

COMPARando cores em português e inglês

Susana Inácio, Diana Santos e Rosário Silva

Linguatca, FCCN e SINTEF ICT

1. Apresentação

Neste artigo discutimos várias questões relacionadas com a cor em português e a sua relação e contraste com o inglês, usando como ferramenta o COMPARA (Frankenberg-Garcia & Santos, 2002), o maior corpus paralelo editado do mundo, e fazendo já algum uso da anotação sintáctica da parte portuguesa do mesmo (Inácio & Santos, 2008).¹

Esta é a nossa primeira incursão na classificação semântica do COMPARA, começando por um assunto relativamente "fácil", a cor. O nosso trabalho teve três objectivos principais: iniciar a anotação semântica do COMPARA, dar a conhecer as potencialidades deste corpus a uma audiência maior, e ilustrar uma metodologia de linguística com corpora. Não podemos aqui demorar-nos numa longa introdução à problemática da cor e aos estudos que conhecemos e que abordaram essa questão, mas convém realçar que a cor tem sido alvo de um sem-número de estudos e argumentos em linguística, quer para fundamentar ou rebater posições filosóficas (Berlin & Kay, 1991; Pinker, 1994; Sampson, 2005) quer para exemplificar e discutir problemas de representação (Gärdenfors, 2000). Para o português, veja-se Jorge et al. (2003) e Correia (2006), entre outros.

Sendo a cor uma forma de representação do universo, estando presente na língua de uma maneira, muitas vezes, subliminar, em expressões tão variadas como *lista negra*, *via verde*, *páginas amarelas*, entre muitas outras, achámos pertinente fazer uma análise da cor nas duas línguas que constituem o corpus, não só na perspectiva das línguas originais, mas também no que se refere à tradução da cor. Neste primeiro estudo, explorámos o COMPARA para identificar qual o seu conteúdo nesta área, tanto a nível lexical, sintáctico e semântico, nas duas línguas, como contrastivo. Como é sabido, há dois tipos de estudos em linguística com corpora, os **exploratórios**, cujo fim é observar os dados, definir categorias, detectar correlações, e os **experimentais**, que tentam corroborar ou refutar uma dada hipótese bem definida (Santos, no prelo). O nosso estudo foi exploratório, como base para mais tarde poder formular hipóteses.

A primeira tarefa a que nos lançámos foi a de marcar as palavras, expressões e frases que ocorrem no COMPARA com a informação de cor. Para que futuros utilizadores possam utilizar (e prosseguir) este trabalho com seriedade, descrevemos detalhadamente as várias opções tomadas em Silva, Inácio & Santos (2008). Muito resumidamente, atribuímos o valor `cor` ao atributo `sem`, especificando alguns casos, tal

¹ O COMPARA é publicamente acessível através do endereço <http://www.linguatca.pt/COMPARA/>. Para os dados publicados neste artigo foi usada a versão 9.2.1, com 640 mil palavras em português original e 797 mil em português traduzido.

como `cor:raça` ou `cor:humana`. Às cores puras, atribuímos ainda um grupo de cor, como BRANCO ou ROSA. Finalmente, as expressões (ou nomes próprios) que contêm palavras de cor com um sentido que já não é primário, foram marcadas como `cor:original`. (Trabalho semelhante foi feito para o inglês, cf. Santos et al. (2008).)

2. A cor no COMPARA em traços largos

De acordo com os nossos critérios, foram ao todo encontradas 4427² palavras denotando cor em português e 4435 em inglês, correspondendo a 227 lemas de cores distintas em português (405 palavras diferentes) e 430 palavras em inglês. Na Tabela 1 apresentamos a sua distribuição por texto traduzido vs. texto original e por variante, para o português e para o inglês.

Tipo de texto	Textos	Cores (lemas)	Ocorrências
Total em português	74(75)	227	4427
Original em português	40	129	2136
Traduzido para português	34	177	2286
Por variante do português			
Brasil total	27 (28)	109	1374
Brasil original	24 (25)	98	1243
Brasil traduzido	3	46	98
Portugal total	39	195	2983
Portugal original	12	80	795
Portugal traduzido	31	171	2188
África original	4	16	65
Tipo de texto	Textos	Cores (formas)	Ocorrências
Total em inglês	73(75)	430	4435
Original em inglês	32	323	2323
Traduzido do português	41	213	2112
Por variante do inglês			
Reino Unido total	37 (39)	311	2645
Reino Unido original	21 (23)	237	1515
Reino Unido traduzido	16	144	1130
Estados Unidos total	32	165	1242
Estados Unidos original	7	76	322
Estados Unidos traduzido	25	124	920
África do Sul	4	94	619

Tabela 1: A cor no lado português e inglês do COMPARA

Analisando os gráficos da Figura 1, que mostram a distribuição por grupos de cor para os autores portugueses e ingleses, podemos observar que, tanto na língua portuguesa como na língua inglesa, o grupo de cor predominante é o BRANCO/WHITE (com 21% e 14,1% respectivamente), seguindo-se o grupo PRETO/BLACK (com 15,4% e

² Não foram incluídas nesta contagem as palavras de cor etiquetadas como `cor:original`. Além disso, e para lidar com a marcação da vagueza (Santos, 1998), foi seguido o seguinte método: quando uma palavra foi marcada como podendo ser cor (mas vaga entre uma interpretação que não o é), foi contada como palavra de cor. Nos casos relevantes, quando uma palavra foi considerada vaga entre cor pura e outro tipo de cor (`cor:raça`, etc.) foi contada metade para cada lado. Os números entre parênteses devem-se ao facto de alguns textos terem mais do que uma tradução no COMPARA, sendo assim apenas contados uma vez como originais, mas duas vezes as suas traduções.

13,2%). Os grupos de cor VERMELHO/RED e AZUL/BLEUE são também muito significativos nas duas línguas; no entanto, podemos constatar que os autores de língua inglesa têm mais preferência pelo grupo de cor BROWN do que os de língua portuguesa. Tal acontece ao invés do grupo NÃO ESPECIFICADA, que ocupa o terceiro lugar nas preferências portuguesas, sendo apenas quinto nas preferências inglesas.

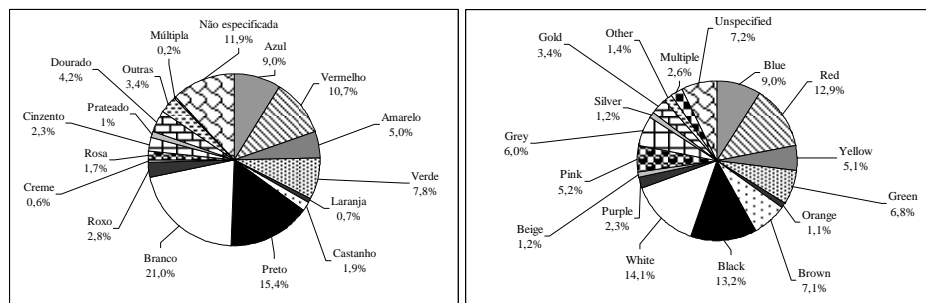


Figura 1: Distribuição por grupo de cor em português original (à esquerda) e em inglês original (à direita) no COMPARA

Apresentamos também uma comparação do grupo de cor em português original e em todo o português (original e traduzido), na Figura 2.

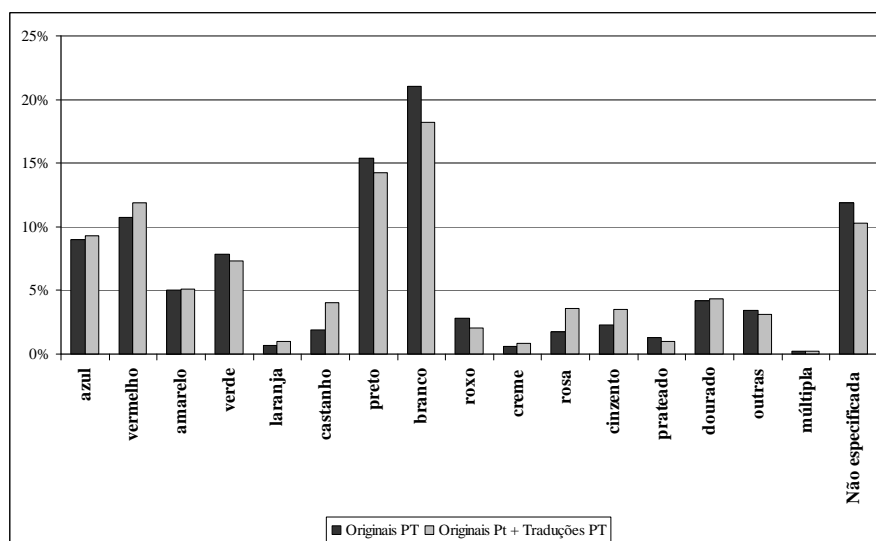


Figura 2: Comparando os grupos de cor em português original e em todo o português (originais e traduções)

3. As cores por autor

A primeira análise a que procedemos dedicou-se a explorar possíveis diferenças no uso da cor por parte dos autores dos textos que compõem o COMPARA³. Analisaremos um total de trinta e seis autores (vinte e dois dos quais de língua portuguesa; os restantes catorze de língua inglesa) e de setenta e dois títulos (quarenta dos quais em língua portuguesa; os restantes trinta e dois em língua inglesa).

A Tabela 2 mostra os autores, o número de obras e as datas de edição de cada obra que foram utilizados para a presente análise, juntamente com as cores usadas por cada autor. #cores indica o número de cores diferentes na língua original, enquanto #palavras conta o número de palavras marcadas como cor na parte do COMPARA escrita por esse autor (se a palavra for vaga entre cor e não cor, conta apenas 0,5). Como a contribuição de cada autor para o COMPARA varia desde um mínimo de 3197 a um máximo de 184377 palavras, apresentamos também a frequência relativa das palavras de cor por cada mil palavras (freq) para facilitar a comparação.

Autor	Textos e datas	# cores	# palavras	freq.
Autores de língua portuguesa				
Aluísio Azevedo	PBAA1 (1890) e PBAA2 (1881)	32	159	4,00
Austran Dourado	PBAD1 (1973) e PBAD2 (1975)	51	364,5	8,91
Camilo Castelo Branco	PPCC1 (1862)	5	8	0,51
Chico Buarque	PBCB1 (1995) e PBCB2 (1991)	31	112,5	5,39
Eça de Queirós	PPEQ1 (1880) e PPEQ2 (1887) e PPEQ3 (1925)	36	255,5	6,96
Jô Soares	PBJS1 (1995)	13	43	2,32
Jorge de Sena	PPJS1 (1978)	23	53	1,25
José Cardoso Pires	PPCP1 (1983)	17	34	2,28
José de Alencar	PBJA1T1 (1865)	10	33	10,32
José Eduardo Agualusa	PAJA1 (1992) e PAJA2 (1990)	9	19	4,00
José Saramago	PPJA1 (1995) e PPJA2 (1989)	24	122,5	1,92
Lídia Jorge	PPLJ1 (1988)	20	76	4,38
Machado de Assis	PBMA1 (1886) e PBMA2 (1881) e PBMA3 (1899) e PBMA4 (1904) e PBMA5 (1876) e PBMA6 (1908)	21	72	0,90
Manuel António de Almeida	PBMAA1 (1852)	8	13	1,15
Marcos Rey	PBMR1 (1986)	19	53	2,87
Mário de Carvalho	PPMC1 (1994)	19	37	1,78
Mário de Sá-Carneiro	PPSC1 (1914) e PPSC2 (1915)	40	188	8,60
Mia Couto	PMMC1 (1987) e PMMC2 (1990)	12	46	2,87
Osman Lins	PBOL1 (1966)	30	103,5	5,81
Patrícia Melo	PBPM1 (1998) e PBPM2 (1995)	27	112,5	3,60
Paulo Coelho	PBPC1 (1988) e PBPC2 (1987) e PBPC3 (1996)	13	29	0,70
Rubem Fonseca	PBRF1 (1988) e PBRF2 (1983)	32	174	2,97
Autores de língua inglesa				
David Lodge	EBDL1 (1995) e EBDL2 (1989) e EBDL3 (1975) e EBDL4 (1980) e EBDL5 (1991) e EBDL6 (1984)	95	534	2,90
Edgar Allan Poe	EUEP1 (1837)	7	20	0,82
Henry James	EUHJ1 (1880) e EUHJ2 (1895) e EUHJ3 (1893)	22	65	1,38
Ian McEwan	EBIM1 (1992) e EBIM2 (1998) e EBIM3 (1978)	44	103	2,47
Joanna Trollope	EBJT1 (1996) e EBJT2 (1993) e EBJT3 (1995)	101	312	3,70
Joseph Conrad	EBJC1 (1902)	19	41	3,37

³ Como tal, os dados utilizados e as consequentes conclusões limitam-se aos excertos e títulos que os autores deram autorização para integrar o corpus. Cada excerto representa, regra geral, 30% da obra total.

Joseph Heller	EUJH1 (1990)	12	25	1,89
Julian Barnes	EBJB1 (1984) e EBJB2 (1989) e EBJB3 (1991)	46	121	2,02
Kazuo Ishiguro	EBKI1 (1995) e EBKI2 (1989)	23	69	0,81
Lewis Carroll	EBLC1 (1871)	8	20	1,80
Mary Shelley	EBMS1 (1818)	6	12	0,57

Tabela 2: Características da cor nos autores do COMPARA

3.1. Primeiras observações

Depois de tratados os dados, pudemos constatar que 49% dos vinte e dois autores de língua portuguesa utilizam o branco como cor preferida (seja este fenómeno consciente ou não), como ilustrado na Figura 3. O autor de língua portuguesa que mais contribui para esta tendência é José Saramago. Do total de cores que utiliza, 46,85% encaixam-se no grupo BRANCO. O autor com menor contribuição para a cor total é Camilo Castelo Branco (ironicamente!) com 0% de BRANCO, sendo a sua cor de eleição o PRETO, 57,14%.

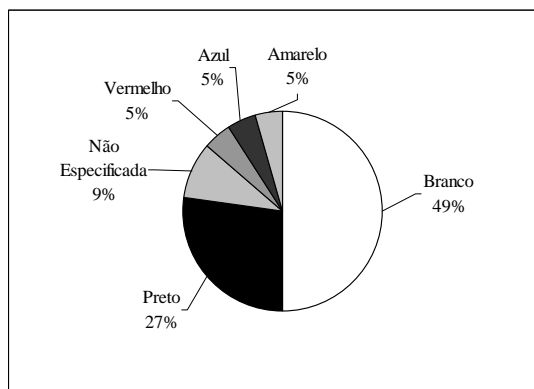


Figura 3: Cores preferidas dos autores de língua portuguesa: cada autor contribui apenas com a sua cor preferida

No que toca à utilização de palavras coloridas de cor, cor:raça, cor:humana e cor:vinho nas suas obras, a Figura 4 mostra-nos a percentagem de palavras coloridas no total de palavras das obras de cada autor de língua portuguesa. Por conseguinte, podemos verificar que o autor que tem uma percentagem maior de palavras coloridas no total de palavras dos seus textos é José de Alencar com 1,03%, correspondendo 0,28% dessa percentagem a palavras cor:raça, 0,06% a palavras cor:humana e 0% a cor:vinho. O autor que menos utiliza palavras de cor nas suas obras é Camilo Castelo Branco, apenas com 0,05%, não estando nenhuma relacionada com raça, atributos humanos ou vinho.

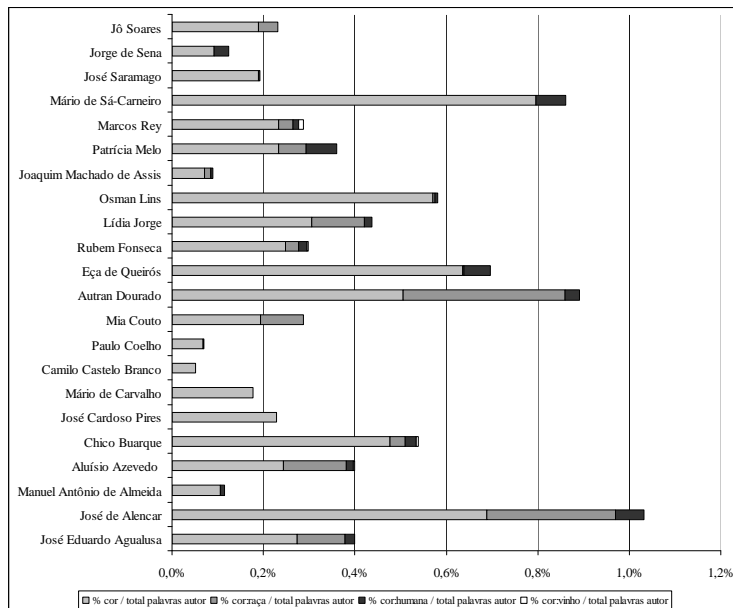


Figura 4: A cor no total de palavras utilizadas por cada um dos autores de língua portuguesa

Vale a pena chamar a atenção para o facto de ser Autran Dourado que apresenta uma percentagem maior de palavras cor:raça na totalidade das palavras coloridas nas suas obras com 39,8%, enquanto Jorge de Sena tem a maior percentagem de palavras cor:humana na totalidade de palavras coloridas da sua obra, com 26,4%.

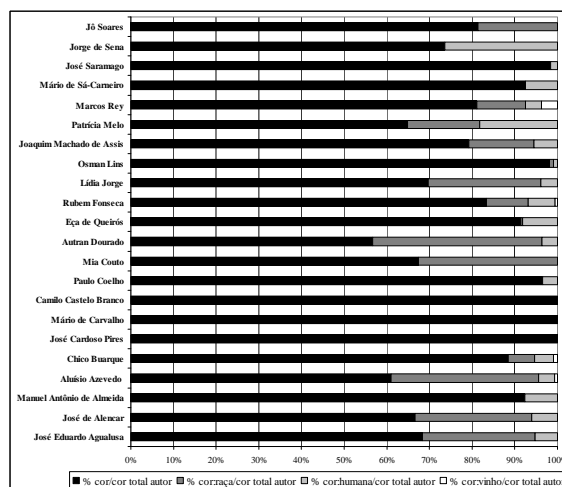


Figura 5: As categorias de cor no total da cor de cada um dos autores de língua portuguesa

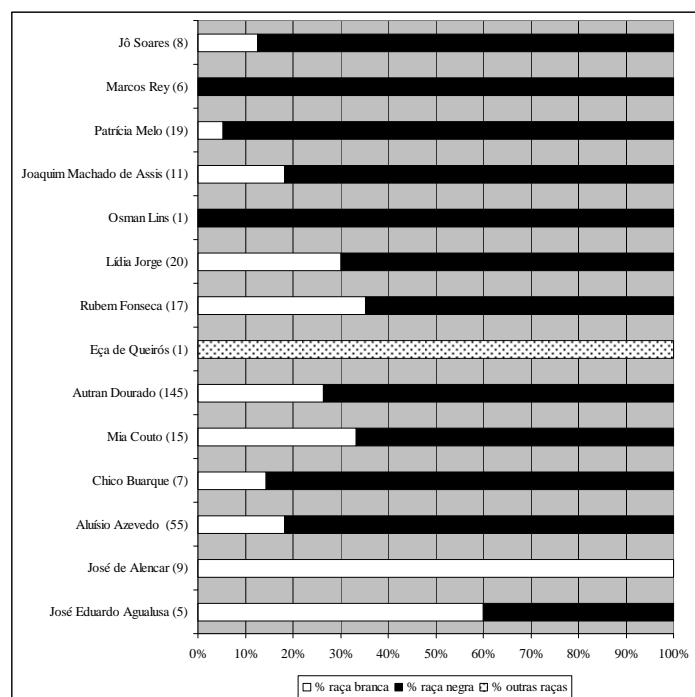


Figura 6: Raças a que correspondem as cores marcadas *cor : raça*: os números entre parênteses à direita dos nomes dos autores portugueses referem-se ao número efectivo de ocorrências

Finalmente, podemos verificar na Figura 6 que, percentualmente, a raça negra é a mais mencionada pelos autores.

4. O léxico da cor no COMPARA português

4.1 Cores não lexicalizadas

A descrição da cor não se resume ao uso de palavras isoladas com esse significado intrínseco. Pelo contrário, o uso da língua é criativo e permite a evocação de cores nunca antes mencionadas (ou pelo menos ainda não padronizadas). Apresentamos aqui uma primeira análise das cores produzidas por justaposição (marcada por hífen) e das obtidas através da expressão *cor de X*.

No COMPARA existem 93 lemas de palavras compostas por justaposição marcada por hífen. Nos três dicionários que consultámos (Houaiss, 2005; Casteleiro, 2001; Ferreira, 2004), apenas nove lemas estão dicionarizados, enquanto os restantes 84 não estão. Como exemplos de palavras do primeiro grupo, veja-se *azul-celeste*, *azul-marinho*, *verde-azeitona* ou *verde-garrafa*. Exemplos de lemas não dicionarizados são: *amarelo-açafrão*, *azul-metálico*, *louro-palha*, *verde-ramagem*, *vermelho-lacre*. Dentro

das palavras de cor compostas por justaposição (marcada por hífen), a figura 7 ilustra os grupos de cor que são mais moldáveis, quer por mistura com outras cores quer pela mistura com adjectivos de intensidade ou nomes referentes a textura, frutos, plantas e outros (ver Jorge et al. para uma proposta de tipologia), sendo eles o VERDE e o AZUL (originais); o VERDE e o CINZENTO (traduções). Por exemplo, o VERDE é composto por 5 lemas diferentes nos originais e 13 lemas diferentes nas traduções (se eliminarmos os lemas repetidos, temos 21 lemas no total, incluindo os 5 casos do grupo Múltipla)⁴.

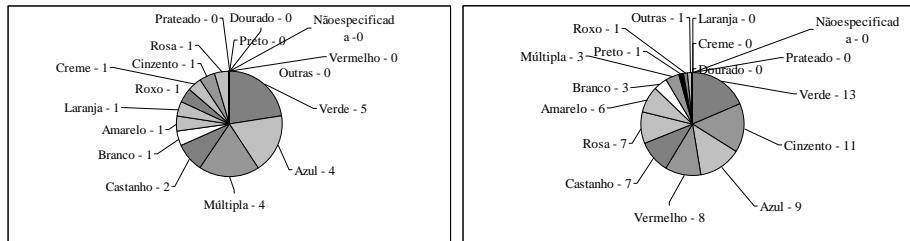


Figura 7: Grupos a que pertencem as palavras compostas por justaposição (português original e traduzido)

Relativamente à análise de cores produzidas pelo uso da expressão *cor de X* – donde provieram as cores *castanho* (de *cor de castanha*), *laranja*, *rosa*, *violeta*, agora já dicionarizadas enquanto cor nos dicionários consultados – podemos observar que o COMPARA contém 51 casos de *cor de X*, dos quais apenas doze (*cor de bronze*, *cor de creme*, *cor de marfim*, *cor de pérola*, *cor de ferrugem*, *cor de palha*, *cor de fogo*, *cor de vinho*, *cor de carne*, *cor de cereja*, *cor de salmão* e *cor de tijolo*) estão dicionarizados em Casteleiro (2001). Entre as 39 expressões não dicionarizadas encontram-se *cor de laranja*, *cor de prata*, *cor de cobre* e *cor de cinza*. Embora a expressão *cor de laranja* não esteja dicionarizada sem hífen, é a expressão *cor de X* que ocorre mais vezes (12) no COMPARA, enquanto *cor-de-rosa* (com hífen) aparece 58 vezes.⁵

Assumimos aqui que cores que aparecem tanto na forma *cor de X*, *cor-de-X* como só *X* são palavras que se encontram em fase de lexicalização, mais especificamente através do processo de adjectivalização denominal. Assim, por exemplo, *rosa* e *laranja* sofreram uma derivação imprópria, parecendo-nos estar, actualmente, lexicalizadas como adjectivos, embora apareçam de ambas as formas no COMPARA, assim como as palavras *cinza*, *cobre*, *bronze*, *malva* e *marfim*. Será assim verosímil que palavras como *fogo*, *vinho*, *carne* ou *tijolo*, venham a afirmar-se enquanto formas designando cor, levando a, por exemplo, *mesa carne* (de *mesa cor de carne*) ou *camisola vinho* (de *camisola cor de vinho*), à semelhança de *fato malva* ou *mosaicos ocre*.

⁴ *verde-acinzentado*, *verde-amarelado*, *verde-amarelo*, *verde-azeitona*, *verde-azul*, *verde-azulado*, *verde-cinza*, *verde-dourado*, *verde-claro*, *verde-brilhante*, *verde-negro*, *verde-escuro*, *verde-garrafa*, *verde-musgo*, *verde-ramagem*, *verde-esmeralda*, *verde-lima*, *verde-pálido*, *verde-pastoso*, *verde-seco* e *verde-vivo*.

⁵ Não nos parecendo ser esta uma diferença fundamental entre os dois termos, notamos no entanto a forte discrepância entre o comportamento dos dicionários, por um lado, e das editoras, por outro.

Parece pois ser comum a neologia no domínio cromático, ou seja, a criação de palavras que representem uma nova tonalidade de cor, através de um novo sentido atribuído ao objecto que ilustra essa cor. Nas palavras compostas, verifica-se uma espécie de "simbiose" linguística, onde a elipse das palavras *cor de* liga simbioticamente, através de um hífen (ou não), o nome modificador à palavra modificada. Assim, *amarelo- (cor de) açafião*, transforma-se em *amarelo-açafrão*, e *castanho-(cor de) mostarda* em *castanho-mostarda*.

Outra forma de formação de palavras novas que designam cor é a que é feita naturalmente através da invenção de termos por parte dos autores do COMPARA. Exemplos disso são *saliva-laranja* (Mia Couto) ou *branco-manhã* (José Saramago). Noutros casos, a própria actividade de tradução leva à criação de novas designações de cor, como acontece em *cor de massa de vidraceiro* ou *cor de colmo bolorento*, respectivamente traduzindo *putty-coloured*⁶ (em Nadine Gordimer) e *colour of mouldy thatch* (em Joanna Trollope).

4.2 Cores que perderam o seu significado primário

A classificação da informação de cor em contexto levou-nos a ter de decidir, em alguns casos, que, embora originalmente uma palavra denotasse cor, no contexto em que se inseria o seu principal sentido não era esse. Foi contudo interessante marcá-la (*cor:original*) para apreciar o potencial metafórico e o peso total no léxico.

Existem 255 casos marcados como *cor:original*, 106 dos quais correspondendo a nomes próprios (*Cruz Vermelha, Danúbio Azul, etc.*). Nos outros 149 casos, o lema mais frequente é *negro* (41 casos), seguido de *branco* e *pardo* (22 e 16 casos). As expressões mais comuns para cada um destes lemas são *buraco negro*, *em branco* e *papel pardo*. A Figura 8 apresenta a distribuição de *cor:original* por várias subcategorias:

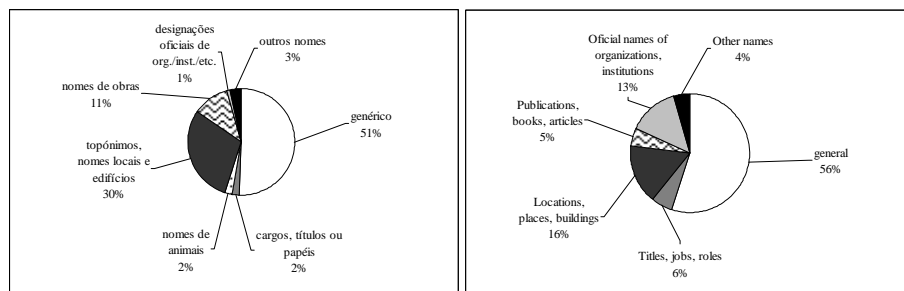


Figura 8: Tipo de expressão de *cor:original* (autores portugueses à esquerda e autores ingleses à direita)

⁶ Em inglês este não é um termo "inventado", mas em português podemos falar de criatividade através do uso da tradução literal, que introduz novas formas de designar tonalidades na nossa língua.

4.3 Outras questões

Mencionamos que a observação feita em Correia (2006) de que a palavra *vermelho* é mais básica como termo de cor, ou seja, mais importante do que *encarnado*, tem plena confirmação no COMPARA. De facto, o lema *vermelho* é utilizado 279 vezes contra apenas onze ocorrências de *encarnado* como cor. Além disso, também não encontramos palavras compostas com *encarnado*, ao contrário de com *vermelho*. Contudo, convém acentuar que a nossa preocupação não é a de Correia ou Berlin & Kay, interessados nas cores básicas, enquanto nós procurámos todas e quaisquer palavras empregues como cor.

5. A sintaxe da cor no COMPARA português

É interessante verificar que, embora a cor seja uma propriedade prototípica, e, como tal, naturalmente realizada por adjectivos, as suas manifestações encontram-se em todas as categorias gramaticais. Na Tabela 3 apresentamos uma descrição global da classificação sintáctica das cores em português, remetendo o leitor para Inácio & Santos (2008) para a explicação e motivação das categorias apresentadas. Visto que as palavras referentes a *cor:raça* poderiam aumentar a proporção de substantivos⁷, apresentamos os dados com e sem as mesmas.⁸

Categoria gramatical	# sem raça	%	# com raça	%
Adjectivos	2940	72,88	3168	67,55
Substantivos	902	22,09	1322	28,03
Nomes próprios	106	2,60	106	2,22
Verbos	90	2,41	90	2,17
Advérbios	1		1	
Vagos entre substantivo e adjectivo (N_ADJ)			27	
Vagos entre adjectivo e verbo (ADJ_V)	17		17	
Adjectivos usados como nome (ADJn)	28		29	
Total	4084		4765	

Tabela 3: Distribuição da cor por categorias sintácticas

Note-se de passagem que a vagueza entre as categorias sintácticas não está necessariamente relacionada com a possível vagueza em relação à cor. Por exemplo, a decisão de considerar, no exemplo *Mas Isidoro, mesmo preto, não era capaz de uma coisa dessas.*, *preto* como *cor* ou *cor:raça* não está dependente, ou associada, à consideração do mesmo como substantivo ou adjectivo no contexto, está sim associada a considerações de natureza sintáctica.

Quanto à cor como verbo, a predominância é para casos de *cor:humana*, descrita por verbos como *corar*, *bronzear* e *ruborizar*. Além disso, a maior parte das ocorrências verbais das cores não humanas em português no COMPARA especificam

⁷ Usamos indistintamente neste artigo *nome* ou *substantivo*.

⁸ Tal como para a cor, a percentagem é calculada contando os casos vagos metade para cada lado, ou seja, 17 ADJ_V contam 8,5 para ADJ e o mesmo para V. ADJn contam como adjectivos.

apenas a presença de cor (tal como *colorir*) deixando para adjuntos a sua particularização.

Não nos é possível fazer aqui um esboço das construções mais frequentes em que adjectivos de cor ocorrem (notamos aliás que a informação de construção sintáctica, ao invés da categoria gramatical, ainda não se encontra revista no COMPARA). Seleccionámos contudo os casos de coordenação, comparação e orações relativas.

5.1 Coordenação de cores

Em relação à coordenação, existem ao todo 134 casos de ocorrência de cor e cor, dezassete deles correspondendo a ainda maior lista de cores: um de seis, dois de cinco, dois de quatro, e doze de três. Podemos apresentar o exemplo de: *cascatas de gerânios, rubros, brancos e cor-de-rosa*. Para cor ou cor, além de três casos unindo três ou mais cores, temos dez casos. Esta quase centena e meia de casos permite-nos algumas observações, mas note-se contudo que esta contabilização não abrange casos também relativos a coordenação de cor como *os pios e ruídos mais estranhos vindos na boca da noite que devorava o verde viçoso e o verde velho empoeirado* ou *o verde, o amarelo ou o castanho dourado que dominavam, consoante a estação*.

Em primeiro lugar, as cores mais coordenadas são o BRANCO (51 casos), o VERMELHO (38), o AZUL (34) e o PRETO (31). Por outro lado, é patente que a posição preferida do BRANCO é final, e a do VERMELHO inicial, enquanto o AZUL não parece ter preferência. É interessante reparar que o VERDE, sendo uma cor tão misturável, é pouco coordenável, com apenas dez e oito casos respectivamente na posição inicial e final. Se retirarmos a expressão *preto e branco*, que ocorre sete vezes relativa a fotografias ou ecrãs, ainda mantemos a ordenação, assim como a apetência destas duas cores para se combinarem (dois casos BRANCO e PRETO,⁹ e nove casos PRETO e BRANCO).

5.2 Comparação de cores

Por outro lado, as cores não são alvos frequentes de comparação ou de explicitação através de comparação: de facto, se encontramos 44 casos de uma palavra considerada cor seguida da palavra *como*, apenas 24 se referem à cor (*branca como a cal da parede*) e não ao substantivo ou sintagma nominal em questão (*mantilhas pretas como as que usavam na Semana Santa*). Ainda mais flagrantemente, dos 31 casos de *mais* seguido de cor, apenas quatro explicitam o termo de comparação em relação à cor¹⁰: *Se você viesse a ter netos queria que eles apanhassem palmatoadas de um professor mais negro que esta batina?*. Nenhum caso há de *menos*, e no contexto de *tão* cor, em nove casos apenas três comparam a cor, mais a oração consecutiva: *um tapete que dá vontade de patinar, tão branco que Ariela só falta pedir ao cliente para tirar os sapatos*. Finalmente a cor aparece apenas uma vez no superlativo (*a mais verde*).

⁹ É interessante notar que em ambos os casos o inglês tem a ordem BLACK seguido de WHITE.

¹⁰ Ou seja, a maior parte das vezes a comparação refere-se a outros participantes na frase, como é evidente por exemplo nas frases seguintes: *o passado parece ter mais cor local que o presente*, ou *uma mecha de cabelo, mais louro, evidentemente, do que nas fotografias*.

5.3 Orações relativas referentes às cores

Quanto à existência de orações relativas, das 153 sequências de uma palavra marcada como cor e a palavra *que*, apenas fomos capazes de detectar os seguintes cinco casos referindo-se a cores concretas: *cabelo louro-claro que já vai ficando branco*; *de um azul que só há nas porcelanas*; *tapete da cor do vermelho que se vê quando se fecham os olhos para evitar a luz do sol*; *uma tez vermelhusca que os invejosos atribuíam à bebida*; *cabelos alvos que fazia gosto*; e mais cinco que se referem precisamente a cores não especificadas, como: *As estranhas cores que coloriam os cabelos*; *a cor que lhe tingira o rosto*; *uma cor que nunca vira num ser humano*.

5.4 O género da cor

Será que as palavras femininas ou masculinas atraem mais alguns tipos de cor do que outros? Ou o género morfológico em português é puramente gramatical e não tem qualquer relação com o significado? No COMPARA a cor aparece 56,5% das vezes associada a palavras masculinas, e apenas 41,5% das vezes associada a palavras femininas. (A distribuição dos géneros no COMPARA em geral – nos casos dos nomes e adjectivos – é de 50,8 % para o masculino e de 46,9% para o feminino.)

A discrepância poderá ser atribuída ao facto de cores abstractas serem descritas no masculino, assim como o conjunto de pessoas de uma dada raça. Se descontarmos os 886 casos de *cor:raça* (que de facto têm 82% de casos no masculino e 18% apenas no feminino), obtemos uma distribuição de género de 52,2% masculino e 45,5% feminino que é praticamente indistinguível do resto da língua.

Restringindo a nossa investigação apenas ao caso em que as cores são adjectivos (1586:1334) tentámos averiguar então se existia uma preferência marcada por um dos géneros, para cada cor (lema). Os lemas mais associados ao masculino foram *castanho-escuro*, *violeta*, *róseo*, *pardo*, *ouro*, *castanho*, *marrom*, e *rosa*, ... enquanto as cores mais associadas ao feminino foram *lilás*, *roxo*, *amarelado*, *acastanhado*, *encarnado*, *dourado*, e *laranja*... Repetindo o processo para o grupo de cor, o mais masculino revelou-se o CASTANHO, seguido do PRATEADO e do AZUL, enquanto os mais femininos (e únicos) foram o ROXO (0,63), o DOURADO (0,55) e o BRANCO (0,54).

6. A tradução da cor no COMPARA

Afinal de contas, o COMPARA como corpus paralelo é o veículo ideal para identificar e cartografar questões de tradução. Em relação à cor, perguntas imediatas poderão ser: há cores que se perdem (na tradução entre as duas línguas em questão)? Há cores mais difíceis de traduzir? Existem estratégias específicas para a tradução da cor? Há cores mais dependentes da cultura do que outras?

Começamos por apresentar as bases para um estudo contrastivo indirecto, através da comparação entre a frequência e diversidade das palavras e expressões de cor em português original e português traduzido (do inglês), na Tabela 4, para indagar se existem, de facto, diferenças significativas entre o uso do léxico da cor nas traduções e

nos originais. E se essas diferenças podem iluminar a questão da riqueza ou pobreza lexical do português em relação ao inglês no caso das cores.

Português	total	original	#freq	traduzido	#freq.
Branco	622	353	55,2	269	33,8
Preto	485	251	39,2	234	29,4
Azul	321	152	23,8	169	21,2
Amarelo	175	83	13	92	11,5
Vermelho	407	176	27,5	231	29
Laranja	33	11	1,7	22	2,8
Verde	237	120	18,8	117	14,7
Roxo	70	47	7,3	23	2,9
Castanho	139	31	4,8	108	13,6
Creme	29	10	1,6	19	2,4
Cinzeno	119	38	5,9	81	10,2
Rosa	122	29	4,5	93	11,7
Dourado	140	60	9,4	80	10
Prateado	33	21	3,3	12	1,5
Outras	119	62	9,7	57	7,2
Múltipla	8	5	0,8	3	0,4
Nãoespecificada	353	197	30,8	156	19,6

Tabela 4: Distribuição entre os grupos de cor em português original e traduzido; #freq é a frequência em cada cem mil palavras

Não podemos aqui mais do que comentar alguns assuntos que nos pareceram especialmente interessantes, deixando mais para trabalho futuro. Em relação à questão original vs. traduzido, parece existir uma maior ocorrência de palavras compostas (por justaposição marcada por hífen) no português traduzido. Vejamos como exemplos os grandes grupos de cor CASTANHO, ROSA e CINZENTO: No grupo CASTANHO, em que o valor de #freq em português traduzido é de 13,6 vs. 4,8 em português original, há sete lemas¹¹ diferentes de palavras compostas em português traduzido, contra apenas *castanho-claro* e *castanho-escuro* em português original. O grupo ROSA mostra a mesma tendência, com quatro lemas diferentes de palavras compostas (*rosa-alperce*, *rosa-baço*, *rosa-acinzentado* e *rosa-shocking*) em português traduzido e nenhum em português original, assim como o grupo CINZENTO: oito lemas diferentes¹² de palavras compostas vs. um lema (*cinzeno-escuro*) em português original.

A tendência para o neologismo cromático através da justaposição na tradução é além disso confirmada mesmo nas cores que manifestam maior número de ocorrências em português original do que traduzido, como é o caso de BRANCO, PRETO e NÃOESPECIFICADA: por exemplo, três lemas diferentes na tradução (*branco-sujo*, *branco-brilhante* e *branco-azulado*) e só o lema *branco-manhã*, em português original.

¹¹ *castanho-escuro*, *castanho-avermelhado*, *castanho-carvão*, *castanho-azeitona*, *castanho-mostarda*, *castanho-rosado* e *castanho-azulado*

¹² *cinzeno-claro*, *cinzeno-esverdeado*, *cinzeno-pálido*, *cinzeno-escuro*, *cinzeno-médio*, *cinzeno-suave*, *cinzeno-aço* e *cinzeno-metalizado*

Focando agora a atenção na tradução propriamente dita, queremos investigar se existe uma abordagem directa na tradução das cores, com simplificação e empobrecimento ou, pelo contrário, os tradutores tentam incorporar na tradução cores diferentes daquelas a que os seus leitores estão habituados? Ou seja, na tradução das cores no COMPARA, consegue-se identificar estratégias de estrangeirização e de domesticação/etnocentrismo (também chamadas de adopção vs. adaptação)?

De acordo com a Tabela 5, uma primeira observação é a de que parece existir maior imprecisão na descrição da cor em português (original e traduzido) vs. maior especificação (do grupo) de cor em inglês (original e traduzido). Assim, vejam-se os casos *cor de gelo*, *cor de fumo*, e *ocre* inseridos no grupo de cor OUTRAS, que ao serem traduzidos por *off-white*, *smoke-gray*, *yellow* levou a que fossem inseridos nos grupos de cor WHITE, GREY e YELLOW, respectivamente.

	Inglês	corresponde	não corresponde
Branco	White	504	118
Preto	Black	399	86
Azul	Blue	300	21
Amarelo	Yellow	154	21
Vermelho	Red	342	65
Laranja	Orange	31	2
Verde	Green	207	30
Roxo	Purple	58	12
Castanho	Brown	124	15
Creme	Beige	23	6
Cinzento	Grey	100	19
Rosa	Pink	113	9
Dourado	Gold	65	75
Prateado	Silver	15	18
Outras	Other	21	98
Múltipla	Multiple	1	7
Nãoespecificada	Unspecified	288	65

Tabela 5: Comparação entre os grupos de cor em português (original e traduzido) e em inglês (original e traduzido): casos em que o grupo corresponde e não corresponde

7. Comentários finais

Este é apenas o primeiro passo de uma exploração apaixonante no universo da cor, nas línguas portuguesa e inglesa e na tradução entre elas, apresentado nesta conferência para permitir que outros efectuem estudos sobre o assunto, e para ilustrar as capacidades de um recurso com esta informação. É importante referir que o COMPARA está aberto à anotação cooperativa de outros campos semânticos sobre os quais haja interesse, enriquecendo assim em conjunto um recurso que está acessível para todos.¹³

¹³ Esta é uma abordagem complementar aos serviços na rede que permitem anotar os corpora privativos, mas não partilháveis, como é o caso do Corpógrafo, www.linguateca.pt/Corpografo ou do VISL, www.visl.sdu.dk.

Muito fica por estudar em relação à cor no COMPARA, mas contamos com os leitores para o fazerem, visto que fornecemos este instrumento à comunidade. Será também interessante comparar as nossas observações em corpora diferentes e maiores, tais como os servidos pelo projecto AC/DC, "irmão" do COMPARA.

Agradecimento. O COMPARA é desenvolvido no âmbito da Linguateca, financiada através dos projectos POSI/PLP/43931/2001 e POSC 339/1.3/C/NAC. Agradecemos as sugestões iniciais da Ana Frankenberg-Garcia, assim como a implementação das funcionalidades de distribuição por grupo de cor no COMPARA e outros serviços desenvolvidos pelo Pedro Sousa.

Referências

- Berlin, Brent & Paul Kay (1991) *Basic Colour Terms: their Universality and Evolution*. Stanford: CSLI [Primeira edição: 1969].
- Casteleiro, João Malaca (org.) (2001) *Dicionário da Língua Portuguesa Contemporânea*. Lisboa: Academia das Ciências de Lisboa/Editorial Verbo.
- Correia, Margarita (2006) Towards a General Description of the Semantic Field of 'Colour' in European Portuguese. In: C.P. Biggam & Christian J. Kay (eds.), *Progress in Colour Studies, 1: Language and Culture*. Amsterdão/Filadélfia: John Benjamins, pp. 111-125.
- Frankenberg-Garcia, Ana & Diana Santos (2002) COMPARA, um corpus português-inglês na Web, *Cadernos de Tradução* 9 (2002/1), pp. 61-79.
- Gärdenfors, P (2000). *Conceptual Spaces: The Geometry of Thought*. Cam MIT Press.
- Ferreira, Aurélio Buarque de Holanda (2004) *Novo Dicionário de Língua Portuguesa*. 3.ª ed., Curitiba: Positivo.
- Houaiss, A. (2005) *Dicionário Houaiss da Língua Portuguesa*. Lx: Temas e Debates,.
- Inácio, Susana & Diana Santos (2008) Documentação da anotação da parte portuguesa do COMPARA. Primeira versão: 9 de Dezembro de 2005. Em actualização constante. <http://www.linguateca.pt/COMPARA/DocAnotacaoPortCOMPARA.pdf>.
- Jorge, Guilhermina (coord.) et al. (2003) As cores preto no branco: uma análise comparativa, *Polifonia* 6, Edições Colibri, Lisboa, pp. 119-133.
- Pinker, Steven (1994) *The language instinct*. Londres: Penguin books.
- Sampson, Geoffrey (2005) *The 'Language Instinct' Debate*. Londres & Nova Iorque: Continuum International.
- Santos, Diana (1998) A relevância da vagueza para a tradução, ilustrada com exemplos de inglês para português, *TradTerm* 5, 1, pp. 41-70.
- Santos, Diana (no prelo) Corporizando alguns desejos. In: Stella Tagnin (ed.) e Oto Vale (eds.), *A Lingüística de Corpus no Brasil: pesquisa e crítica*. São Paulo: Editora Humanitas.
- Santos, Diana, Rosário Silva & Susana Inácio (2008) What's in a colour? Studying and contrasting colours with COMPARA. In: Proceedings LREC 2008.
- Silva, Rosário, Susana Inácio & Diana Santos (2008) Documentação da anotação relativa à cor no COMPARA. Em actualização constante. 1 versão: 27 de Novembro de 2007. <http://www.linguateca.pt/COMPARA/DocAnotacaoCorCOMPARA.pdf>