

Uso de técnicas de text mining para o estudo da natureza da resposta estética a pinturas de Paula Rego em adolescentes

Alexandra Santos Silva¹ & Marta Castel-Branco²

¹Faculdade de Medicina da Universidade de Lisboa

²StartFactor, statistical consulting and training.
CEAUL, Centro de Estatística e Aplicações da Universidade de Lisboa.

8 Outubro 2015



Objetivo



Cruzar a intencionalidade da pintora aquando da criação da obra *O Crime do Padre Amaro* com a natureza da resposta à experiência estética por parte dos adolescentes

- partindo das cadeias para $n = 5$, obtém-se *MFFMFMM*, *MFMFMM*, e *MFFFFMM*.

Concluímos assim que o número N_n de cadeias com n crias e com as características requeridas verifica a relação recursiva

$$N_n = N_{n-1} + N_{n-2}, \quad n \geq 2.$$



Figura 3.41: Paula Rego. Coelha grávida a dizer aos pais (1982).

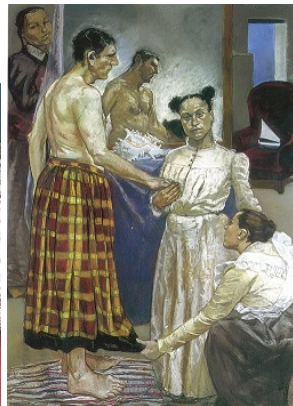
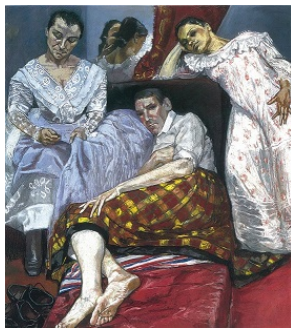
Mas a coelha pretende saber muito mais: quer calcular a probabilidade p_n de ser necessário ter n filhos para conseguir aquele feito. Obviamente, espera-se que o

*in Pestana D. e Velosa S. (2002).
Introdução à Probabilidade e à Estatística*



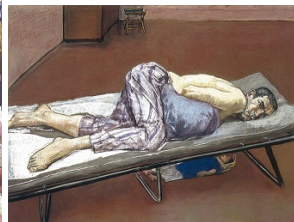
- Extraordinária contadora de histórias.
- Utilização de linguagem não verbal como forma de expressão.
- Justiça.

As figuras



Figuras (1) "O Anjo"; (2) "Entre as Mulheres"; (3) "A Mãe"

As figuras



Figuras (4) "A Capoeira"; (5) "O Embaixador de Jesus"; (6) "A Cella

Recolha de dados

- Realização de entrevistas individuais presenciais a 30 adolescentes, 15 rapazes e 15 raparigas, com idades compreendidas entre os 13 e os 16 anos ($\bar{X} = 14.8$; $SD = 0.75$).
- As entrevistas foram conduzidas em 3 escolas: QM - Escola Quinta de Marrocos (n=10), NP - Escola Navarro Paiva (n=10) e EA - Escola Alemã (n=10). Em cada escola foram entrevistados 5 rapazes e 5 raparigas.
- A cada adolescente foram apresentadas as seis reproduções da temática *O Crime do Padre Amaro* da pintora Paula Rego e foram feitas 6 questões acerca de cada imagem.

Recolha de dados

Questões:

- Q1: O que vê aqui?
- Q2: O que é que vê que o faz ver isso?
- Q3: Que história poderia contar a partir desta imagem?

- Q4: Qual a imagem que mais gostou? Porquê?
- Q5: Qual a imagem que menos gostou? Porquê?
- Q6: Qual o impacto/sentimento que este conjunto de imagens teve em si?

Texto como *input de informação*

"Descrição objetiva, sistemática e quantitativa do conteúdo manifesto da comunicação" (Berelson, 1952)

"A análise de conteúdo dos textos escritos não serve apenas a descrição mas também a inferência no sentido de passagem da descrição à interpretação, quando da atribuição de sentido às características do material que foram levantadas, enumeradas e organizadas." (Bardin, 1979)

Text mining

Transformação do texto num formato estruturado com base na frequência de termos e aplicação de técnicas convencionais de análise de dados.

- IMPORTAÇÃO DO TEXTO - CORPUS
- PRÉ-PROCESSAMENTO
- MATRIZ DE TERMOS
- ANÁLISE DE OCORRÊNCIAS
- ANÁLISE DE CLUSTERS



CORPUS: coleção de documentos de texto, organizados por imagem e questão.

Exemplo: I6:"A Cela"; Q1: O que vê?

[[1]] Um senhor que ficou sozinho que provavelmente perdeu a mulher. Num quarto sozinho. Uma senhora que pode ser a mulher ou a filha debaixo da cama.

(...)

[[9]] Um homem que parece estar doente ou triste.

(...)

[[25]] Homem deitado numa cama com uma cara um bocado rabugenta, com uma almofada debaixo da barriga. Uma sala grande mas um bocado vazia, um armário e uma mesa. Debaixo da cama uma boneca.

(...)

[[30]] Um homem deitado na cama que parece estar com dores de barriga. Em baixo pode ser a imagem do homem em pequeno.

Pré-processamento

- Eliminação de espaços em branco;
- Transformação de maiúsculas em minúsculas;
- Remoção de pontuação e de números;
- Remoção de "stopwords" (palavras comuns, cujo valor de informação é muito baixo) - "o", "a", "e", "uma", "se", "ou", "de", "cima", "baixo", "não", "sim" (..) e determinados verbos - "fazer", "ser", "ver", "parecer", "desenhar" (...);
- Tratamento de sinónimos - identificação de palavras distintas com o mesmo significado e transformação dos sinónimos numa mesma palavra - "senhora" e "mulher", "senhor" e "homem" (...).



Exemplo: I6:"A Cela"; Q1: O que vê?

[[1]] homem sozinho perdeu mulher quarto sozinho mulher mulher filha cama
(...)

[[9]] homem doente triste
(...)

[[25]] homem deitado cama rosto rabugenta almofada barriga sala grande vazia armario
mesa cama boneca

(...)

[[30]] homem deitado cama dores barriga imagem homem pequeno

MATRIZ DE TERMOS: Transformação do texto pré-processado num formato estruturado.

id	almofada	(...)	cama	homem	doente	(...)	sozinho	triste	zangado
1	0	(...)	1	1	0	(...)	2	0	0
(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)
9	0	(...)	0	1	1	(...)	0	1	0
(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)
25	1	(...)	2	1	0	(...)	0	0	0
(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)	(...)
30	0	(...)	1	2	0	(...)	0	0	0

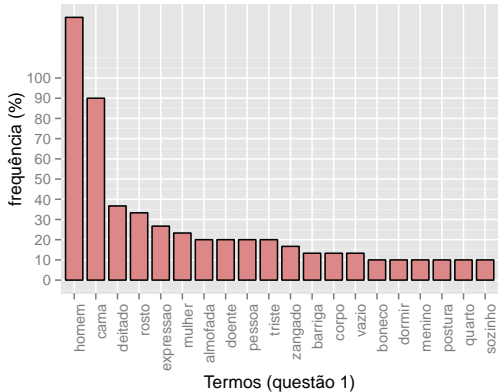
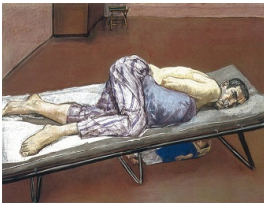
Análise de ocorrências

Termos	Frequência
Homem	39
Cama	27
Deitado	11
Rosto	10
Expressao	8
Mulher	7
Almofada	6
Doente	6
Pessoa	6
Triste	6
Zangado	5
Barriga	4
Corpo	4
Vazio	4
Boneco	3
Dormir	3
Menino	3
Postura	3
Quarto	3
Sozinho	3



Análise de ocorrências

Hipótese implícita: quanto maior for o interesse do emissor por um dado objeto (termo) maior será a sua frequência de ocorrência no respetivo discurso.



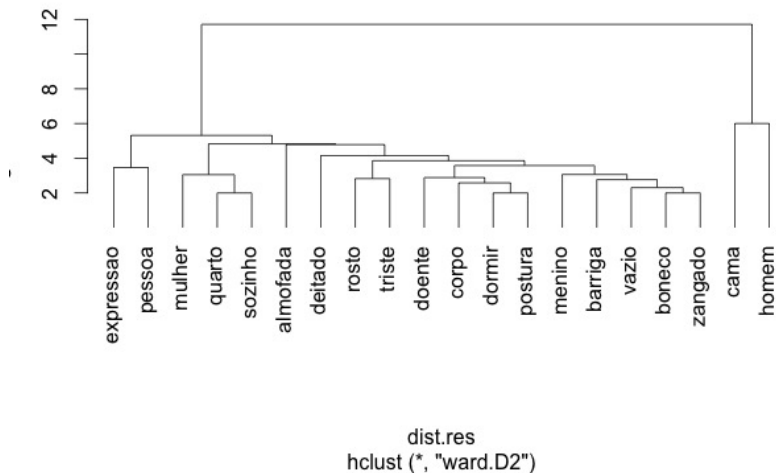
Análise de clusters - Etapas

OBJETIVO: Agrupar os termos em sub-grupos homogêneos, na tentativa de encontrar grupos com significado.

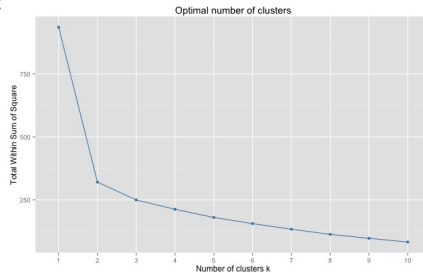
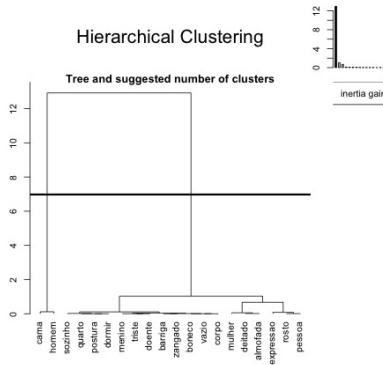
- Seleção de uma medida de semelhança/dissemelhança: distância Euclidiana como medida de proximidade (dissemelhança) entre os termos;
- Escolha do método de formação dos clusters: método hierárquico aglomerativo;
- Escolha do critério usado para definir as distâncias entre os grupos: critério de Ward (minimização da variância intra-classe);
- Representação dos dados (dendograma) e escolha do número de grupos;
- Atribuição de significado aos grupos.

Dendrograma

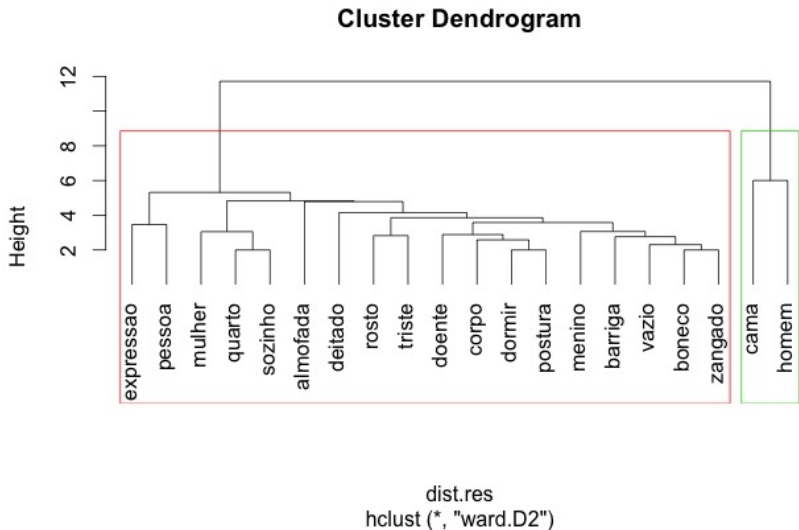
Cluster Dendrogram



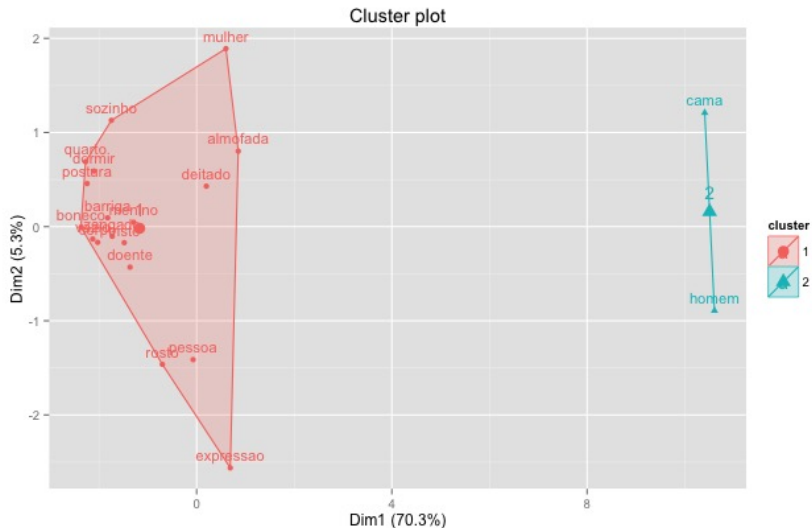
Número de clusters



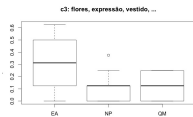
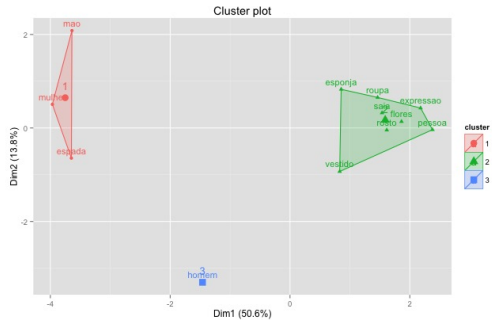
Dendrograma



"Cluster plot"

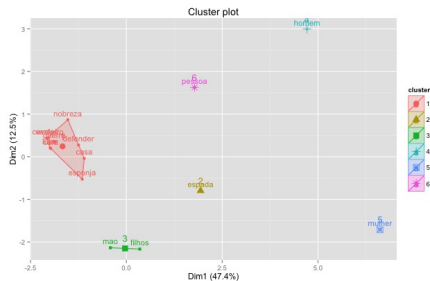


O Anjo - Q1: "O que vê?"



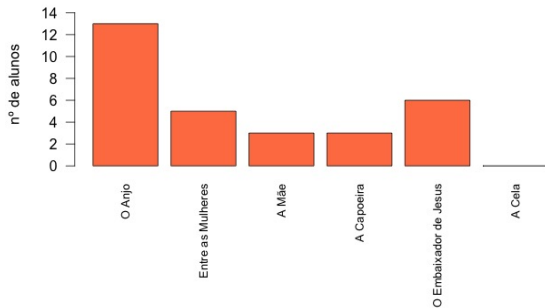
O Anjo - Q3: "Que história poderia contar a partir desta imagem?"

Termos	Frequência
Mulher	17
Espada	14
Homem	14
Pessoa	10
Casa	6
Mão	6
Filhos	4
Guerra	4
Vestes	4
Cara	3
Cavaleiro	3
Defender	3
Esponja	3
Lutar	3
Nobreza	3



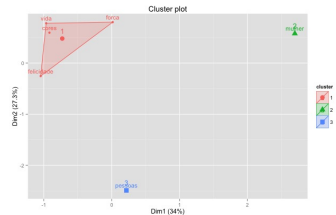
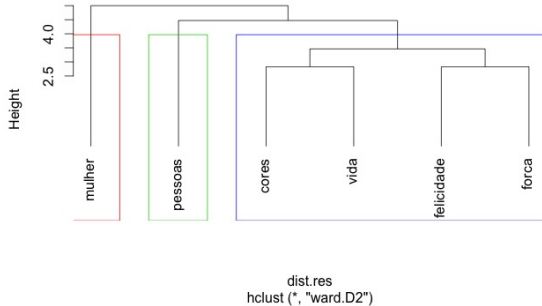
"Ela é ao mesmo tempo um Anjo da guarda e um Anjo vingador. A sua missão é proteger e vingar. Traz os símbolos da Paixão, a espada e a esponja. Ela apareceu, ganhou forma e não sabemos o que se lhe seguirá." Paula Rego

Q4: Qual a imagem que mais gostou?

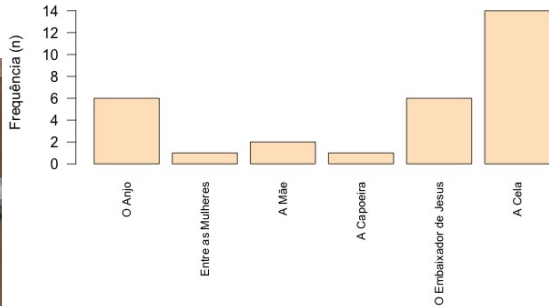


Q4: Qual a imagem que mais gostou? Porquê?

Cluster Dendrogram

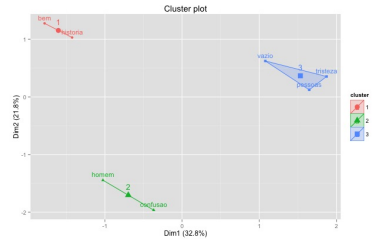
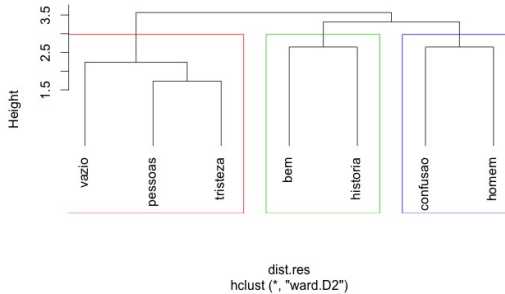


Q5: Qual a imagem que menos gostou?



Q5: Qual a imagem que menos gostou? Porquê?

Cluster Dendrogram



Q6: Qual o impacto/sentimento que este conjunto de imagens teve em si?

A word cloud visualization of sentiment analysis results. The words are arranged in a vertical stack. 'confusao' is at the top in green. Below it is 'vida' in blue. Then 'estranheza' in blue. At the bottom is 'tristeza' in orange. The words are of varying sizes, with 'tristeza' being the largest.

confusao
vida
estranheza
tristeza

Questões 4, 5 e 6

Termos Q4	Frequências
Mulher	9
Pessoas	6
Cor	5
Vida	5
Felicidade	3
Força	3

Termos Q5	Frequências
Confusão	5
História	4
Homem	4
Tristeza	4
Bem	3
Pessoas	3
Vazio	3

Termos Q6	Frequências
Tristeza	7
Confusão	5
Estranheza	3
Vida	3

Conclusões



A intencionalidade da autora cruza-se com a experiência à resposta estética, ou seja, que os adolescentes tornaram-se co-criadores.