Key aspects
- COMPARA
- Colouring COMPARA
- English-speaking authors
- Contrastive analysis (Portuguese vs. English)
- Morphosyntactic patterns in Portuguese and in English
- Originals vs. Translations

COMPARA
- An English-Portuguese publicly available parallel corpus (since 2001)
- developed at Linguateca, project leaders: Ana Frankenberg-Garcia & Diana Santos
- manually revised texts, alignment and grammatical annotation (from PALAVRAS and CLAWS, in progress)
- several innovative features

COMPARA contents
- 75 pairs, corresponding to 72 fiction works
- 1.436 million words in Portuguese (Africa, Brazil and Portugal)
- 1.549 million words in English (Great-Britain, America and South Africa)
- 97,728 alignment units
- see http://www.linguateca.pt/COMPARA/

Colour research by:
- Language
  - Portuguese
  - English
- Language variety
  - Portuguese (Portugal, Brazil, Mozambique and Angola)
  - English (United Kingdom, United States and South Africa)

Several outputs available (concordances and distributions)

Colour research by:
- Text
- Author
- Translated or Original text(s)
- Date of publication
Why COLOUR?
- We wanted to start semantic annotation of COMPARA
- An easy to grasp and relatively-contained lexical category
- The subject of many arguments in linguistics
- A frequent category, both literally and metaphorically

What is COLOUR? – Criteria
- All straightforward words conveying colour, regardless of word class
- Words that convey colour but are not in themselves a colour
- Words that have colours in them but have gone beyond the mere colour reference
- The word “colour” and its derivations.

COLOUR categories and groups
- Colour
  - Beige
  - Black
  - Blue
  - Brown
  - Gold
  - Green
  - Grey
  - Orange
  - Pink
- colour:race
- colour:human
- colour:wine
- colour:original

Colour categories - examples
- colour: amber-coloured; blackened; lime-green; red-and-white; watery-blue; white; yellowing...
- colour:race: whites; blacks; white-skinned; coloureds...
- colour:human: blond; blushed; brunette; red-haired; rosy-cheeked; suntanned; white-bearded...
- colour:wine: white; red.
- colour:original: blacklist; bluebirds; brownies; whitewash; eggwhite...

Quantitative data
- Instances marked as colour in COMPARA: 4,609 in PT and 4,548 in EN
  - 0.0032%
  - 0.0029%
- Number of different colours in COMPARA: 409 in PT (209 lemmas) and 427 in EN

ENGLISH-SPEAKING AUTHORS IN COMPARA
- Julian Barnes
- Lewis Carroll
- Joseph Conrad
- Nadine Gordimer
- Joseph Heller
- Kazuo Ishiguro
- Henry James
- David Lodge
- Ian McEwan
- Mary Shelley
- Joanna Trollope
- Oscar Wilde
- Richard Zimler
### English-speaking authors

<table>
<thead>
<tr>
<th>English Authors</th>
<th>Texts and Dates</th>
<th># types</th>
<th># words</th>
<th>Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carrol</td>
<td>EBLC1 (1871)</td>
<td>71</td>
<td>1</td>
<td>0.99</td>
</tr>
<tr>
<td>Conrad</td>
<td>EBJC1 (1902)</td>
<td>19</td>
<td>41</td>
<td>3.57</td>
</tr>
<tr>
<td>Heller</td>
<td>EUJH1 (1990)</td>
<td>12</td>
<td>25</td>
<td>1.89</td>
</tr>
<tr>
<td>Ishiguro</td>
<td>EBKI1 (1995) e EBKI2 (1989)</td>
<td>23</td>
<td>69</td>
<td>0.81</td>
</tr>
<tr>
<td>James</td>
<td>EUHJ1 (1880) e EUHJ2 (1895) e EUHJ3 (1893)</td>
<td>22</td>
<td>65</td>
<td>1.38</td>
</tr>
<tr>
<td>Poe</td>
<td>EUEP1 (1837)</td>
<td>72</td>
<td>0</td>
<td>0.82</td>
</tr>
<tr>
<td>Shelley</td>
<td>EBMS1 (1818)</td>
<td>61</td>
<td>2</td>
<td>0.57</td>
</tr>
<tr>
<td>Wilde</td>
<td>EBOW1 (1890)</td>
<td>52</td>
<td>166</td>
<td>6.47</td>
</tr>
</tbody>
</table>

### Most colourful authors (no. of different colour types)

1. **Camilo Castelo Branco**
   - negro, verde, amarelo, azul, cor, preto.

2. **Agualusa**
   - black, blue, green, white, red, blueness.

3. **Joanna Trollope**
   - azul, verde, amarelo, azul, cor, preto.

### Least colourful authors (no. of different colour types)

1. **Mary Shelley**
   - negro, verde, amarelo, azul, cor, preto.

2. **Joanna Trollope**
   - azul, verde, amarelo, azul, cor, preto.

### Authors’ favourite colour groups

- **Black**: 54%
- **White**: 47%
- **Red**: 7%

(Each author contributed only with his/her favourite colour group)
Author with the highest individual percentage of the colour:

**English-speaking authors**

- **Lewis Carroll**
- **Mary Shelley**
- **Henry James**
- **J. M. Barrie**

**Portuguese-speaking authors**

- **José de Alencar**
- **Camilo Castelo Branco**
- **Mia Couto**
- **José Eduardo Agualusa**

Favourite colour groups

(Each author contributed only with his/her favourite colour group)

**English-speaking authors**

- White 43%
- Black 36%
- Red 14%
- Blue 7%

**Portuguese-speaking authors**

- White 43%
- Black 36%
- Red 14%
- Blue 7%

All colour groups

(The authors contributed with all their colours)

**English-speaking authors**

- Blue 15%
- Red 13%
- Yellow 6%
- Green 9%

**Portuguese-speaking authors**

- Blue 15%
- Red 13%
- Yellow 6%
- Green 9%

Races (colour:race)

9 out of 14
Does colour quantity change with time?
Number of colour types per text dates

Does colour quantity change with time?
Number of colour types per authors' birth date

Does colour quantity change with time?
Number of colour types per text dates

Does colour quantity change with time?
Number of colour types per authors' birth date

Grammatical Category of Colours

- ADJECTIVE: 3,721 words
  - (red-hot needle, white image, blue eyes,...)

- NOUN: 932 words
  - (...came back in the afternoon gold.)
    - (hand-painted in orange.)

- VERB: 108 words
  - (...she smiled and blushed.)
    - (eyes blackened and engorged by fatigue...)

- ADVERB: 7 word
  - (...she smiled rosily and said...)
Grammatical Category of Colours

- **ADJECTIVE**: 3,193 words (161 lemmas)
  - (carrow verde, fato azul)

- **NOUN**: 1,349 words (93 lemmas)
  - (...)o verde era uma cor que... (...estou em pleno azul, suspenso...)

- **VERB**: 107 words (15 lemmas)
  - (...muito além daquela serra, que ainda azula no horizonte...)
  - (...as saladas esverdinhavam à hora do jantar.)

- **ADVERB**: 1 word (1 lemma)
  - (púrpureamente)

- **ADJn**: 29 words (16 lemmas)
  - (-- No entanto gosto do azul -- prosseguiu, a cliente -- e o verde é muito bonito. )

- **ADJ_N**: 27 words (5 lemmas)
  - (Mas Isidoro, mesmo preto, não era capaz de uma coisa dessas.)

- **ADJECTIVE**: 3,193 words (161 lemmas)
  - (carro verde, fato azul)

Morphosyntactic Patterns in English

**COLOUR + NOUN (1,138)**

- hair 49 (brown hair, reddish hair, wine-red hair...)
- eye 45 (red eye, watery-blue eyes,...)
- silk 22 (crimson silk, amber-coloured silk,...)
- paper 16 (yellowing paper, brown paper,...)
- light 16 (purple light, greyish-blue light,...)
- velvet 11 (black velvet, purple velvet,...)
- sky 11 (indigo sky, grey sky,...)
- plastic 11 (pale-grey plastic, white plastic,...)
- bear 10 (brown bear, black bear,...)

Morphosyntactic Patterns in Portuguese

**COLOUR + NOUN (914)**

- cabelo 41 (cabelos brancos, cabelo negro,...)
- olho 32 (olhos azuis, olhos castanhos, olhos verdes,...)
- luz 21 (luz vermelha, luz anilada,...)
- seda 20 (seda verde, seda rosa,...)
- papel 19 (papel colorido, papel amarelo,...)
- mármore 13 (mármore negro, mármore verde,...)
- vestido 10 (vestido azul, vestido prateado,...)
- céu 9 (céu azul, céu cinzento, céu negro,...)

**COLOUR + "and" + COLOUR (98)**

- pink and (colour) 17
- black and (colour) 10
- red and (colour) 9
- green and (colour) 7
- white and (colour) 6
- orange and (colour) 6
- blue and (colour) 6

**COLOUR + "e" + COLOUR (32)**

- branco e (colour) 7
- azul e (colour) 5
- preto e (colour) 3
- ouro e (colour) 2
- verde e (colour) 2
- vermelho e (colour) 2
- amarelo e (colour) 2
Morphosyntactic Patterns using Colour

**COLOUR (hyphenated words)**

- colour-[borrowed from something] 85 (cherry-red,...)
- colour-[hue modifier] 66 (deep-blue,...)
- colour-[applied to something] 66 (green-shirted,...)
- colour-["colo(u)red"] 81 (carrot-colored,...)
- colour-[colour] 46 (pink-brown,...)
- colour-[hair-related] 44 (red-haired,...)
- colour-[skin-related] 20 (olive-skinned,...)
- colour-[eye-related] 8 (gray-eyed,...)

*Colour can appear on both sides of the hyphen*

**Example: Julian Barnes**

<table>
<thead>
<tr>
<th>Originals vs. Translations</th>
<th>Colour:human</th>
<th>Julian Barnes</th>
<th>Translator (JB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Julian Barnes</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Translator (JB)</td>
<td>4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Translations fall into other colour categories

- No translation

---

**Morphosyntactic Patterns using Colour**

**COLOUR (hyphenated words)**

- [hue modification] 104 (verde-vivo, vermelho-escuro,...)
- cor-de-... 62 (cor-de-rosa, cor-de-ralo,...)
- [borrowed from something] 46 (verde-musgo,...)
- [several colours] 6 (verde-azul, rosa-e-azul,...)
- [applied to something] 1 (saliva-laranja)
Example: Julian Barnes

<table>
<thead>
<tr>
<th></th>
<th>Red</th>
<th>Julian Barnes</th>
<th>Translator (JB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>16</td>
<td>red</td>
<td>red</td>
</tr>
<tr>
<td>Translit</td>
<td>9</td>
<td>red</td>
<td>red</td>
</tr>
</tbody>
</table>

No translation

Translations fall into other colour categories

Which language uses more colour?

COLOUR DENSITY (per 1,000 words)
- English: 3.041
- Portuguese: 3.315

COLOUR DIVERSITY (per 1,000 words)
- English: 12.569
- Portuguese: 7.980

Which language uses more...?

COLOUR: RACE (per 10,000 words)
- English: 5.026
- Portuguese: 4.997

COLOUR: WINE (per 10,000 words)
- English: 0.145
- Portuguese: 0.079

COLOUR: ORIGINAL (per 10,000 words)
- English: 2.055
- Portuguese: 1.414

We presented a large repository of aligned coloured texts available to every researcher interested in Portuguese or English.

COMPARA is also open to the semi-automatic annotation of other semantic fields, in collaboration with interested researchers who volunteer to revise the final annotation.

Acknowledgment

- This work was done in the scope of Linguateca, contract no. 339/1.3/C/NAC, project jointly funded by the Portuguese Government and the European Union.