motivado sobretudo pela ideia de fluência verbal (SINCLAIR, 1991; MANNING; SCHÜTZE, 1999), e os termos "expressão idiomática*"* e "expressão cristalizada*"* privilegiam a opacidade semântica e a fixidez (XATARA *et al*., 2002, por exemplo). No *Esqueleto*, nomeamos tais combinações de EVP (expressões com várias palavras).[10]

Se consideramos a dimensão da convencionalidade, temos uma EVP como "conhecer=como=a=palma=da=mão", pois a vasta maioria dos usos de "como a palma da mão" acontece com "conhecer". Pelo critério do sentido único, podemos argumentar que "conhecer como a palma da mão" engloba duas ideias, com a especificação do sentido de "conhecer". Assim, pelo critério do sentido único, a EVP é apenas "como=a=palma=da=mão" e essa é a opção do *Esqueleto*. Considerando a enorme quantidade de expressões que incluem alguma palavra do

[10] O nome EVP também foi escolhido para que pudéssemos diferenciar nossas combinações das combinações já marcadas no *corpus* pelo PALAVRAS, que são chamadas de MWE.

léxico do corpo humano, limitamos o escopo das combinações apenas às expressões que correspondem a uma unidade de sentido.[11] Do mesmo modo, em "lembrar de cabeça", "contar de cabeça" e "saber de cabeça", consideramos apenas uma EVP, "de=cabeça". Nos exemplos [27-30] indicamos com = exemplos do que consideramos EVPs.

27. daqui ninguém **arreda=pé**
28. E tem gente que ainda **fica=na=mão**?
29. Uma delas seria o **braço=direito** do traficante Djavan
30. Ela é quem vai **ganhar=o=coração** do personagem Pedro

No entanto, o critério "sentido único" também é escorregadio em certos casos [31-34].

31. Tenha coração, não use peles.
32. Ele é bom, tem um coração enorme e me trata como uma princesa.
33. O que importa é que fulano tem bom coração.
34. Hoje, ter princípios é como ter bom coração.

É possível entender "ter coração" e variações como "ter (bons) sentimentos", e neste caso não haveria EVP, apenas a atribuição de [sema="corpo:sentimento"] a "coração", ou podemos entender como *ser* "generoso", e neste caso temos uma EVP do tipo [sema="corpo:opiniao"].

O mesmo ocorre em "doente da cabeça" [35]. Uma leitura possível atribui a "doente da cabeça" o sentido de louco – um sentido único, portanto. Nesse caso, "doente da cabeça" seria considerado EVP do tipo "outros". Outra leitura atribui apenas a "cabeça" um sentido não convencional – ficar doente das ideias/faculdades mentais ou algo parecido. Mantém-se o sentido de doente, com o sintagma seguinte

[11] O critério da unidade de sentido é do ponto de vista do falante da língua em análise. Assim, por exemplo, não consideramos "ponta dos dedos" uma EVP, ainda que a combinação corresponda a uma única palavra / sentido em inglês (*fingertips*).

especificando a doença. Nessa leitura, não temos a marcação de uma EVP, apenas a anotação de *cabeça* como [sema="corpo:faculdade"].

35. Inclusive ela ficou até doente da cabeça por causa dele.

Idealmente, em ambos os casos ("bom coração" e "doente da cabeça") deveríamos poder considerar todas as leituras possíveis, sem privilegiar alguma delas, mas isso tornaria tecnicamente o processo de anotação de expressões mais complexo do que já é. Nossa preferência, em todos esses casos [36-39], foi por considerar a leitura EVP, privilegiando o sintagma maior.

As EVPs, sobretudo verbais, com frequência admitem variações no formato. Nesses casos, atribuímos um lema único para os casos, isto é ver=com=bons=olhos em [36-37] e abrir=os=olhos=para em [38-39].

36. Ver com bons olhos

37. Ver com muito bons olhos

38. Ter olhos abertos para

39. Abrir os olhos para

Ainda outro ponto quanto à anotação das expressões no *Esqueleto* é a marcação em dois níveis: na EVP e nas palavras do corpo que compõem a EVP. Em [40], "dor de cotovelo*"* é anotada como uma EVP com o sentido de *sentimento*. Nessa EVP, a palavra do corpo, "cotovelo", também é anotada, e recebe [sema="corpo:outros"]. Já em [41], atribuímos o sentido de "sentimento" à EVP (um sentimento de calma; foi para casa acalmar-se), mas "cabeça" recebe [sema="corpo:faculdade"].

40. (...)s quatro últimos emprestam suas vozes em faixas que fazem o ouvinte entrar numa **dor de cotovelo** gostosa

41. (...) foi pra casa **esfriar a cabeça**

Por fim, lembramos que a anotação é sempre feita em contexto. Os trechos [42-44] ilustram três casos com a palavra "olho". As palavras ligadas por = correspondem a uma EVP.

42. O **olho** [sema="corpo:lugar_parte"] do furacão Gonzalo tocou a terra nas ilhas Bermudas.
43. No **olho** [sema="corpo:lugar_parte"]**=do=furacão** [sema="corpo:outros"], FHC faz que não é com ele.
44. Josias é que foi para o **olho** [sema="corpo:lugar"]**=da=rua**[sema="corpo:outros"]

A motivação da anotação em dupla camada é tornar as pesquisas com o léxico do corpo humano mais ricas. Assim, temos a possibilidade de facilitar a procura, em contexto, de expressões do corpo que envolvem um determinado sentido de uma palavra (por exemplo, todas as EVPs que contêm a palavra "boca"*)* ou de buscar as palavras do corpo que participam de expressões (pode-se querer investigar a existência de relação entre palavras do corpo e a classe semântica das EVPs em que participam, por exemplo), como será visto na seção 4.

Como sabemos que nossas interpretações não são definitivas, mas somos forçados, pela anotação, a tomar decisões, lembramos que todas as combinações consideradas EVPs estão listadas na página do projeto ou podem ser recuperadas em uma busca pelo AC/DC. Assim, para além da anotação em *corpus*, as EVPs identificadas estão listadas em um arquivo específico, em que para cada EVP, além da informação da classe semântica, é atribuída uma classe gramatical.

3.2.3 Estudo e validação das classes propostas

Ao longo do processo de anotação e estabilização dos semas, algumas classes nos pareceram especialmente delicadas, como ilustramos no início desta seção.

A fim de verificar o quão consensuais eram as análises propostas, por um lado, e o quão clara era a documentação, por outro, utilizamos a ferramenta Rêve (SANTOS *et al.*, 2015), desenvolvida justamente para permitir a revisão e comparação de diferentes análises. Assim, utilizamos o Rêve, não para aferir uma concordância entre anotadores, mas para, sobretudo, revisar e discutir os casos considerados complicados e, como consequência, fornecer uma medida de nossas escolhas e decisões.

Criamos dois exercícios de anotação distintos, com cerca de 100 frases cada, distribuídas de maneira equilibrada pelas diferentes classes. As frases foram escolhidas de acordo com o grau de dificuldade, isto é: escolhemos preferencialmente frases que, ao longo do processo de anotação, suscitaram dúvidas. Nem todas as frases eram frases difíceis, mas escolhemos uma amostra representativa das dúvidas e discordâncias que tivemos ao longo do processo de anotação. Além disso, evitamos escolher frases em que o sema seria atribuído a uma EVP, para evitar que a discordância na segmentação embaralhasse classificação de semas e identificação da EVP. Em alguns casos isso foi quase impossível, pois há semas que comparecem quase exclusivamente em expressões. Nesses casos, a opção foi por EVPs consensuais, que dificilmente dariam margem a dúvidas quanto à segmentação, como "torcer o nariz" ou "jogar na cara".

Em cada exercício, testamos grupos distintos de classes, mas que, segundo nossa experiência na anotação, continham sobreposições. Além disso, cada grupo incluiu também duas classes em comum, com frases repetidas.

No *Esqueleto* 1, testamos 100 frases distribuídas pelas classes corpo, sentimento, opinião, faculdade, centralidade e outros (além das possibilidades de classificação múltipla).

No *Esqueleto* 2, testamos 92 frases, distribuídas pelas classes corpo, lugar, parte de objeto, centralidade, posição, sentimento, e outros (além das possibilidades de classificação múltipla).

Não consideramos as classes "animal", "vegetal", "medida" e "grupo", por não terem gerado qualquer discussão ao longo do processo de anotação, o que seria indicativo de seu pouco potencial para discordância.

Os participantes do exercício tinham perfis distintos: participaram quatro anotadores: duas com experiência no *Esqueleto* e autoras deste artigo, e dois que tomaram contato com o *Esqueleto* pela primeira vez com o exercício.

Após a primeira rodada de anotação, e observando as discordâncias, refinamos algumas explicações e reanotamos. A alteração do nome "parte de objeto" para apenas "parte", por exemplo, foi consequência dessa discussão posterior.

Apenas o *Esqueleto* 2 passou por uma segunda rodada de anotação. A concordância foi de cerca de 87%, com 15 frases discordantes. Dessas 15, três eram claramente casos em que todas as alternativas / interpretações propostas eram possíveis. A anotação foi então revista e corrigida.

Dos 12 demais casos de discordância, todos envolviam a atribuição da categoria "outros" por alguma das anotadoras, o que, por um lado, revela que nossas intuições quanto à dificuldade dos exemplos selecionados estava correta e, por outro, sugere que os casos em questão eram pouco típicos para a sua classe, e bem poderiam compreender a classe "outros" (o que foi feito, e indicado na documentação). Também nesses casos de discordâncias (com "outros"), boa parte das frases envolvia a atribuição do sentido de *centralidade*.

Em termos gerais, e considerando a variedade de classes, ficamos satisfeitas com os resultados do exercício de anotação / revisão, que indicaram que houve consenso entre a classificação proposta: atingimos cerca de 87% de concordância, sendo que nos casos discordantes não houve propostas radicalmente diferentes, mas antes a atribuição da classe "curinga".

Quanto às concordâncias (79 frases), notamos que em 21 delas havia vagueza envolvida, isto é: se mais de uma classe havia sido proposta por uma participante, e as demais escolheram uma única classe, mas uma classe que já havia sido prevista na classificação múltipla, consideramos que não houve discordância. Nas restantes 58 frases, a concordância foi absoluta, mesmo quando havia mais de uma classe proposta.

3.2.4 Os grupos do corpo

Além das classes semânticas, às palavras do corpo também são atribuídos grupos, de acordo com a zona do corpo a que pertencem. Só recebem informação de grupo as palavras ou expressões do corpo do tipo [sema="corpo"], ou seja, não atribuímos grupo aos usos "especiais".

Atualmente, temos dez grupos, por ordem alfabética: Braços, Cabeça, Cabelos, Interno, Ossos, Pele, Pernas, Sangue, Sexual e Tronco.

É possível também que uma palavra pertença a mais de um grupo. A palavra ombro, por exemplo, pertence aos grupos Tronco e Braço.

A motivação para atribuir uma "zona" corporal proveio de já existir essa possibilidade para os campos cor e vestuário (não sendo, portanto, necessária nenhuma programação adicional) e por ser relativamente fácil essa atribuição (quando as palavras se referem ao corpo humano). Daí poderíamos oferecer aos usuários ainda outro tipo de procuras. Assim, pode-se investigar se existem áreas mais mencionadas que outras, ou áreas que têm mais ou menos usos vinculados ao corpo, por meio da informação dos grupos.

4 Resultados

Embora seja difícil considerar finalizado um projeto com essas dimensões, acreditamos que já temos material suficiente para divulgar resultados sobre a estruturação do léxico do corpo humano.

Revisamos dois *corpora* com características bastante distintas, entrevistas pessoais (Museu da Pessoa, 1.421.677 palavras e 93.479 frases) e obras literárias (OBras, 1.204.436 palavras e 38.011 frases), e os demais foram parcialmente revistos. Nas explorações a seguir, além do material integralmente revisto, relatamos também resultados que levam em conta o *corpus* Floresta (FREITAS *et al*. 2008), um *corpus* majoritariamente jornalístico, e o *corpus* Todos, a união de todos os *corpora* disponibilizados pelo AC/DC.

As Tabelas 1 e 2 apresentam a distribuição dos usos das palavras do corpo por *corpus* e por sentido. A diferença é que, na Tabela 1, não levamos em conta a classe gramatical das palavras do corpo, e na Tabela 2 consideramos apenas as palavras do corpo que são substantivos.

Os resultados são bastante parecidos em todos os campos, exceto em dois aspectos: quando consideramos apenas os substantivos (TABELA 2), a proporção de palavras do corpo mais que quadruplica, o que está de acordo com nossa estratégia de anotação, que privilegiou substantivos. Nesse contexto, a proporção de palavras do corpo no OBras é muito maior que nos demais *corpora*: cerca de 5% de todos os substantivos do *corpus* referem-se a palavras do corpo.

De maneira geral, os dados nas Tabelas 1 e 2 nos permitem inferir que a distribuição das palavras do corpo, nos *corpora*, segue a seguinte ordem: o material mais corporal é o Obras, seguido pelo Floresta, Museu da Pessoa e Todos. No entanto, é importante lembrar que tanto o Floresta como o Todos foram apenas parcialmente revistos.

Tabela 1 – Distribuição de palavras do corpo pelo total de palavras, sem considerar a classe gramatical

Corpus	Total de palavras do corpo	Distribuição dos semas do corpo	
		Tipo de sema	quantidade
OBras	**1,04%**	corpo	**85%** (10649 ocorrências)
		corpo:xxx	**15%** (1931 ocorrências)
MP	**0,17%**	corpo	**50,40%** (1218 ocorrências)
		corpo:xxx	**49,50%** (1198 ocorrências)
Floresta	**0,2%**	corpo	**80%** (12948 ocorrências)
		corpo:xxx	**20%** (3159 ocorrências)
Todos	**0,2%**	corpo	**87%** (2575969 ocorrências)
		corpo:xxx	**13%** (385794 ocorrências)

Tabela 2 – Distribuição de palavras do corpo pelo total de palavras, considerando apenas a classe gramatical dos substantivos

Corpus	Total de palavras do corpo	Distribuição dos semas do corpo	
		Tipo de sema	quantidade
OBras	**5%**	corpo	**85%** (10565 ocorrências)
		corpo:xxx	**15%** (1921 ocorrências)
MP	**1%**	corpo	**50,40%** (1170 ocorrências)

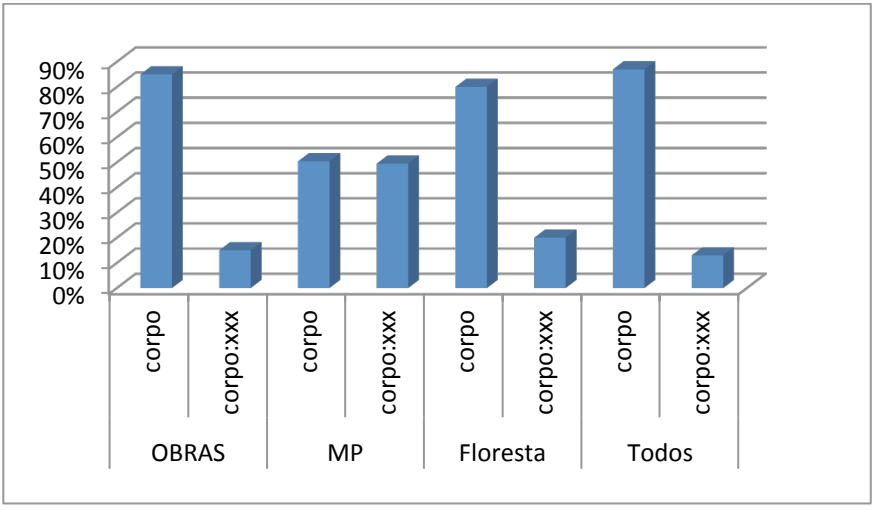
		corpo:xxx	**49,50%** (1146 ocorrências)
Floresta	**1,27%**	corpo	**80%** (12215 ocorrências)
		corpo:xxx	**20%** (3113 ocorrências)
Todos	**0,8%**	corpo	**87%** (2059946 ocorrências)
		corpo:xxx	**13%** (384880 ocorrências)

Outro dado interessante visível nas tabelas é a constância da proporção corpo / corpo:xxx, com cerca de 85% de palavras do corpo para usos corporais. A exceção é o MP, em que a distribuição corpo / corpo:xxx é equilibrada, com 50% das ocorrências para cada um dos usos. O Gráfico 1 apresenta apenas essa distribuição.

Considerando agora apenas o material completamente revisto, temos dois cenários bastante distintos. No OBras, apenas 15% das palavras do corpo não se vincula ao corpo, o que corrobora a ideia de forte presença de descrição nos textos literários analisados.[12] Já no Museu da Pessoa a situação é bem diferente: apenas metade das palavras do corpo se refere ao corpo. No Floresta e no Todos, a proporção de corpo / corpo:xxx é parecida: cerca de 80% das palavras do corpo fazem referência ao corpo, o que nos surpreende um pouco - esperávamos um uso maior de corpo:xxx. No entanto, esses *corpora* não passaram por uma revisão completa, e é possível que esse quadro se altere.

Gráfico 1 – Distribuição dos tipos de sentido das palavras do corpo por *corpus*

[12] Vale mencionar ainda que as obras com maior ocorrência de palavras do corpo são romances brasileiros do chamado período realista/naturalista, especificamente os romances *O Mulato* e *O Cortiço*, de Aluísio Azevedo, *Quincas Borba* e *Dom Casmurro*, de Machado de Assis, e o *Ateneu*, de Raul Pompéia.



O Gráfico 2 apresenta a distribuição dos semas corpo:xxx (pelo total de semas corpo:xxx), considerando apenas o material totalmente revisto.[13] Considerando apenas o OBras, vemos que o sentido mais frequente é o de sentimento – impulsionado pelos usos de "coração"–, seguido de outros e de posição, este último também típico de descrições.

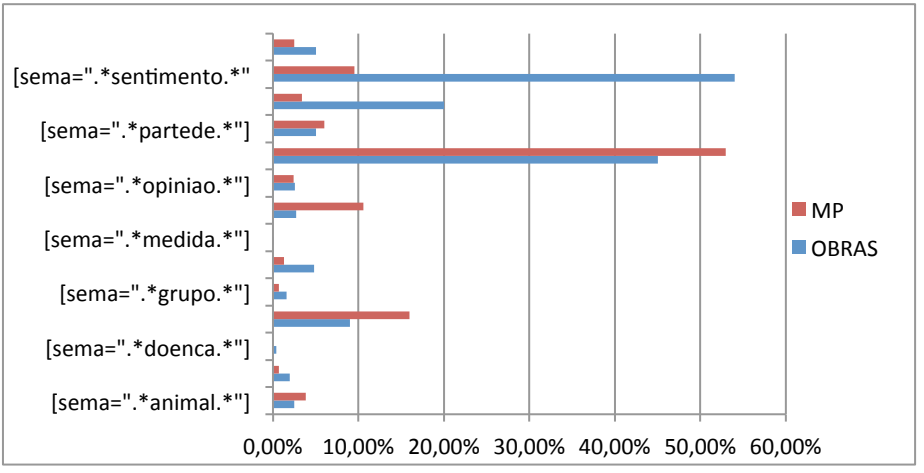
Chamou-nos a atenção o relativamente frequente uso de corpo:vegetal no OBras, e percebemos que a imensa maioria refere-se à palavra *tronco*. No entanto, dessas, boa parte se refere ao tronco em que os escravos eram castigados, o que aparece em obras como *Escrava Isaura*, *O Mulato* e *O Cortiço*. No Museu da Pessoa, o uso mais frequente é [sema="corpo:outros"], seguido de [sema="corpo:faculdade"]; [sema="corpo:movimento"] e [sema="corpo:sentimento"].

No Quadro 2, em uma abordagem qualitativa, apresentamos os lemas que tomam parte em alguns dos semas corpo:xxx. Para o quadro, consideramos o material do OBras, MP e também da Floresta. Como é

[13] Para simplificar, as palavras consideradas vagas entre várias classificações contam em cada uma delas, ou seja, se uma palavra estiver marcada [sema="corpo:opiniao_sentimento"], conta nesta tabela uma vez por corpo:opiniao e outra por corpo:sentimento. As expressões de procura utilizadas foram: [sema=".*movimento.*"], [sema=".*sentimento.*"] etc.

possível observar, há palavras do corpo especialmente maleáveis quanto ao sentido, que participam de todos os semas (ou quase todos), como "pé", "boca" e "mão". No Quadro 2, os lemas estão listados por ordem alfabética, e não por frequência. É importante notar também que, no Quadro 2, estamos considerando apenas os lemas, dissociados das EVPs de que fazem parte. Assim, por exemplo, "dente" integra a EVP "com unhas e dentes". A ideia do quadro é tão somente apresentar a variedade de palavras do corpo utilizada nos diferentes sentidos. É interessante perceber que a ideia de importância / centralidade, que normalmente associaríamos apenas a cabeça / cérebro, também pode estar associada ao coração – que normalmente associaríamos apenas ao sentimento – e ao umbigo.

Gráfico 2 – Distribuição dos semas corpo:xxx no OBras e no Museu da Pessoa



Especificamente quanto às expressões, e ilustrando as possibilidades da anotação em dupla camada, a Tabela 3 apresenta a distribuição dos sentidos das EVPs,[14] ou seja, sem considerar as palavras do corpo que as compõem.

[14] As expressões de procura foram [sema=".*corpoEVP"], [sema=".*corpo:sentimentoEVP"], etc.

Nos três *corpora*, expressões com sentidos que não conseguimos sistematizar – o que corresponde ao "outros" – são as mais frequentes, consistindo em cerca de metade dos sentidos das expressões. Comparando OBras e MP, expressões com o sentido de corpo são muito mais recorrentes no OBras, o que, novamente, é compatível com a ideia de que, neste material, as descrições físicas são bastante usuais. Outro dado interessante é que a distribuição dos sentidos das EVPs segue, de maneira geral, a distribuição dos sentidos total.

Nas Tabelas 4 e 5 apresentamos a distribuição das palavras do corpo nas EVP. Na Tabela 4, listamos, por frequência, as palavras do corpo que mais comparecem em expressões, mas mantendo o sentido de palavra do corpo, independentemente do sentido da expressão (e excluindo as expressões com sentido de corpo humano)[15]; na Tabela 5, fazemos o inverso: listamos as palavras do corpo que mais comparecem em expressões, mas perdem o sentido de corpo.[16]

Quadro 2 – Lista de lemas por semas considerando apenas OBras, MP e Floresta

SEMA	LEMAS
[sema=".*centralidade.*"]	cabeça; coração; cérebro; regaço; seio; umbigo
[sema=".*faculdade.*"]	boca; cabeça; coração; cérebro; língua; mão; nervo; olho; pulmão; orelha; ouvido
[sema=".*lugar.*"]	boca; coração; costas; estômago; face; fronte; olho; pé; seio
[sema=".*movimento.*"]	pé
[sema=".*opiniao.*"]	boca; barriga; cabeça; cara; coração; cotovelo, desmiolado; estômago; língua; mão; nervo; olho; osso; pé; saco
[sema=".*parte.*"]	boca; braço; cabeça; corpo; costas; dente; dorso; espádua; goela; membro; olho; peito; perna; punho; pé; seio

[15] Com a expressão <mwe> []* @[sema=".*corpo_.*EVP"] []* </mwe> conseguimos todas as expressões em que há uma palavra do corpo com o sentido de corpo. Como excluímos as expressões cujo sentido global é corpo, a expressão de busca usada foi <mwe> []* @[sema=".*corpo_.*EVP" & sema!=".*corpo_.*corpoEVP"] []* </mwe>.

[16] Para essa busca, a expressão usada foi <mwe> []* @[sema=".*EVP" & sema!=".*corpo_.*EVP" & sema!=".*corpo_.*corpoEVP"] []* </mwe>.

[sema=".*posicao.*"]	braço; cabeça; cara; costas; face; ilharga; joelho; punho; pé; punho
[sema=".*sentimento.*"]	barriga; boca; cabelo; cabeça; cara; coração; corpo; costas; cotovelo; dedo; dente; estômago; garganta; mão; nariz; nervo; olho; ombro; orelha; ouvido; peito; pele; pé; queixo; sangue; sobrolho; tripa; tropinha; unha; venta
[sema=".*outros.*"][17]	artéria; barriga; boca; braço; busto; cabeça; cara; carne; celular; coração; corpo; costas; célula; dedo; dente; embrionário; embrião; esqueleto; face; franja; língua; manual; mão; olho; osso; ouvido; palma; peito; perna; pulso; pé; rabo; sangue; seio; tronco; umbigo; unha; veia

Tabela 3 – Distribuição das expressões com várias palavras (EVP) nos corpos, sobre o total de EVP

Classificação	**OBras**	**MP**	**todos**
corpo	**8,8%** (76)	2% (11)	**14%** (23354)
corpo:animal	0	0,2% (1)	0,1% (155)
corpo:doenca	0,2% (2)	0 (0)	0,2% (458)
corpo:faculdade	0	0,4% (2)	0,04% (76)
corpo:lugar	0,3% (3)	0,2% (1)	0,07% (123)
corpo:movimento	3% (26)	**12%** (64)	1% (1619)
corpo:opiniao	2,9% (25)	4% (24)	1,2% (2111)
corpo:outros	**45%** (385)	**59%** (301)	**60%** (99275)
corpo:posicao	**28%** (241)	**8%** (40)	**15%** (25378)
corpo:sentimento	**10%** (85)	**11%** (59)	**5%** (9416)

É possível constatar a variação nos lemas. "Cabeça", por exemplo, quando comparece em expressões, costuma a perder o sentido do corpo. "Pé", por outro lado, parece sofrer o efeito inverso: tende a manter o seu sentido de corpo, mesmo quando em expressões. Já as

[17] Como ilustração, e considerando a variedade de lemas em cada corpus consideramos apenas as 25 primeiras ocorrências de cada corpus, sugerindo ao leitor interessado repetir a busca.

palavras "mão" e "olho" participam de expressões de ambas as maneiras.[18]

Tabela 4 – Distribuição das palavras do corpo que mantêm o sentido corporal em expressões

MP		OBras		Floresta	
pé	38,8% (97)	pé	35% (159)	mão	32% (281)
mão	26,4% (66)	mão	22% (100)	pé	21% (189)
cara	7,6% (19)	olho	12% (54)	olho	7% (61)
olho	4,4% (11)	braço	5% (26)	cara	6% (57)
boca	3,2% (8)	boca	4% (21)	palma	3% (33)
braço	2,8% (7)	cara	4% (20)	saco	3% (28)
corpo	2,4% (6)	cabeça	3% (13)	boca	2% (21)
saco	2% (5)	palma	2% (9)	costas	2% (20)
joelho	2% (5)	costas	1% (8)	língua	2% (18)
pele	1,6% (4)	língua	1% (7)	cabeça	2% (17)

Tabela 5 – Distribuição das palavras do corpo que não mantêm o sentido corporal em expressões

MP		OBras		Floresta	
cabeça	27% (63)	joelho	20% (56)	cabeça	17% (109)
mão	19% (45)	olho	9% (25)	olho	16% (101)
olho	11% (26)	mão	8% (22)	mão	14% (90)
cara	5% (12)	cabeça	7% (21)	cara	10% (62)
boca	5% (12)	ouvido	7% (20)	ouvido	8% (53)
coração	5% (11)	coração	6% (17)	corpo	5% (30)

[18] Notamos que a anotação de expressões em dupla camada permite essas e ainda outras possibilidades de busca na interface. Para encontrar expressões que contêm uma palavra específica, por exemplo, mão, a expressão deve ser <mwe> []* @[lema="mão"] []* </mwe>, que devolverá EVPs como *deitar a mão*, *mão na massa*, *de mão beijada*, de *segunda mão*, *abrir mão* etc. Para especificar não apenas a palavra mas também o sentido da EVP, a expressão deve ser <mwe> []* @[lema="mão" & sema=".*sentimento.*"] []* </mwe>.

[19] Expressão de pesquisa: [sema=".*corpo.*"] e pedido de Distribuição de grupo

costas	2% (6)	dente	5% (16)	pé	4% (29)
pé	2% (5)	punho	4% (12)	boca	4% (27)
corpo	1% (4)	boca	4% (12)	punho	3% (20)
sangue	1% (4)	ilharga	3% (10)	dedo	2% (12)
peito	1% (4)	costas	3% (9)	coração	1% (10)

Levando em conta agora apenas as palavras do corpo em seus usos corporais, a Tabela 6 ilustra, para MP, OBras e Floresta, a distribuição de lemas do corpo sobre o total de palavras do corpo (não consideramos corpo:xxx). Consideramos apenas os 10 lemas mais frequentes de cada *corpus*.

Há diferenças significativas entre as ordenações, sobretudo entre um *corpus* literário e um de linguagem oral (que, além disso, são separados por um século: os escritores do OBras são fundamentalmente do século 19; enquanto os entrevistados são todos dos séculos 20-21). Debruçando-nos agora sobre algumas das diferenças mais gritantes: a palavra "olho" (ou "olhos") é muito mais usada no OBras devido, novamente, à descrição de personagens em textos literários – os "olhos" (e também as "mãos") atuam como elemento de descrição física (olhos castanhos), mas também psicológicas (olhos astutos; mãos trêmulas). No Floresta, um *corpus* majoritariamente jornalístico, a ampla frequência da palavra "corpo" refere-se aos óbitos, que optamos por deixar anotado como [sema="corpo"]. Nos três, chama a atenção o grande comparecimento de "mão".

Tabela 6 – Distribuição dos lemas de corpo por *corpus*

MP		OBras		Floresta	
mão	14%	olho	12%	mão	10%
pé	10%	mão	10%	corpo	8%
cabeça	6%	cabeça	6%	olho	7%
perna	5%	coração	5%	cabeça	7%
cabelo	5%	braço	5%	pé	5%
braço	4%	corpo	4%	cara	4%
costas	3%	pé	4%	coração	4%
olho	3%	rosto	3%	cabelo	3%
cara	3%	boca	3%	boca	3%
mama	3%	cabelo	2,6%	braço	2,6%

Finalmente, observamos a distribuição por grupos das palavras de corpo que identificamos, na Tabela 7. Convém indicar que palavras referentes a algo que não diz respeito a uma parte do corpo, como a própria palavra "corpo", estão marcadas com zona (grupo) Não especificada.

Inesperadamente, o grupo Interno é extremamente frequente, na mesma ordem de grandeza que a Cabeça, e o Tronco tem quase o dobro de ocorrências que os membros (quer Braço quer Perna), enquanto que Pele e Cabelo são dos menos usados. Será preciso estudar com mais cuidado todos os gêneros que constituem o corpo total para podermos compreender melhor a causa desses resultados quantitativos. Para um estudo inicial dos gêneros em questão, veja-se também Santos (no prelo).

Tabela 7 – Distribuição das palavras referentes a corpo humano por grupos (zonas)[19]

Grupo	**OBras**	**MP**	**todos**
Braço	2330	345	355828
Cabeça	5130	392	739374
Cabelo	655	84	61770
Interno	1414	173	744269
Osso	94	38	97741
Não especificada	592	49	341575
Pele	90	15	73504
Perna	1331	350	204311
Sangue	262	29	103380
Sexual	81	45	64653
Tronco	1795	238	549171

Por fim, o último levantamento tem como objetivo explorar a ideia do corpo (e do corpo, na língua) como mais uma maneira de

[19] Expressão de pesquisa: [sema=".*corpo.*"] e pedido de Distribuição de grupo

elaborar identificações. Na Figura 1 ilustramos as caracterizações (adjetivos) atribuídas às palavras do corpo, apenas quando estas se referem ao corpo humano.[21] Destacam-se – as palavras maiores – dimensões físicas: espaço (direito, esquerdo), tamanho (baixo, enorme, comprido, curto, pequeno, grande) e cor (branco, vermelho, negro, moreno, negro, castanho, preto). O passo seguinte, que não fizemos aqui, é encontrar, para cada atributo / classificação, as partes do corpo caracterizadas. Nesta breve exploração, consideramos apenas o *corpus* OBras.

Figura 1 – Adjetivos que caracterizam as palavras com sentido de corpo no OBras

[21] A expressão de busca utilizada foi [sema="corpo|corpo_corpoEVP"] @[pos="ADJ.*" & func="<PRED|N<.*"], com pedido de distribuição por lemas.

5 Considerações finais

Apresentamos aqui um estudo voltado ao léxico do corpo humano em língua portuguesa, que tomou como metodologia a anotação de *corpus* e que oferece, como resultado, não apenas análises e interpretações mas também um vasto material, publicamente disponível, para outras explorações.[22]

No decorrer do processo de anotação, criamos classes e categorizamos fenômenos, sempre de um ponto de vista que privilegia o sentido. Ao longo deste texto, explicitamos nossas decisões e motivações. Não temos a ilusão de interpretações definitivas, mas acreditamos ter sido suficientemente explícitos em nossas explicações, permitindo a continuação do debate, o que é facilitado pelo compartilhamento de um mesmo material – especificamente, do mesmo material anotado. Afinal, o corpo humano, e o modo pelo qual nos referimos a ele, interessa a estudos que se alinham a diferentes perspectivas teóricas.

À pergunta inicial do projeto: como se distribui, quanto ao sentido, o léxico do corpo humano em português?, oferecemos a resposta provisória de 13 diferentes sentidos sistemáticos, excetuando-se o próprio sentido corpo e os sentidos que não conseguimos sistematizar, indicados com "outros".

Quanto ao vínculo entre sentimento, opinião e corpo, mostramos que, sem dúvida, é forte e presente na língua portuguesa (como nas demais línguas), mas consideramos as possibilidades dos demais sentidos um rico caminho para explorações futuras, assim como a relação entre cada um dos sentidos e as palavras a eles associadas.

Podemos, então, nos debruçar sobre a segunda pergunta: como falamos do corpo humano em português?, o que só pode ser feito (ou é imensamente facilitado) depois de termos separado as palavras / sentidos do "corpo" dos demais sentidos que também se utilizam das palavras do corpo. Tomando por base os resultados do MP, por exemplo, vemos que

[22] O estudo apresentado aqui foi feito com a versão 2.1 do Obras e versão 5.7 do MP. Pode ser que esse material se amplie no futuro e, com isso, os números precisem ser atualizados.

essa primeira separação é fundamental, visto metade das palavras e expressões do corpo referirem a outros tipos de sentido.

Dissemos no início deste artigo que anotar é uma forma de estudo. A anotação motivada pelo sentido, atividade que precisa compatibilizar sentido e forma, nos força a delimitações que não nos são "naturais", mas antes necessidades decorrentes da própria atividade de anotação. No caso das expressões, altamente frequentes no estudo do léxico do corpo humano, o estudo da anotação opera como uma lente de aumento para a multiplicidade de nuanças que devem ser levadas em conta quando o que está em jogo é o processamento automático da língua e a sua descrição.

Quais as palavras do corpo mais frequentes em expressões? Leitão de Almeida *et al*. (2009) referem, em sua pesquisa, "cabeça", "mão" e "pé". Nossos resultados vão na mesma direção, ainda que apresentem uma variedade de outras palavras também comuns em expressões.

Todo o material criado, bem como os léxicos (listas de palavras e regras) utilizados na anotação, estão disponíveis na página do projeto.[23] Ao longo do artigo, indicamos nas notas de rodapé as expressões que utilizamos na interface de busca, com o intuito de motivar a/os leitoras/es a realizar novas pesquisas e / ou análises. Assim, os interessados nos variados sentido do corpo na língua têm à disposição um material rico e preciso, bastando para isso fazer uso dos diferentes tipos de sema. Para os especialmente interessados nos usos metafóricos, não sistemáticos, a grande quantidade de "outros" indica que há muito ainda por explorar. Para os interessados em como, na língua, identificamos pessoas, quer do ponto de vista físico, quer psicológico, o espaço de exploração é muito vasto.

23 http://www.linguateca.pt/acesso/Esqueleto/

Referências

ARROJO, Rosemary; RAJAGOPALAN, Kanavillil. Noção de literalidade: metáfora primordial. In: ARROJO, Rosemary (Org.). *O signo desconstruído*. São Paulo: Pontes, 1992. p. 47-56.

BAPTISTA, Jorge. Body-part nouns and local grammars. In: DISTER, Anne. (Ed.). *Révue d'Informatique et Statistiques en Sciences Humaines*, v. 36, p. 53-66, 2000.

BICK, Eckhard. *The Parsing System "Palavras"*: Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework. Aarhus, Denmark: Aarhus University Press, 2000.

BUTLER, Judith. Corpos que pesam. In: LOURO, Guacira Lopes (Org). *O corpo educado*: Pedagogias da sexualidade. Belo Horizonte: Autêntica, 2000. p.110-125.

CARVALHO, Paula; SARMENTO, Luis; TEIXEIRA, Jorge; SILVA, Mário J. Liars and Saviors in a Sentiment Annotated Corpus of Comments to Political Debates. In*:* ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS: HUMAN LANGUAGE TECHNOLOGIES, 49, 2011, Stroudsburg, PA, USA. *Proceedings…,* v. 2. Stroudsburg: Association for Computational Linguistics, 2011. p. 564-568.

COSTA, Luís; SANTOS, Diana; ROCHA, Paulo. Estudando o português tal como é usado: o serviço AC/DC. In: BRAZILIAN SYMPOSIUM IN INFORMATION AND HUMAN LANGUAGE TECHNOLOGY – STIL, 7, 2009, São Carlos. *Proceedings…* São Carlos: Universidade de São Paulo, 2009.

DODDINGTON, George; MITCHELL, Alexis; PRZYBOCKI, Mark; RAMSHAW, Lance; STRASSEL, Stephanie; WEISCHEDEL, Ralph The Automatic Content Extraction (ACE) Program: Tasks, Data, and Evaluation. In: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION, 4, 2004, Lisboa. *Proceedings of LREC'2004*. Ed. M. T. Lino; M. F. Xavier; F. Ferreira; R. Costa; R. Silva, Lisboa, Portugal: Universidade Nova de Lisboa, 2004. p. 837-40.

DURAN, Magali S.; ALUÍSIO, Sandra M. Propbank-Br: a Brazilian Treebank annotated with semantic role labels. In: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION, 8, 2012, Istanbul. *Proceedings of LREC'2012*. Ed.: N. Calzolari; K. Choukri; T. Declerck; M. U. Dogan; B. Maegaard; J. Mariani; J. Odijk; A. Moreno; S. Piperidis. Istanbul: Lüfti Kirdar Convention & Exhibition Centre, 2012. p. 1862-1867.

ECO, Umberto. *Semiótica e filosofia da linguagem*. São Paulo: Ática, 1991.

FREITAS, Cláudia. *Esqueleto: anotação das palavras do corpo humano*. Primeira edição: 15 nov. 2013. Disponível em: <http://www.linguateca.pt/acesso/Esqueleto/Esqueleto.html>

FREITAS, Cláudia; MOTTA, Eduardo; MILIDIÚ, Ruy L.; CÉSAR, Juliana. Sparkling Vampire... lol! Annotating Opinions in a Book Review Corpus. In: ALUÍSIO, Sandra; TAGNIN, Stella E. (Ed.). *New Language Technologies and Linguistic Research*: A Two-Way Road. Cambridge Scholars Publishing, 2014, p. 128-146.

FREITAS, Cláudia; SANTOS, Diana; SILVA, Rosario. (2012). Corpos e cores: colorindo a descrição da língua portuguesa. In: ENCONTRO DE LINGUÍSTICA DE CORPUS: ASPETOS METODOLÓGICOS DOS ESTUDOS DE CORPORA, 10, 2012, Belo Horizonte. *Anais*... Ed.: D. P. Dutra; H. R. Mello. Belo Horizonte: Faculdade de Letras da UFMG, 2012. p. 76-99.

FREITAS, Cláudia, ROCHA, Paulo; BICK, Eckhard. Um mundo novo na Floresta Sintá(c)tica – o treebank para Português. *Calidoscópio*, v. 6.3, p. 142-148, 2008. DOI: <http://dx.doi.org/10.4013/cld.20083.03>

GREINER, Christine. *O Corpo* – Pistas para estudos indisciplinares. São Paulo: Ed. Annablume, 2005.

KINGSBURY, Paul; PALMER, Martha, MARCUS, Mitch. Adding Semantic Annotation to the Penn TreeBank. In: THE HUMAN LANGUAGE TECHNOLOGY CONFERENCE, 2002, San Diego. *Proceedings*... San Diego, CA, USA, 2002.

KURIYAMA, Shigehisa. *The Expressiveness of the Body, and the Divergence of Greek and Chinese Medicine*. New York: Zone Books, 1999.

LAKOFF, George; JOHNSON, Mark. *Metaphors we Live By*, Chicago: The University of Chicago Press. 1980.

LEITÃO DE ALMEIDA, M. L. *et al*. (Org.) A hipótese de corporificação da categorização e do léxico. In: LEITÃO DE ALMEIDA, Maria Lúcia *et al*. (org.). *Linguística Cognitiva em foco: morfologia e semântica do português*. Rio de Janeiro: Publit, 2009, p. 187-204.

MAALEJ, Zouheir A.; YU, Ning (Ed.). *Embodiment via Body Parts*: Studies from Various Languages and Cultures. *Human Cognitive Processing*, v. 31. Amsterdam and Philadelphia: John Benjamins, 2011. DOI: <http://dx.doi.org/10.1075/hcp.31>

MAIA, Belinda; SANTOS, Diana. Emotions in Language, PhD-course, University of Oslo, 1-5 Jun. 2015. (Mimeo)

MANNING, Christopher; SCHÜTZE, Hinrich. *Foundations of Statistical natural language processing*. Cambridge, MA: The MIT Press, 1999.

MARTINS, Helena. Wittgenstein, the body, its metaphors. *D.E.L.T.A.*, São Paulo, v. 26, p. 479 – 501, 2010. (Edição Especial)

MATTOSO CÂMARA JR, J. *Dicionário de Linguística e Gramática*. Editora Vozes, 1984.

MOTA, Cristina; SANTOS, Diana (Ed.). *Desafios na avaliação conjunta do reconhecimento de entidades mencionadas*: o segundo HAREM. Linguateca, 2008. Disponível em: <http://www.linguateca.pt/LivroSegundoHAREM/>

MOTA, Cristina; SANTOS, Diana. Emotions in natural language: a broad-coverage perspective. *Linguateca*, Jan. 2015. Disponível em: <http://www.linguateca.pt/acesso/EmotionsBC.pdf>.

ORSI, Vivian; ZAVAGLIA, Claudia. Expressões idiomáticas interditas: uma proposta lexicográfica bilíngue. *Linguasagem Revista Eletrônica de*

Popularização Científica em Ciências da Linguagem, v. 11, p. 1-17, 2010.

PALMER, Martha; GILDEA, Dan; KINGSBURY, Paul. The Proposition Bank: A Corpus Annotated with Semantic Roles. *Computational Linguistics Journal*, v. 31, n. 1, p. 71-106, 2005.

PIAO, S.; ARCHER, D.; MUDRAYA, O.; RAYSON, P.; GARSIDE, R.; McENERY, A.; WILSON, A. A large semantic lexicon for corpus annotation. In: CORPUS LINGUISTICS CONFERENCE, 2005, Birmingham. *Proceedings*... Series on-line e-journal, v. 1, n. 1, Birmingham, UK, July 14-17, 2006. ISSN 1747-9398.

PINTO, Joana P. Conexões teóricas entre performatividade, corpo e identidades. *D.E.L.T.A.*, São Paulo, v. 23, n. 1, p. 1-26, 2007.

RASO, Tomaso; MELLO, Heliana. (Org.). C-ORAL-BRASIL I. *Corpus de referência do português brasileiro falado informal*. Belo Horizonte: Editora UFMG, 2012.

SANTOS, Diana. Podemos contar com as contas? In: ALUÍSIO, Sandra Maria; TAGNIN, Stella E. O. (Ed.). *New Language Technologies and Linguistic Research*: A Two-Way Road. Cambridge Scholars Publishing, 2014, p. 194-213.

SANTOS, Diana. Comparando corpos orais (transcritos) e escritos na Gramateca. In: PARLER LES LANGUES ROMANES / PARLARE LE LINGUE ROMANZE / HABLAR LAS LENGUAS ROMANCES / FALANDO LÍNGUAS ROMÂNICAS. Napoli. *Atti del convegno internazionale GSCP 2014*. Ed.: Camilla Bardel; Anna De Meo. Napoli: Università di Napoli L'Orientale, Il Torcoliere, no prelo.

SANTOS, Diana. Gramateca: corpus-based grammar of Portuguese. In: In: INTERNATIONAL CONFERENCE – PROPOR – COMPUTATIONAL PROCESSING OF THE PORTUGUESE LANGUAGE, 11, Oct. 6-8, 2014. *Proceedings*... Ed. J. Baptista; N. Mamede; S. Candeias; I. Paraboni; T. Pardo; Maria das Graças V. Nunes. São Carlos: Springer, Heidelberg, 2014. p. 214-219.

SANTOS, Diana; MOTA, Cristina. A admiração à luz dos corpos. In: SIMÕES, A.; BARREIRO, A.; SANTOS, Diana; SOUSA-SILVA, R.; TAGNIN, Stella E. O. (Ed.) *Linguística, Informática e Tradução*: Mundos que se Cruzam. Homenagem a Belinda Maia, *OSLa*, v. 7, n. 1, p. 57-77, 2015.

SANTOS, Diana; SILVA, Rosário; FREITAS, Cláudia. Pluralidades na cor: contrastando a língua do Brasil e de Portugal. In: SOARES DA SILVA, Augusto; TORRES, Amadeu; GONÇALVES, Miguel. (Ed.). *Línguas Pluricêntricas*: Variação Linguística e Dimensões Sociocognitivas. [Pluricentric Languages: Linguistic Variation and Sociocognitive Dimensions.] Braga: Aletheia, Publicações da Faculdade de Filosofia da Universidade Católica Portuguesa, 2011. p. 555-572.

SANTOS, Diana; MARQUES, R; FREITAS, Cláudia; SIMÕES, A.; MOTA, Cristina. Comparando anotações linguísticas na Gramateca: filosofia, ferramentas e exemplos. *Domínios de Lingu@gem*, v. 9, n. 3, 2015.

SANTOS, Diana; MOTA, Cristina. Experiments in human-computer cooperation for the semantic annotation of Portuguese corpora. In: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION (LREC), 7, 2010, Valletta, Malta. *Proceedings*... Ed. N. Calzolari; K. Choukri; B. Maegaard; J. Mariani; J. Odjik; S. Piperidis; M. Rosner; D. Tapias (Ed.). Valletta, Malta: Mediterranean Conference Centre, 2010. p. 1437-1444.

SILVA, Rosário; SANTOS, Diana. *Arco-íris*: notas sobre a anotação do campo semântico da cor em português. 16 ago. 2012. Disponível em: <http://www.linguateca.pt/acesso/ArcoIris.pdf>.

SINCLAIR, J. *Corpus, concordance, collocation*: Describing English language. Oxford: Oxford University Press, 1991.

SOARES DA SILVA, Augusto. Metáfora, Metonímia e Léxico. *Diacrítica*, v. 7, p. 313-330, 1992.

VALE, Oto. As opiniões nas expressões e a expressão da opinião. In: LAPORTE, Éric; SMARSARO, Aucione; VALE, Oto. (Org.). *Dialogar*

é preciso: Linguística para processamento de línguas. Vitória: PPGEL/UFES, 2013. p. 259-267.

VILLENA ROMÁN, J.; GARCÍA MORERA, J.; MARÍNEZ CÁMARA, E.; JIMÉNEZ ZAFRA, S. M. TASS 2014 – The Challenge of Aspect-based Sentiment Analysis. *Procesamiento del Lenguage Natural*, v. 54, p. 61-68, marzo 2015.

WIEBE, Janyce; WILSON, Theresa; CARDIE, Claire. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, v. 39, n. 2-3, p. 165-210, 2005.

XATARA, Claudia M.; RIVA, Huelinton. C.; RIOS, Tatiane H. C. As dificuldades na tradução de idiomatismos. *Cadernos de Tradução*, Florianópolis, NUT, v. 8, p. 183-194, 2002.