

Corpos e cores: colorindo a descrição da língua portuguesa

Cláudia Freitas (Linguatca & PUC-Rio)
Diana Santos (Linguatca & Univ. de Oslo)
Rosário Silva (Linguatca)

Motivação

Categorização das cores

universalismo

relativismo

Cores básicas são universais

NÃO universalidade da conceptualização da cor

Experiência com informantes
Comparação entre as línguas

E a descrição linguística das cores?

Trabalhos com base em corpos
→ Comportamento das cores nas línguas

Objetivos

- Como se relacionam cor e categoria gramatical nas variantes PP e PB?
 - Exploração da distribuição de cores por categoria gramatical: cores mais verbais? cores mais nominais?
 - continuação das explorações de Santos et al. 2011

(... E também discutir um pouco a complexidade do trabalho de anotação semântica.)

Contexto da exploração: o AC/DC

Acesso a corpos / Disponibilização de corpos

- ❑ Desenvolvido e mantido pela Linguateca / liderado por Diana Santos
- ❑ 20 corpos em português (publicamente disponíveis – www.linguateca.pt/ACDC)
- ❑ Cerca de 380 milhões de palavras / 17 milhões de frases
- ❑ Variantes do Brasil e de Portugal
- ❑ Diferentes tipos de texto (jornalístico, literário, técnico, blogs..)
- ❑ Anotados automaticamente pelo PALAVRAS (Bick, 2000) (ANÁLISE MORFOSSINTÁTICA)
- ❑ Anotados automaticamente com campos semânticos de **COR** e de roupa (Corte-e-Costura (Mota & Santos 2009)) – REVISÃO DA ANOTAÇÃO EM ANDAMENTO

AC/DC – galeria dos corpos

Jornalístico geral

- CETEMPúblico
- NILC/São Carlos
- CHAVE

Jornalístico regional

- NatMinho
- DiaCLAV

Jornalístico Específico

- Esportes : CONDIVport
- Moda: CONDIVport
- Saúde: CONDIVport
- Política: Avante!
- Ciências: CorTradjorn

99 milhões de palavras (F. S. Paulo (PB) /Público (PP)).

Literário

- Vercial
- ClassLPPE
- ENPCpub



Transcrição de fala

- Museu da Pessoa
- ECI-EBR falado
- Selva falado



Textos Técnicos

- ECI-EE
- NILC/São Carlos
- Selva Ciência



Email

- Listas: ANCIB
- SPAM: CoNE



Recursos de Avaliação

CDHAREM



Blog

- Amazônia



5 milhões de palavras - revisto quanto às cores

Linguateca

[Estrutura](#)
[Equipa](#)

Apresentação

Acesso a recursos

- AC/DC
- - Procura
- - Frequência
- - PJR e exemplos
- - Corte-e-costura
- - Ensinador
- - VARRA
- CETEMPúblico
- CETENFolha
- CHAVE
- COMPARA
- Corpógrafo
- CorTrad
- Esfinge
- Floresta Sintá(c)tica
- GeoNetPT
- PAPEL
- ReEscreve
- REPENTINO
- Repositório
- WebSpell
- WPT

Catálogo de recursos

Catálogo de

ferramentas

Catálogo de atores

Catálogo de

publicações

Informação

interessante

Fórum

Avaliação conjunta

Perguntas já

respondidas

English

Norsk



Os apontadores na janela do lado direito não actualizam o lado esquerdo do ecrã. Sugerimos, pois, que naveguem

Acesso a corpos de português: Projecto AC/DC

[Linguateca](#)

[Information in English](#)

O projecto AC/DC (Acesso a corpos/Disponibilização de corpos), iniciado em 1999, surgiu da necessidade de juntar os poucos recursos disponíveis num único ponto na rede e dessa forma facilitar a comparação e a reutilização do material, permitindo ao mesmo tempo acesso a uma ferramenta poderosa de interrogação de corpos, o sistema [CWB](#) (versão nova do [IMS corpus workbench](#)), para o qual desenvolvemos esta interface.

Desde 2000, a anotação dos corpos tem sido feita automaticamente pelo [PALAVRAS](#) de Eckhard Bick, e convertida para o "formato AC/DC", descrito pormenorizadamente na página de [Anotação](#).

Uma descrição quantitativa inicial dos corpos servidos presentemente pelo AC/DC encontra-se na tabela abaixo. **Clique num dos corpos para o interrogar.** Para cada corpo, pode pedir concordâncias, distribuição e frequências simples e complexas, veja os [exemplos](#). Se é a primeira vez que visita o AC/DC e quer apenas experimentar, [procure no corpo Vercial](#).

Breve descrição dos corpos

Corpo	Tamanho (unidades)	Tamanho (palavras)	Tamanho (frases)	Variante(s)	Breve descrição
AmostRA-NILC	127.832	98.786	4.965	BR	AmostRA-NILC
ANCIB	1.690.376	1.258.764	80.992	BR	Correio electrónico correspondente ao tráfego na lista ANCIB
Avante!	7.766.309	6.501.146	204.414	PT	Semanário político Avante!, 1997-2002
CD HAREM	290.001	225.766	12.558	PT BR	Colecção dourada do HAREM
CETEMPúblico	232.543.379	189.575.095	7.665.410	PT	Jornal PÚBLICO, dividido em extractos, 1991-1998
CHAVE	123.936.528	99.355.321	4.740.448	PT BR	Jornais PÚBLICO e Folha de São Paulo, 1994-1995
CONDIVport	7.089.872	5.577.161	328.214	PT BR	Jornais desportivos e revistas de moda e saúde
CoNE	925.262	685.244	31.562	PT BR	Mensagens de correio electrónico não-endereçadas
DiaCLAV	7.758.469	6.651.523	232.152	PT	Diário de Coimbra, Diário de Leiria, Diário de Aveiro, Viseu Diário
ECI-EBR	917.124	724.008	44.381	BR	Texto do corpo Borba-Ramsey, compilado pelo ECI
ECI-EE	32.034	27.138	839	PT	Texto de chamada do programa europeu ESPRIT
ENPCPUB (parte em português)	92.688	72.375	4.371	PT BR	Literatura traduzida do inglês proveniente do ENPC
Floresta	7.252.252	5.950.072	327.179	PT BR	Floresta Sintá(c)tica
FrasasPB	23.313	19.162	653	BR	Frasas em português do Brasil
FrasasPP	20.048	16.232	594	PT	Frasas em português de Portugal
Museu da Pessoa	517.740	375.149	27.288	PT BR	Entrevistas realizadas pelo Museu da Pessoa
Natura/Minho	2.156.280	1.749.068	68.910	PT	Jornal regional Diário do Minho, antes da revisão
NILC/São Carlos	42.519.799	32.303.761	1.954.909	BR	Texto do corpo NILC, contendo maioritariamente texto jornalístico, mas também cartas comerciais e textos didácticos
todos juntos	18.851.915	14.313.624	596.869	todas	Todos os corpos juntos
Vercial	18.851.915	14.313.624	596.869	PT BR	Clássicos da literatura portuguesa, séculos XVI a XX
Total	473.363.136	379.793.019	16.923.577	PT BR	todos os corpos

ANOTAÇÃO DAS CORES: classes

Classes de cor

cor → cor “pura”, representação de atributos visuais

cor:humana → louro, moreno, castanho, ruivo...

cor:vinho → branco, tinto, verde, rosé

cor:raça → branco, preto, negro, amarelo, pardo

cor:política → verde, laranja

cor:equipa → encarnado, tricolor, rubro-negro

cor:original → usos não convencionais; não representativos de atributos visuais

ANOTAÇÃO DAS CORES: desafios (1)

CLASSE

cor:original

Usos não diretamente relacionados a propriedades visuais:

*Culpado ou não, acabará sempre no **vermelho**;
o período **negro** do Plano Collor;
uma concordata **branca***

- uso metafórico de uma cor, separado ou englobado em expressões fixas → *por um unha **negra**; em **branco***
- termos que originalmente contêm uma palavra de cor mas que são considerados unidades, e que não se referem a cor → *buraco **negro**; elefante **branco**; caixa **preta***
- termos às vezes descritivos; às vezes usados metaforicamente: → *cartão **vermelho**; sinal **verde** (≠ encontrei um cartão vermelho na mesa com um poema)*

ANOTAÇÃO DAS CORES: grupos

BRANCO

branqueamento,
embranquecer,
alvura

PRETO

pretinhas, negro,
enegrecer

AZUL

azul-marinho,
turquesa, anil

ROXO

lilás, roxidão,
violáceo

CASTANHO

chocolate
acastanhado
café

VERMELHO

vermelhidão
encarnado

LARANJA

alaranjado
laranja-claro

CINZENTO

plúmbeo
acinzentar

VERDE

esverdear
pistache
verde-paste

ROSA

framboesa
rosé

CREME

bege
perolado
marfim

AMARELO

amarelar
amarelo-limão

PRATEADO

metalizado
prata

MÚLTIPLA

alvinegra
lilás-rosa

incolor
translúcido
transparente

AUSÊNCIA

ouro
dourar

DOURADO

tijolo
ocre
pardo

OUTRAS

multicolor
supercolorido
de cor

DESCONHECIDA

Documentação detalhada no Arco-íris

ANOTAÇÃO DAS CORES: desafios (2)

GRUPO

Outras

expressões cuja cor pareceu impossível identificar/definir:

Cor de apoplexia;

Cor de crime e traição

Cor de burro quando foge

Cor de ostra

Cor de treva

Cor de ferro velho

Cor de mármore

Cor de morte

...

não é possível decidir (e portanto anotar) sobre a inclusão de uma cor em um determinado grupo

PARDO: é castanho/marrom ou cinzento?

(discordância entre anotadores..)

Tijolo?

Ocre?

Cor de telha?

ANOTAÇÃO DAS CORES: criação dos grupos - Motivação

uso variado e frequente de um termo para uma cor/grupo de cor → de alguma maneira esse grupo/cor é relevante na língua

a busca por um “grupo” é mais prática que uma busca por lema.. (busca pelos lemas de todos os tipos de vermelho, ou de azul..)

Creme: bege; bege-areia; bege-caramelo; bege-carregado; bege-claro; bege-dourado; bege-escuro; bege-esverdeado; bege-marfim; bege-nacarado; bege-rato; bege-rosado; bege-suave; creme; creme-bege; marfim; perolado; pérola

Ainda estamos tentando descobrir categorias/grupos de cores!!!

Interesse apenas nos grupos de cores “pouco óbvios”?
[sema="cor.*" & grupo="Outras.*"]

Interesse apenas nas cores tijolo e telha?
[sema="cor.*" & lema="tijolo|telha"]

ANOTAÇÃO DAS CORES: desafios (3)

*Sem a ajuda soviética e vítima do bloqueio americano, o sonho do socialismo **moreno**, que tanto encantou a esquerda brasileira, parecia estar fadado a um final infeliz.*

Moreno → cor:humana

DÚVIDA!!

socialismo **moreno** = a socialismo **verde** ??
= a socialismo **vermelho**??

cor:política?

cor:original?

- *Se definem como socialistas (ou «**socialismo moreno**», como diz Brizola) , mas tem forte apelo populista .*
- *Ainda que tingido do ` **socialismo moreno**» do autor de “O Povo Brasileiro” , o livro ...*
- *Impressionado com o bronzeado dos colegas, o baiano Jaques Wagner (PT-BA) , desbotado pelas viagens à Alemanha, Coréia do Sul e Inglaterra, disse: ` O **socialismo moreno** do Brizola não deu certo, mas pelo visto o neoliberalismo mulatinho do Fernando Henrique vai bem» .*

ANOTAÇÃO DAS CORES: desafios (4)

*...vomitam sobre nós indecentemente uma ideologia **parda**, sem cor.*

...uma ideologia parda, isto é, sem cor.

... uma ideologia sem interesse, chata, insossa...

cor:ausência

cor:original??

cor:política??

VAGUEZA: quando não conseguimos escolher entre uma ou mais classificações, atribuímos todas as classificações → sema="cor_0"

*A área tem muito **verde**, é arborizada e quieta.*

[sema="cor_0" & cor="Verde_0"]

Anotação semântica semi-automática no AC/DC

1. Escolha das etiquetas semânticas (a partir das ocorrências nos corpos)
2. Anotação de textos com as etiquetas escolhidas
3. Criação de regras de anotação a partir de informação lexical
4. Aplicação das regras
 - regras gerais, independentes de assuntos ou gêneros
 - regras específicas para determinados corpos (assuntos/gêneros)
5. Melhoria das etiquetas e das regras
6. Documentação das escolhas subjacentes à anotação

Exemplos de regras do Corte-e-costura

Mota & Santos 2009

Linguatca > Acesso a recursos > AC/DC > Corte-e-Costura

a:[lema="branco"] b:[lema=","] c:[lema="mestiço"] >> a:[sema="cor:raça"]

a:[word="dos | os"] b:[lema="branco | negro | preto" & pos="N"]
[pos="PU | ADJ | V.*"] >> b:[sema="cor:raça"]

a: [lema="sacola"] b: [lema="pardo" & sema="raça"] >> b:[sema="cor"]

Distribuição das cores por categoria gramatical

	CHAVE		CONDIV	
	ADJ cor	Verbos cor	ADJ cor	Verbos cor
BR	0,6%	0,2%	2,2%	0,06%
PT	0,6%	0,2%	3,4%	0,06%

[pos="ADJ.*" & sema="cor.*" & variante="... "]

[pos="V.*" & sema="cor.*" & variante="... "]

Verbos coloridos

- Grupo Vermelho → avermelhar
- Grupo Laranja → alaranjar
- Grupo Amarelo → amarelar; amarelecer
- Grupo Roxo → arroxear
- Grupo Preto → enegrecer
- Grupo Verde → esverdecer; reverdecer; verdejar
- Grupo Cinza → acinzentar
- Grupo Rosa → rosar
- Grupo Dourado → dourar/doirar
- Grupo Branco → branquear; embranquecer
- Grupos Castanho, Creme, Prateado e Azul → sem verbos

- no CHAVE e no CONDIV
- **sem participio**

[sema="cor.*" & grupo="Vermelho" & temcagr!="PCP" & pos="V.*"]

DOURAR/DOIRAR...

... QUEM doura/doira?

O QUÊ?

mestres; estatutos; virtudes;
retórica; jornal, PSD,
instalações...

PT/BR

carne, grill sol, luz

BR

1... a pílula (quase 50% em PT e BR)

PT

Imagem;
perspectiva;
brasão; asa; sala;
perdão

nenhum uso relacionado
diretamente à cor!

BR

alho, bacon; assado; cebola;
manteiga; alho; frango
ignorância; fisiologia; imagem;
década; personalidade, jogadas,
triunfo
paisagem; cidade; luz, gente

[func="SUBJ>.*"] [pos!="V.*"]* [pos="V.*" & lema="do[ui]rar.*" & sema="cor.*" & temcagr!="PCP"] within s

[pos="V.*" & lema="do[ui]rar.*" & sema="cor.*" & temcagr!="PCP"] [pos!="V.*"] [func="<ACC"]within s

++/ ++/ ++/ ++

BRANQUEAR

... QUEM BRANQUEIA?

O QUÊ?

PT

dinheiro!

imagem

PIDE

passado

comportamento

situação

acção

fascismo

lucro

regime

política

90 ocorrências

BR

roupa

celulose

rolha

4 ocorrências

CONDIV

PT: pele; dente; hálito;
marfim; perna; mão;
aspecto

BR: pescoço

dente (1)
roupa (1)
casa (1)

PT

PSD; relatório;
falsificação; ganância;
deputado; dirigente;
garantia; tráfico...

BR

*"diziam que o Brasil não tinha um problema com os negros, pois iríamos **branqueando** gradativamente com os casamentos inter-raciais..."*

EMBRANQUECER

TODOS os corpos

branquear

embranquecer

esbranquiçar

PT	391 (27 Vercial)
BR	12

PT	45 (31 Vercial)
BR	24

PT	1
BR	2

O QUÊ / QUEM?

PT

BR

cabelo

pele

mão

barba

cabeça

limão

camisa

principalmente cor:humana

cor

Esbranquiçar:

corolas (PT)

espinha; nação (BR)

BR: cor:raça | cor: humana (?)

“uma mídia que busca **embranquecer** tudo e todos”

“o mulato, buscando ascender socialmente, trata desesperadamente de **embranquecer**,...”

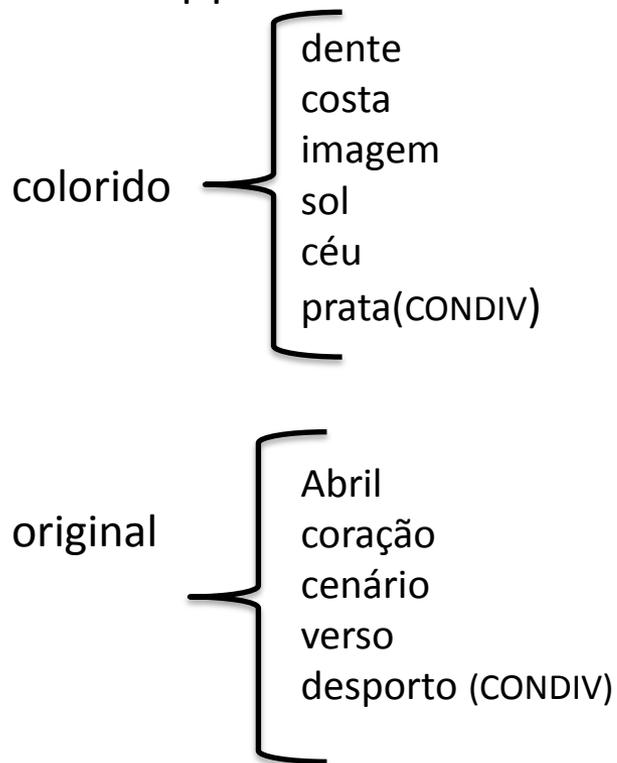
ENEGRECER

“enegrecer” em TODOS os corpos →

PT	117 (63Vercial)
BR	6

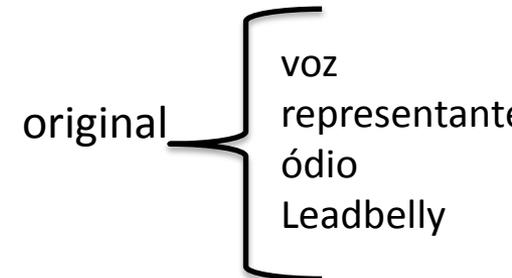
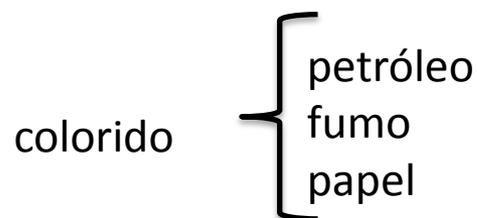
O QUÊ?

PT



... QUEM ENEGRECE?

PT



só 2 ocorrências em BR (CHAVE);
Nenhuma cor “pura” :
As incertezas **enegreceram** o cenário ;
...enquanto os americanos se preocupavam, pois acreditavam
que iriam **enegrecendo** por meio dos mesmos cruzamentos
entre negros e brancos

AMARELAR AMARELECER

“amarelar” em TODOS os corpos →

BR	16
PT	14 (-6 Vercial)

“amarelecer” em TODOS os corpos →

PT	65 (- 33 Vercial)
BR	3

AMARELAR

PT / BR

original

jogadores;
remorso;
pessoas;
“amarelar nos
momentos decisivos”
...

humano

olhos
pele
unha

AMARELECER

PT

cor

relva
páginas
frutas
tintas
coletâneas
zonas
poluição
mármore
livros
...

humano

pele

original

sorriso
amarelecer de inveja

só 2 ocorrências em BR (CHAVE);

Tomou amarelo, que, somado ao segundo, **amareleceu** o espetáculo de domingo e o desta noite.

Aparecem coceiras na pele, que **amarelece** e ganha **feridas**.

**ESVERDEAR
REVERDECER
VERDEJAR**

esverdear

PT	24 (8 Vercial)
BR	2

reverdecer

PT	69 (65 Vercial)
BR	6

verdejar

PT	84 (78 Vercial)
BR	2

ESVERDEAR...

PT

Nenhum uso colorido!

...as indústrias; o sistema; projectos ; a actividade económica

VERDEJAR

PT

colorido?

verdejar e florir das suas cidades e vilas; áreas irrigadas, que **verdejam** como ilhas de prosperidade A paisagem ainda verdeja até Caruaru...

REVERDECER

BR

colorido!

várzea

colorido?

“Passeio Público (BR) reverdece e perfuma o centro do Rio”

“a vida é como a palmeira: murcha quando tudo **reverdece**”

“e novas esperanças **reverdecem** na gente”

original

original

“a esperança verdejou na alma”(BR)

“neblinas e fragâncias, dessas que, (...) continuam sempre a verdejar”(PT)

Outras cores verbais (sem considerar o participio)

AVERMELHAR

	CHAVE	CONDIV	Todos
PT	5	4	47*
BR	0	1	14

*23 no Vercial

ALARANJAR

	CHAVE	CONDIV	Todos
PT	1	0	15*
BR	0	0	2

*2 no Vercial

ARROXEAR

	CHAVE	CONDIV	Todos
PT	0	0	5*
BR	1	0	1

*4 no Vercial

ROSAR

	CHAVE	CONDIV	Todos*
PT	2	0	366
BR	0	0	194

*muitos erros do PALAVRAS (Rose)

ACINZENTAR

	CHAVE	CONDIV	Todos
PT	5	0	19*
BR	3	0	5

*2 no Vercial

Adjetivos coloridos

- no CHAVE e no CONDIV

- GRUPOS de cores mais divergentes entre variantes: cinzento; castanho e laranja.
- Casos em que a propriedade de colorir está mais distante?
 - “conta laranja”; “deputado laranja”; (BR)
 - “eminência parda”; “molho pardo”..
- Pardo → está perdendo a sua força de cor (em ambas as variantes)?
 - EMV: eminência parda, molho pardo; papel /envelope pardo
 - Em BR: cor:raça

CONSIDERAÇÕES FINAIS (1)

- Como esperado, verbos se prestam pouco a representar o processo de “colorização” das coisas.
- *BRANQUEAR* e *EMBRANQUECER* têm comportamentos diferentes não apenas quanto à frequência... branqueamos *dinheiro e imagem; cabelos e pele* embranquecem.
- *ENEGRECER* é muito mais usado em PT que em BR, e também é bastante usado de maneira original.
- *DOURAR* , em ambas as variantes, é mais usado de maneira original que de maneira “descritiva” . Em BR, tem amplo uso em culinária.

CONSIDERAÇÕES FINAIS... E outras questões...

- ? Nossa busca excluiu locuções verbais: usamos mais “ficar vermelho” do que “avermelhar”?
- ? O que nos diria a análise das cores no PCP?
- ? Será que, em termos gerais, os verbos associados a cores – em formas finitas - são usados sobretudo de maneira original? (e deixam de lado o vínculo com a cor?)

CONSIDERAÇÕES FINAIS (2)

- Para comparar termos de cor entre duas ou mais línguas é importante uma caracterização detalhada das cores em cada uma das línguas contrastadas.

...e também entre variantes

- A anotação das cores é uma tentativa de enriquecer os corpos (do AC/DC) com informação *analisada interpretada por pessoas*, o que implica na escolha de critérios e opções que nem sempre podem / têm de ser consensuais...
 - ... mas que podem ser criticados, mudados e postos em dúvida, visto que se encontram documentados...

Integração com a anotação semântica de outras áreas

Embora a anotação da cor tenha sido a primeira área levada a cabo no AC/DC, outras áreas estão em andamento:

- vestuário (Santos et al., em edição permanente)
- emoções (Maia e Santos, 2011)

(com a utilização dos mesmos mecanismos (e atributos))

Expressões metafóricas ou fixas envolvendo roupa & cor

- *colarinho branco, de ponto em branco..*

É possível estudar a relação entre as várias questões, procurando por [sema="roupa.*"] [sema="cor.*"]
, ou seja, roupa modificada por expressões de cor...

Referências

Ana Frankenberg-Garcia & Diana Santos. "COMPARA, um corpus paralelo de português e de inglês na Web". Cadernos de Tradução IX.1 (2002), pp. 61-79. Universidade de Santa Catarina. ISSN: 1676-7047. <http://www.linguateca.pt/documentos/Frankenberg-GarciaSantosCadTrad.pdf>

Susana Inácio, Diana Santos & Rosário Silva. "COMPARando cores em português e inglês". In Sónia Frota & Ana Lúcia Santos (eds.), Artigos seleccionados do XXIII Encontro da Associação Portuguesa de Linguística (APL) (Évora, 1-3 de Outubro de 2007), APL/Colibri, pp. 271-286. Slides <http://www.linguateca.pt/Diana/download/InacioSantosSilvaAPL2008.pdf>

Cristina Mota & Diana Santos. "Corte e costura no AC/DC: auxiliando a melhoria da anotação nos corpos".. Setembro de 2009. <http://www.linguateca.pt/aceso/corte-e-costura.pdf>

Diana Santos, Rosário Silva & Cláudia Freitas. "Pluralidades na cor: contrastando a língua do Brasil e de Portugal". In Augusto Soares da Silva, Amadeu Torres & Miguel Gonçalves (orgs.), 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, pp. 555-572. Slides: <http://www.linguateca.pt/Diana/download/PluralidadesCor2010.pdf>

Diana Santos, Stella E.O. Tagnin & Elisa Duarte Teixeira. "Colours, clothing and food in CorTrad: why corpus-based translation studies are revealing". Presentation at ICAME 2011 (Oslo, 1-5 de Junho de 2011).

Guy Deutscher. Through the language glass: Why the World Looks Different in Other Languages. Metropolitan: Henry Holt, 2010.

Agradecimentos

- O presente trabalho é financiado pelo governo Português, UMIC, FCCN e União Europeia (FEDER e FSE), por meio de POSC/339/1.3/C/NAC (Linguateca)

