

# Análisis de textos desde la estilometría

Francisco Javier Blasco Pascual  
y Cristina Ruiz Urbón



# ÍNDICE

Nota de los autores.....	13
I. ¿Qué es la estilometría?.....	15
I.1. El futuro de la estilometría en los estudios literarios.....	23
I.2. Nuestros objetivos.....	27
I.3. Idiolecto.....	29
II. El tratamiento de los textos y la obtención de datos.....	43
II.1. Tratamiento del corpus.....	43
II.2. La cosecha de datos y su operatividad.....	47
II.3. Tareas para la estructuración de la información.....	48
III. Análisis de caracteres.....	51
a) <i>Simple conteo</i> .....	52
b) <i>Las cadenas de Markov</i> .....	55
c) <i>El paquete 'stylo'</i> .....	58
IV. Análisis léxico.....	61
IV.1. Agrupaciones y clasificaciones 'stylo'.....	63
a) <i>'Classify'</i> .....	68
b) <i>'General imposters'</i> .....	73
c) <i>'Rolling.classify'</i> .....	74
IV.2. Análisis con LIWC.....	76

IV.3. Palabras de función.....	79
a) <i>Hugh Craig (PCA)</i> .....	80
b) <i>La ley de Zipf</i> .....	81
IV.4. La función tf-idf.....	84
IV.5. Corpus de referencia.....	86
IV.6. Análisis cualitativos.....	90
a) <i>Ngram Viewer</i> .....	90
b) <i>Diccionarios de emociones</i> .....	91
c) <i>Un ejemplo: Cervantes / Lope de Vega</i> .....	93
V. Morfología.....	101
V.1. Clases de palabras.....	101
a) <i>'Udpipe'</i> .....	102
b) <i>Análisis de componentes principales</i> .....	109
V.2. Usos verbales: prefijación y derivación.....	112
VI. Sintaxis: la frase y sus componentes.....	115
VI.1. Longitud de frase medida en palabras.....	118
VI.2. <i>N-grams</i> de palabra.....	121
VI.3. Colocaciones (o <i>verbatim</i> ) y coocurrencias.....	122
VI.4. Marcadores del discurso.....	131
VI.5. Clases de palabras dentro de la frase.....	133
VI.6. Nombre de persona y dependencias (LSD).....	138
VI.7. Tipo de oración.....	139
VI.8. Análisis cualitativos de interés.....	140
a) <i>Conjunciones</i> .....	141
b) <i>Codificación de los argumentos</i> .....	143
c) <i>Datos adicionales resultantes de la 'close reading'</i> .....	144
VII. Semántica.....	147
VII.1. Comportamiento de un conjunto de textos ante diversos campos semánticos.....	148
a) <i>Tonalidad emocional</i> .....	149
b) <i>Temas de trabajo y ocio</i> .....	151
c) <i>Clases de palabras</i> .....	152

VII.2. Palabras clave y TF_IDF.....	153
VII.3. Tópicos y contenido de un texto.....	158
VII.4. Palabras clave y palabras asociadas.....	165
Conclusiones.....	169
Bibliografía.....	173
Herramientas utilizadas.....	191
Corpus de referencia.....	193
Índice de textos analizados.....	195
Glosario.....	201
Índice onomástico y de temas.....	207