

# Arco-íris: notas sobre a anotação do campo semântico da cor em português

Rosário Silva e Diana Santos

Versão 5.2

Última actualização: 16 de Agosto de 2012

(Primeira versão: 25 de Junho de 2009)

## 1. Introdução

Este documento é uma extensão de Silva *et al.* (2008c), documento que relatava a anotação do campo semântico da cor para o COMPARA (Frankenberg-Garcia & Santos, 2002), relativo à aplicação e eventual melhoria dos critérios sobre outros corpos do português, nomeadamente os do projecto AC/DC (Santos & Sarmento, 2003, Costa *et al.*, 2009, Santos, 2011).

Ao contrário de Silva *et al.* (2008c), não apresentamos uma panorâmica exaustiva dos dados encontrados, mas tentamos documentar as opções tomadas e os problemas existentes.

## 2. Questões técnicas e de utilização

Em conjunto com os atributos da análise sintática, podem ser utilizados nas pesquisas: o campo semântico *sema* (com que podem ser pesquisadas as várias classes da cor) e o campo com a informação sobre a cor chamado *grupo* (com que podem ser pesquisados os diversos grupos de cor, e chamado *cor* no COMPARA). Repare-se que estes atributos podem ser partilhados com outra informação, por exemplo outros campos semânticos como roupa ou emoção, como referiremos mais abaixo.

Este documento relata um trabalho em progresso, visto que, para alguns corpos, ainda nos encontramos na fase da deteção de candidatos adicionais para anotação automática, enquanto que para outros, já tratámos da revisão dos casos em contexto.

### SÍNTESE:

**Classes da cor:** *cor*, *cor:humana*<sup>1</sup>, *cor:vinho*, *cor:raça*, *cor:original*, *cor:política* e *cor:equipa*. Por agora, as classes marcadas ainda não foram totalmente revistas.

**Grupos de cor:** BRANCO, PRETO, AZUL, AMARELO, VERMELHO, LARANJA, VERDE, ROXO, CASTANHO, CREME, CINZENTO, ROSA, DOURADO, PRATEADO e também AUS\*ENCIA, OUTRAS.\*, MÚLTIPLA.\*, NÃOESPECIFICADA e DESCONHECIDA.<sup>2</sup>

Desta forma, os utilizadores têm à disposição pesquisas tais como:

```
[sema="cor.*"]
```

<sup>1</sup> Há dois tipos de *cor:humana*, o segundo marcado como *cor:humana2*, como explicado na secção correspondente.

<sup>2</sup> É importante esclarecer que só existe classificação de grupo de cor se a ocorrência for classificada *cor*, ou alguns casos de *cor:humana*, e não em todos os casos. Assim *branca* como *raça* não aparece no grupo Branco, tem sim um valor de 0 ou vazio no grupo. (Já agora, tanto 0 como nada têm o mesmo sentido no AC/DC.)

```
[sema="cor" & grupo="Azul"]3
[sema="cor:original"]
[grupo="Múltipla.*"]
<mwe> [sema="cor.*"] []* </mwe>
[lema!="amarelo" & grupo="Amarelo"]
```

### 3. Critérios linguísticos

#### 3.1 Opções de marcação da cor

No seguimento da metodologia utilizada em Silva et al. (2008c) para o português, complementamos e/ou adicionamos novas premissas:

- As expressões com mais de uma palavra, caso não sejam hifenizadas, e incluam já uma forma de cor, tais como *verde bilhar* e *azul gelo*, só têm marcadas com a indicação de `sema="cor"` as palavras *verde* e *azul*, ou seja, a primeira palavra da expressão. Mas encontram-se também marcadas como `<mwe>`, o mesmo se verificando para as expressões *cor de ...*. Contudo, nos casos em que a expressão `<mwe>` não começa pela palavra de cor, como em *buraco negro* ou *sangue azul* (ambas pertencentes à classe `cor:original`), é a primeira palavra da expressão que aparece marcada, tendo esta de ser considerada como um todo.
- Quando são utilizados objectos, materiais, etc. para indicar determinada cor, mas esta última não é especificada, somente a palavra *cor* é marcada. Por exemplo:

*Nem pelo formato, nem pela cor do papel, nem pela impressão é o Oslobodenje que conhecem há quase cinquenta anos.*

- Marcámos as ausências de cor no grupo Ausência, uma vez que fazem parte do campo semântico da cor<sup>4</sup>, ainda que como uma espécie de antónimos. Por exemplo: *hipertransparente, incolor, translúcido*. Contudo, ainda não revimos os casos em que essa ausência é utilizada no sentido figurado, por exemplo, e em que pretendemos remover a indicação de cor:

*par=2181: Afirmando a esperança, dizemos disponibilidade para olhar e ouvir; especialmente as sugestões ou as críticas transparentes.*

- Nos casos que interpretámos como anafóricos, considerámos (e anotámos) expressões (`<mwe>`) de tamanho 1. Assim, na frase seguinte, marcámos tanto *marinho* como *ultramarino* como cor, considerando-os como casos de expressão com várias palavras em que o primeiro elemento, *azul*, foi elidido, ou seja criámos regras de forma a marcar `<mwe lema="azul=marinho"> marinho </mwe>` e `<mwe lema="azul=ultramarino"> ultramarino </mwe>` nesse contexto.

*A cor dominante é o azul – marinho ou ultramarino, conforme a sensibilidade de cada um, mas sempre apelando à água em tempos de calma.*

Seguindo a mesma lógica, e devido à especificidade da estrutura gramatical utilizada, também marcámos *marfim* como se tivesse havido elisão do primeiro elemento da expressão de cor, neste caso *branco marfim* (`<mwe lema="branco=marfim"> marfim</mwe>`) no exemplo seguinte.

*par=moda5660: As cores vão da gama dos verdes, aos brancos óptico e marfim, aos pretos mates e brilhantes e aos vermelhos vivos e bordeaux.*

<sup>3</sup> Os nomes dos grupos têm a letra inicial maiúscula, para indicar que correspondem a uma classificação e não às formas ou lemas respectivos.

<sup>4</sup> Ver nota anterior.

Note-se que, no exemplo anterior, temos uma cor no plural e duas especificações dessa cor no singular. Nestes casos a primeira cor (*óptico*) é marcada como correspondendo a *brancos óptico*.

### 3.2 Diferenças em relação à anotação da parte portuguesa do COMPARA

Especificamente em relação às opções tomadas na anotação do COMPARA, tomámos uma decisão diferente em relação aos seguintes casos:

- Actualizámos a ortografia dos grupos NÃOESPECIFICADA (NAOESPECIFICADA no COMPARA) e MÚLTIPLA (MULTIPLA no COMPARA), acrescentando-lhes os respectivos acentos.
- O grupo MÚLTIPLA não recebeu por enquanto a informação de quais as cores envolvidas. Poderá eventualmente vir a ser incluído o nome dos grupos a que corresponde, ou seja, *branco-e-negro* ficará, tal como no COMPARA, como grupo="Múltipla:Branco-Negro" mas, por agora, essa questão não foi ainda implementada.
- Não marcámos como `cor` palavras estrangeiras que indicassem cor noutras línguas (*amarillo, azurra, azurro, blanco, blue, white, black-tie, white-metal*); contudo, marcámos aquelas que foram utilizadas propositadamente para dar ideia de uma cor (*bordeaux, fraise, fuchsia, royal, candy*).
- Aquando da actualização dos grupos de cor, considerámos pertinente efectuar algumas alterações pontuais, apresentadas na tabela 1.

Expressão de cor	Grupo no COMPARA	Grupo actual
cor de milho	Outras	Amarelo
cor de salmão	Laranja	Outras
cor de palha	Outras	Amarelo
cor de enxofre	Outras	Amarelo
cor de aço	Outras	Cinzento

Tabela 1: Alterações pontuais nos grupos de cor de algumas expressões

- Alargámos o âmbito de marcação da classe `cor:humana`, como explicamos na secção 3.3.2, adicionando `cor:humana2`.
- As expressões com várias palavras (<mwe>) foram classificadas gramaticalmente, nomeadamente como adjectivos ou nomes.

### 3.3 A questão da identificação da expressão de cor

É pertinente perguntar, em casos em que uma cor não é representada por uma palavra só, ou em casos em que estamos a falar de expressões com várias palavras cujo sentido envolve alguma conotação ou referência a cor, como é que decidimos marcar a cor.

Isto é importante tanto em relação à interpretação do que está marcado e à sua contagem, como à sua procura nos corpos.

Como veremos, muitas das opções são discutíveis e não encontramos por ora a forma ideal como deveriam ser tratados todos os casos.

Como já mencionado acima, em geral outras propriedades visuais de uma cor, como brilhante, metálico, esbatido, não são incluídas na expressão de cor.

#### 3.3.1 O caso de `cor:original`

A questão de mais difícil resolução, contudo, refere-se à atribuição da marca `cor:original`, que engloba todos os usos que não nos pareçam sobretudo

descritivos de uma questão visual, e que engloba de facto pelo menos dois tipos de questões:

- uso metafórico de uma cor, separado ou englobado em expressões fixas, tal como *por um unha negra* ou *em branco*, em que apenas a palavra de cor tem a marcação `cor:original`
- e termos que na sua génese contêm uma palavra de cor mas que são considerados unidades, e que por si não se referem a cor, tal como *buraco negro* ou *sorriso amarelo* e que são analisados como expressões de várias palavras recebendo a primeira unidade a marca de `cor:original`

Entre os dois temos ainda a questão do uso de termos que por vezes são descritivos mas que podem eles próprios ser usados metaforicamente, como é o caso de *cartão amarelo* e *sinal verde*. Optámos por os marcar sempre como `cor:original` (do segundo tipo) e não separar as duas questões. (Isso não se refere, naturalmente, a casos em que as palavras não constituem uma unidade, como no exemplo fictício *encontrei um cartão amarelo no chão com o desenho de um esquilo*, em que *amarelo* seria marcado simplesmente como `cor`).

Além disso, consideramos como parte integral de `cor:original` o uso metonímico da cor em casos como *dia cinzento* ou *manhã cinzenta*, em que a cor está a caracterizar uma parte (no caso o céu). Em casos em que *cinzento* pode referir-se tanto à cor como à conotação de sombrio, pesado ou chato, não tivemos outra alternativa senão marcar como vago entre `cor` e `cor:original`, ou [veja, cor:original](#).

### 3.3.2 O caso de `cor:humana`

Embora os casos prototípicos de cor humana sejam claros, como *loiro*, *moreno*, *ruivo*, *grisalho*, ou em casos como *olhos negros* ou *olhos verdes*, existe uma quantidade de expressões convencionalizadas referentes à aparência de certos atributos humanos (como pele, cara, cabelos e olhos) que são difíceis, mais uma vez, de fixar como `cor:humana` ou simplesmente `cor`.

No AC/DC, sempre que uma palavra de cor estivesse associada a atributos da aparência humana, tal como cabelo, olhos ou pele, marcámos como `cor:humana`.  
*par=66: A promotora fitou-o calmamente, com seus olhos verdes-cinzas.*

Mas deparámo-nos, naturalmente, com muitos outros casos: tentámos dividir entre expressões mais ou menos frequentes e mais ou menos irrealistas: *ficar vermelho* ou *roxo* não é exatamente literal,<sup>5</sup> *olhos cinzentos* também não são exatamente da cor das cinzas, mas considerámo-los `cor:humana` (possível). Já quando se pinta o braço de azul ou o cabelo de verde, considerámos simplesmente `cor` (porque totalmente artificial). Estamos contudo conscientes do alto grau de subjetividade da tarefa, e sugerimos que, nestes casos, os utilizadores usem procuras mais específicas (por exemplo `[lema="cabelo" [sema="cor.*"]]`), ou amalgamadas (`[sema="cor|cor:humana"]`), se especialmente interessados num atributo qualquer.

O problema principal é naturalmente os casos menos claros, como *enegrecer* no exemplo seguinte, que considerámos `cor:humana`.

*par=6: Não tinha as pernas rapadas e o cabelo, pintado, começava a enegrecer na raiz.*

---

<sup>5</sup>Note-se que *ficar vermelho de fúria* é um uso metafórico, e nesse caso é marcado como `cor:original`.

Após um período de indecisão, decidimos aceitar a proposta da Cláudia Freitas, inserida aqui:

A cor humana é especial porque temos dois tipos de informação: por um lado, há uma referência direta e explícita à propriedade colorida (cabelos louros; olhos azuis; pele morena); por outro, há, como nos outros exemplos, um uso que não é exclusivamente colorido, e que remete a sub-especificações: mulher loura/ruiva/morena. (a mulher loura não é dourada, mas o cabelo louro é dourado). Por isso, me parece que estamos juntando duas informações importantes em uma única etiqueta, e acho que é isso que sempre me traz dúvidas. Se considerarmos então os casos do tipo [cor:humana] como apenas aqueles em que o uso de uma propriedade de cor serve para informar uma subespecificação, teríamos cor:humana apenas em situações do tipo

enfermeira ruiva; homem grisalho; anjo louro (mas "anjo" seria humano?)

Com isso, perderíamos a informação - valiosa e que já temos anotada - do vasto uso das cores para fazer referência ao campo semântico humano: lábios roxos, bebê roxo (acho aqui que há sim referência à cor, não um roxo prototípico, mas uma cor arroxeadada); bochechas rosadas etc. Minha sugestão, nesses casos, seria incorporar a informação de Grupo nesses casos, sem perder a informação de sema.

A proposta consiste em separar suavemente os dois casos que estavam a ser cobertos pela categoria cor:humana, através da atribuição de grupo de cor a apenas um deles, que passou a ser chamado cor:humana2:

- os casos marcados lexicalmente, que se referem à aparência humana como um todo, como *louro* ou *ruivo*
- os casos que se referem a cores de atributos humanos, e que passam a ter também grupo de cor

Ainda em relação à classe cor:humana, os casos que incluem palavras de cor da lista desta classe, mas que não se referem a atributos humanos, foram reclassificados simplesmente para a classe cor (pura). Um exemplo disso é o seguinte caso de *louro*, considerado aqui no grupo Amarelo:

par=32044: *Se, em lugar de um romancista, fora qualquer um de nós a conceber tais considerações, não faltaria quem nos apelidasse de pequenos jordanos brunos, semeadores de joio daninho, em bons e louros trigais sem ventos...*

### 3.3.3 O caso de cor:raça

Da mesma forma, nem sempre é possível separar casos de menção da cor da pele de uma atribuição implícita ou explícita de raça. Nesses casos tivemos mais uma vez de recorrer a um marcação múltipla, de cor:raça\_cor:humana2, como no seguinte exemplo:

id="4409" titulo="cinema em preto branco take-3": *A Mulata Sedutora (Uma espécie de mulher-objeto cor de chocolate, desejada por todas as raças)* .

Nos casos que em que a palavra *branca* não está directamente relacionada com a raça, mas sim com a alvura da pele, classificámos as respectivas ocorrências como cor:humana2, no grupo Branco.

### 3.3.4 O caso de cor:vinho

Esta classificação é claramente uma subespecificação da cor para uma categoria específica, a do vinho, e também da cerveja. Existe pois cerveja loura e preta, e nesses

casos também foi marcada como `cor:vinho`. Poderia ser mais correto usar `cor:bebida`, mas não nos pareceu necessária essa correção.

### 3.4 Casos que podem suscitar dúvidas

Aqui relatamos as decisões que foram vindo a ser tomadas durante o processo de anotação mas cujo fundamento é fraco, e que podemos portanto reconsiderar no futuro.

- Decidimos não marcar os gentílicos cujos países, cidades, etc. de onde derivam têm alguma componente de cor, uma vez que nos pareceu já estarem demasiado distanciados do campo semântico da cor. Contudo, fica aqui a referência aos casos encontrados: *amarelejenses*, *verdense*, *caboverdeano*, *caboverdiano*, *cabo-verdiano*, *castanheirense*, *monteverdiano*, *vilaverdense*, *vilaverdenses*, *ouropretano*, *rosarenses*.
- Decidimos também não marcar alguns casos que podem fazer parte do campo semântico da cor, mas num contexto mais abrangente: *bem-pintado*, *bicromático*, *corante-revelador*, *despigmentadas*, *despigmentado*, *escurecer*, *garrir*, *iridiscente*, *matizar*, *mesclar*, *monocromo*, *pigmentado*, *pigmentar*, *pigmentário*, *pintadinho*, *pintalar*, *pintalgar*, *pintar*, *repintar*, *tingir*, *transparecer*, *variegar*. Contudo, convém referir que (ainda) não retirámos os lemas equivalentes encontrados anteriormente no COMPARA, ficando esta questão em aberto, para futuro tratamento. A dificuldade prende-se com o estabelecimento de limites para o que é ou não cor, o que transmite ou não cor, o que remete ou não para cor, e se devemos marcar todas essas nuances ou não.
- Na sequência da não marcação de cores numa língua estrangeira, não marcámos expressões como: *black-propaganda*, *blackmail*, *green-mail*, *red-devils*, *whiteano*. Contudo, convém referir que, por vezes, algumas palavras em inglês poderão estar incorrectamente marcadas como cor devido ao facto de terem uma grafia igual a uma palavra de cor portuguesa. Assim, a palavra inglesa *core*, igual à palavra portuguesa *core* (do verbo corar), poderá aparecer nas pesquisas que envolvam a classe `cor:humana` – enquanto todo o material não tiver sido revisto. Por exemplo:

*dt=2003-06-26: Equipamentos, softwares e processos de produção desenvolvidos aqui, com base na experiência do telejornalismo e do videomaking, podem e devem ser transferidos às indústrias que disponham de core competence para implantá-los em escala', defendeu Lage.*

- Palavras que podem pertencer a mais do que um campo semântico: Os casos *bi-tonal*, *bitonal*, *não-tonal* tanto podem fazer parte do campo semântico da cor como da música. Contudo, como (ainda) não identificámos ocorrências destas palavras nos corpos com significado de cor, não marcámos estes casos.
- Erros no corpo: Considerámos que seria preferível corrigir os erros em vez de marcar palavras inexistentes, mas isso significa que podem ainda existir nos corpos ocorrências de cor, mal grafadas, que não se encontrem marcadas.
- Nos casos em que surgiram dúvidas na atribuição de grupos por motivo de haver várias cores associadas à expressão, tanto por a) incapacidade de identificar a cor de algo como por b) existir efectivamente mais do que uma cor para uma dada coisa, optámos por atribuir o grupo Outras. No caso c) que engloba expressões cuja cor nos pareceu impossível de determinar, usámos o grupo Desconhecida. No primeiro caso, surgiram dúvidas quando à expressão *cor de leão*, dado que não foi possível estabelecer se um leão é castanho-amarelado ou amarelo-acastanhado. Como exemplo do segundo caso temos *cor de ameixa*, dado que as

ameixas podem ser amarelas, avermelhadas, arroxeadas ou negras. No último caso (cor Desconhecida), temos o exemplo de *cor de apoplexia*, ou *rachei*.

- Embora expressões como *cartão vermelho*, *cartão amarelo*, tivessem sido inicialmente marcadas como pertencendo à classe `cor:equipa` devido à sua relação com o futebol, decidimos posteriormente que, por não se referirem à cor de uma equipa, mas sim à cor de um objeto com uma função bem definida e também já usado metaforicamente em outros domínios, a sua classificação correta seria a classe `cor:original`.
- Por enquanto, ainda não tratámos consistentemente o caso das raças de animais, sendo que existem várias denominações que envolvem palavras de cor, tal como *urso pardo*, *vacas pardas*, ou mesmo casos mais complexos, como o exemplo abaixo, mas por agora estamos inclinadas a considerar `cor:original`.

*Da mesma forma, existem dois tipos de atletas, os pardos (os pássaros jovens, até cinco anos em média)*

### 3.5 Casos relativos à classificação gramatical

Em vários casos não nos foi difícil optar por uma classificação semântica de cor, mas tivemos dúvidas na classificação gramatical desse tipo de ocorrências, que fica assim aqui documentada para os utilizadores do AC/DC poderem saber como estão classificadas estes casos:

- em expressões *de* seguido de palavra de cor (como *de negro*), ou *em* seguido de palavra de cor, assumimos que a palavra de cor era adjetivo

*id=«El-Rei Junot Prosa RB=1903»: Onde tudo isto já cediço vai, com trejeitos fora de moda, risos, o famoso David Peres a ensinar-lhes árias, e serões, galas, homens vestidos **de rosa e verde** e agaloados de prata! ...*

*Cidades e povoados também surgem em **rosa, lilás e tons de vermelho***

- em expressões em que as palavras de cor eram usadas predicativamente, também as considerámos adjetivos

*id=«Alba Plena Poesia AGil=1916»: Todo ele é **oiro e carmesim e rosa**, Noiva que vai casar, sorrindo airosa...*

### 3.6 Casos pontuais

- A frase seguinte foi considerada um caso especial, uma vez que tudo o que se encontra mencionado se refere inequivocamente a cores. Desta forma, para além das marcações normais, também marcámos como cor «tons de especiarias» e «pastéis» (este último no pressuposto de que se tratava da alusão a cores pastéis).

*P940123-103: As cores são as harmonias quentes, chocolate, camelo, tons de especiarias e os **pastéis**.*

- Quando as cores correspondem a grupos, tal como *cores de especiarias*, em que interpretamos a expressão como referindo-se a várias cores, marcamos a expressão total como cor, de grupo Nãoespecificada.

No caso da expressão *cores vermelhas* na frase seguinte, adoptámos o mesmo critério, marcando a totalidade da expressão, só que desta vez no grupo Vermelho.

*E055-PT-83: Aquelas fardas bizarras, que aquilo é fantástico tudo de veludo a **cores vermelhas**, os chapéus.*

- Quando há uma sequência de especificação de cores, tal como em *Rímel **colorido azul marinho** ou castanho na ponta dos cílios também dão cor*

embora na sequência apenas uma cor esteja explícita, consideramos que o verbo *colorir* também contribui com a especificação cor de grupo NãoEspecificada.

- Embora não tenhamos conseguido descobrir o significado da expressão socialismo moreno na frase seguinte, optámos por incluir a palavra **moreno** na classe cor:política, dado que surge num contexto político.

*par=8084: 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*

Uma solução diferente, nomeadamente cor:ausência, foi por agora escolhida, devido ao resto da frase, para o seguinte caso de *parda*, embora não seja evidente se não devêssemos considerar como cor:política ou como cor:original (considerando *pardo* como significando simplesmente chato, sensorão, ...)

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

### 3.7 Diminutivos, aumentativos e superlativos

Os diminutivos, aumentativos e superlativos foram anotados com essa informação, mas constam dos lemas de raiz (ou seja, azulinho é lema azul, diminutivo), à exceção daqueles que nos pareceu serem itens lexicais diferentes e para os quais fizemos regras específicas para serem identificados separadamente ou não aplicámos regras de análise. Os casos que mantivemos como novos lexemas são:

- cinzentão
- amarelinho
- vermelhão
- marronzinho
- amarelinha
- amarelão
- verdão

Em alguns casos de superlativo, entramos numa zona cinzenta em que mais uma vez não sabemos se é uma nova cor ou se é apenas a modificação ou superlativização de uma característica, tal como em:

*«Rose Brume» (nº 18 -- rosa acastanhado médio) ; «Mandarine» (nº 11 -- cor de laranja claro) ; «Bossa Nova» (nº 2 -- **laranja claríssimo**) ; «Safari Red» (nº 13 -- bronzeado) ; «Rose Pastel» (nº 3 -- rosa) e «Rose Orange» (nº 10 -- rosa alaranjado) .*

Por agora considerámo-lo o superlativo de *laranja claro*. Este caso ilustra contudo claramente como a decisão de não incluir modificadores livres tal como escuro e claro levanta problemas: se por um lado a marcação com hífen indica que a pessoa considera uma única cor e não uma gradação de cor, por outro lado as normas ortográficas aceitam as duas grafias, e poder-se-ia defender que *verde claro* e *azul escuro*, por exemplo, são outras cores, mais específicas.

De momento, um utilizador terá de procurar [sema="cor"] [lema="claro|escuro"] para obter esses casos.

## 4. Questões associadas a géneros textuais diferentes

Em diferentes corpos é possível que nos surjam formas diferentes de realizar e exprimir a cor, dependentes do género textual. Assim, fazemos aqui algumas observações que

nos pareceram pertinentes quanto a esse assunto, à medida que os primeiros corpos foram sendo anotados e cuidadosamente considerados.

#### 4.1 CONDIVport, parte da moda

No campo específico da moda existem vários termos que são cor nesse contexto mas não em outros, e sobretudo existem formas que em contexto são compreendidas como cores embora possam necessitar de esclarecimento noutros contextos, como o caso de *marinho*. Como exemplos, veja-se os seguintes casos:

*par=moda1182: Saia de lã pied-de-poule com bolsos aplicados em viés e cinto de cabedal **turquesa**.*

*Par=moda22472: Crepe cristal suíço, desenhos em tons de **marinho**, preto, vermelho, verde e **azul-royal** destacados sôbre fundo branco...*

*par=moda12720: O casaco e o vestido são guarnecidos com incrustações de cores contrastantes, branco, **gerâneo** e cor de vinho.*

*Par=moda22914: Encantadoras cores lisas: ciclame, champanhe, lilás, coral, rosa, **maravilha**, salmon, royal, verde-jade, resedá, turquesa, cinza, fraise, pérola, marrou, preto e brancos.*

#### 4.2 CONDIVport, parte do futebol

No CONDIVport referente ao futebol existe um número muito inferior de lemas diferentes de cor comparado com o registado no CONDIVport moda.

Pudemos no entanto constatar que, no caso do futebol, é comum a utilização das cores dos clubes ou das camisolas dos clubes para referir a própria equipa ou os seus elementos, nomeadamente *encarnado*, *rubro-negro*, *alvinegro*, *rubro-verde*, *verde-branco*, etc. Por exemplo:

*par=fut1945: Os **encarnados** chegaram ao intervalo a perder por 1-0, golo obtido por Sérgio, aos 42 m., na marcação de um canto.*

*Par=fut6981: Da conjugação destas circunstâncias não foi difícil aos **alvinegros** transformarem um zero-dois em dois-dois.*

#### 4.3 CONDIVport, parte da saúde

Existem mais termos de cor do que estávamos à espera neste contexto, descrevendo secreções ou aparência de situações de doença. São exemplos disto:

*par=saude3104: É pela coloração **rósea** da mucosa bucal e conjuntival que se procura geralmente observar o grau de anemia...*

*par=saude7850: A lesão começa sob forma papular, evolui para ulceração de aspecto serpiginoso, de superfície **avermelhada**.*

Outros casos correspondem a cor:original, tal como em *febre amarela*, já encontrado (uma vez) no COMPARA mas que neste domínio tem naturalmente uma frequência muito superior. Veja-se este e outros exemplos:

*par=saude7944: O autor inicia com o relato das doenças da época: varíola, febre **amarela**, disenterias...*

*par=saude106: A própria palavra – poliomielite – é misteriosa e vem do grego, significando massa **cinzenta**.*

*Par=saude249: também mortos pela peste **branca** .*

*par=saude12913: Baixa da taxa de homoglobias (glóbulos **vermelhos**) no sangue .*

#### 4.4 ENPCPUB

Este corpo, sempre correspondendo a português traduzido do inglês, contém o mesmo tipo de texto que o COMPARA (e, de facto, duas das obras são comuns), por isso não há muitas diferenças a mencionar. É contudo interessante indicar que mesmo assim novos termos foram encontrados.

## 5. Exemplos de procuras

Aqui apresentamos algumas procuras mais complexas para explicar como ter acesso à informação de cor presente no AC/DC.

Por exemplo, para procurar casos de expressões com várias palavras que signifiquem cor, e visto que é sempre a primeira palavra da expressão que está marcada com os atributos sema e grupo, uma forma possível de procurar será especificar:

```
<mwe> [sema="cor.*"] []* </mwe>
```

Enquanto que, para procurar uma expressão específica, é possível usar

```
<mwe> "azul" "eléctrico" </mwe>
```

Para ver todas as expressões com mais de uma palavra num corpo que envolvam cor, em contexto, a expressão de procura ideal será, por outro lado:

```
[sema="cor.*"] within mwe expand to mwe
```

Se apenas se pretende expressões com mais de uma palavra que elas próprias sejam cor, então a procura será

```
<mwe> [sema="cor"] expand to mwe
```

Para ver todas as ocorrências incluídas no grupo OutrasCores, a procura será

```
[sema="cor" & grupo="OutraCores.*"]
```

## 6. Listagem das palavras de cor em português

Todas as palavras e expressões de cor que identificámos como tal nos corpos do projecto AC/DC, assim como as que já tinham sido encontradas no COMPARA, encontram-se disponíveis em <http://www.linguateca.pt/acesso/cores.txt> (tanto lemas de palavras simples como de expressões com mais de uma palavra).

Alternativamente, apresentamos no ficheiro [http://www.linguateca.pt/acesso/todas\\_cor.txt](http://www.linguateca.pt/acesso/todas_cor.txt) o mesmo material, mas com a informação adicional, para cada expressão, de a que grupo(s) pertence.

### 6.1 Critérios relativos à identificação de palavras simples de cor

Nas listas referidas acima, as palavras encontram-se listadas através dos respectivos lemas. Convém a esse respeito lembrar que, no caso dos adjectivos, o lema aparece na sua forma masculina singular (por exemplo: o lema de *amarelas* é *amarelo*), ao passo que, no caso dos substantivos, o lema aparece apenas na sua forma singular, podendo ser masculino ou feminino (por exemplo: o lema de *brancos* é *branco* e o lema de *brancas* é *branca*).

No que se refere aos aumentativos, diminutivos, superlativos, também os lemas destes aparecem na forma base (a título de exemplo: o lema de *vermelhinhas* quando adjectivo é *vermelho*, o lema de *louríssimas* quando substantivo é *loura*).

Visto que a análise sintáctica pode não estar correcta, ou ser susceptível de melhorias que, por enquanto, não estamos a efetuar, deixámos alguns casos dúbios (como lemas) nas listas que aqui apresentamos. Sobretudo a questão da vagueza entre possíveis participios passados ou adjectivos não foi ainda tratada, e por isso mantivemos casos como *aloiado* e *aloiar* na lista. Nos casos em que nos pareceu que o

aumentativo poderia ter uma acepção ou uma conotação de cor distinta (tal como em *amarelão*, *cinzentão*), não considerámos como lema *amarelo*, e sim o próprio *amarelão*. Exemplos disso são:

*par=ext494318-nd-95a-1: Fernando Nogueira é cinzento, tremendamente cinzentão .*

*par=7362: Deve ser lembrado que a propaganda de cuidados individuais, para evitar o amarelão (ancilostomíase) e outras enfermidades causadas por vermes do solo...*

Um caso singular com que nos deparámos foi a palavra *verduras*, para a qual detetámos três acepções diferentes. A primeira acepção, em que *verduras* equivale a zonas verdes, está a ser marcada automaticamente como `cor`. A segunda, em que *verduras* equivale a hortaliça, não iremos marcar como `cor`. A terceira acepção, em que *verduras* equivale a algo jovem, imaturo, etc., decidimos marcar como `cor:original`.

## 6.2 Critérios relativos à identificação de expressões com várias palavras de cor

Assim como no caso das palavras simples, também as expressões com várias palavras são identificadas pela sequência dos respectivos lemas. Assim teremos o elemento *cor de rosa* em vez de *cores de rosa* na lista.

## 6.3 Grupos de cor

Os grupos de cor correspondentes à presente documentação encontram-se acessíveis em [http://www.linguateca.pt/acesso/Grupos\\_cor.txt](http://www.linguateca.pt/acesso/Grupos_cor.txt), e vão sendo actualizados à medida que o trabalho prossegue.

Devem ser contudo lidos e interpretados da seguinte forma: se a palavra em questão é uma palavra que representa cor no contexto em que é empregue, então o grupo a que pertence é dado pela lista que consta no ficheiro.

Ou seja, a listagem dos grupos não significa que todas as palavras indicadas como pertencendo ao grupo lhe pertençam **sempre**.

## 7. Regras de anotação condicional

De forma a tornar o processo de anotação mais célere, iniciámos em paralelo um processo de criação de regras automáticas de anotação de cor para os casos de palavras ou expressões em contexto. Exemplos são a questão da desambiguação de sentidos através dos outros elementos da frase ou através do tema, ou regularidades que, não sendo uma regra absoluta da língua, foram encontradas nos próprios corpos que pretendemos anotar.

Para pôr em prática essas regras foram criados cinco ficheiros distintos, que passamos a explicar em seguida.

O primeiro caso desta questão são as expressões com várias palavras, tal como *cor de ameixa*, *azul noite* ou *peito de rola*, que marcamos como `<mwe>`, tal como fizemos em geral na análise sintática (ver Inácio & Santos, 2008).

```
<mwe lema="peito=de=rola" sema="cor" pos="ADJ" grupo="Outras">
peito
de
rola
</mwe>
```

Como sabemos que a anotação sintáctica foi feita automaticamente e inclui, portanto, alguns erros que prejudicam a anotação da cor, criámos também algumas regras de correcção da sintaxe anteriores à marcação da cor.

No ficheiro `regras_corr_PALAVRAS_cor.txt`, colocam-se as regras que pretendem corrigir alguns erros comuns na própria análise sintáctica, tanto em termos de lemas como em termos de categoria gramatical (pos). Por exemplo, palavras como *Marinho* ou *Branco* correspondentes a apelidos de pessoas, muitas vezes consideradas como adjectivos pelo PALAVRAS, devem ser transformadas em nomes próprios para evitar a sua marcação como cor. Uma situação semelhante é a das palavras que apenas indicam cor quando adjectivos, tendo um sentido completamente diferente como substantivos, como é o caso de *alvo* ou *alvos*. Nesse casos e visto que a atribuição da semântica de cor é dependente da categoria gramatical, convém corrigir antes de colorir. Um exemplo de correcção apenas de lema é a palavra *ultravioleta*, a que o PALAVRAS por vezes atribui erradamente o lema *ultravioleto*. Nesse caso é o valor do lema que tem de ser corrigido, ao invés de adicionar lemas "erróneos" às nossas listas.

No ficheiro `regras_positivas_cor.txt` tratamos das palavras que na grande maioria das suas ocorrências não são cor, mas que em contextos particulares se verifica serem exemplos de cor. Por exemplo, se numa frase encontrarmos referência a um *vestido tom de café*, neste caso *café* será cor, pelo que terá de ser marcado.

O ficheiro `regras_negativas_cor.txt` é usado nos casos opostos, ou seja, para palavras cuja grande maioria das ocorrências seja cor, mas em que existem alguns casos que o não é. Vejamos o caso de alguém que «colhe os louros»: regra geral, o lema *louro* está a ser marcado automaticamente como cor, mas neste caso não é, pelo que é preciso retirá-lo dos casos de cor.

Desde o início que separámos as palavras que eram sempre cor daquelas que apenas o seriam se usadas, por exemplo, adjectivamente, como é o caso de *laranja*, *rosa* ou *cinza*. Contudo e como já discutido em Inácio *et al.* (2008), no processo de lexicalização dessas novas cores existem já alguns casos em que são cores e são substantivos (embora notemos que mudaram de género). Por isso pode convir que sejam criadas regras específicas para, nos casos em que a palavra é substantivo, também seja cor.

Em seguida, empregamos o ficheiro `regras_especialização_cor.txt` para os casos marcados inicialmente como cor, mas que necessitam de ser alterados ou especializados, quer sejam casos de vagueza, raça, original ou outros.

Vejamos o que acontece quando um lema de cor pertence a duas classes de cor distintas, por exemplo, `cor` e `cor:raça`. Tomemos como exemplo o lema *branco*. Este lema é marcado automaticamente como cor, contudo, sabemos que em muitos casos essa cor se refere especificamente a raça. Assim, criámos a seguinte regra: o lema *branco* (e *preto*) pertence à classe `cor:raça` quando é precedido por um dos lemas "homem|mulher|criança|rapaz|rapariga|raça|pele|pessoa|..." ou quando é o nome predicativo do sujeito numa primeira pessoa.

Demo-nos conta de que por vezes, à semelhança do que acontece com as regras de cor não especializadas, é preferível formular uma regra geral e retirar depois algumas poucas exceções. Criámos por isso ainda mais uma iteração, o ficheiro `regras_desespecialização_cor.txt` para esses casos. Assim, se tivéssemos marcado todas as ocorrências do lema *negro* como raça, poderíamos aqui passar para cor (pura) os casos que tivessem sido erroneamente especializados no passo anterior.

Finalmente, ainda existe um ficheiro adicional, `regras_recurativas_cor.txt` para os casos de regras recursivas, ou seja, de

regras que recorrem às próprias classificações (de cor) para adicionar mais classificações, e que portanto podem ser aplicadas várias vezes. Um exemplo é o da conjunção de cores.

Este conjunto de regras (identificadas a partir da análise dos próprios corpos do AC/DC) permite ter uma maior correcção nos casos anotados automaticamente, resultando assim em menos casos para uma revisão humana.

Além disso, e para garantirmos uma anotação perfeita, também reunimos “regras” específicas para cada corpo que tratam dos últimos casos mais idiossincráticos: há casos muito específicos que não faz sentido estarem incluídos nas regras gerais, dada a baixa probabilidade de se repetirem nos outros corpos. As regras exclusivas estão divididas da mesma forma que as gerais, ou seja, existem ficheiros de regras exclusivas *corr\_PALAVRAS\_cor*, *positivas\_cor*, *negativas\_cor*, *especialização*, *desespecialização* e *recursivas*. Foi necessário seguir esta metodologia para que a aplicação das regras produzisse o efeito desejado.

Para mais informação sobre a sintaxe das regras e a forma da sua aplicação, veja-se Mota & Santos (2009) e Santos & Mota (2010).

## 8. Resumo quantitativo

Aqui indicamos, por corpo, quantas palavras de cor foram obtidas, à medida que os diversos corpos foram sendo tratados e anotados, separando a questão de quantos termos de cor (por lema) do número real de ocorrências correspondente à cor.

A Tabela 2 apresenta, por corpo<sup>6</sup>, quantas palavras de cor existem (Palavras), correspondentes a quantos lemas diferentes (Tipos). De forma a ser possível comparar diferentes corpos, também os seguintes índices são apresentados: **Índice<sub>pal</sub>** que descrever o número de palavras de cor por número de palavras do corpo (x 10.000) e **Índice<sub>tip</sub>** que corresponde ao número de tipos diferentes por número de palavras diferentes do corpo (x 10.000).

Corpo	Palavras	Tipos	Índice <sub>pal</sub>	Índice <sub>tip</sub>	Data
AMOSTRA	203	57	20,608	52,164	4 Ago 2010
ANCIB	249	37	1,978	5,227	6 Ago 2010
AVANTE	2.667	121	4,102	13,058	21 Mai 2010
CDHAREM	213	47	9,435	22,733	5 Ago 2010
CETEMPUBLICO	141.040	626	7,440	4,689	29 Ago 2009
CHAVE	82.571	586	8,300	6,094	29 Mai 2010
COMPARA (port)	4.829	305	30,348	76,024	31 Dez 2008
CONDIV	20.061	607	35,968	53,993	1 Nov 2010
CONE	247	31	3,605	8,321	25 Mai 2010
CorTrad culinário	977	39	64,006	79,931	23 Jun 2010
CorTrad jornalístico	650	83	8,381	25,594	23 Jun 2010
CorTrad literário <sup>7</sup>	392	55	50,947	82,422	28 Jun 2010
DIACLAV	4.490	119	6,750	13,532	22 Mai 2010
ECI-EBR	1.191	164	16,450	46,140	6 Ago 2010
ECI-EE	2	1	0,737	3,622	3 Ago 2010

<sup>6</sup> Para efeitos de comparação, embora imperfeita, colocamos também os (antigos) valores calculados para o COMPARA.

<sup>7</sup> No caso do CorTrad literário, os números que nos parecem relevantes são os relativos à tradução final.

ENPCPUB	257	64	35,509	80,140	3 Nov 2010
FLORESTA	6.748	214	11,341	12,775	24 Jul 2010
FRASESPB	20	13	10,437	30,559	4 Ago 2010
FRASESPP	22	16	13,553	44,053	4 Ago 2010
MUSEUDAPESSOA	166	28	4,425	22,136	30 Out 2010
NATMINHO	733	71	4,191	12,163	Ago 2010
NILC/SAOCARLOS	28.302	410	8,751	8,754	6 Jun 2010
VERCIAL	25.059	330	22,817	12,699	18 Set 2010
TOTAL	328.633	1092			

Tabela 2: Distribuição da cor pelos corpos. A coluna "Data" indica a última vez que cada corpo foi anotado com cor e isso se refletiu em mudanças nos valores da tabela

Para alguns corpos, para ter uma visão mais clara é preciso separar entre os diversos domínios ou géneros. Assim, apresentamos, na Tabela 3, a distribuição da cor pelos três domínios temáticos do CONDIV. A Tabela 4 ilustra a distribuição pelos diversos géneros do corpo NILC/São Carlos.

Parte	Tam. pal	Tam. tipo	Palavras	Tipos	Índice <sub>pal</sub>	Índice <sub>tip</sub>
moda	1104337	33795	13880	490	125,686	144,992
futebol	3005675	61015	3980	134	13,242	21,962
saúde	1806126	46996	2201	157	12,186	33,407

Tabela 3: Distribuição da cor nos três temas do CONDIV, a 1 de Novembro de 2010

Parte	Tam. pal	Tam. tipo	Palavras	Tipos	Índice <sub>pal</sub>	Índice <sub>tip</sub>
Didáctico	425690	21586	469	54	11,017	25,016
ENCiclopédico	286550	22956	312	52	10,888	22,652
ENSaio	2169720	58910	1615	94	7,443	15,957
EPistolar	3332	904	0	0	-	-
Jornalístico	29073909	400603	23523	362	8,091	9,036
LEgal	111772	34090	65	8	0,582	2,347
Literário	920930	26124	2143	139	23,270	53,208
REvista	152366	13585	175	51	11,486	37,541

Tabela 4: Distribuição da cor pelos géneros do NILC/São Carlos, a 6 de Junho de 2010

Vemos assim que há géneros e tipos de texto mais ricos em palavras de cor (tal como o literário ou o técnico referente a moda) e outros que são muito pobres, tal como o legal, ou a linguagem falada. Pode também ter interesse comparar a proporção entre os tipos de cor (*cor:equipa*, *cor:política*, *cor:original*, etc.) nos diferentes corpos, ou comparar as variantes do português envolvidas.

Para este último objetivo, criámos as seguintes coleções de expressões relativas à cor que temos vindo a apurar como apenas pertencendo a uma variante, respetivamente nos ficheiros *cores\_PP.txt* e *cores\_PB.txt*, e para o qual a contribuição da Cláudia Freitras tem sido valiosíssima.

## 9. Análise das cores marcadas como pertencentes ao grupo Outras

No que se segue, tentámos classificar as cores marcadas como pertencentes ao grupo Outras, após termos em vão procurado novos grupos de cor que merecessem identificação. O resultado foi a seguinte separação por origem da nova cor.

### **9.1 OUTRAS frutas e legumes**

alperce, ameixa, azeitonado, beringela, cor de abrunho, cor de ameixa, cor de amêndoa, cor de amora escura, cor de amoras escuras, cor de azeitona, cor de azeitona, cor de batata, cor de beringela, cor de citrina, cor de cuia, cor de damasco, cor de gengibre, cor de ginja, cor de mamão, cor de melão, cor de melão, cor de oliva, cor de pêssego, cor de pistácia, cor de sopa de beterrabas, cor-de-ameixa, damasco, goiaba, melão, melão pálido, oliva, uva.

### **9.2 OUTRAS metais e minerais**

âmbar, ambarino, areia, bronze, brônzeo, cobre, cobrear, cor de âmbar, cor de âmbar claro, cor de areia, cor de areia, cor de areia ou terra queimada, cor de areia quente, cor de argila, cor de bronze, cor de bronze, cor de cobre, cor de cobre, cor de esmalte, cor de ferro velho, cor de granito, cor de greda, cor de jade, cor de jaspe, cor de mármore, cor de ocre, cor de opala, cor de pedra, cor de terra, cor de topázio, marmóreo, marmorizado, ocre, ocre-carregado, ocre-claro, terra, terroso, turmalina.

### **9.3 OUTRAS conceitos impalpáveis**

cor de «acid trip», cor de alegria, cor de apoplexia, cor de cólera, cor de crime e traição, cor de fada, cor de inferno, cor de luz, cor de morte, cor de saúde.

### **9.4 OUTRAS flores**

cor de cravo, cor de flor de alecrim, cor de gerânio, cor de glicínia, cor de malva, cor de malva, cor de malva rosa, cor de malva-rosa, cor de tília, cor de tília, cor do miosótis, cor-de-papoila, gerâneio, malva.

### **9.5 OUTRAS animais**

cor de burro, cor de burro quando foge, cor de burro-quando-foge, cor de camurça, cor de castor, cor de leão, cor de leopardo, cor de ostra, cor de pombo, cor de salmão, cor de salmão cozido, cor de veado, peito de rola, salmão, salmon.

### **9.6 OUTRAS fenômenos naturais**

aurora boreal, cor da noite escura, cor da treva, cor de fogo, cor de fogo, cor de fogo e bosque, cor de fumo, cor de lua, gelo.

### **9.7 OUTRAS ainda sem classificação**

adamascado, brique, candy, caqui, cáqui, champanhe, cor de açúcar, cor da açúcar queimado, cor de barbante, cor de boca, cor de borra de vinho, cor de borra de vinho no fundo, cor de cana, cor de caqui camuflado, cor de carne, cor de carne, cor de cera, cor de cidra, cor de Coca-Cola, cor de colmo bolorento, cor de coral, cor de coral, cor de ferrugem, cor de ferrugem, cor de ferrugem metalizada, cor de ferrugem-carregado, cor de gelado, cor de gelo, cor de gema de ovo, cor de granada, cor de iodo, cor de jaca passada, cor de lama e lodo, cor de lodo, cor de luvas cirúrgicas, cor de massa de vidraceiro, cor de moringa, cor de ossos, cor de pão queimado, cor de pão queimado, cor de parma, cor de pó de tijolo, cor de poeira, cor de praline laranja, cor de rebuçado, cor de rebuçado, cor de sebo, cor de sulfato de cobre, cor de tabaco, cor de tabaco, cor de tabaco de Havana, cor de tabaco loiro, cor de teia de aranha, cor de telha, cor de tijolo, cor de tijolo, cor de torrado, cor-de-caca, empardecer, ferrugem, fulvo, havana, maravilha, transparose.

## 10. Integração com a anotação semântica de outras áreas

De notar que embora a anotação da cor tenha sido a primeira área levada a cabo no AC/DC, estamos a trabalhar em outras questões, como o vestuário (Santos et al., em edição permanente) e as emoções (Maia e Santos, 2011), e esse trabalho leva a que os mesmos mecanismos (e atributos) sejam usados.

Notamos assim que existem expressões metafóricas ou fixas envolvendo tanto roupa como cor, e que poderão ou deverão estar marcadas como pertencendo a ambos os campos, tal como *colarinho branco, de ponto em branco, camisola amarela, roupa branca*, etc.

Além disso, é possível estudar a relação entre as várias questões, procurando por exemplo casos de [sema="roupa.\*"] [sema="cor.\*"], ou seja, roupa modificada por expressões de cor, ou mesmo casos de medo classificado por expressões de cor, [sema="medo"] [sema="cor.\*"].

## 11. Tarefas ainda por executar

Anotamos aqui algumas das tarefas que é preciso realizar consistentemente para todos os corpos do AC/DC e que ainda não foram iniciadas ou pelo menos consistentemente levadas a cabo:

Tratar dos diminutivos, aumentativos e superlativos, de forma a que os lemas das palavras de cor estejam corretamente indicados (já tratámos de alguns, mas ainda não exaustivamente).

Tratar de obter automaticamente os grupos das cores múltiplas, de forma a ter, por exemplo para *preto-e-branco*, **ambos** os grupos Preto e Branco...

## 12. Agradecimentos

Agradecemos à Cláudia Freitas as observações muito pertinentes sobre este texto e sobre a anotação semântica em geral e sobre as cores em particular. É apenas por uma questão de simplificação no processo de documentação que não a adicionamos como co-autora, mas referimos as suas opiniões e contribuições sempre que necessário.

Também estamos obviamente gratas à Cristina Mota pelo desenvolvimento do corte-e-costura (de que é a principal autora, em conjunto com Diana Santos), e por várias observações sobre a anotação semântica.

Embora seja óbvio, fica sempre bem indicar que, dado que este trabalho é a continuação lógica da anotação da cor no COMPARA, estamos gratas a todo esse trabalho preliminar e ao resto da equipa que conosco o efetuou, muito especialmente a Susana Inácio.

Este texto é escrito no âmbito da Linguateca, cuja responsabilidade administrativa pertence à FCCN, e que tem sido financiada por vários organismos do governo português e da administração pública portuguesa, a saber: a ADI, o POSI, o POSC, a UMIC, o MCTES, a FCCN e a FCT.

## Referências

[Costa et al. 2009]

Luís Costa, Diana Santos & Paulo Alexandre Rocha. "Estudando o português tal como é usado: o serviço AC/DC", *STIL 2009, The 7th Brazilian Symposium in Information and Human Language Technology* (São Carlos, Brasil, 8-11 de Setembro de 2009).

[Frankenberg-Garcia & Santos 2002]

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.

[Inácio & Santos 2008]

Susana Inácio & Diana Santos. "Documentação da anotação morfossintáctica da parte portuguesa do COMPARA". Versão 3.4, 31 de Dezembro de 2008. <http://www.linguateca.pt/COMPARA/DocAnotacaoPortCOMPARA.pdf>

[Inácio et al. 2008]

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), 2008, pp. 271-286.

[Maia & Santos 2011]

Belinda Maia & Diana Santos. "Who is afraid of... what?: Fear in English and Portuguese". Apresentação no *ICAME2011*, Univ. de Oslo, 1 de Junho de 2011.

[Mota & Santos 2009]

Cristina Mota & Diana Santos. "Corte-e-costura no AC/DC: auxiliando a melhoria da anotação nos corpos". Versão 1.0, 17 de Setembro de 2009. <http://www.linguateca.pt/acesso/corte-e-costura.pdf>

[Santos 2011]

Diana Santos. "Linguatca's infrastructure for Portuguese and how it allows the detailed study of language varieties". *OSLa: Oslo Studies in Language 3.2* (2011), pp. 113-128. [Volume edited by J.B.Johannessen, *Language variation infrastructure*].

[Santos & Mota 2010]

Diana Santos & Cristina Mota. "Experiments in human-computer cooperation for the semantic annotation of Portuguese corpora". In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner & Daniel Tapias (eds.), *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2010)* (Valletta, Malta, 17-23 de Maio de 2010), European Language Resources Association, pp. 1437-1444.

[Santos & Sarmiento 2003]

Diana Santos & Luís Sarmiento. "O projecto AC/DC: acesso a corpora/disponibilização de corpora". In Amália Mendes & Tiago Freitas (eds.), *Actas do XVIII Encontro Nacional da Associação Portuguesa de Linguística (APL 2002)* (Porto, 2-4 de Outubro de 2002), Lisboa : APL, pp. 705-717.

[Santos et al. 2008]

Diana Santos, Rosário Silva & Susana Inácio. "What's in a colour? Studying and contrasting colours with COMPARA". In *Proceedings of the 6th International*

*Conference on Language Resources and Evaluation (LREC 2008)*  
(Marraqueche, 28-30 de Maio de 2008), European Language Resources  
Association (ELRA), s/pp.

[Santos et al. em edição permanente]

Diana Santos, Augusto Soares da Silva & Cristina Mota. "Guarda-fatos: notas  
sobre a anotação do campo semântico do vestuário em português". Primeira  
edição: 26 de Outubro de 2009.  
<http://www.linguateca.pt/acesso/GuardaFatos.pdf>

[Silva et al. 2008a]

Rosário Silva, Susana Inácio & Diana Santos. "Colouring COMPARA:  
contrastive and monolingual colour studies in English and Portuguese".  
Apresentação na conferência da *American Association for Corpus Linguistics*  
(Provo, UT, EUA, Março de 2008).

[Silva et al. 2008b]

Rosário Silva, Susana Inácio & Diana Santos. "Colouring COMPARA:  
contrastive and monolingual colour studies in English and Portuguese".  
*Workshop on Language Resources for Teaching and Research* (Faculdade de  
Letras da Universidade do Porto, 23 de Abril de 2008).

[Silva et al. 2008c]

Rosário Silva, Susana Inácio & Diana Santos. "Documentação da anotação  
relativa à cor no COMPARA". Primeira versão: 27 de Novembro de 2007.  
Última versão (4.4): 31 de Dezembro de 2008.  
<http://www.linguateca.pt/COMPARA/DocAnotacaoCorCOMPARA.pdf>

## **Anexo 1: Casos de fronteira entre duas cores**

Estes casos estão, por agora, marcados no grupo Outras, mas correspondem claramente a casos fronteira entre duas cores, ou mesmo a um novo grupo emergente:

*cor de leão*: grupo Castanho ou grupo Amarelo?

*cor de gema de ovo*: grupo Amarelo ou grupo Laranja? Se o ovo é cru, é claramente laranja, se é cozido, a gema é claramente amarela. O que fazer, dar ambos os grupos?

*cor de telha e cor de tijolo e cor de tijolo*: grupo Laranja ou grupo Vermelho? Ou já estamos em presença de um grupo novo com individualidade própria?

## **Anexo 2: Explicação de outras dúvidas (no grupo Outros)**

Por outro lado, damos conta aqui dos casos classificados em Outras em que, embora o referente possa ter muitas cores, há uma que nos parece fundamental:

*cor de granito*: o granito pode aparecer em muitas cores (branco, cinzento, preto?), embora o cinzento seja a mais comum: grupo Cinzento?

*cor de terra*: a terra pode ter várias cores; castanho, amarelo, vermelho, preto, cinzento, mas o castanho é a mais comum: devemos optar pelo grupo Castanho?

## **Anexo3: Classificação de palavras denotando cor:humana em grupos de cor**

Conforme referido, a cor:humana não teve inicialmente grupos, mas depois, devido ao seu alargamento, isso passou a fazer sentido no caso de cor:humana2.

Aqui apresentamos – ainda em fase preliminar – uma proposta de atribuição de grupos. Basicamente, os termos específicos mantêm-se como Outras, devido às gamas diferentes que podem cobrir:

Outras:loiro porque loiro pode ser visto como Dourado, Amarelo, ou mesmo Prateado.

Outras:ruivo, porque ruivo pode ser visto como Vermelho ou Laranja.

Rosa: rosado

Outras:pálido, porque pode ser visto como Branco, Creme, ou Ausência

Branco: branco

Cinzento grisalho

Outras:moreno, (no caso da pele) porque é demasiado difícil de definir

Outras:pardo, idem

Vermelho; corado

## Índice

1. Introdução.....	1
2. Questões técnicas e de utilização.....	1
3. Critérios linguísticos.....	2
3.1 Opções de marcação da cor.....	2
3.2 Diferenças em relação à anotação da parte portuguesa do COMPARA.....	3
3.3 A questão da identificação da expressão de cor .....	3
3.3.1 O caso de cor:original.....	4
3.3.2 O caso de cor:humana.....	4
3.3.3 O caso de cor:raça.....	5
3.3.4 O caso de cor:vinho.....	6
3.4 Casos que podem suscitar dúvidas.....	6
3.5 Casos relativos à classificação gramatical .....	7
3.6 Casos pontuais.....	7
3.7 Diminutivos, aumentativos e superlativos.....	8
4. Questões associadas a géneros textuais diferentes.....	9
4.1 CONDIVport, parte da moda.....	9
4.2 CONDIVport, parte do futebol.....	9
4.3 CONDIVport, parte da saúde.....	9
4.4 ENPCPUB.....	10
5. Exemplos de procuras.....	10
6. Listagem das palavras de cor em português .....	10
6.1 Critérios relativos à identificação de palavras simples de cor .....	10
6.2 Critérios relativos à identificação de expressões com várias palavras de cor.....	11
6.3 Grupos de cor.....	11
7. Regras de anotação condicional.....	11
8. Resumo quantitativo.....	13
9. Análise das cores marcadas como pertencentes ao grupo Outras.....	15
9.1 OUTRAS frutas e legumes.....	15
9.2 OUTRAS metais e minerais.....	15
9.3 OUTRAS conceitos impalpáveis.....	15
9.4 OUTRAS flores.....	15
9.5 OUTRAS animais.....	15
9.6 OUTRAS fenómenos naturais.....	15
9.7 OUTRAS ainda sem classificação .....	15
10. Integração com a anotação semântica de outras áreas.....	16
11. Tarefas ainda por executar.....	16
12. Agradecimentos.....	16
Referências.....	18
Anexo 1: Casos de fronteira entre duas cores.....	20
Anexo 2: Explicação de outras dúvidas (no grupo Outros).....	20
Anexo3: Classificação de palavras denotando cor:humana em grupos de cor .....	20