

Esqueleto

Investigação do léxico do corpo humano com base em corpus

Cláudia Freitas (PUC-Rio e Linguateca)

Diana Santos (Univ. Oslo e Linguateca)

Raphaela Santos (PUC-Rio)

Heidi Jansen (Univ. Oslo)

Cristina Mota (Linguateca)

Bruno Carriço (PUC-Rio)



Gramateca



Motivação

- Relação entre opiniões e palavras do corpo (Freitas et al., 2014; 2012)
 - O livro é um *soco no estômago/tapa na cara..*
- Dependência cultural das expressões
 - Ensino de português a estrangeiros
- Alta frequência do léxico do corpo na língua
 - MP e Floresta: cerca de 1% de todos os **N** são corpo
 - OBRAS: 5% de todos os **N** são corpo
- Muitos usos não físicos

Objetivo

- Estudar o léxico do corpo humano na LP a partir da sua ocorrência *em corpora*
 - Quando usamos palavras do léxico do corpo humano, de que mais falamos (e que não é o nosso corpo)?
 - Como se descreve a aparência física em português?



Gramateca

Com a mão_[corpo:outros]=na=massa: Qual corpus?

- **Corpus:** projeto AC/DC
 - Livre acesso
 - Documentado
 - Anotação morfossintática prévia
 - Anotação semântica concluída/em andamento
 - Cor; emoções; vestuário
 - Tamanho e variedade

“Genre/register/mode” assigned: written

Political newspaper Local newspaper

Global newspaper Book reviews

by students Thematic newspaper

Thematic mailinglist Blogs Magazines/journals

Cookbook **Web pages** (Mail) spam Encyclopedia

Unedited local newspaper Legal text

Literary works Letters to the editor

Translations Essay Academic writing

Technical

Santos, 2014a

Acesso a corpos de português: Projeto AC/DC

[Linguateca](#)

O projeto AC/DC (Acesso a corpos/Disponibilização de corpos), iniciado em 1999, surgiu da necessidade de juntar de corpos, o [Open CWB](#) (versão nova do *IMS corpus workbench*), para a qual desenvolvemos esta interface.

Desde 2000, a anotação dos corpos tem sido feita automaticamente pelo [PALAVRAS](#) de Eckhard Bick, e convertida para o formato [XML](#).

Uma descrição quantitativa inicial dos corpos servidos presentemente pelo AC/DC encontra-se na tabela abaixo. Para uma [série de serviços especializados](#) para consultar os corpos de forma mais complexa.) Se é a primeira vez que visita esta página, clique aqui para ir para a página de início.

Breve descrição dos corpos

Corpo	Tamanho (unidades)	Tamanho (palavras)	Tamanho (frases)	Variante(s)	
AmostRA-NILC	128.203	98.633	4.931	BR	AmostRA-NILC
ANCIB	1.707.758	1.257.109	83.509	BR	Correio electrónico de Coimbra
Avante!	8.191.783	6.762.503	212.183	PT	Semanário político Avante!
Corpus Brasileiro	1.203.011.797	977.944.871	43.755.641	BR	Corpus Brasileiro, um milhão de palavras
CD HAREM	290.001	225.766	12.558	PT BR	Colecção dourada do CD Harem
CETEMPúblico	240.470.163	191.274.451	8.128.620	PT	Jornal PÚBLICO, div. de Coimbra
CHAVE	124.095.306	97.884.763	4.682.363	PT BR	Jornais PÚBLICO e FOLHA
Colonia	6.434.590	5.027.890	216.857	PT BR	Obras dos séculos XV e XVI
CONDIVport	7.159.365	5.576.544	318.776	PT BR	Jornais desportivos e de notícias
CoNE	922.984	675.462	31.361	PT BR	Mensagens de correio electrónico
C-Oral-Brasil	464.501	263.396	30.647	BR	C-Oral-Brasil, português brasileiro
DiaCLAV	7.723.448	6.551.019	210.373	PT	Diário de Coimbra, Diálogo
Diáspora TL-PT	29.365	21.855	1.038	TL	Diáspora TL-PT, entre outros
ECI-EBR	922.510	723.995	44.381	BR	Texto do corpo Borba
ECI-EE	32.133	27.130	839	PT	Texto de chamada do corpo
ENPCPUB (parte em português)	93.164	72.374	4.371	PT BR	Literatura traduzida de Portugal
Floresta	7.252.611	6.046.536	327.050	PT BR	Floresta Sintá(c)tica
FrasesPB	23.355	19.162	653	BR	Frases em português brasileiro
FrasesPP	20.222	16.236	676	PT	Frases em português europeu
Moçambula	71.398	58.264	2.321	MO	Textos de leitores de jornais moçambicanos
Museu da Pessoa	1.837.113	1.421.396	93.466	PT BR	Entrevistas realizadas pelo Museu da Pessoa
Natura/Minho	2.208.576	1.748.636	69.099	PT	Jornal regional Diário do Minho, antes da revisão
OBras	1.522.320	1.201.787	41.536	BR	Obras brasileiras
ReLi	196.773	151.062	9.121	BR	ReLi, corpo de resenhas de livros
NILC/São Carlos	42.915.206	32.461.799	1.988.621	BR	Texto do corpo NILC, contendo maioritariamente texto jornalístico, mas também cartas comerciais e textos didácticos
todos juntos	1.570.432.753	1.267.976.310	57.036.509	todas	Todos os corpos juntos
Tycho Brahe	2.792.351	2.391.744	77.555	PT BR	Texto do corpo Tycho Brahe
Vercial	20.608.290	14.712.121	950.980	PT	Clássicos da literatura portuguesa, séculos XVI a XX
<i>Total (incluindo repetições)</i>	<i>3.251.558.039</i>	<i>2.622.592.814</i>	<i>118.336.035</i>	PT BR	<i>todos os corpos</i>

Projeto AC/DC: corpo Museu da Pessoa

[AC/DC](#) : [Linguateca](#)

O corpus **Museu da Pessoa** é um corpus de cento e sete entrevistas transcritas pelo [Núcleo de Estudos em Linguística](#)

Procurar:

OK

Resultado:

- Concordância
- Distribuição das formas ([word](#))
- Distribuição dos lemas ([lema](#))
- Distribuição da categoria gramatical (PoS) ([pos](#))
- Distribuição do tempo verbal e/ou do caso pronominal ([temcagr](#))
- Distribuição de pessoa e/ou número ([pessnum](#))
- Distribuição do género morfológico ([gen](#))
- Distribuição da função sintáctica ([func](#))
- Distribuição por entrevista ([ent](#))
- Distribuição por variante do português ([variante](#))
- Distribuição por campo semântico ([sema](#))
- Distribuição por grupo (de cor, roupa, etc.) ([grupo](#))

Opções

- Resultados por ordem alfabética (só distribuições)
- Ignorar maiúsculas/minúsculas (não admite parâmetros)
- Mostrar texto original nas correções

Amostra aleatória de linhas.

Com a mão_[corpo:outros]=na=massa: Como anotar?

- Palavras/expressões do corpo humano que **se referem ao corpo humano vs “resto”** (palavras e expressões do corpo humano que se referem a outras coisas)
 - “*Resto*”: **subclasses** que organizam as palavras/expressões do corpo humano por outros campos semânticos

Anotação
SEMÂNTICA

CLASSES INICIAIS

corpo
corpo : opinião
corpo : sentimento
corpo : lugar
corpo : outros

Por que anotar?

- Forma de estudar a língua
 - Pôr em prática hipóteses sobre certos fenômenos
 - Um corpus anotado é insumo para pesquisa
- Corpus anotado e esquemas de anotacao interessam ao PLN

Como *exatamente*

a:[word="coração"] [word="vermelho"] [word="do"] [word="incêndio"] > a:[sema="corpo:lugar"]

Regras exclusivas de cada corpus

Regras gerais

Léxico inicial

- Palavras ou expressões
- Regras positivas ou negativas

penca junta

[lema="coluna"] [lema="social"] > a:[sema="o"]

- Palavras ou expressões
- **Classes semânticas iniciais**

a:[lema="veia"] [lema=".*ico|.ista"] > a:[sema="corpo:outros"]

```
céu da boca[]
cordão umbilical[]
costas[] da mão[]

barriga[] de aluguel
braço[] armado
testa[] de ferro
de mão[] beijada
sem pé[] nem cabeça[]

botar a boca[] no trombone
encher o ouvido[]
quebrar o cara[]
saber de cabeça[]
saltar[] aos olhos

pé[] de valsa

de cortar o coração[]
em pé[] de guerra
```

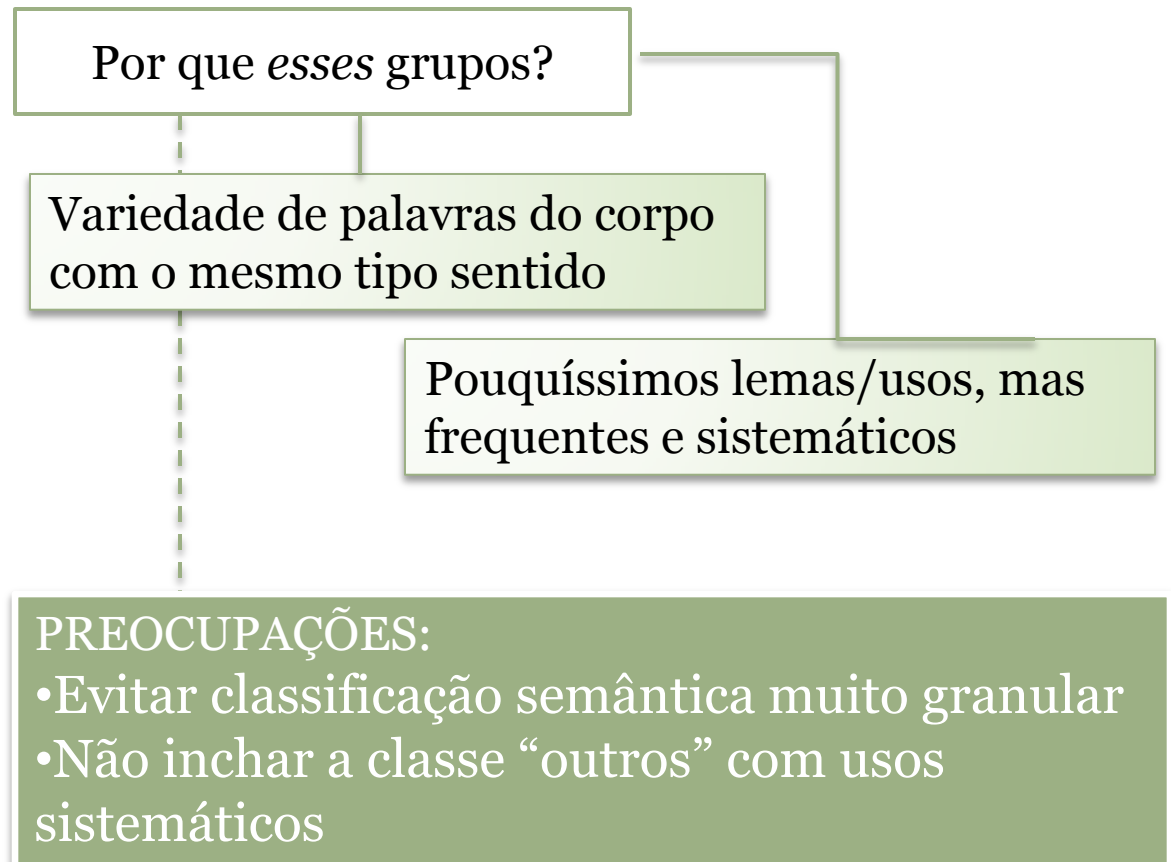
```
cintura
clavícula
colhão
colo
coluna
compleição
cona
coração
corpanzil
corpo
corporal
corpóreo
costas
costela
cotovelo
```

Em que pé_[corpo:outros] estamos

MP: revisto
OBRAS: em andamento

- Quando usamos palavras do léxico do corpo humano, de que mais falamos (além do corpo)?

- Sentimento
- Vegetal
- Parte de objeto
- Lugar
- Doença
- Opinião
- Posição
- Movimento
- Faculdade
- Medida
- Grupo
- Animal
- Outros



As classe semânticas do corpo

- SENTIMENTO – *com o **coração** apertado*
- VEGETAL – ***dente** de alho; **pé** de laranja*
- PARTE DE OBJETO - ***boca** do fogão*
- LUGAR - *A localização fez com que a batizassem de Te Pito o Te Henua, o **umbigo** do mundo*
- DOENÇA - *não tenho medo do **pé de atleta***
- OPINIAO – *ele é um **bundão**; tem sempre um **orelhudo** na conversa..*
- POSICAO - *suplicou **de joelhos**; dormiu **em pé***
- MOVIMENTO – *ir **a pé**; assim que **pôs os pés** na cidade*
- FACULDADE – *uma provocação plástica para **olhos e ouvidos** livres*
- MEDIDA – *dois **dedos** de pinga; onda de 3 **pés***
- GRUPO - ***corpo** docente; **corpo** de baile*
- ANIMAL – ***orelha** de porco*
- OUTROS – *o **pulmão** da cidade; uma **veia** pop*

Quebrando a cabeça₅ [corpo:faculdade] (1)

- LUGAR vs PARTE DE OBJETO

- Partedeobjeto: quando as fronteiras são bem definidas

- **boca** da garrafa → parte de objeto
- **boca** do rio → lugar
- **coração** da cidade → lugar

*Uberlândia: da **boca[lugar]** do sertão à cidade jardim.*

BOCA: 'ENTRADA' (LUGAR) → (TEMPO) 'início'

Na **boca da safra**, as commodities estão perdendo o fôlego .

A conversão no setor acontecerá na **boca da entressafra**, quando a oferta ...

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.

corpo:outros

Se houver mais casos, podemos criar novo SEMA

Quebrando a cabeça_; [corpo:faculdade] (2)

- **SENTIMENTO vs OPINIAO**

- Opinião: se o próprio termo ou expressão se refere a algo já com a indicação de opinião

- Ele é um **pé de valsa** → bom dançarino

OPINIAO

- Ele é um **bundão** → um covarde

OPINIAO

- Meu **coração** partiu quando ele se foi

SENTIMENTO

Fundindo a cabeça₃ [corpo:faculdade]

EVP=expressões com várias palavras

- Posição conservadora:

- *nó na cabeça*
- *boca do rio*
- *boca da noite*
- *dor=de=cotovelo*
- *nó=na=garganta*

Consideramos EVP apenas combinações opacas / não levamos em conta convencionalidade

O **olho[corpo:lugar]** do furacão Gonzalo tocou a terra nas ilhas Bermudas.

No **olho=do=furacão[corpo:outros]**, FHC faz que não é com ele.

Josias é que foi para o **olho[lugar]=da=rua [corpo:outros]**

Fundindo a cabeça_[corpo:faculdade]

EVP verbais

- EVPs variáveis
 - Ver com (muitos) bons olhos
 - lema: ver=com=bons=olhos
 - Ter olhos abertos para; abrir os olhos para (#prestar atenção)
 - lema: abrir=os=olhos=para
- EVP com lemas parecidos
 - de=mãos=abanando ; com=as=mãos=abanando
 - botar=a=boca=no=trombone ;
pôr=a=boca=no=trombone;
meter=a=boca=no=trombone;
colocar=a=boca=no=trombone

Verificação das classes criadas

- Rêve

Rêve

Partes do corpo

Classifique a palavra segundo o seu sentido:

- referindo-se ao corpo humano (corpo);
- usando uma palavra do corpo para outros fins:
 - o corpo:faculdade
 - o corpo:outros
 - o corpo:animal
 - o corpo:partedeobjeto
 - o corpo:sentimento
 - o corpo:vegetal
 - o corpo:grupo
 - o corpo:doença
 - o corpo:posição
 - o corpo:emoção
 - o corpo:lugar
 - o corpo:movimento
- ou não se referindo de todo ao corpo humano: 0

Se a palavra fizer parte de uma expressão maior também relacionada com o corpo, indique essa conexão no campo de comentário, com o sufixo EVP, por exemplo `corpoEVP` ou `corpo:lugarEVP`.

Quando considerar que mais de uma classificação é possível, indique ambas no comentário ligadas por _.

Mostrar Revisões

Guardar

E001-PT-332: E na aula recebem-me com uma salva de **palmas** .

corpo:faculdade ▾

E001-PT-905: Eu andei com caixotes às **costas** , todos nós.

corpo ▾

E003-PT-24: A ferramenta é esta que tenho na **mão** : a colmadeira, o lareiro e a estaca.

corpo ▾

Mais explicações; mais exemplos..

- **Esqueleto** (v. 7) – anotação, exemplos, fundamentação linguística
- **Ortopedista** – como tratar do Esqueleto – as regras
- Exemplos de procuras

<http://www.linguateca.pt/acesso/Esqueleto/>

O Corpo e as metáforas

- Metáforas: longa tradição nas reflexões sobre a linguagem
 - literal vs figurado
 - verdade vs retórica
 - significado vs uso

A história da discussão sobre a metáfora é a história de uma série de variações em torno de poucas tautologias, talvez uma só: «A metáfora é esse artifício que permite falar metaforicamente.»

Se apenas se entende por metáfora tudo aquilo que dela foi predicado ao longo dos séculos, torna-se claro que tratar a metáfora significa no mínimo tratar também (e a lista está incompleta): símbolo, ideograma, modelo, arquétipo, sonho, desejo, delírio, rito, mito, magia, criatividade, paradigma, ícone, representação – e também, como é óbvio, linguagem, signo, significado, sentido (Eco, 1994, p.20).

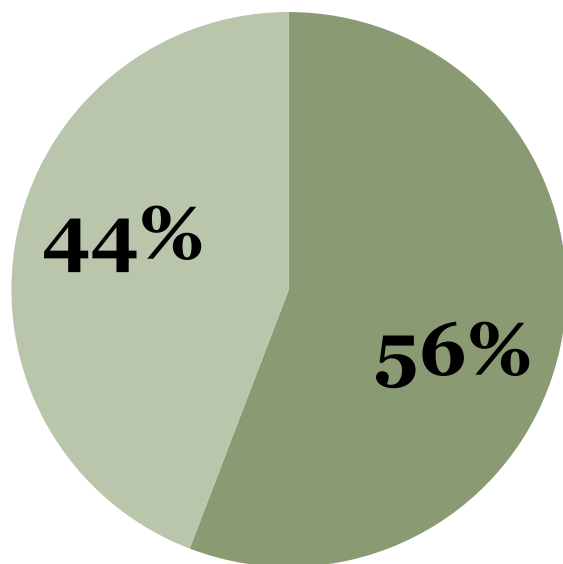
O que salta aos olhos

- Corpus Museu da Pessoa

1.4 milhões de palavras
2.338 palavras do corpo humano

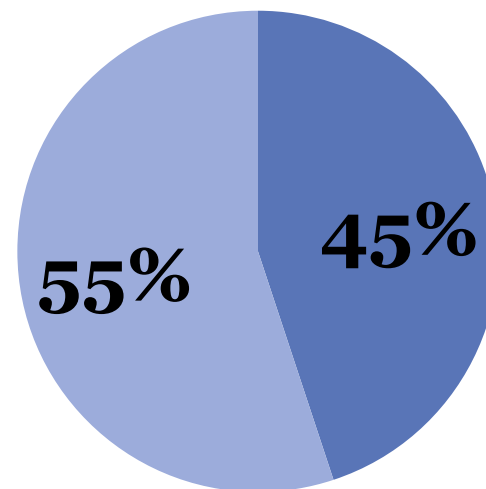
Distribuição das palavras de corpo no MP

■ corpo ■ corpo:xxx

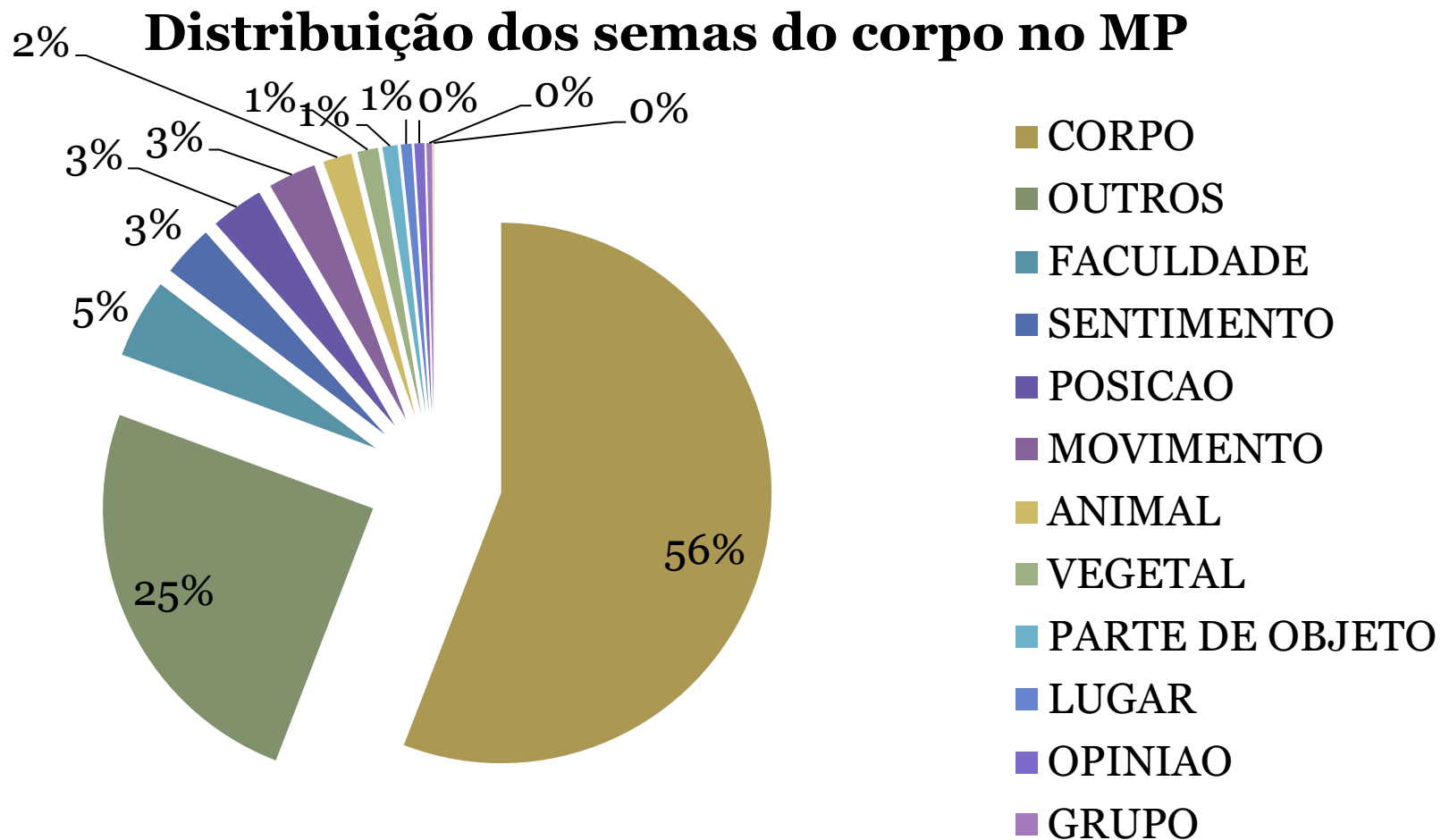


Distribuição das unidades lexicais com sema=corpo:xxx

■ corpo:xxxEVP ■ corpo:xxx



O que salta aos olhos



Tue Nov 4 21:00:08 WET 2014

Procura: [sema="corpo:.*"]

Distribuição de lema

Corpo: Museu da Pessoa v. 5.6

850 casos.

Distribuição
de corpo:xxx
por lema

Distribuição

Houve **62** valores diferentes de lema.

pé	179
cabeça	159
mão	100
cara	59
coração	58
boca	33
olho	29
manual	26
tronco	18
braço	18
corpo	17
sangue	14
rabo	8
perna	8
ouvido	7

Tue Nov 4 21:11:04 WET 2014

Procura: [sema=".*sentimento.*"]

Distribuição de lema

Corpo: Museu da Pessoa v. 5.6

73 casos.

“sentimento”

Distribuição

Houve **13** valores diferentes de lema.

coração	40
cabeça	9
olho	8
cara	3
pé	3
pele	2
mão	2
queixo	1
garganta	1
cotovelo	1
tripa	1
boca	1
dedo	1

Tue Nov 4 21:07:35 WET 2014

Procura: [lema="boca"]

Distribuição de sema

Corpo: Museu da Pessoa v. 5.6

65 casos.

Distribuição
de boca por
sema

Distribuição

Houve **12** valores diferentes de sema.

corpo	23
corpo:outros_corpo:outrosEVP	10
corpo:lugar	7
corpo_corpo:outrosEVP	7
corpo:partedeobjeto	5
corpo:outros	5
corpo:faculdade_corpo:outrosEVP	2
corpo:animal	2
corpo:animal_corpo:animalEVP	1
corpo:faculdade	1
corpo_corpo:sentimentoEVP	1
0	1

Trabalhos relacionados

- Expressões verbais do corpo para indicar opinião (Vale, 2013)
- Metáforas com *cabeça*, *pé* e *mão* (Almeida et al., 2009)
- Relação de meronímia entre partes do corpo (Baptista et al., 2014)
- Metáforas do corpo nas diferentes línguas (Maalej & Yu, 2011)

Referências

- Arrojo, R. e RAJAGOPALAN, K. 1992. Noção de literalidade: metáfora primordial. In Arrojo, R. **O signo desconstruído**. São Paulo: Pontes.
- Baptista, J. 2000. **Body-part nouns and local grammars**. Révue d'Informatique et Statistiques en Sciences Humaines. V. 36
- Costa, L., Santos, D., Rocha, P.A.: **Estudando o português tal como é usado: o serviço AC/DC**. STIL 2009
- Eco, U. **Semiótica e filosofia da linguagem**. São Paulo: Ática, 1991.
- Freitas, C; Motta, E. Milidiú, r.L & César, J. **Vampiro que brilha... rá! Desafios na anotação de opinião em um corpus de resenhas de livros**. ELC 2012.
- Lakoff, G & Johnson, M. *Metaphors we Live By*, Chicago: The University of Chicago Press. 1980.
- Leitão de Almeida, M. L. *et alii* (org.) **A hipótese de corporificação da categorização e do léxico**. In: LEITÃO DE ALMEIDA, Maria Lúcia *et alii* (org.). *Linguística Cognitiva em foco: morfologia e semântica do português*. Rio de Janeiro: Publit, 2010.
- Maalej, Z. e Yu, N. (eds.). 2011. *Embodiment via Body Parts: Studies from Various Languages and Cultures* (Human Cognitive Processing, Vol. 31). Amsterdam and Philadelphia: John Benjamins.
- Santos, D. "First steps of Gramateca: a corpus-based grammar initiative for Portuguese, driven by Linguateca", presentation at the Corpus Linguistics Group, University of Oslo, 20 February 2014.
- Santos, D. "Gramateca: corpus-based grammar of Portuguese". PROPOR'2014
<http://www.linguateca.pt/Diana/download/gramateca.pdf>
- Santos, D., Mota, C.: **Experiments in human-computer cooperation for the semantic annotation of Portuguese corpora**. LREC 2010
- Vale, O. 2013. As opiniões nas expressões e a expressão da opinião. In: Laporte, E., Smarsaro, A., Vale, O. (orgs). **Dialogar é preciso: Linguística para processamento de línguas**. PPGEL/UFES, 2013, pp.259-267.