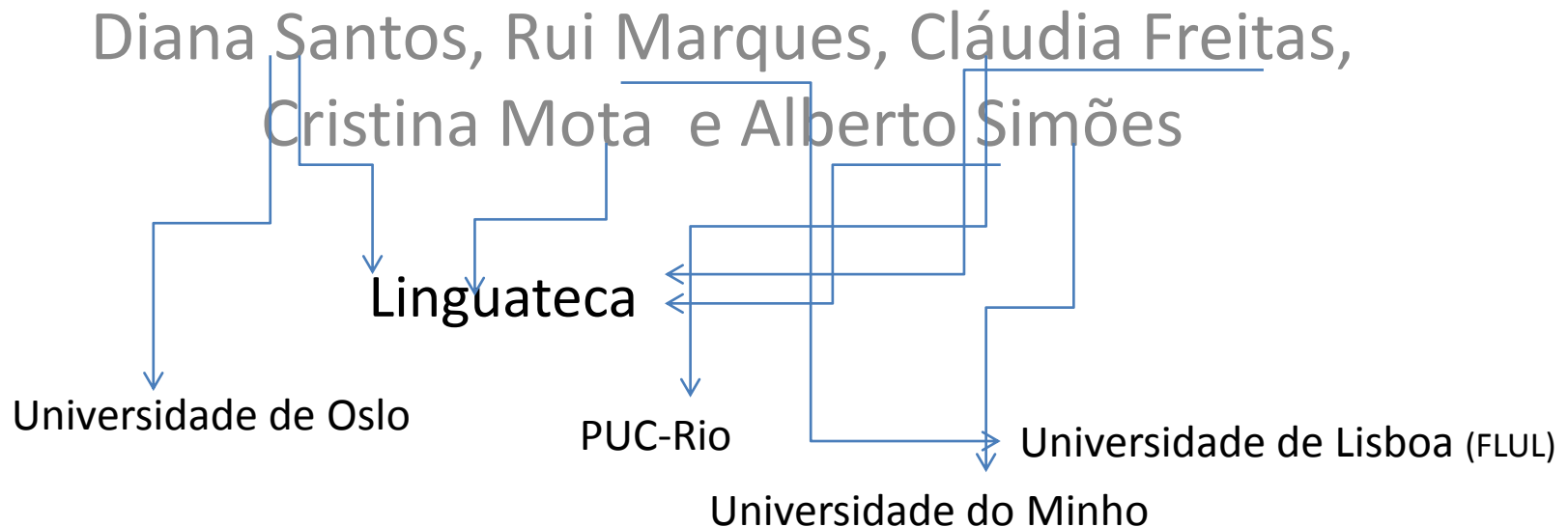




Gramateca

# Comparação de anotações na Gramateca





Gramateca

# Resumo

- Breve motivação da Gramateca e do Rêve
- Conectores condicionais
  - Teoria
  - Classificação em amostras de corpos do AC/DC
- Uso de palavras de corpo humano
  - Emoção ou opinião?
- Emoções
  - O caso da admiração: surpresa ou respeito?



Gramateca

# Enquadramento

- Textos anotados → pesquisa linguística
  - como encontrar informação em um (grande) corpus anotado?
- Linguateca → AC/DC: Acesso a Corpos; Disponibilização de Corpos
- A partir da infraestrutura Linguateca/AC/DC → Gramateca
  - Estudos da língua portuguesa com base em corpos
  - Inspiração: Biber et al., 1999 - *The Longman grammar of spoken and written English*

Gramateca: não só permitir a repetição de uma experiência (o que é uma das propriedades exigidas à metodologia científica) mas também partilhar diferenças de interpretação de um mesmo material.



Grar

## O que é a/uma gramática?

Como todas as palavras de uma língua, a sua definição não é simples. E, como em todos os ramos do conhecimento, há diversas perspetivas e opiniões e delimitações. É mais fácil saber usar do que saber definir. E o que é uma definição?

Para os efeitos da Gramateca, *gramática* é a forma como a língua funciona. Por isso pode incluir questões por outros chamadas pragmáticas, estudos de género textual, ou culturais.

E como é que se pode estudar e compreender a língua?

Também a este respeito, há opiniões e posições variadas. A resposta da Gramateca é

- através do estudo de grandes quantidades de texto
- através da interpretação dos casos de comunicação em contexto

Santos, 2014



Gramateca

# O que é a Gramateca

- Um laboratório para o estudo da língua portuguesa, que oferece:
  - todos os corpos disponibilizados pelo AC/DC
  - anotação automática desses corpos
  - ferramentas de visualização e de exploração dos corpos
  - **anotação manual de subconjuntos dos mesmos**
  - **uma plataforma de revisão e de comparação de diferentes análises**



Gramateca

# Laboratório

## Comparador

Procurar: [func="SUBJ.\*"] [pos="DET.\*|ADV.\*"]\* [lema="tocar" &

Corpo: CHAVE

Distribuir por: lema

Procurar: [func="SUBJ.\*"] [pos="DET.\*|ADV.\*"]\* [lema="tocar" &

Corpo: CONDIV

Distribuir por: lema

fundir numa única tabela

Limite mínimo de frequência

procurar

que	1172
se	118
ele	112
eu	78
eles	62
telefone	59
banda	58
quem	34
sino	33
você	28
orquestra	26
músico	25
ela	24
ninguém	23
nós	20
gente	19
música	18
grupo	16
bola	15
nai	12

que	55
bola	16
ele	7
telefone	3
atacante	3
esférico	3
ela	3
mão	3
Oséas	2
joelho	2
casaco	2
perna	2
adversário	2
anca	1
Vandick	1
cáli	1
Fabão	1
Marcelo=Santos	1
Juninho=Paulista	1
Rison	1



Gramateca

# Laboratório - Ferramentas

## Comparador

Procurar: [func="SUBJ.\*"] [pos="DET.\*|ADV.\*"]\* [lema="marcar"]

Corpo:

Distribuir por:

Procurar: [func="SUBJ.\*"] [pos="DET.\*|ADV.\*"]\* [lema="tocar" &

Corpo:

Distribuir por:

fundir numa única tabela

Limite mínimo de frequência

procurar

que	145
ele	27
quem	27
roberto	17
se	16
juiz	14
1	13
Benfica	12
árbitro	11
Sporting	10
Pelé	8
este	7
flamengo	6
Vasco	6
Académica	6
Estoril	6
golo	6
adversário	6
jogador	6

que	55
bola	16
ele	7
telefone	3
atacante	3
esférico	3
ela	3
mão	3
Oséas	2
joelho	2
casaco	2
perna	2
adversário	2
anca	1
Vandick	1
cáli	1
Fabão	1
Marcelo=Santos	1
Juninho=Paulista	1



Gramateca

# Laboratório - Ferramentas

## Distribuidor

Expressão de pesquisa:

Corpo:

Tipo de resultado:

lema	Frequência Total	pessnum	Frequência Parcial	%
pé	179	P	7	3,91%
		S	172	96,09%
cabeça	159	P	5	3,14%
		S	154	96,86%
mão	100	P	9	9,00%
		S	91	91,00%
cara	59	P	5	8,47%
		S	54	91,53%
coração	58	P	1	1,72%
		S	57	98,28%
boca	33	P	4	12,12%
		S	29	87,88%
olho	29	P	13	44,83%
		S	16	55,17%
manual	26	P	5	19,23%
		S	21	80,77%
braço	18	P	6	33,33%
		S	12	66,67%
tronco	18	P	4	22,22%
		S	14	77,78%





Gramateca

# Gramateca

- Usar a infra-estrutura dos corpos anotados da Linquateca para estudar gramática da língua portuguesa:  
<http://www.linquateca.pt/Gramateca/>
- Permitir a inspeção por outros **linguistas** das classificações efetuadas, através da ferramenta **Rêve**: <http://www.linquateca.pt/Reve/>
- Um projeto de todos: qualquer pessoa se pode agregar



Gramateca

# Rêve – O que é possível

- Criar subconjuntos de casos a partir de uma pesquisa no AC/DC;
- Disponibilizar uma interface para um utilizador **anotar** ou rever uma dada **anotação**, sobre um ou vários constituintes do subconjunto obtido anteriormente;
- Exportar uma quantificação simples sobre os resultados;
- Permitir a análise e a criação de novas revisões sobre os dados já presentes e analisados por outros;
- Quantificar as diferenças existentes entre anotações de diferentes pessoas.



Gramateca

# Rêve – Interface Geral

Linguatca

## Rêve

Novo Projeto de Rêvisão.

## Rêvisões em Curso

- Material Condicional
  - Ferramentas: [Rever anotação](#), [Anotar de novo](#), [Exportar Revisões](#), [Estatísticas](#)
  - Revisões disponíveis: [Rui Marques](#), [Diana](#),
- Partes do corpo
  - Ferramentas: [Rever anotação](#), [Anotar de novo](#), [Exportar Revisões](#), [Estatísticas](#)
  - Revisões disponíveis: [MPversão5](#), [claudia](#), [Bruno](#),
- Admiração: respeito ou espanto?
  - Ferramentas: [Rever anotação](#), [Anotar de novo](#), [Exportar Revisões](#), [Estatísticas](#)
  - Revisões disponíveis: [Diana](#), [claudia](#), [Cristina](#),



Gramateca

# Conectores condicionais

## Questão de partida:

Apurar informação sobre se os conectores condicionais do português "se", "caso", "no caso de" e "a" (cf. Peres *et al.* 1999) estão em variação livre ou que factor(es) condiciona(m) a sua distribuição.

Exemplos de construções condicionais com estes conectores:

- (1) a. *Se* a estrada for arranjada, haverá menos acidentes.
- b. *Caso* haja um feriado nacional, a maior parte das lojas está fechada.
- c. *No caso de* ocorrer alguma anomalia, devem contactar imediatamente os serviços.
- d. *A* não ser possível usar papel reciclado, evitar-se-ão textos muito extensos.



Gramateca

# Conectores condicionais

## Enquadramento prévio:

- Há diferentes tipos de orações condicionais, sendo observada a distinção, de Lopes 2009, entre “*condicionais canónicas*” vs “*condicionais não canónicas*” (cf., e.g., (1)) e dentro destas a individualização de “*condicionais de enunciação*” (cf., e.g., (2)):
  - (1) Se o teu filho é irrequieto, o meu é hiperactivo.
  - (2) Se tiveres sede, há cerveja no frigorífico.
- No que respeita às “*condicionais canónicas*”, assume-se que:
  - As orações condicionais são restritores de um quantificador (que pode estar implícito) sobre mundos possíveis (cf. Kratzer 1991, von Stechow 1994, e.o.).
  - Há dois tipos de base modal (Kratzer 1991): epistémica e circunstancial.



Gramateca

# Conectores condicionais

## Primeiros resultados

- Foi criado um conjunto de 100 casos aleatórios no CETEMPúblico, [ProcuraConectores100CP.html](#), anotados em [Anotacao100casos.html](#), que serviu para **identificar algumas categorias relevantes**.
- Foi criado um conjunto de mais 100 casos para cada conector, agora no conjunto de todos os corpos do AC/DC:
  - [100 casos de "se", sua anotação](#)
  - [100 casos de "a", sua anotação](#)
  - [100 casos de "caso", sua anotação](#)
  - [100 casos de "no/em caso de", sua anotação](#)
- Em suma, encontram-se actualmente 500 casos anotados.



# Conectores condicionais

Gramateca

## Categorias usadas na anotação:

- **EPIS - base modal epistémica**

“A julgar pelo número de telefonemas que tenho recebido, muitas das pessoas que compraram os seus computadores há três ou quatro anos estão a descobrir as limitações do seu equipamento.” (*par=PUB950218-1485*)

- **CIRC - base modal circunstancial**

“A Folha apurou na Fazenda que a nova moeda poderá chegar já em abril, caso a implantação da URV seja bem sucedida e o novo índice adquira confiança rápida.” (*par=FSP940117-179*)

- **ENUN - condicional de enunciação**

“Nosso objetivo não é reenviar mensagens para pessoas que não tenham interesse em recebê-las, portanto caso queira mais receber nossos info[r]mativos, por favor clique aqui para remover seu e-mail de nosso cadastro.”

- **CONT - valor contrastivo**

“O Benfica, se não foi superior, como conjunto, ao Girondins equivaleu na toada defensiva (sempre firme) e conseguiu criar mais momentos de glória e aberturas para remates vitoriosos.” (*par=fut5103*)

- **OUTR - não é conector condicional**

- erro - análise automática errada

“O governo se debate com impostos daqui e de lá para pagar juros de ficção científica.”

- compl - completiva (com “se”)

“O Acarosol, como o nome já diz, é um acaricida antes de usá-lo é preciso saber se sua composição não irrita os alérgicos; além disso, pode manchar tapetes e estofados.” (*par=FSP940807-2449*)

- outsub - outra subordinada, no caso de “a” + infinitivo

“A combatê-lo estiveram 30 bombeiros, apoiados por oito viaturas, conseguindo extinguir as chamas perto das 20 e 10 horas.” (*par=VD-N2099-3*)

- prog - marcador de construção progressiva

“Além disso, o montante do desfalque continua a ser uma incógnita, que começou no «boato» dos 100 mil contos, passou para os «cerca de 250 mil» e, segundo fontes próximas da administração, poderá chegar aos 350 mil ou até mais.” (*par=ext865309-eco-93b-2*)



Gramateca

# Rêve – (Nova) Anotação

*par=ext31050-nd-94a-1: 5* -- Mas, **se** se pretende lançar movimentos de opinião contra a localização destes empreendimentos estatais em Monsanto, gostaria que se debatesse também a questão de fundo de saber se queremos continuar a ter as universidades em Lisboa, se queremos ter novos tribunais, hospitais e outros estabelecimentos públicos de que a cidade carece e, querendo tudo isso, onde e como vamos -- Estado e autarquia -- disponibilizar os solos necessários e adequados para a sua localização .

*par=ext61992-nd-92a-1:* Se lhe tocamos e depois agarramos um objecto virtual no mundo virtual, *podemos* deslocá-lo e manipulá-lo como se de um objecto real se tratasse .

*par=ext76708-pol-98a-3:* «Nestas duas questões seremos intransigentes», garantiu. E **se** nada mudar ? «A estabilização dos preços dos combustíveis» tem como objectivo «consolidar a desinflação», afirmou ontem o ministro da Economia, Augusto Mateus, reconhecendo assim que o Governo está a sacrificar receitas fiscais para evitar que a inflação aumente .

*par=ext95065-soc-96a-2:* **A justificar** esta reivindicação está o facto de os concelhos do Porto, de Vila Nova de Gaia e de Matosinhos, respectivamente, serem os que mais pessoas têm actualmente inscritas nos centros de emprego (cerca de 42 mil, no total) .





Gramateca

# Rêve - Reanotação

*par=ext6881-pol-98b-2*: Apoiar as empresas com facturação até 240 mil contos seria admissível, afirma o secretário de Estado, se não houvesse limites para a despesa fiscal .

ENUN ▾

*par=ext13678-clt-96a-2*: Depois desta saída no espaço, a NASA decidirá se o Endeavour vai regressar ou não à Terra um dia mais cedo, na próxima sexta-feira, por causa da formação de gelo no sistema de arrefecimento vaivém .

ENUN ▾

*par=ext31050-nd-94a-1*: 5 -- Mas, se se pretende lançar movimentos de opinião contra a localização destes empreendimentos estatais em Monsanto, gostaria que se debatesse também a questão de fundo de saber se queremos continuar a ter as universidades em Lisboa, se queremos ter novos tribunais, hospitais e outros estabelecimentos públicos de que a cidade carece e, querendo tudo isso, onde e como vamos -- Estado e autarquia -- disponibilizar os solos necessários e adequados para a sua localização .

OUTR\_compl ▾

*par=ext61992-nd-92a-1*: Se lhe tocamos e depois agarramos um objecto virtual no mundo virtual, *podemos* deslocá-lo e manipulá-lo como se de um objecto real se tratasse .

CIRC ▾

*par=ext76708-pol-98a-3*: «Nestas duas questões seremos intransigentes», garantiu. E se nada mudar ? «A estabilização dos preços dos combustíveis» tem como objectivo «consolidar a desinflação», afirmou ontem o ministro da Economia, Augusto Mateus, reconhecendo assim que o Governo está a sacrificar receitas fiscais para evitar que a inflação aumente .

CIRC ▾



Gramateca

# Rêve – Revisões disponíveis

---

*par=ext6881-pol-98b-2*: Apoiar as empresas com facturação até 240 mil contos seria admissível, afirma o secretário de Estado, se não houvesse limites para a despesa fiscal . CIRC

---

*par=ext13678-clt-96a-2*: Depois desta saída no espaço, a NASA decidirá se o Endeavour vai regressar ou não à Terra um dia mais cedo, na próxima sexta-feira, por causa da formação de gelo no sistema de arrefecimento vaivém . OUTR\_compl

---

*par=ext31050-nd-94a-1*: 5 -- Mas, se se pretende lançar movimentos de opinião contra a localização destes empreendimentos estatais em Monsanto, gostaria que se debatesse também a questão de fundo de saber se queremos continuar a ter as universidades em Lisboa, se queremos ter novos tribunais, hospitais e outros estabelecimentos públicos de que a cidade carece e, querendo tudo isso, onde e como vamos -- Estado e autarquia -- disponibilizar os solos necessários e adequados para a sua localização . ENUN

---

*par=ext61992-nd-92a-1*: Se lhe tocamos e depois agarramos um objecto virtual no mundo virtual, *podemos* deslocá-lo e manipulá-lo como se de um objecto real se tratasse . CIRC

---

*par=ext76708-pol-98a-3*: «Nestas duas questões seremos intransigentes», garantiu. E se nada mudar ? «A estabilização dos preços dos combustíveis» tem como objectivo «consolidar a desinflação», afirmou ontem o ministro da Economia, Augusto Mateus, reconhecendo assim que o Governo está a sacrificar receitas fiscais para evitar que a inflação aumente . CIRC



# Rêve - Estatísticas

Gramateca

Rêve

	Diana	Rui Marques
OUTR_outsub	5	5
????	2	2
CIRC	52	53
ENUN	11	10
OUTR_compl	11	11
CONT	4	4
OUTR_erro	13	13
OUTR_prog	2	2

## Discordâncias

- *par=ext6881-pol-98b-2*: Apoiar as empresas com facturação até 240 mil contos seria admissível, afirma o secretário de Estado, se não houvesse limites para a despesa fiscal .
  - **CIRC**: Rui Marques
  - **ENUN**: Diana
- *par=ext13678-clt-96a-2*: Depois desta saída no espaço, a NASA decidirá se o Endeavour vai regressar ou não à Terra um dia mais cedo, na próxima sexta-feira, por causa da formação de gelo no sistema de arrefecimento vaivém .
  - **OUTR\_compl**: Rui Marques
  - **ENUN**: Diana
- *par=ext31050-nd-94a-1*: 5 -- Mas, se se pretende lançar movimentos de opinião contra a localização destes empreendimentos estatais em Monsanto, gostaria que se debatesse também a questão de fundo de saber se queremos continuar a ter as universidades em Lisboa, se queremos ter novos tribunais, hospitais e outros estabelecimentos públicos de que a cidade carece e, querendo tudo isso, onde e como vamos -- Estado e autarquia -- disponibilizar os solos necessários e adequados para a sua localização .



Gramateca

# Esqueleto – léxico do corpo humano

- CORPO
- **CORPO : SENTIMENTO**
- CORPO : VEGETAL
- CORPO : PARTE DE OBJETO
- CORPO : LUGAR
- CORPO : DOENÇA
- **CORPO : OPINIAO**
- CORPO : POSICAO
- CORPO : MOVIMENTO
- CORPO : FACULDADE
- CORPO : MEDIDA
- CORPO : GRUPO
- CORPO : ANIMAL
- CORPO : OUTROS

Nós anotamos TODOS os corpora e compartilhamos a anotacao com todos no Esqueleto

Emoção e Opiniao:  
Pedir a anotadores independentes para anotarem um subconjunto de frases selecionadas



Gramateca

# Emoções em português

- Pressuposto: a língua é um veículo de referência a emoções
- A forma de lhes referir depende de cada língua
- A distinção entre opiniões, julgamentos de valor e emoções não é pacífica
  - Estudo de *admiração* e *respeito*: o mesmo campo ou dois distintos?



Gramateca

# Exemplo de ADMIRAR

- várias palavras cuja raiz é ADMIRAR:
    - sentidos emocionais: *supresa, respeito*
    - atividade de *olhar*: com *olhos respeitosos* ou *olhos apreciativos*
  - Isto é por acaso?
    - homonímia que não acontece em outras línguas?
- OU
- Nos permite exprimir ambos os sentimentos ou uma atividade prática com um sentimento?
    - vagueza?



Gramateca

# Teste de ADMIRAR

- Com o Rêve, quisemos investigar se havia de facto essa percepção -- vários sentidos distintos, em particular relativos a dois campos semânticos ou emoções -- que não por acaso, se podem referir, em termos físicos, a "ficar de boca aberta"
  - Palavras que, fora de contexto, podem ter as duas ou mais interpretações: *admirar; surpreender; embasbacar; maravilhar; reverenciar.*
  - Seleção de casos de *surpresa* e *respeito* nos corpos MUSEUDAPESSOA e OBRAS (129)
  - Classificação dos casos com as categorias RESPeito, SURPresa, GOSTARDEVER, ambos ou nenhum.
  - Anotação por três de nós: 40 divergências.



Gramateca

## Rêve

	claudia	Cristina	Diana
RESP	36	33	40
NENHUM	5	12	5
AMBOS	13	8	10
GOSTARDEVER	23	20	19
SURP	31	35	34

### Discordâncias

- 117953: Esse Hilário tinha na montra da oficina dele , em 1945 , um carro feito por ele que quem passasse por lá **admirava** aquela arte .
  - **GOSTARDEVER:** claudia, Cristina
  - **AMBOS:** Diana
- 140378: E muita gente se **admirou** como é que eu estava a fazer aquilo ao Teixeira Rainha .
  - **AMBOS:** Diana
  - **SURP:** claudia, Cristina
- 965583: Tem uma cabeça **admirável** .
  - **AMBOS:** claudia, Cristina
  - **NENHUM:** Diana
- 1136715: Eu **admiro** essas meninas hoje que casam com tudo , têm tudo o que existe de mais moderno ... aperta um botão a água tá ali , ; aperta um botão a água tá caindo ...
  - **AMBOS:** Diana, claudia
  - **NENHUM:** Cristina





Gramateca

# Rêve

- Permite compartilhamento de anotações no âmbito da Gramateca
- Forma de pôr à prova uma anotação\_resultado de uma pesquisa (com base em corpus)
  - Maneira de compartilhar os resultados de uma pesquisa, na forma de anotação.



Gramateca

# Qual a relevância da comparação de anotações?

- Anotação é um acréscimo valioso para o corpus/a pesquisa
  - Anotação de: atos de fala; relações retóricas; modalização; polaridade; metáforas; papéis semânticos; erros de aprendizes; mwes...
- Anotação é (sempre!) fruto de interpretação
  - Os marginais precisaram arrombar o portão da casa dela.
- AS TECNOLOGIAS NÃO SÃO NEUTRAS



Gramateca

# Conectores condicionais: Referências

- Costa, Sueli. "Entre o deôntico e o epistêmico: o caráter camaleônico do verbo modal 'poder' ", *Letra Magna* **5**, 11, 2009.
- Geis, Michael & Arnold Zwicky, "On Invited Inferences", *Linguistic Inquiry* **2**, 1971, 561-566.
- Gomes, Gilberto & Priscila Mattos Monken. "Postura epistêmica e parafraseabilidade diferencial em condicionais", *Rev. Est. Ling.* **19**, 2, jul/dez 2011, Belo Horizonte, 127-140.
- Kratzer, Angelika. "Modality." In A. von Stechow & D. Wunderlich (eds.), *Semantics*, 1991, 639- 650.
- Lopes, Ana Cristina Macário. "Contributos para o estudo de construções condicionais não-canônicas em Português europeu contemporâneo", *Diacrítica, Ciências da Linguagem* **23** (1), 2009, 149-169.
- Lopes, Ana Cristina Macário. "Contributos para uma análise semântico-pragmática das causais de enunciação no Português europeu contemporâneo", *Alfa* **56**, (2), 2012, 451-468.
- Peres, João Andrade. "Notas sobre Conexões Interproposicionais", *Linguística do Texto* (aulas) 2009.
- Peres, João Andrade, Telmo Mória & Rui Marques. "Sobre a Forma e o Sentido das Construções Condicionais em Português", in I. H. Faria (org.), *Lindley Cintra, Homenagem ao Homem, ao Mestre e ao Cidadão*, Lisboa: Ed. Cosmos / FLUL, 1999, 627-653.
- Sweetser, Eve. *From Etymology to Pragmatics, Metaphorical and Cultural Aspects of Semantic Structure*, Cambridge University Press, 1990.
- von Fintel, Kai. "Restrictions on Quantifier Domains", Ph.D. Diss., GLSA, UMass Amherst, 1994.



Gramateca

# Corpo humano: Referências

- Arrojo, R. e RAJAGOPALAN, K. 1992. Noção de literalidade: metáfora primordial. In Arrojo, R. **O signo desconstruído**. São Paulo: Pontes.
- 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.
- Lakoff, G & Johnson, M. *Metaphors we Live By*, Chicago: The University of Chicago Press. 1980.
- 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., Mota, C.: **Experiments in human-computer cooperation for the semantic annotation of Portuguese corpora**. LREC 2010



Gramateca

# Emoções: Referências

- Freitas, Cláudia, Eduardo Motta, Ruy Luiz Milidiú & Juliana César. "Vampiro que brilha... rá! Desafios na anotação de opinião em um corpus de resenhas de livros". In *XI Encontro de Linguística de Corpus - ELC 2012* 13-15 de Setembro.
- Justeson, John S. & Slava M. Katz. "Redefining Antonymy: The Textual Structure of a Semantic Relation". *Literary and Linguistic Computing* **7**(3), 1992, 176-184.
- Maia, Belinda. "A Contribution to the Study of the language of Emotion in English and Portuguese". Porto: FLUP. 1994. Versão revista (1996):  
<http://web.lettras.up.pt/bhsmaia/belinda/pubs/thesis.htm>
- Pang, Bo & Lillian Lee. "Opinion mining and sentiment analysis". *Foundations and Trends in Information Retrieval* Vol. 2, No 1-2 (2008), 1–135.
- Santos, Diana & Cristina Mota. 2014. "Emotions in natural language: a broad-coverage perspective". 2014. Em apreciação.
- Santos, Diana & Cristina Mota. 2015. "A admiração à luz dos corpos." No prelo.



Gramateca

# Obrigada!

<http://www.linguateca.pt/Gramateca/>