

¿Qué es la estilometría?

Usos y aplicaciones



José Manuel Fradejas Rueda

Departamento de Lengua Española
Universidad de Valladolid

josemanuel.fradejas@uva.es | @JMFraDeRue

DH@MADRID SUMMER SCHOOL 2019

Creando un proyecto de Humanidades
Digitales usando el modelado de datos y el
procesamiento de textos

Madrid, del 1 al 3 de julio de 2019

Sigue el curso de forma presencial u online

Sala de Grados. Facultad de Educación. UNED. Calle de
Juan del Rosal, 14. (Ciudad Universitaria)



European
Research
Council



POSTDATA
Poetry Standardization
and Linked Open Data



JRR Tolkin en su despacho

Fuente: <http://time.com/4941811/hobbit-anniversary-1937-reviews/>



1981



Sinclair ZX81

Procesador Z80 @ 3.25 MHz

RAM 1 KB = 1024 bytes

ROM 8 KB

Pantalla: 24 líneas x 32 caracteres

UHF blanco y negro

Audio cassette 250 baudios

David Lodge

El mundo es un pañuelo



COMPACTOS ANAGRAMA

... fui con él a su Centro de Informática. ... Dempsey llevó el peso de la conversación. «Supongo que habrá oído hablar de nuestro

Centro para la Estilística

Informatizada», me dijo. «No — contesté—. ¿Dónde está?»

«¿Dónde? Pues aquí, creo yo —me dijo—. Quiero decir que soy yo, y por tanto está allí donde estoy yo.

Es decir, allí donde esté yo cuando me dedico a la **estilística**

informatizada, que es tan solo uno de mis intereses investigadores



La vieja filología ayuda a la justicia: Crítica textual, lingüística forense y humanidades digitales

José Manuel Fradejas Rueda
Universidad de Valladolid

La lengua a juicio: el peritaje lingüístico forense
Valladolid, 4 de marzo de 2016





LEER (COMO ESCRITOR)

EL RITMO DE ESTE PÁRRAFO
ES MAGNÍFICO

ESTE EPILOGO SOBRE
LA NEOLENGUA
VA A CAUSARME
UN ORGASMO

EL BALANCE ENTRE
DEFECTOS Y VIRTUDES
ES LO QUE HACE A
ESTE PERSONAJE TAN
UNICO Y HUMANO

BRILLANTE
USO DE
SUBORDI-
NADAS

EL AUTOR USA
OBJETOS Y ESCENAS
COTIDIANAS PARA
MOSTRAR LA MISERIA EN
LA QUE VIVEN ESTOS
PERSONAJES SIN
DECIRLO EXPLÍCITAMENTE

¿CUANTAS VECES REESCRIBO ESTO?

NO LEÍA NADA TAN
BUENO DESDE EL
PRÓLOGO DE
LOLITA

LOS DIÁLOGOS SON TAN
FLUIDOS Y REALISTAS...
VEAMOS QUE PALABRAS
ESCOGIÓ EL AUTOR PARA
RETRATAR CADA PER-
SONAJE Y QUE
TIPO DE INFOR-

PUN-
TUA-
CIÓN
PERFECTA





LEER
COMO ORDENADOR

Estilometría





Wincenty Lutosławski

(1863–1954)

estilometría
estilema

«Principes de stylométrie appliqués à la chronologie des œuvres de Platon», 1898

https://www.persee.fr/doc/reg_0035-2039_1898_num_11_41_5847

Estilometría

Análisis estadístico del estilo literario

The background features several light gray, curved, brushstroke-like shapes on the right side, and a thin horizontal line at the bottom.

Estilometría

Trata de

- Identificar **semejanzas y diferencias** entre textos y
- **Agrupar** los textos de acuerdo con su características lingüísticas

para

- **Detectar señales** en los textos (autoría, géneros, género (sexo), origen, estilo, etc.)

Estilometría

Federalist Papers

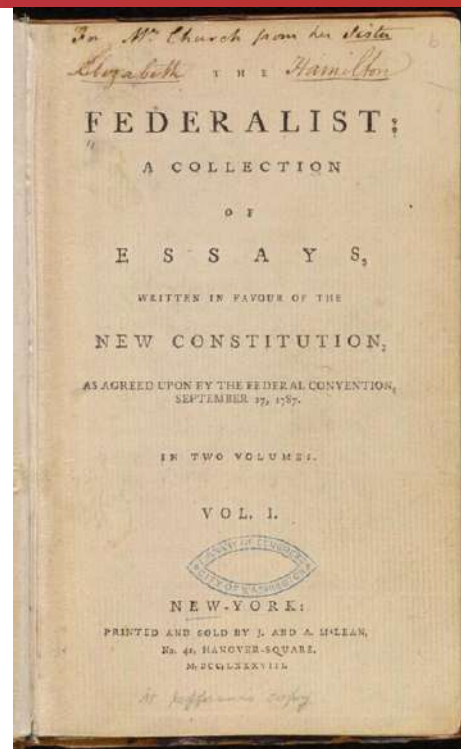
85 artículos

Publius

Alexander Hamilton

James Madison

John Jay



Estilometría

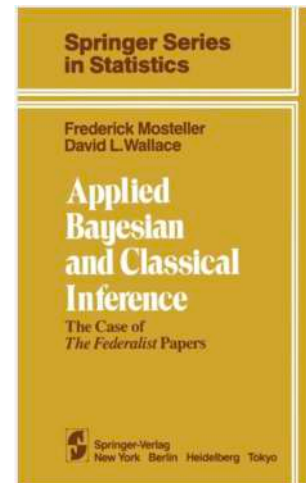
TABLE 1.4-1
INCIDENCE: NUMBERS OF PAPERS IN WHICH WORD OCCURRED
AT LEAST ONCE

	<i>enough</i>	<i>while</i>	<i>whilst</i>	<i>upon</i>	Number of papers examined
Hamilton	14	10	0	23	23
Madison	0	0	13	4	19
Disputed	0	0	5	1	12
Joint	1	0	2	2	3

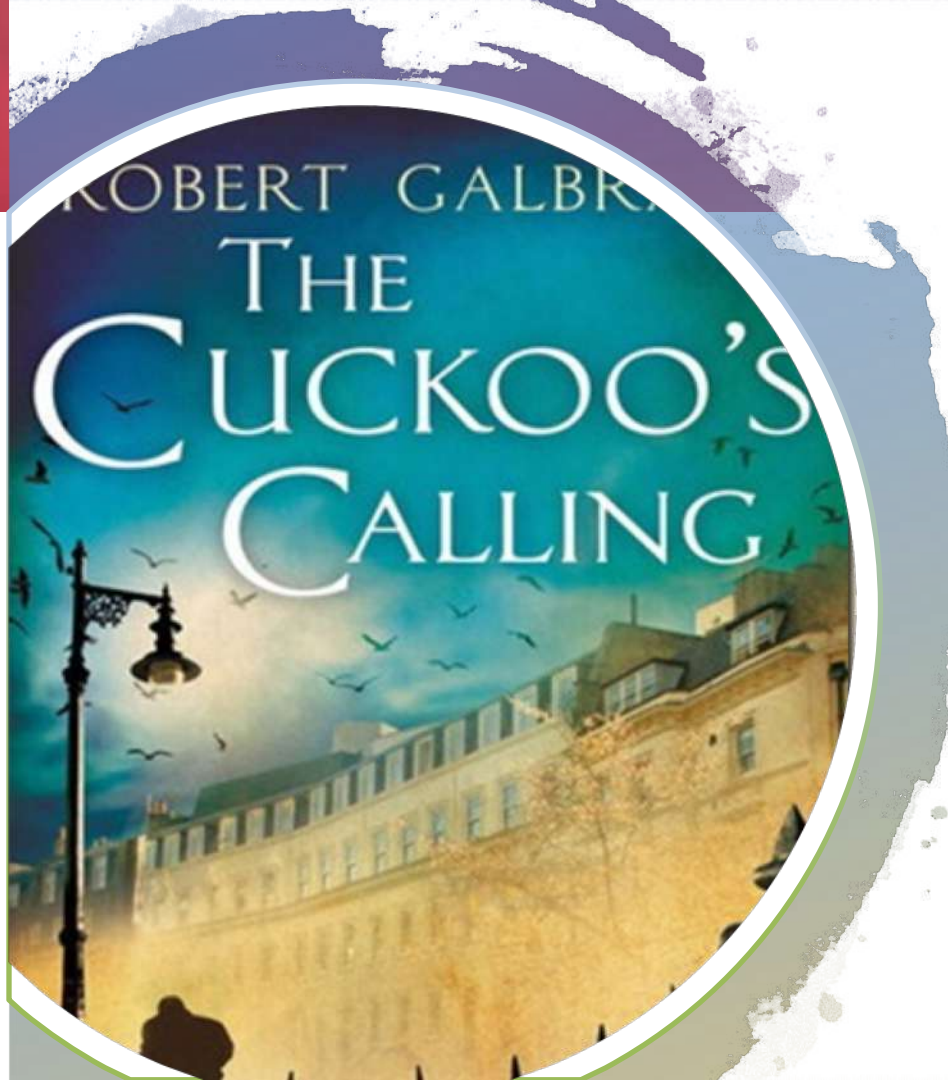
TABLE 1.4-2
RATES PER 1000 WORDS

	<i>enough</i>	<i>while</i>	<i>whilst</i>	<i>upon</i>	Total words in 1000's
Hamilton	0.59	0.26	0	2.93	45.7
Madison	0	0	0.47	0.16	51.0
Disputed	0	0	0.34	0.08	23.9
Joint	0.18	0	0.36	0.36	5.5
					126.1

Mosteller & Wallace 1964



Estilometría

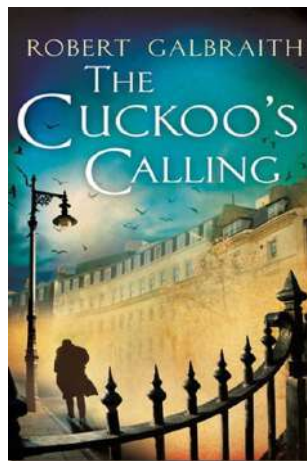


Estilometría

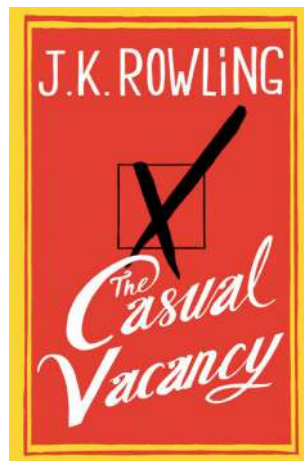


Patrick Juola

¿



=



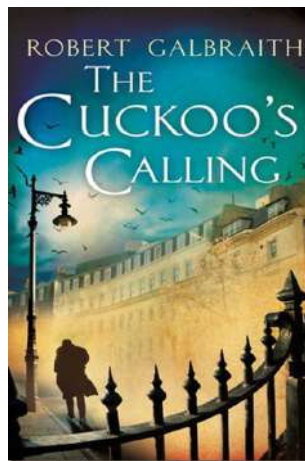
?

Estilometría

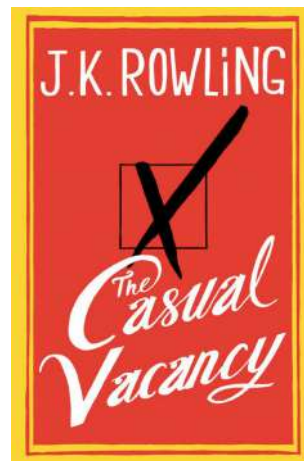


Patrick Juola

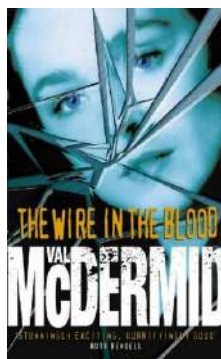
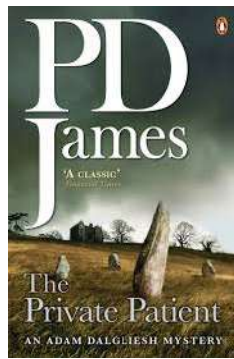
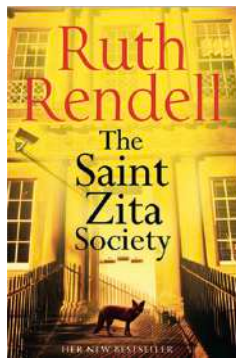
¿



=



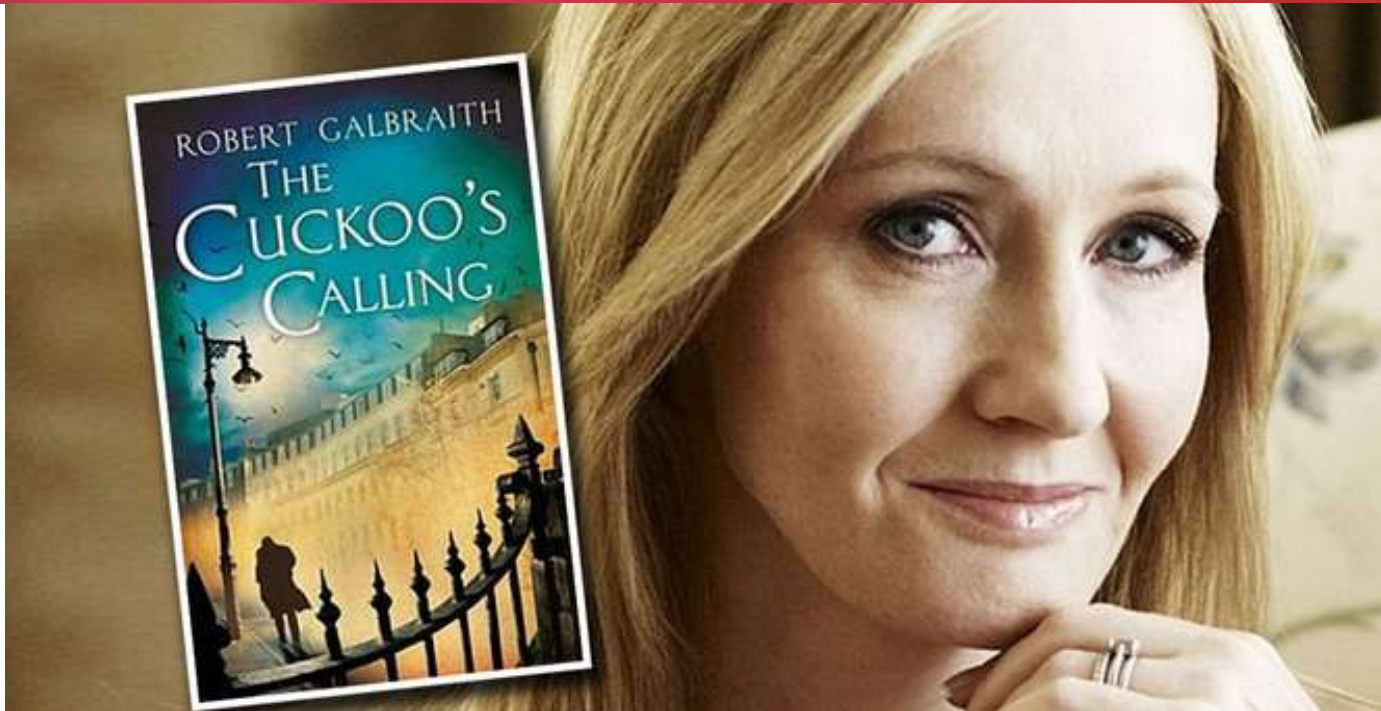
?



Estilometría

- Distribución de la longitud de las palabras
- Uso de las 100 palabras más comunes
- Distribución de 4-grams
- Distribución de bigramas

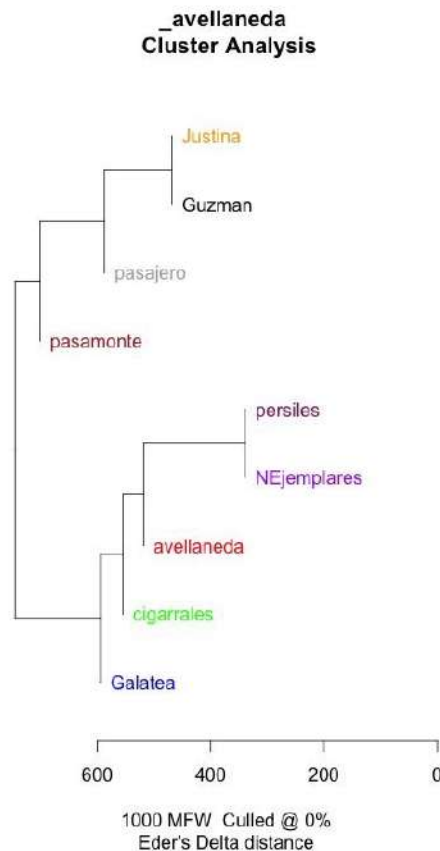
Estilometría



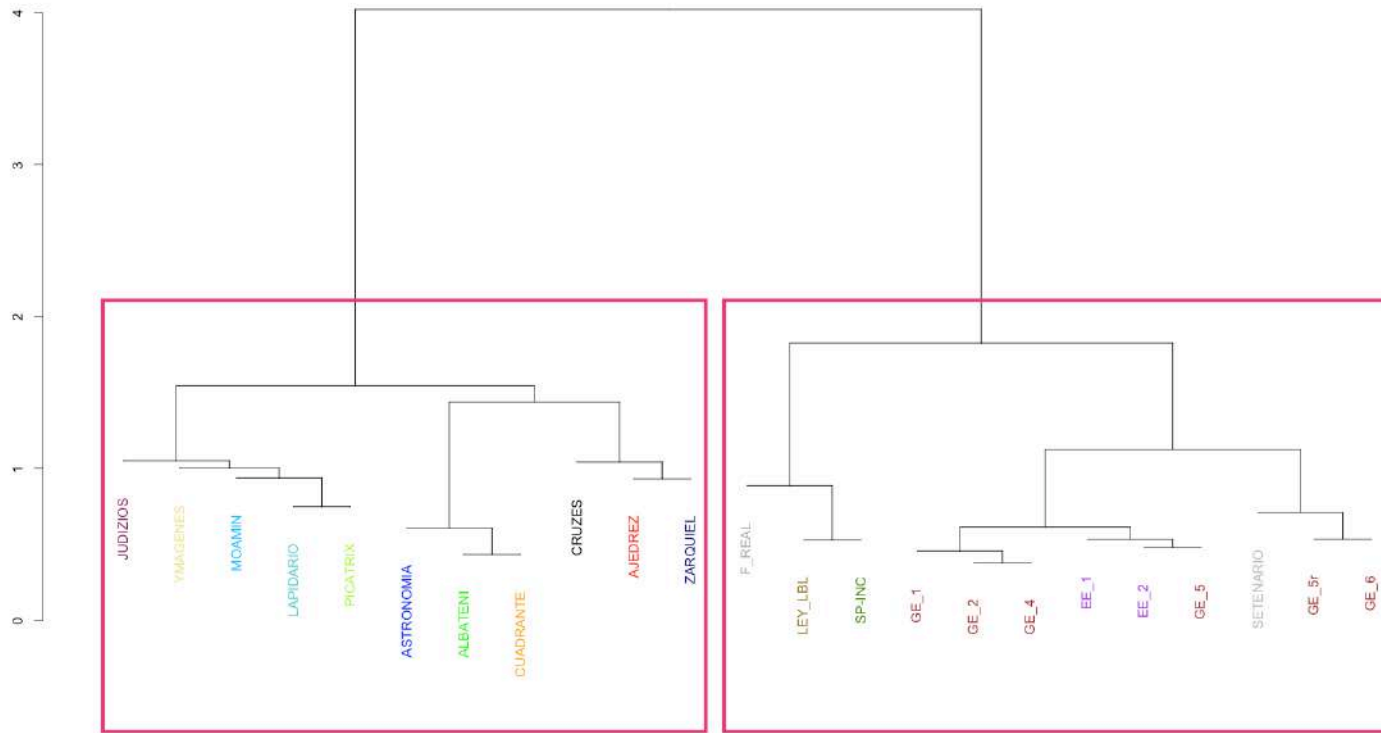
“The only person consistently suggested by every analysis was Rowling, who showed up as the winner or the runner-up in each instance”

Problemas de autoría

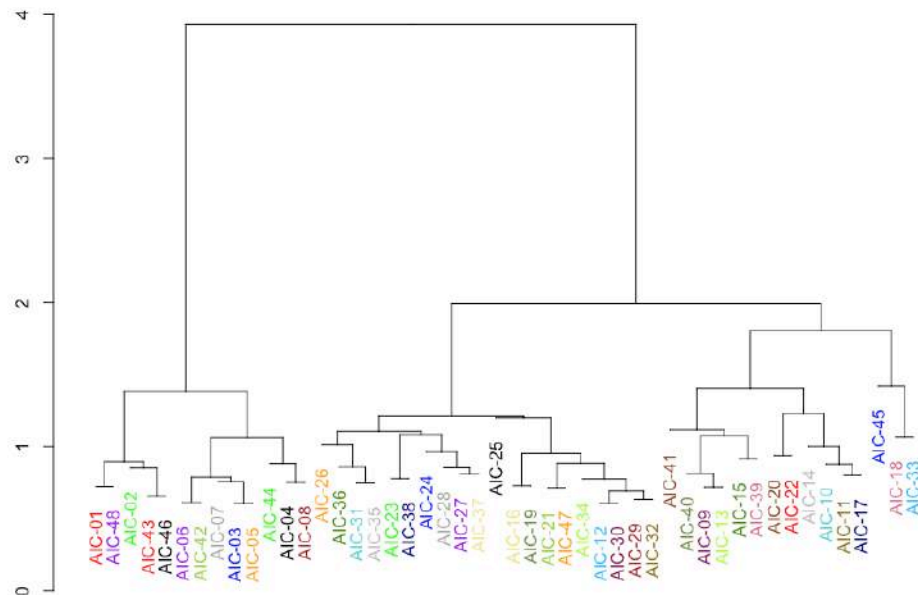
Quijote de Avellaneda



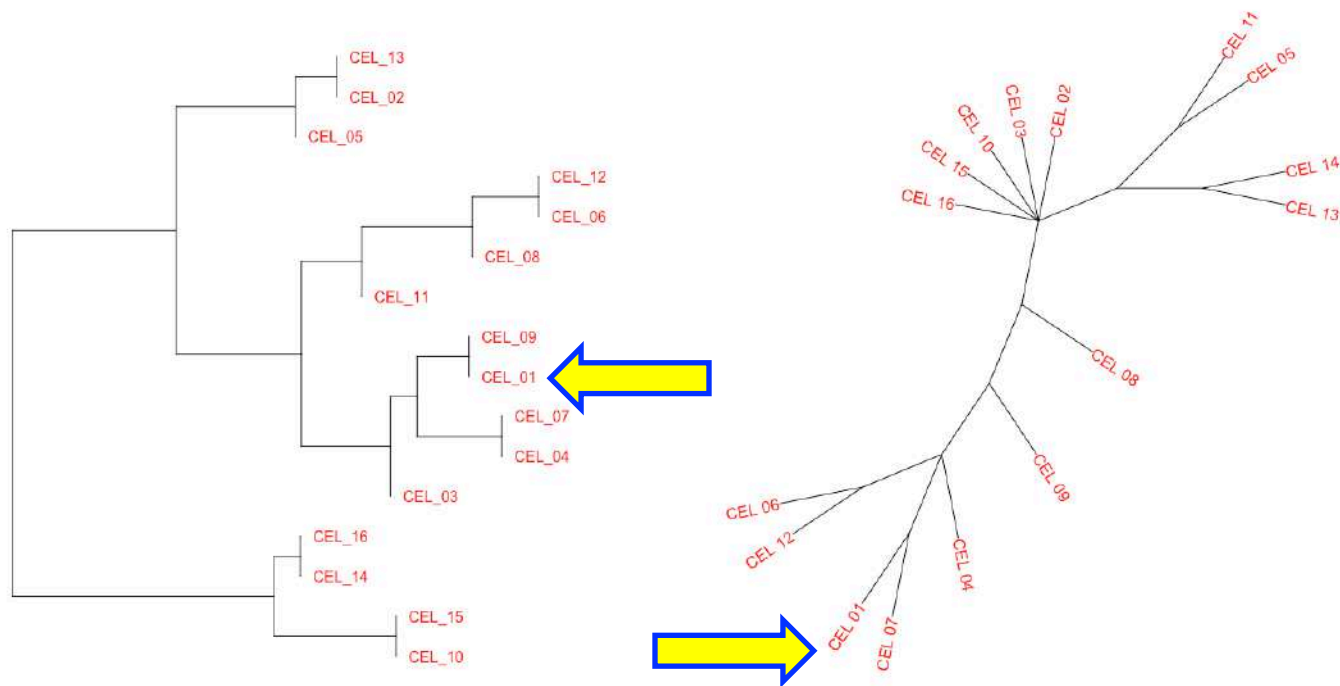
Alfonso X el Sabio



Libro de la caza de las aves



La Celestina (1499)





María de Zayas

¿Alonso Castillo Solórzano?



Matèria: Filologia
Format: 13 x 20 cm
Pàgines: 128
ISBN: 978-84-9168-288-2
Data d'edició: 2019

María de Zayas y otros heterónimos de Castillo Solórzano

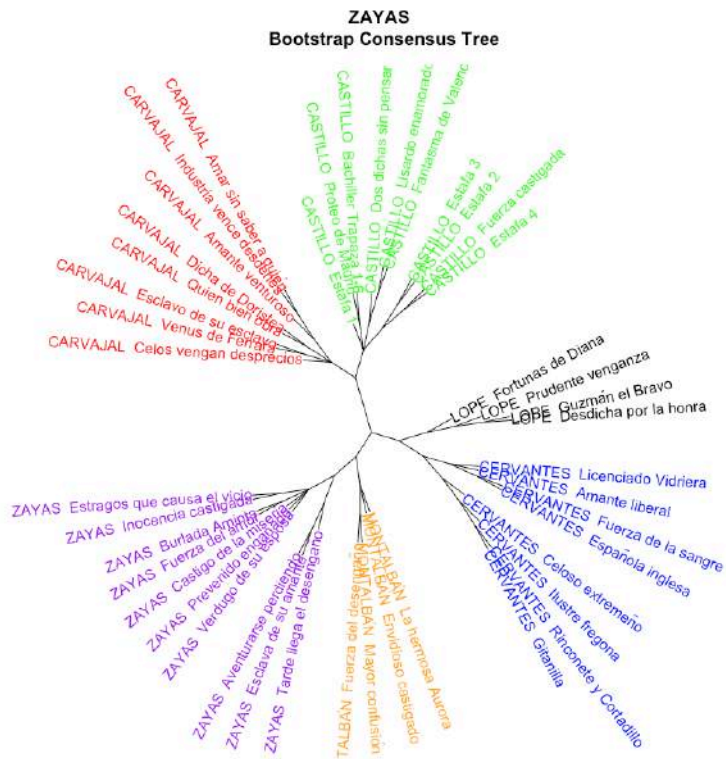
Rosa Navarro Durán

Investigar, para un científico o para un filólogo, supone descubrir algo nuevo. Y puede suceder que, al enfocar con la lupa zonas oscuras de los textos literarios, se encuentren cosas nunca antes oídas ni vistas. Este ensayo ofrece a los lectores un ejemplo: es uno de los frutos —cosecha de 2018— de los muchos años que Rosa Navarro Durán ha dedicado a la investigación, paralela a la docencia en la Universidad de Barcelona. En él están las pruebas que demuestran que la más famosa novelista española del siglo XVII, María de Zayas, no existió: fue solo uno de los heterónimos que inventó Alonso de Castillo Solórzano para escribir algunas de sus muchas novelas breves, curiosamente las más nombradas.

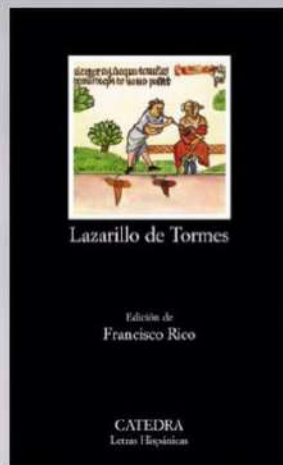
Altres títols de la col·lecció

<https://elcultural.com/quien-se-esconde-tras-maria-de-zayas>

María de Zayas es María de Zayas



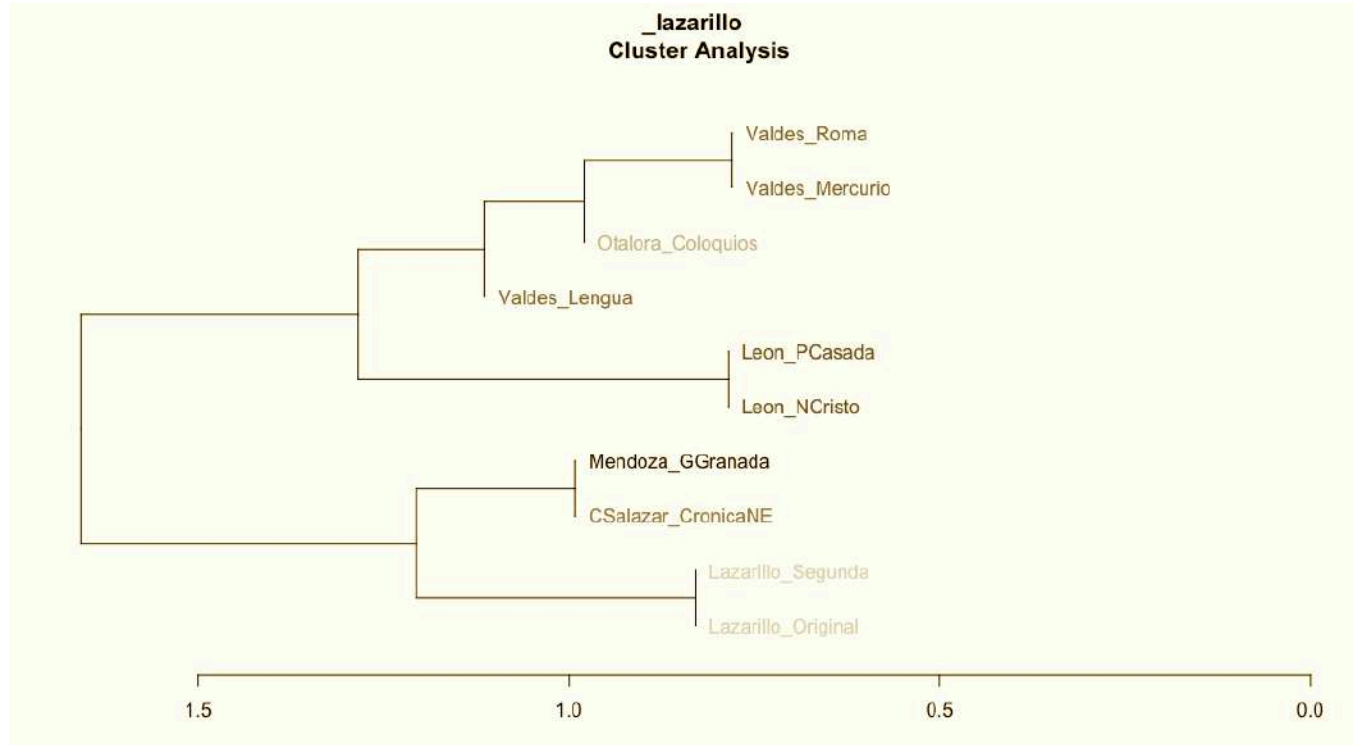
Lazarillo de Tormes



¿AUTOR?



Lazarillo de Tormes

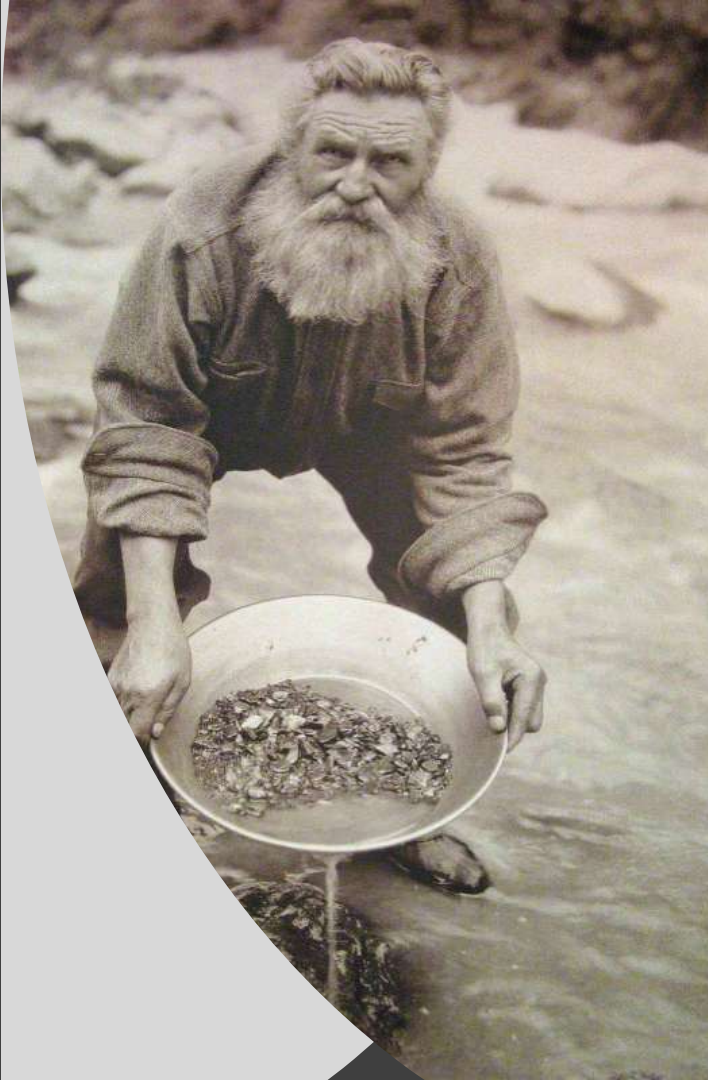


<http://linhd.uned.es/estilometria-un-ambito-por-descubrir/>

8/11/2016

Estilometría

es algo más que desentrañar
problemas de autoría



MINERÍA de DATOS



MINERÍA de TEXTOS





GRATIS TOTAL



Devon Cantwell
@devon_cantwell



Turns out that learning [#R](#) is just being really good at Googling things.

[Traducir Tweet](#)

3:28 · 19/3/19 · [Twitter Web Client](#)

Quantitative Methods in the Humanities
and Social Sciences

Matthew L. Jockers

Text Analysis with R for Students of Literature

 Springer





Part-of-speech tagging with R

BY LÁSZLO SZAKÁCS · THURSDAY, 10 APRIL 2017 · PUBLISHED IN DATA ANALYSIS R



One of the major tasks in natural language processing (NLP) is the part-of-speech (POS) tagging of sentences, i.e. categorizing the words according to grammatical properties. Common parts of speech are noun, verb, article, adjective, preposition, pronoun, adverb, conjunction and interjections.

With the help of a recent R package **RDPOSTagger** now one can perform POS tagging within R on more than 40 languages, including English, Hungarian, French, German, Hindi, Italian, Thai, Vietnamese and many more. Below we present a brief introduction to this topic via a simple R script.

Revolutions

Daily news about using open source R for big data analysis, predictive modeling, data science, and visualization since 2008

Using R to Gain Insights into the Emotional Journeys in War and Peace | Main | Because it's Friday, Borderflyover

December 03, 2018

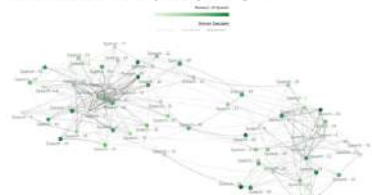
Stylometry: Identifying authors of texts using R

Few people expect politicians to write every word they utter themselves: reliance on speechwriters and spokespersons is a long-established political practice. Still, it is interesting to know which statements are truly the politician's own words, and which are driven primarily by advisors or influencers.

Recently, David Robinson established a way of figuring out which tweets from Donald Trump's Twitter account came from him personally, as opposed to from campaign staff, which he verified by comparing the sentiment of tweets from Android vs iPhone devices. Now, Ali Anslan Kazmi has used stylometric analysis to investigate the provenance of speeches by the Prime Minister of Pakistan.

By looking at the aspects of linguistic style (word/sentence length, frequency of word pairings, use of punctuation, etc.) of the speeches of the Prime Minister Nawaz Sharif, Ali found suggestions of at least 2 authors (and possibly more) behind the speeches. This is particularly apparent in this consensus network of appearances of 3-character sequences in speeches, which divides them into five clusters (of possibly differing authorship).

Consensus Network of the Prime Minister's Speeches using Character-grams



SEARCH ...

RECENT POSTS



Linux Monitoring Tool

Recently I found myself in a position where I...



Using a CDN to Speed Up Static Content Delivery

Introduction Modern websites and applications rely on...



Text Mining South Park

19 FEBRUARY 2018 on R, datasets, text mining, log likelihood

Peer Christensen

Home Posts Publications Projects Contact CV

Fair is foul, and foul is fair: a tidytext sentiment analysis of Shakespeare's tragedies

Jun 7, 2018 · 11 min read · 0 Comments · R, Text mining

Sentiment analysis can be used for many purposes and applied to all kinds of texts. In this exploratory analysis, we'll use a tidytext approach to examine the use of sentiment words in the tragedies written by William Shakespeare. I've previously used Python for scraping and mining texts. However, I recently stumbled upon the tidytext R package by Julia Silge and David Robinson as well as their excellent [book](#) and [resource](#) on combining tidytext with other tidy tools in R. This approach has made my life so much easier!

Loading packages

```
if(!require("pacman")) install.packages("pacman")
pacman::p_load(tidyverse, tidytext, gtsummary, gridExtra, zoo, data.table, gucenbergr)
```

Search

Back to Tutorials



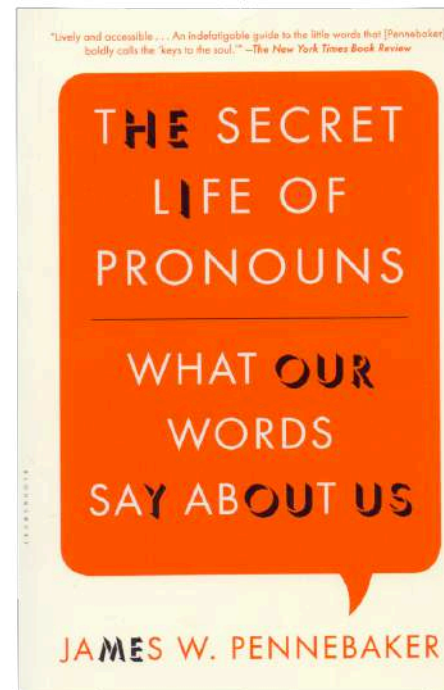
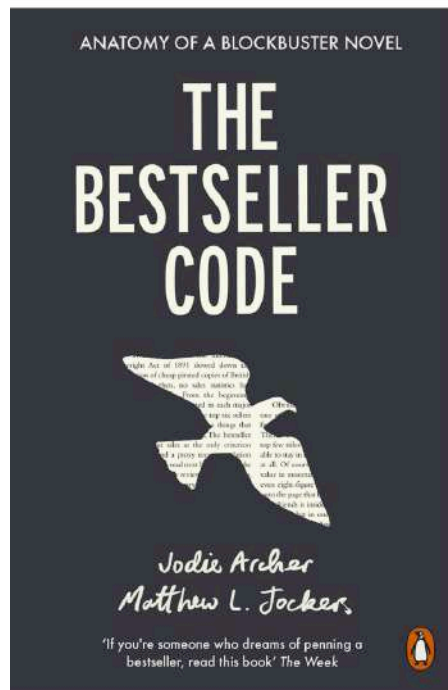
Debbie Liske
March 27th, 2018

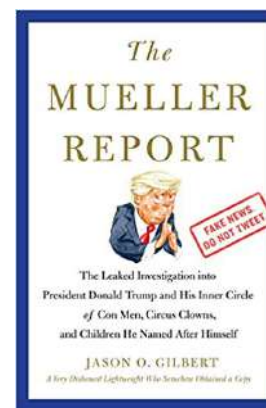
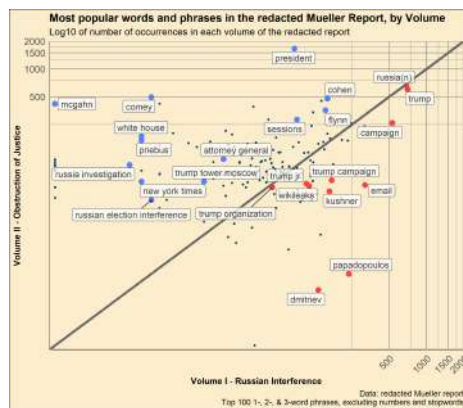
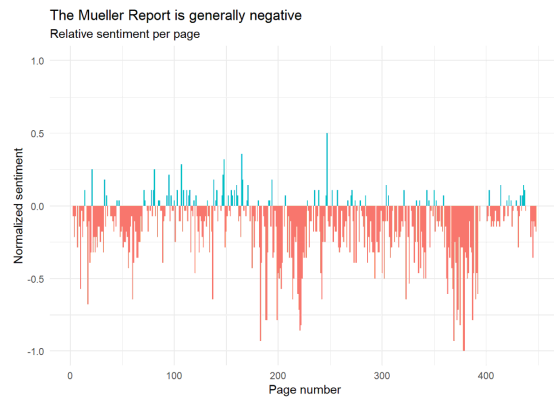
MUST READ · FUN · +3

Tidy Sentiment Analysis in R

Learn how to perform tidy sentiment analysis in R on Prince's songs, sentiment over time, song level sentiment, the impact of bigrams, and much more!

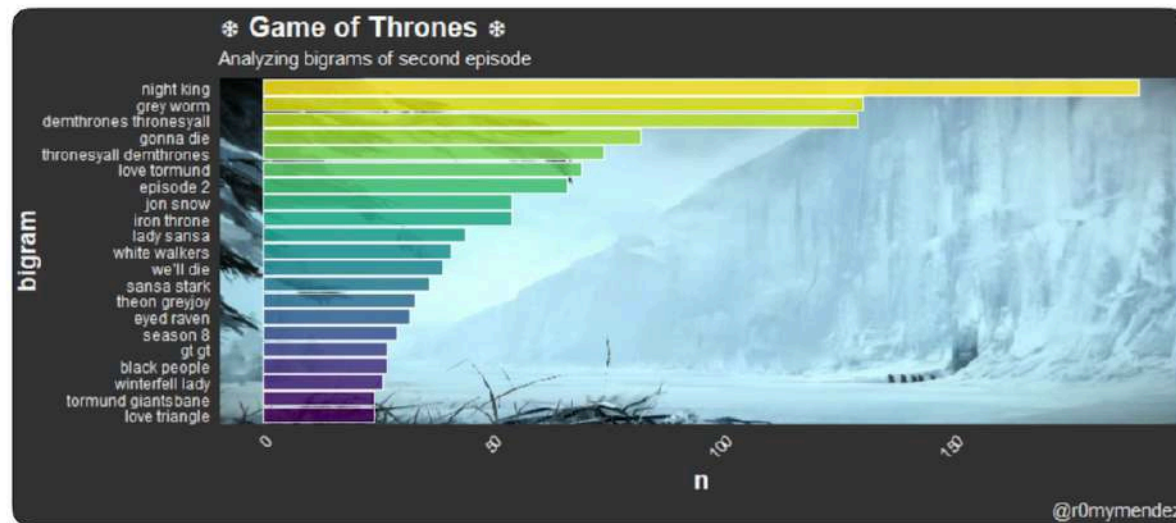
Take a Sentimental Journey through the life and times of Prince, The Artist, in part Two-A of a three part tutorial series using sentiment analysis with R to shed insight on The Artist's career and societal influence. The three tutorials cover the following:





📺 Game of thrones episode II 🤖 and this night... I analyzed the most used terms pairs in tweets with #GameofThrones 🌲 hashtag 📄
#rstats #r #hbo #tidyverse #wordcloud2 #tidytext #rtweet #data #tv
#rstudio #text #textmining #mining #got 🌲 #Episode #season8
#finaleseason #FinalHasCome

Traducir Tweet



4:38 · 22/4/19 · Twitter Web Client

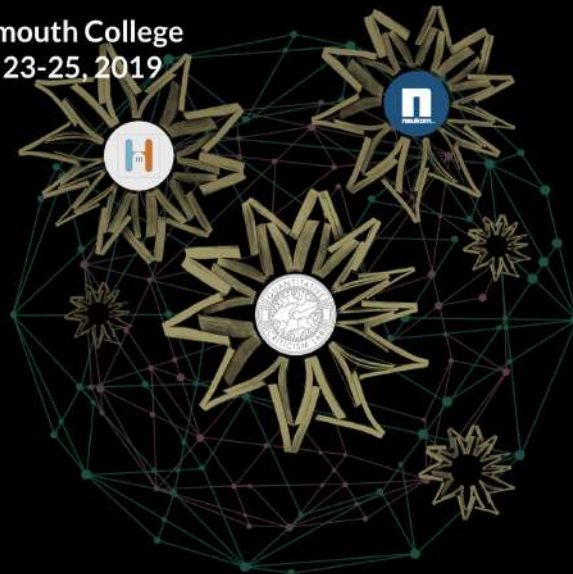


Digital Humanities Beyond Modern English

Computational Analysis of Premodern and Non-Western Literature

Dartmouth College

April 23-25, 2019



ANÁLISIS DE TEXTOS Y ESTILOMETRÍA CON R

INTRODUCCIÓN A R PARA FILÓLOGOS

JOSÉ MANUEL FRADEJAS RUEDA (Universidad de Valladolid)

Introducción a R.
Instalación y primeros pasos

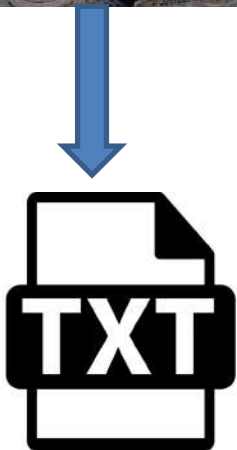


Versión 2019 rev. 01



1. Introducción a R. Instalación y primeros pasos
2. Preparar los datos
3. Cosechar textos de la red. Primeros pasos en la programación
4. Primeros Pasos en el análisis textual con R
5. Primeros Pasos en el análisis textual con R (2)
6. Profundizando en el análisis textual con R
7. De las palabras a las clases de palabras. Introducción al etiquetado morfológico automático
8. Colocación, coocurrencia y redes léxicas
9. *Topic Modelling*
10. Iniciación a la estilometría. El paquete `stylo`
11. Análisis de sentimientos. La forma de las historias





Ministerio del Tiempo



Ministerio del Tiempo

1 EXT/ AMANECER CAMPOS DE FLANDES [1569]

Rótulo: Flandes, 1569.

Vemos un soldado de los tercios yacer en tierra... los colores de su uniforme, su cara, sus manos tienen ahora el color del barro. Ampliamos el campo de visión y vemos que no yace solo: cientos de hombres descansan muertos junto a él.

POR CORTE.

2 INT/DÍA. SALA OFICIALES [1569]

ALONSO llega escoltado a la sala (NO SE LE VE: SÓLO SE VEN LAS BOTAS HANCHADAS DE BARRO). Allí, de espaldas, le espera el CAPITÁN, que se gira al oírles llegar.

CAPITÁN

¡Atacasteis antes de tiempo! ¿Por qué lo hicisteis?

Ahora, vemos a ALONSO (treinta y pico, con ropas de soldado de los Tercios), serio y demacrado.

ALONSO

Porque vos lo ordenasteis.

CAPITÁN

¿Tenéis testigos de lo que decís?

ALONSO

Si los muertos pudieran hablar, los tendría.

CAPITÁN

¡Mentís!

El capitán coge una vara y va a atizarle. Alonso agarra firme la vara según le viene, antes de que le cruce la cara y, mirando amenazante al capitán.

ALONSO

Yo nunca miento.

El capitán le mira como si estuviera loco.

CAPITÁN

¡Prenedle!

Pero antes de que los soldados puedan hacerlo, Alonso estalla. Y se lanza por el capitán.

ALONSO

¡¡Todos han muerto por vos!!!
¡¡¡Todos!!!

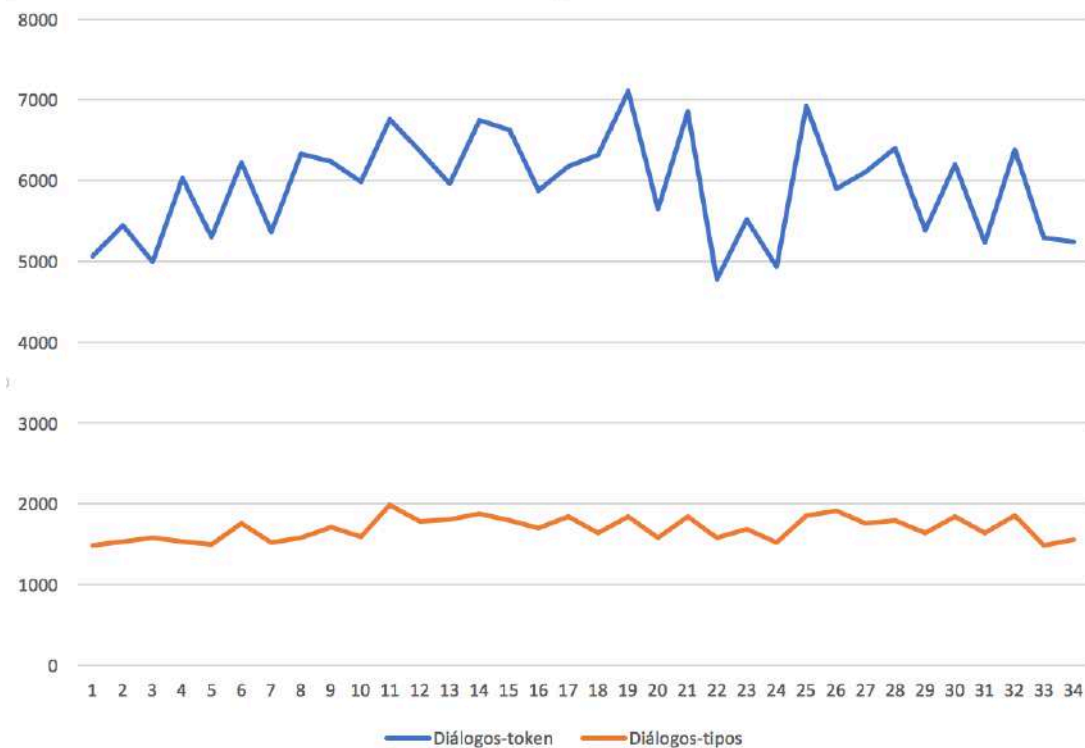


```
07-PalabrasClasePalabras-2-Ulha...  NMT.txt
1 temporada episodio titulo personaje texto
2 1 1 EL tiempo es el que es CAPITÁN "¡Atacasteis antes de tiempo! ¿Por qué lo hicisteis?"
3 1 1 EL tiempo es el que es ALONSO "Porque vos lo ordenasteis."
4 1 1 EL tiempo es el que es CAPITÁN "¿Tenéis testigos de lo que decís?"
5 1 1 EL tiempo es el que es ALONSO "Si los muertos pudieran hablar, los tendría. ¡Mentís!"
6 1 1 EL tiempo es el que es ALONSO "Yo nunca miento."
7 1 1 EL tiempo es el que es CAPITÁN "¡Prenedle!"
8 1 1 EL tiempo es el que es ALONSO "¡¡Todos han muerto por vos!!! ¡¡¡Todos!!!"
9 1 1 EL tiempo es el que es BLANCA "¿Por qué le atacaste? ¡Era tu superior!"
10 1 1 EL tiempo es el que es ALONSO "¡Como si es San Pedro! A veces, un hombre ha de hacer lo que ha de hacer."
11 1 1 EL tiempo es el que es BLANCA "Maldito orgullo... Por él, mañana, tu capitán estará en su cama, y tú en la horca."
12 1 1 EL tiempo es el que es ALONSO "Pues en la horca estará un hombre con honor y en la cama un mamarracho. Blanca, no llores... he tenido una buena vida... he visto mundo... he amado... he luchado por mi patria... No tengo quejas. Quiero que hagas una cosa por mí. Sigue tu vida, no mires atrás."
13 1 1 EL tiempo es el que es BLANCA "Alonso... Yo... Pero..."
14 1 1 EL tiempo es el que es CARCELERO "¡Se acabó el tiempo! ¡Fuera!"
15 1 1 EL tiempo es el que es ALONSO "¡Oídme, he de os hablar!"
```


Ministerio del Tiempo

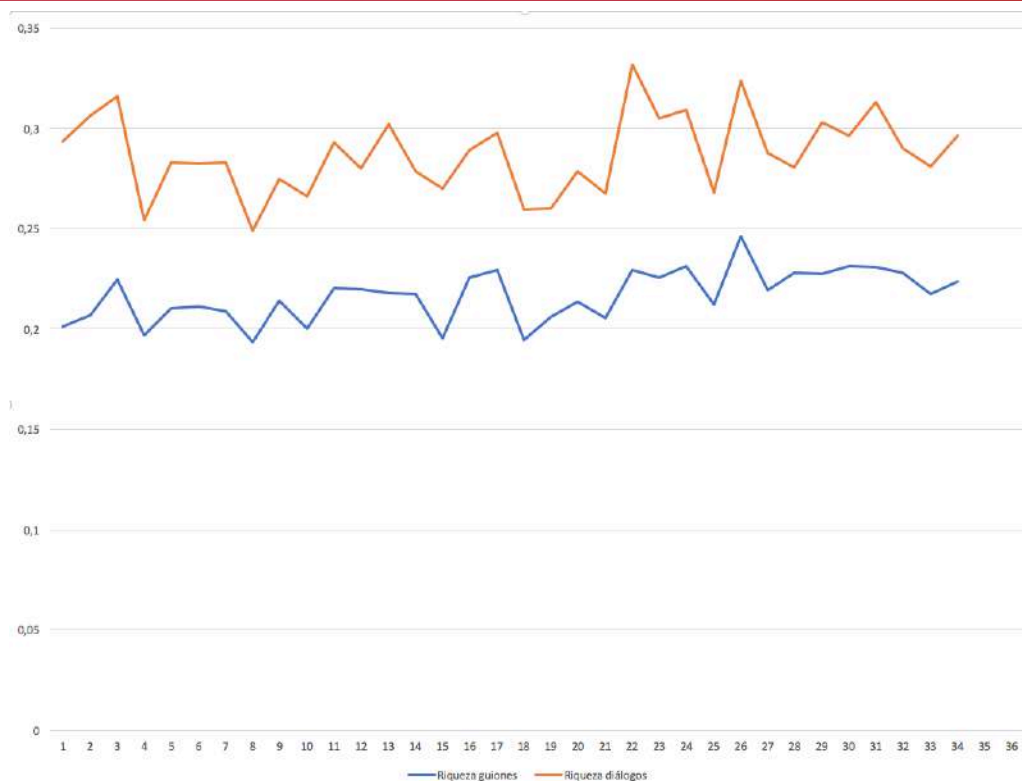


Algunas posibilidades...



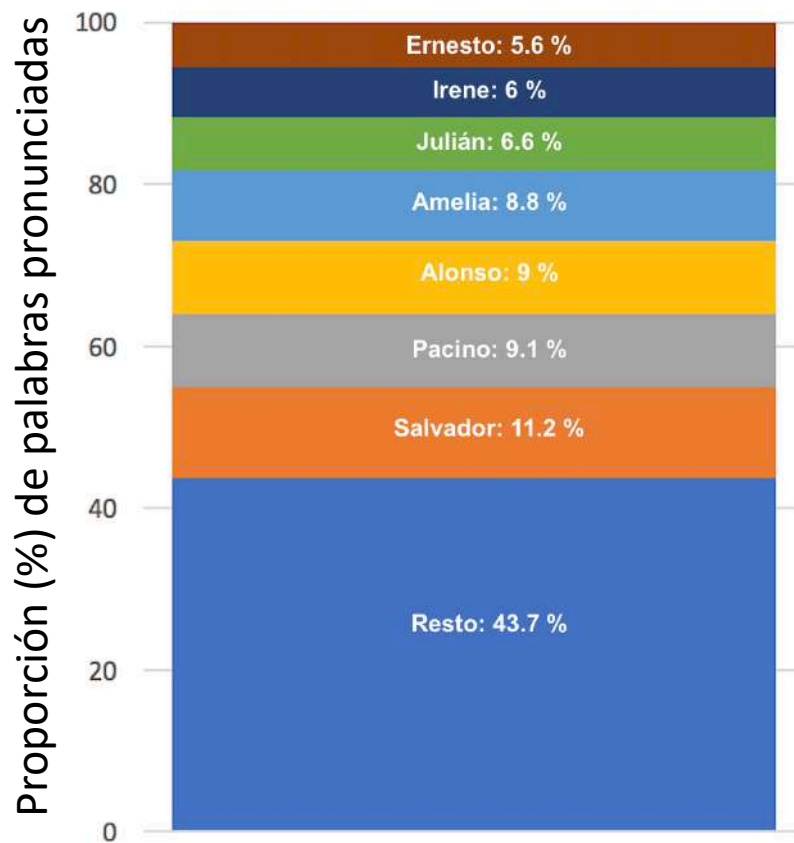
Recuento léxico de los diálogos de *MdT*

Algunas posibilidades...



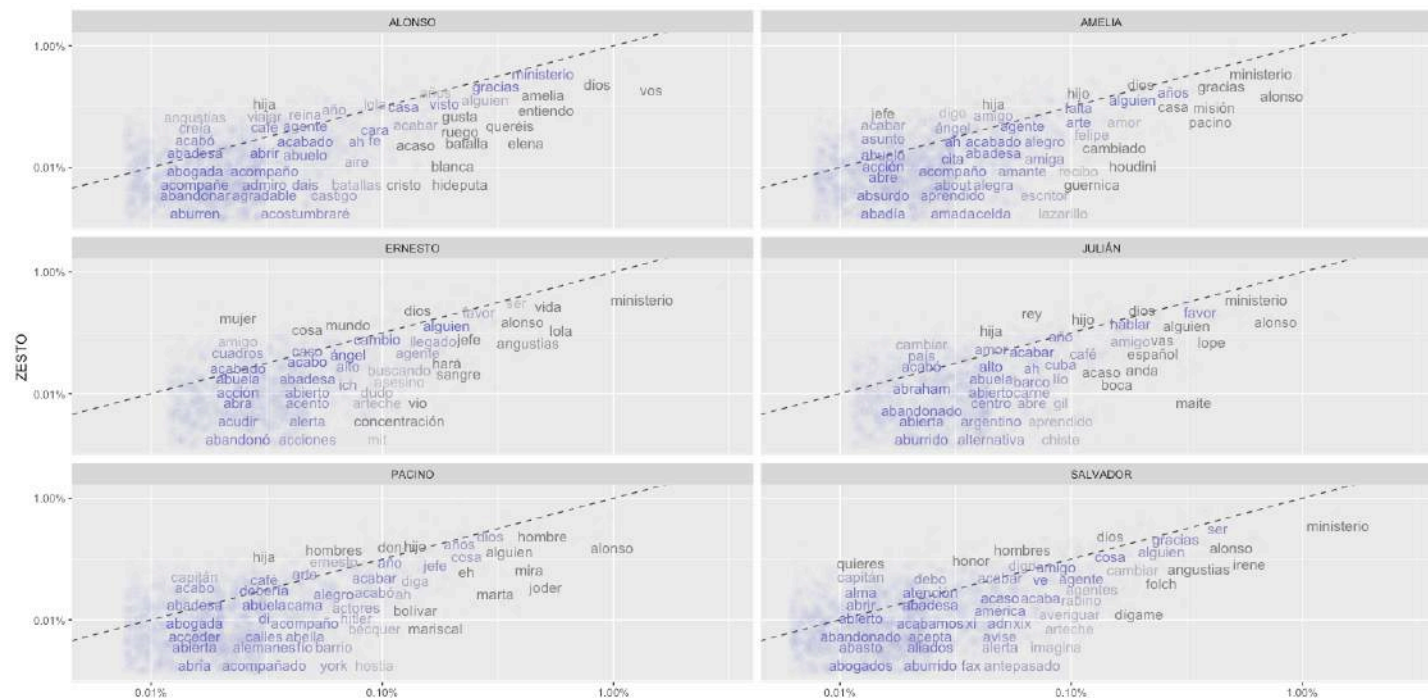
Riqueza léxica de los diálogos de *MdT*

Algunas posibilidades...

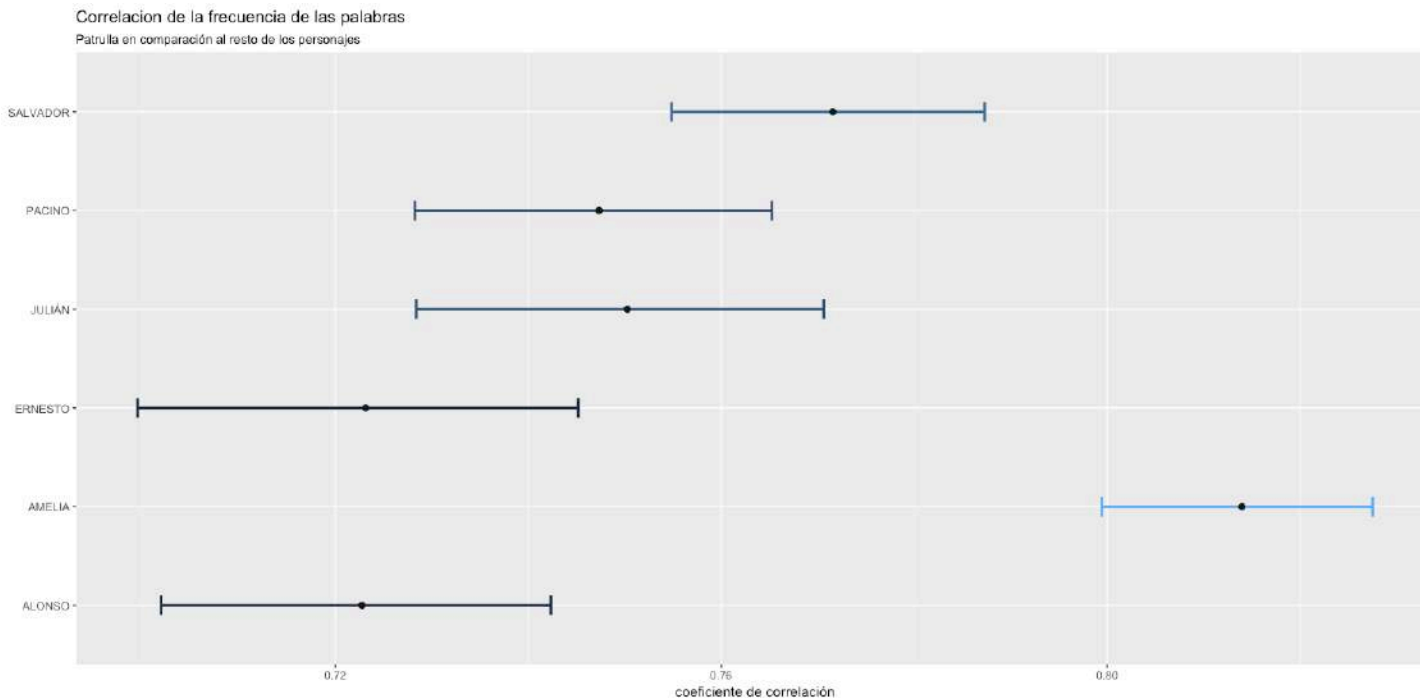


Salvador es el que más habla

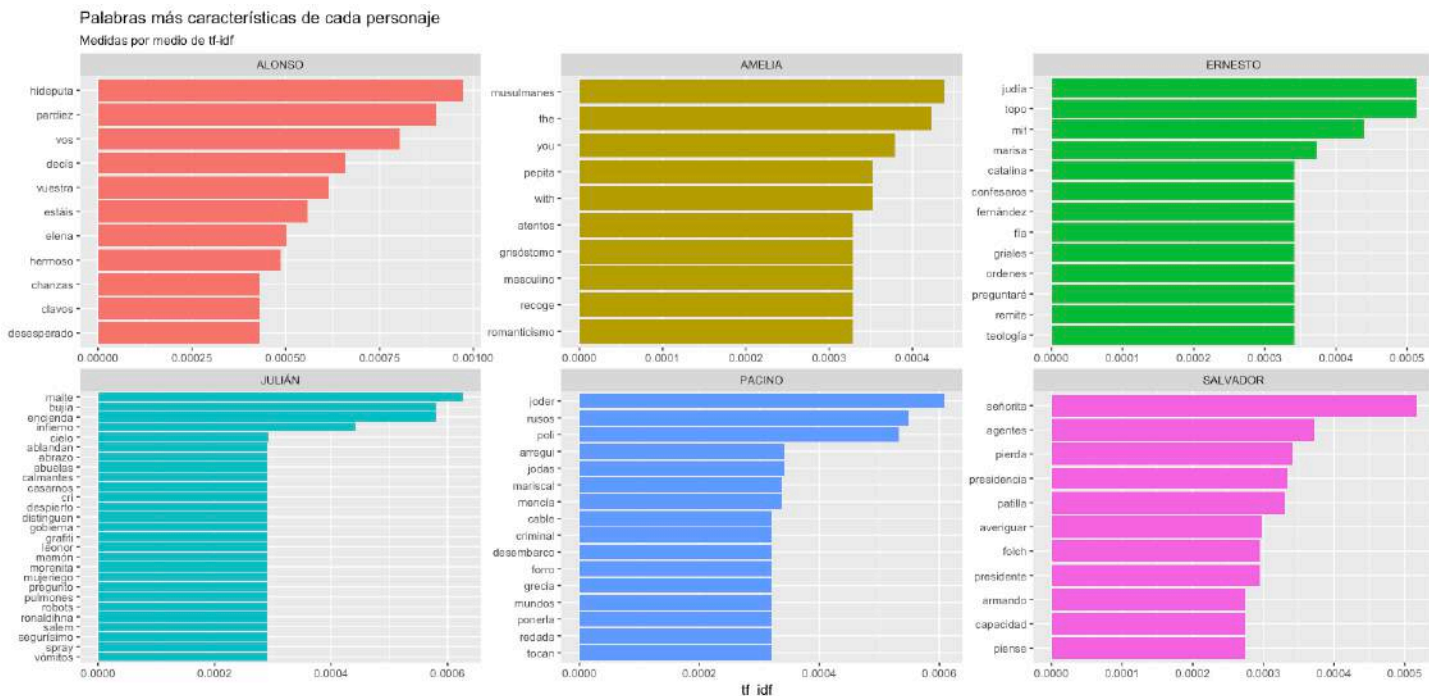
Algunas posibilidades...



Algunas posibilidades...



Algunas posibilidades...



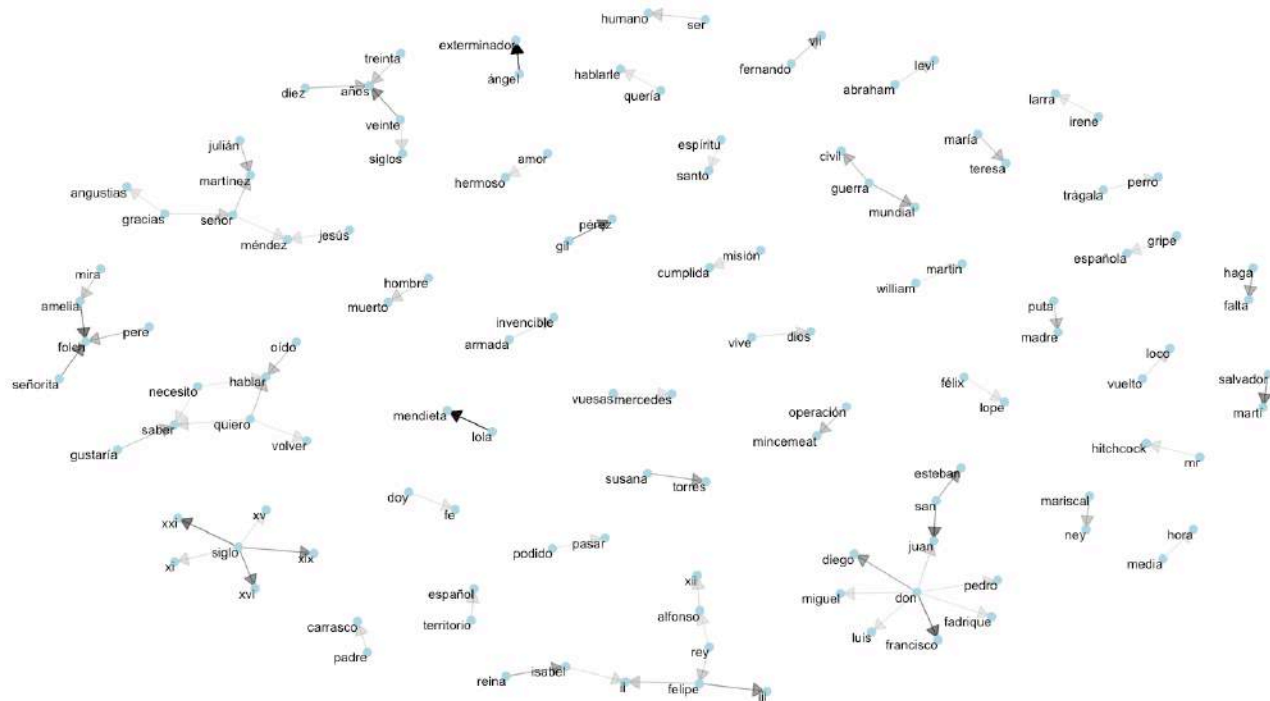
Palabras más características de cada personaje

Algunas posibilidades...

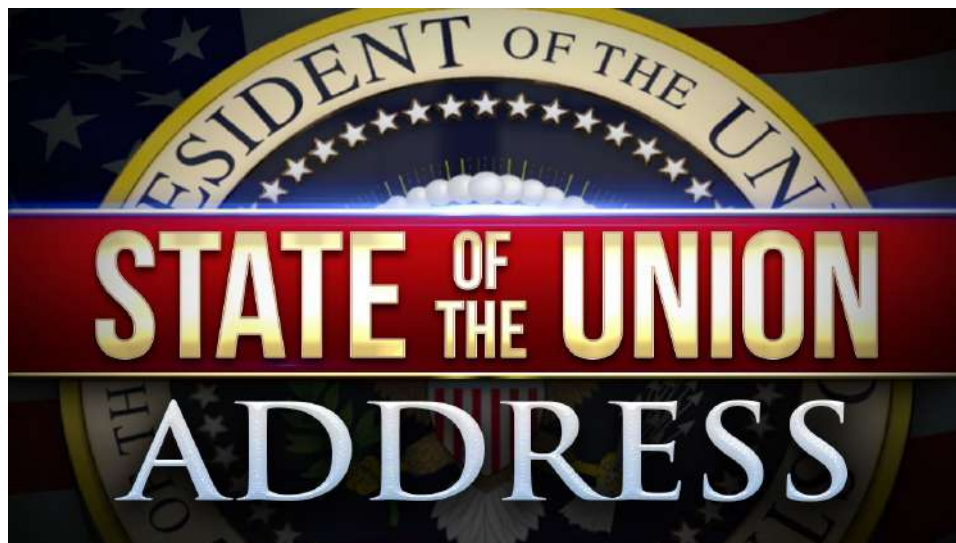
episodio	titulo	palabra	n	total	tf_idf
1	El tiempo es el que es	1808	6	5016	0.0034
2	Tiempo de gloria	esteban	13	5453	0.0068
3	Cómo se reescribe el tiempo	montserrat	12	4979	0.0085
4	Una negociación a tiempo	torquemada	17	6051	0.0080
5	Cualquier tiempo pasado	recibo	17	5310	0.0078
6	Tiempo de pícaros	lázaro	18	6227	0.0082
7	Tiempo de venganza	leiva	35	5381	0.0125
8	La leyenda del tiempo	dalí	24	6331	0.0092
9	Tiempo de leyenda	cid	34	6232	0.0105
10	El tiempo en sus manos	armario	19	6005	0.0068
11	Tiempo de hidalgos	manuscrito	9	6756	0.0047
12	El monasterio del tiempo	abadesa	19	6367	0.0105
13	Un virus de otro tiempo	virus	17	5974	0.0069
14	Tiempo de magia	argamasilla	39	6746	0.0164
15	Tiempo de valientes I	teniente	13	6629	0.0056
16	Tiempo de valientes II	teniente	11	5862	0.0053
17	Óleo sobre tiempo	cuadros	20	6163	0.0040
18	Separadas por el tiempo	vampira	22	6325	0.0123
19	Tiempo de lo oculto	lombardi	32	7105	0.0159
20	Hasta que el tiempo nos separe	constanza	23	5631	0.0116
21	Cambio de tiempo	argüelles	13	6880	0.0067
22	Con el tiempo en los talones	hitchcock	36	4758	0.0162
23	Tiempo de espías	mincemeat	14	5505	0.0090
24	Tiempo de hechizos	mencia	21	4945	0.0150
25	Tiempo de ilustrados	inclusa	19	6872	0.0097
26	Tiempo de esplendor	lerma	23	5910	0.0110
27	Tiempo de esclavos	marqués	21	6103	0.0121
28	Tiempo de censura	viridiana	21	6425	0.0115
29	Tiempo de conquista	aguilar	15	5403	0.0098
30	El cisma del tiempo	papa	24	5248	0.0079
31	Refugiados en el tiempo	moriscos	18	6505	0.0078
32	Tiempo de verbena	verbena	20	6376	0.0089
33	Contratiempos	suárez	16	5313	0.0085
34	Entre dos tiempos	ureña	29	5264	0.0194

Palabra más característica
de cada episodio

Algunas posibilidades...



Algunas posibilidades...



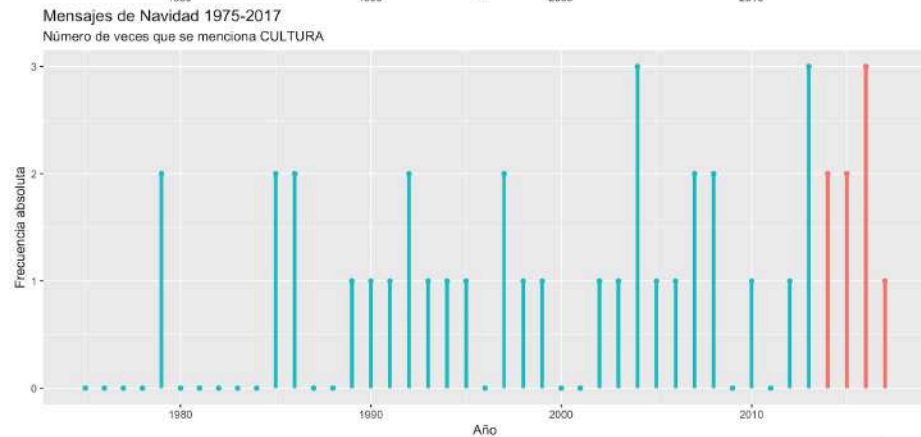
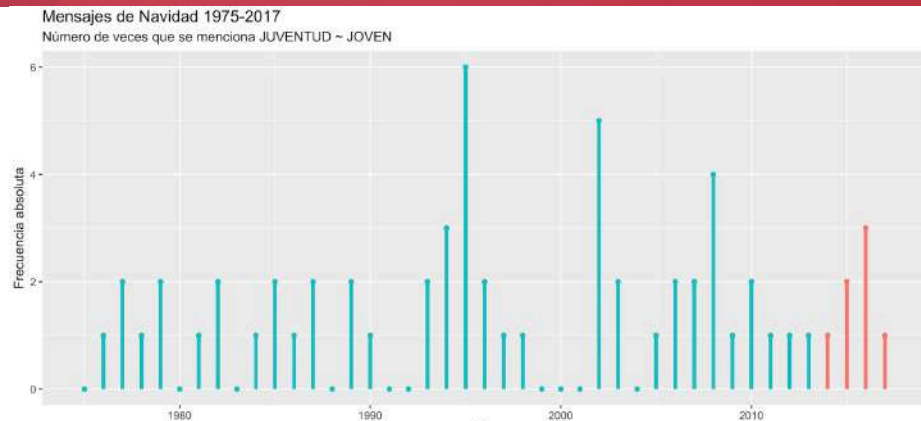
1790-2019

Discurso de Nochebuena



LDA beta (importancia de cada palabra en el tópico)

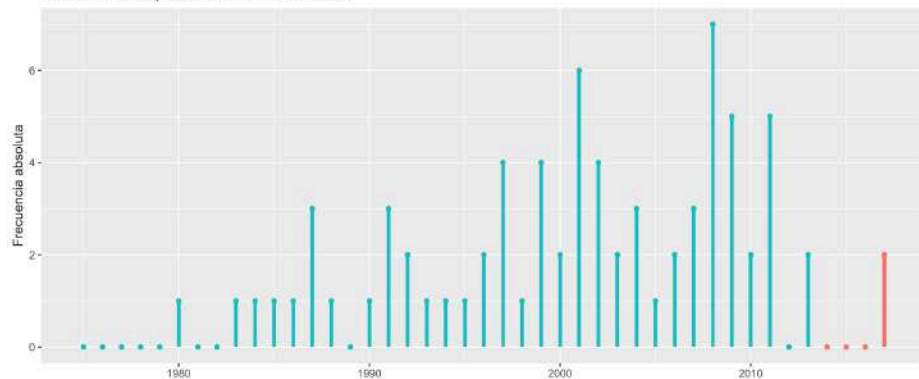
Discurso de Nochebuena



Discurso de Nochebuena

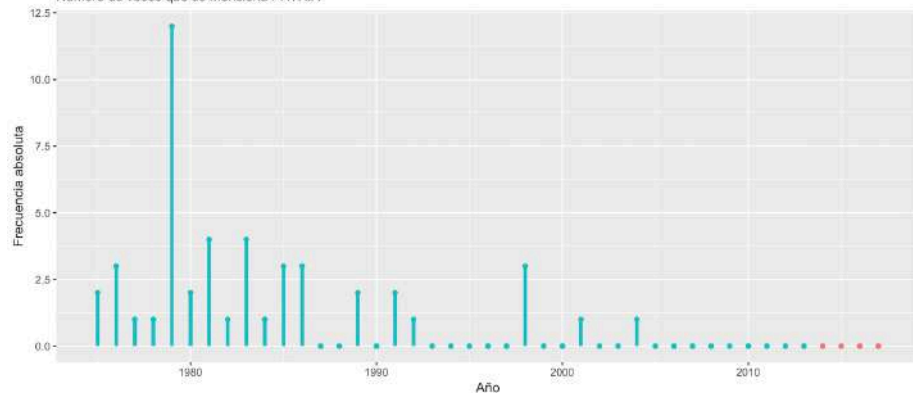
Mensajes de Navidad 1975-2017

Número de veces que se menciona TERRORISMO



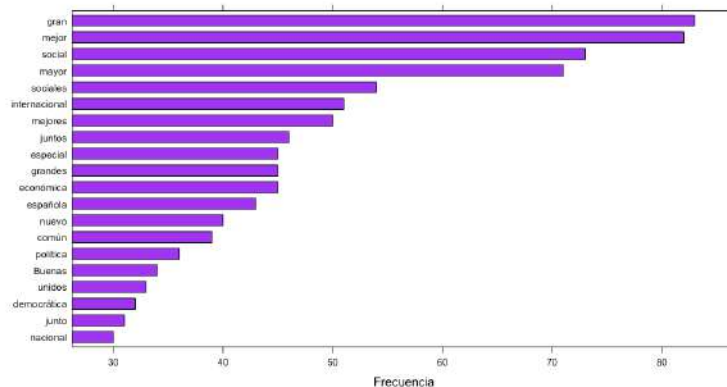
Mensajes de Navidad 1975-2017

Número de veces que se menciona PATRIA

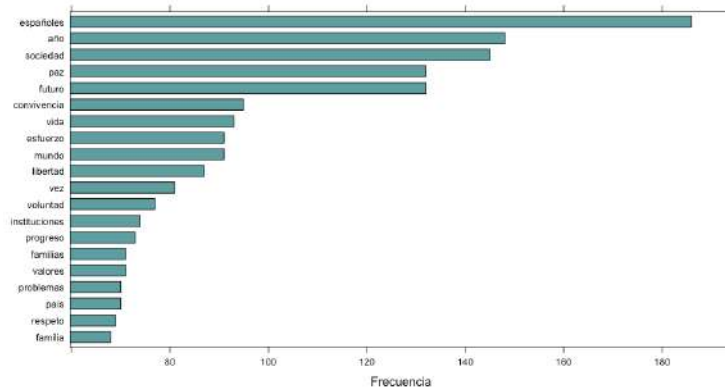


Discurso de Nochebuena

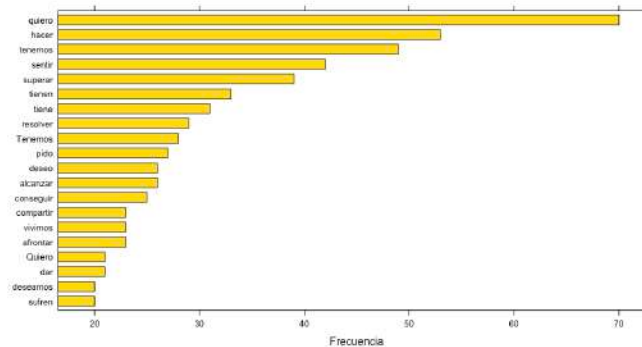
Adjetivos más frecuentes



Sustantivos más comunes

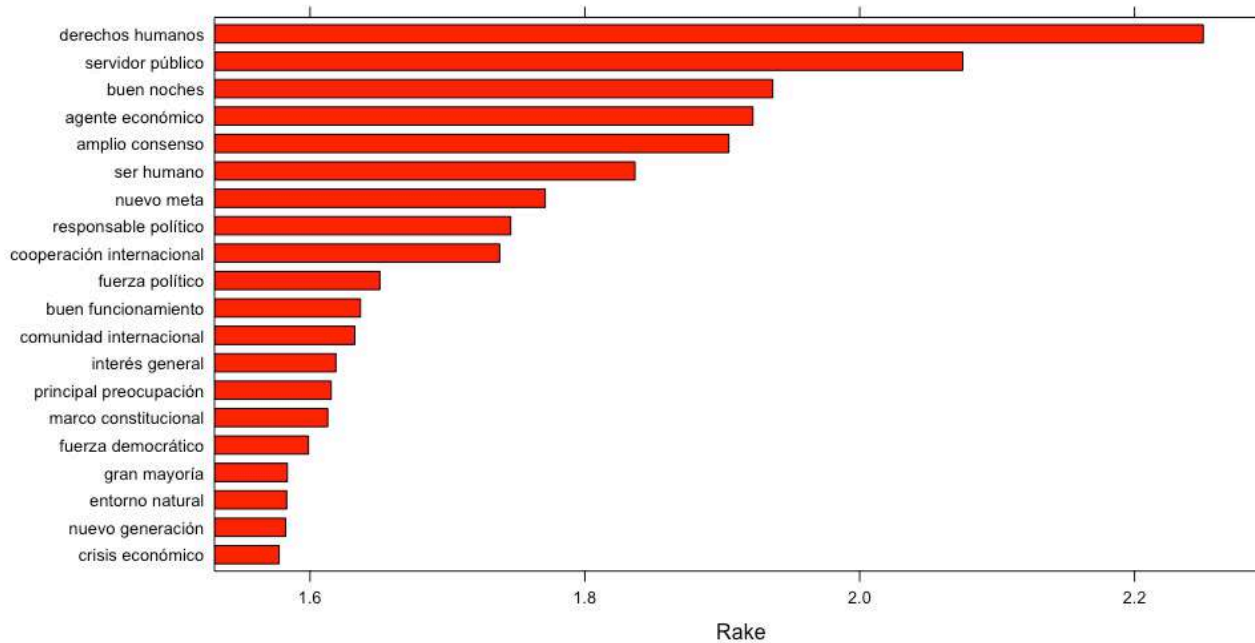


Verbos más frecuentes



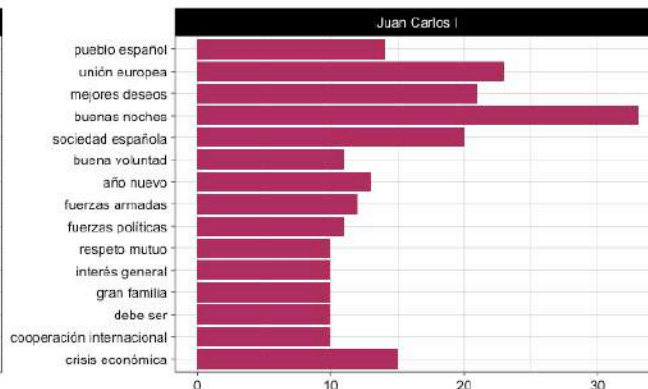
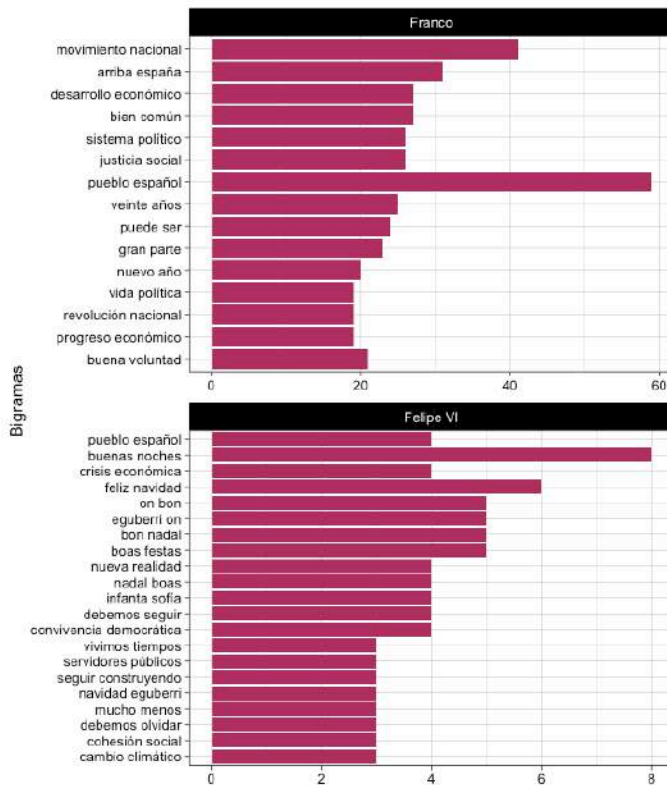
Discurso de Nochebuena

Secuencia de SUST + ADJ más frecuentes
Identificados por RAKE

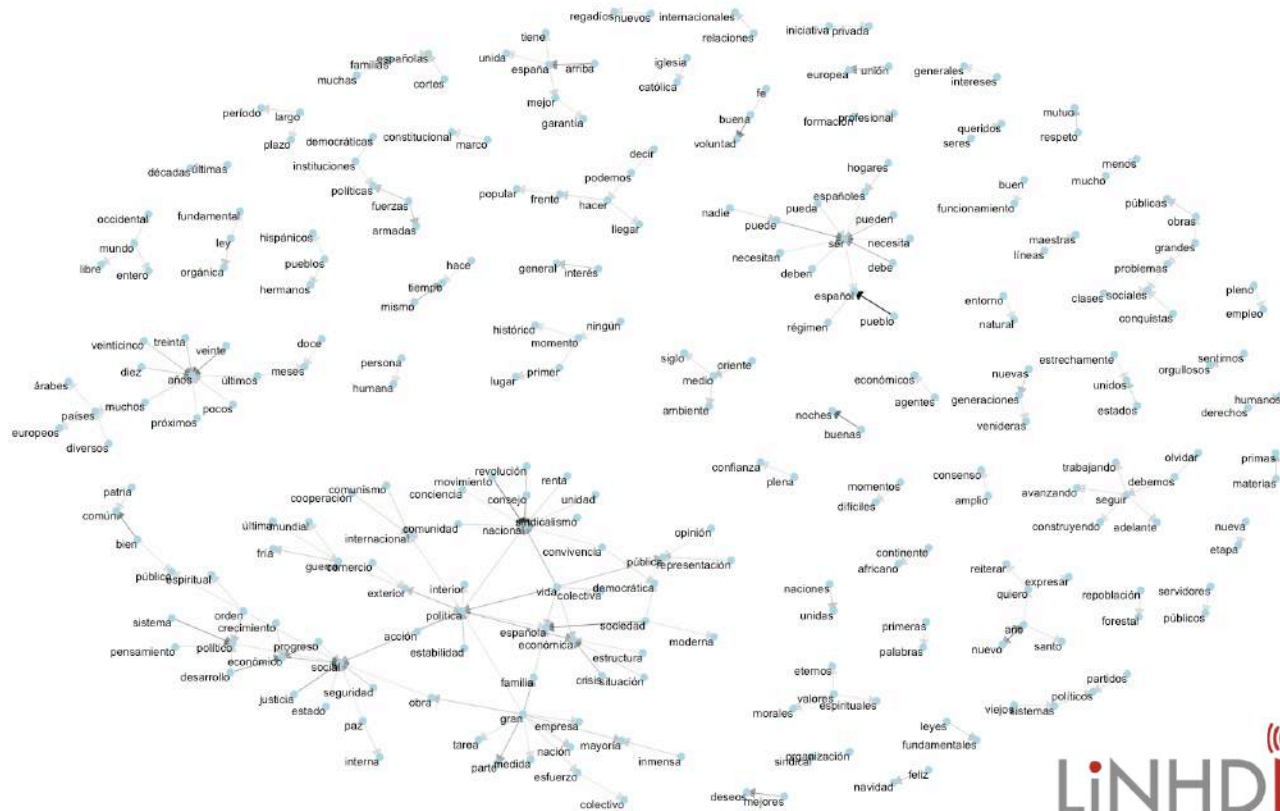


Discurso de Nochebuena

Bigramas más frecuentes de cada J.E.
1937-2018



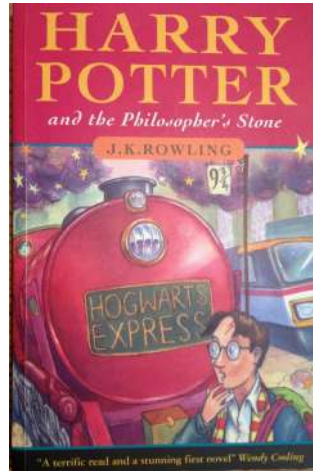
Discurso de Nochebuena





2017

20 ANIVERSARIO

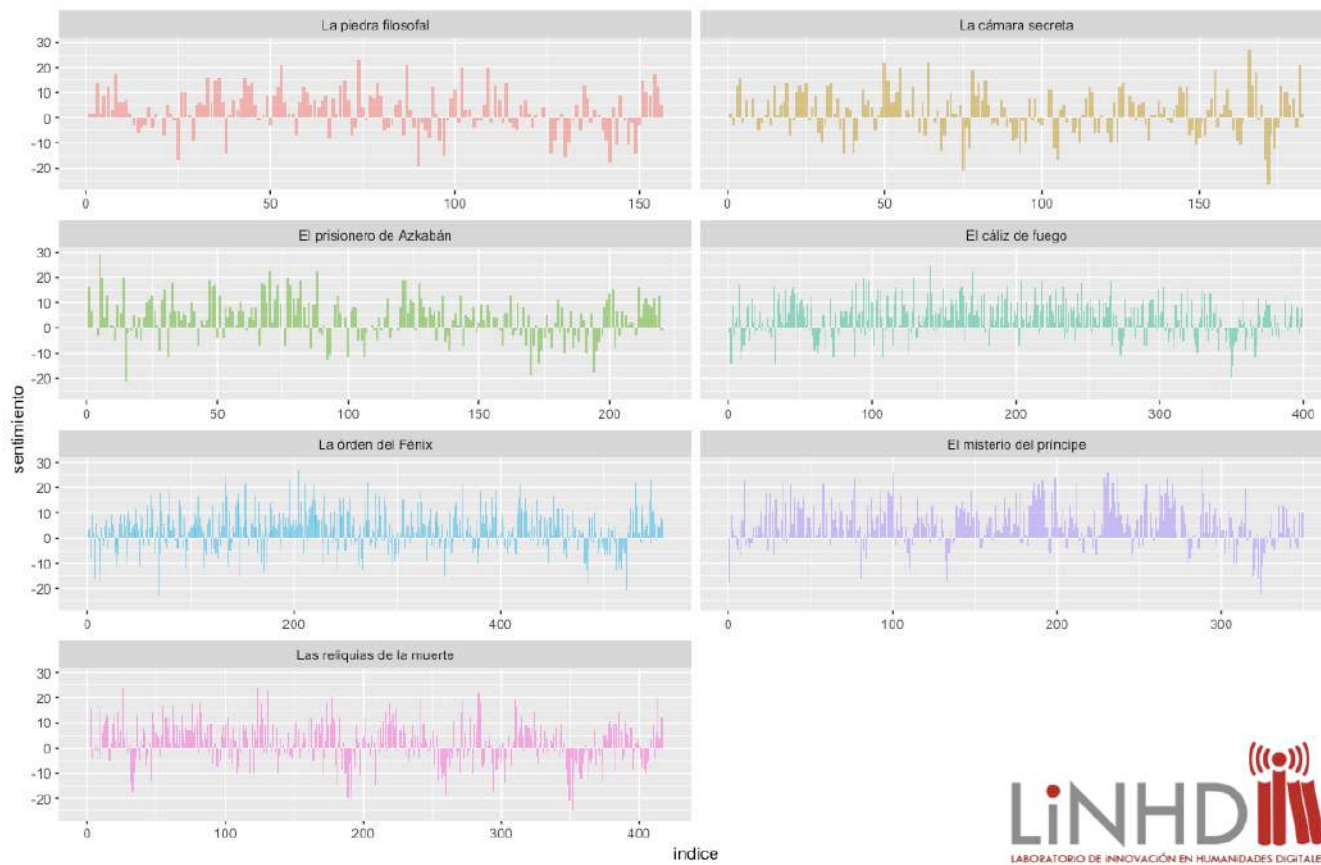


Harry Potter

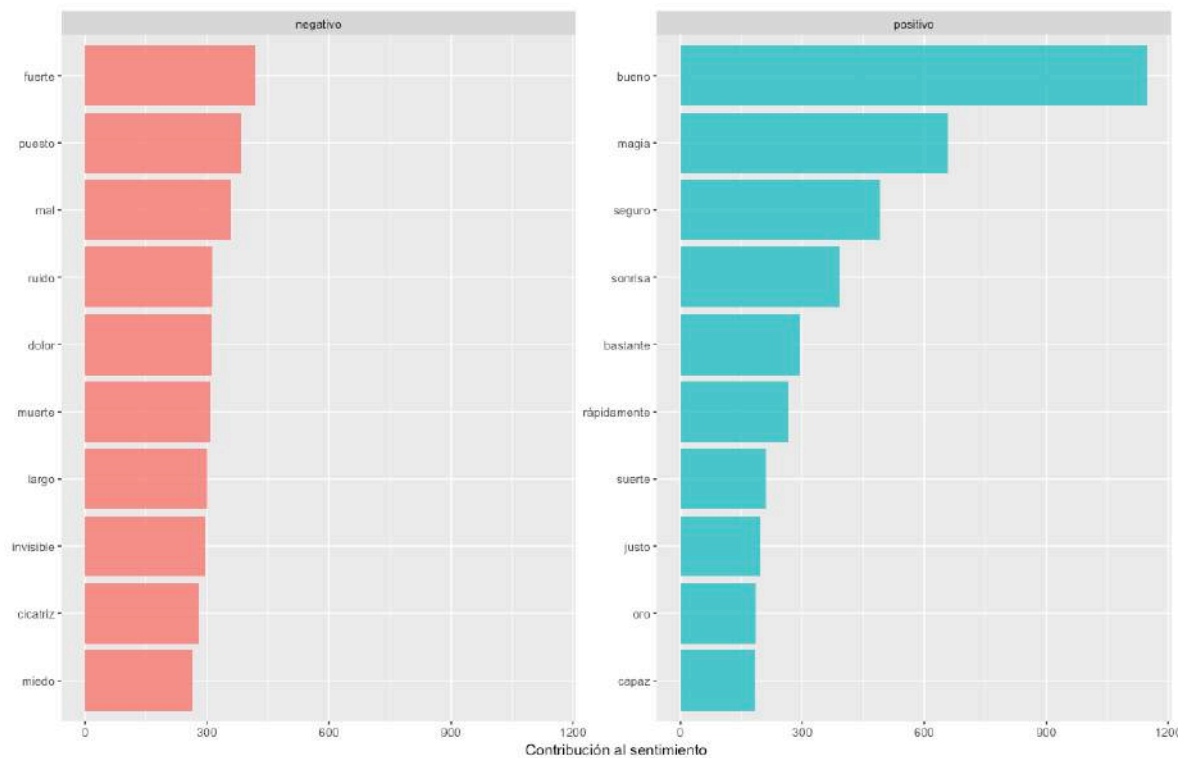
sentimiento	n
1 positivo	45387
2 negativo	37679
3 confianza	30489
4 premonición	21889
5 miedo	21596
6 tristeza	19203
7 ira	15569
8 alegría	15508
9 disgusto	15204
10 asombro	13073

```
# A tibble: 10 x 2
  sentiment      n
  <chr> <int>
1 negative 56579
2 positive 38324
3 sadness  35866
4 anger   32750
5 trust   23485
6 fear    21544
7 anticipation 21123
8 joy     14298
9 disgust 13381
10 surprise 12991
```

Harry Potter



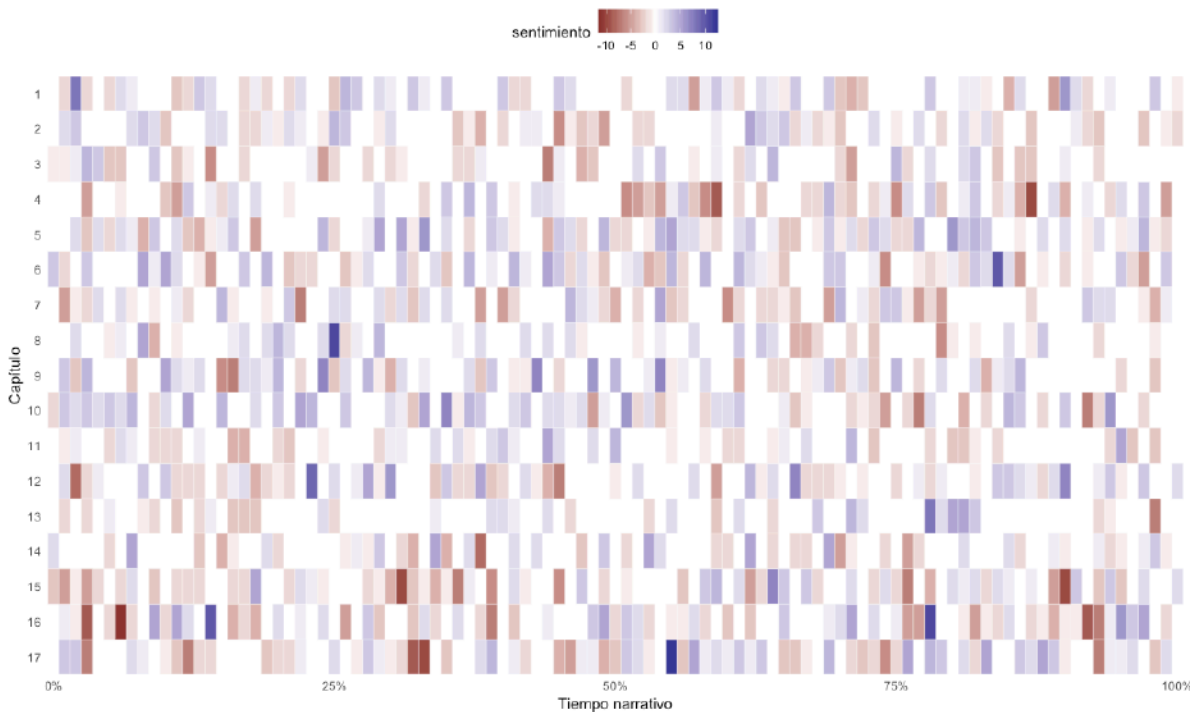
Harry Potter



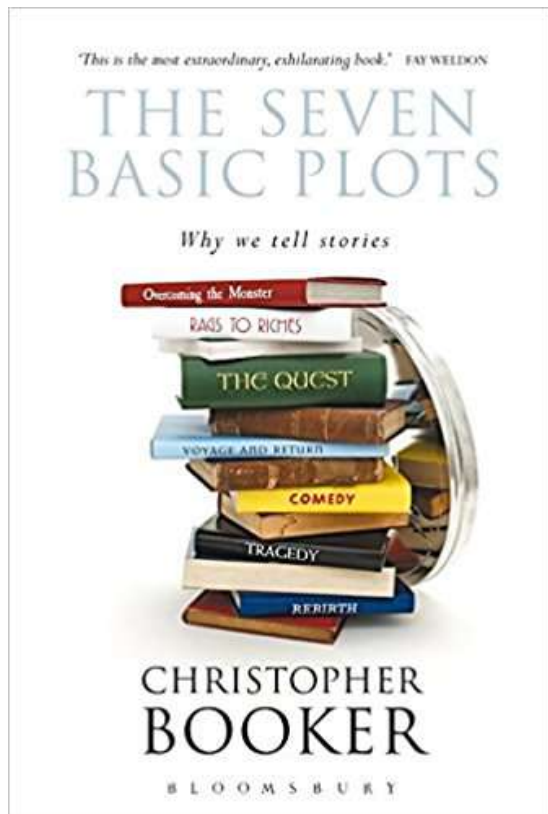
Harry Potter

Sentimiento en "La piedra filosofal"

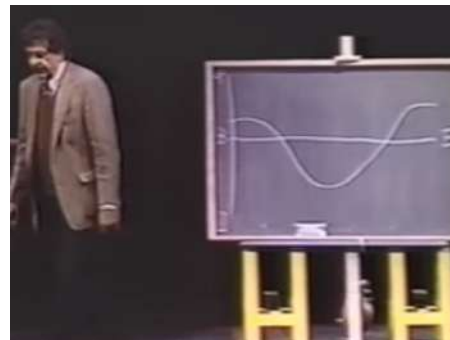
Resumen de la puntuación neta del sentimiento según progresa cada capítulo



La forma de las historias



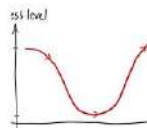
Kurt Vonnegut On the Shapes of Stories



<https://www.youtube.com/watch?v=oP3c1h8v2ZQ>

The Shapes of Stories

November 7, 2016 by Andy Reagan



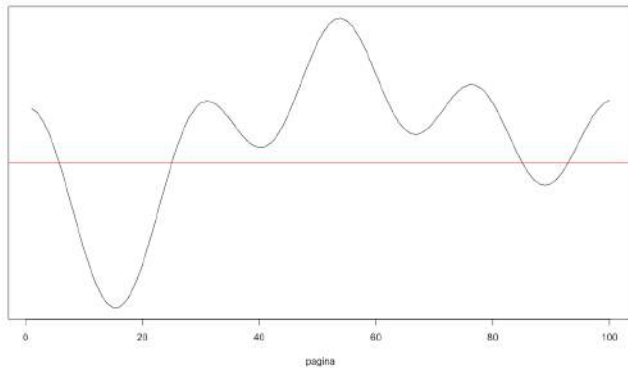
Stories help us encode and understand our collective existence, underpin cultures, and help frame the possible. Describing the ecology of all human stories is an essential scientific enterprise. With the advent of the internet and massive digitization this vital work has become, in part, a data-driven one. There are many aspects of stories to characterize and here we take ... [\[Read more...\]](#)

Dan Brown

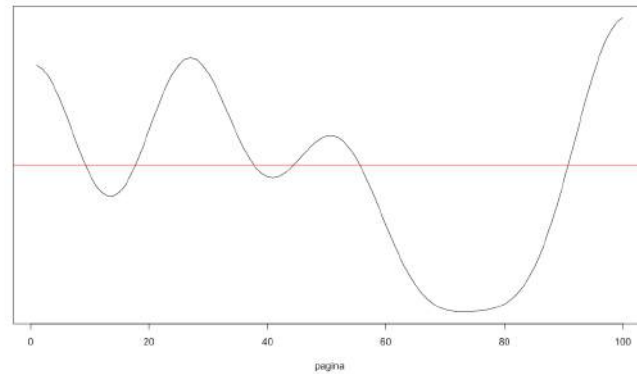


Dan Brown

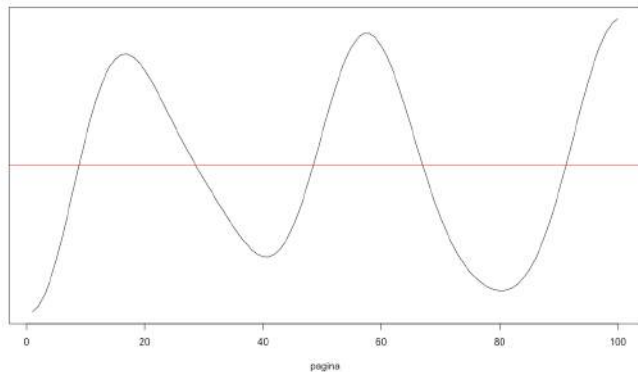
La forma de la historia de
El código Da Vinci



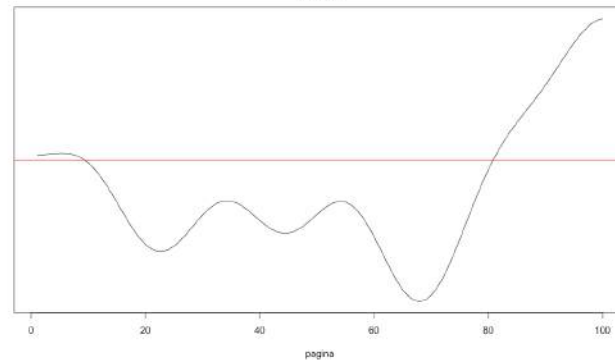
La forma de la historia de
Ángeles y demonios



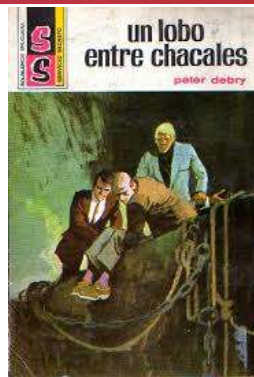
La forma de la historia de
Inferno



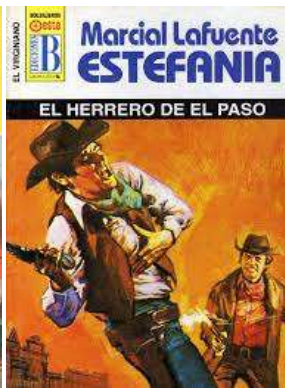
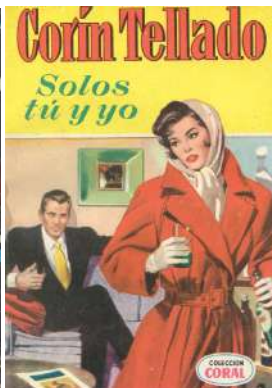
La forma de la historia de
Origen



Trabajo futuro...

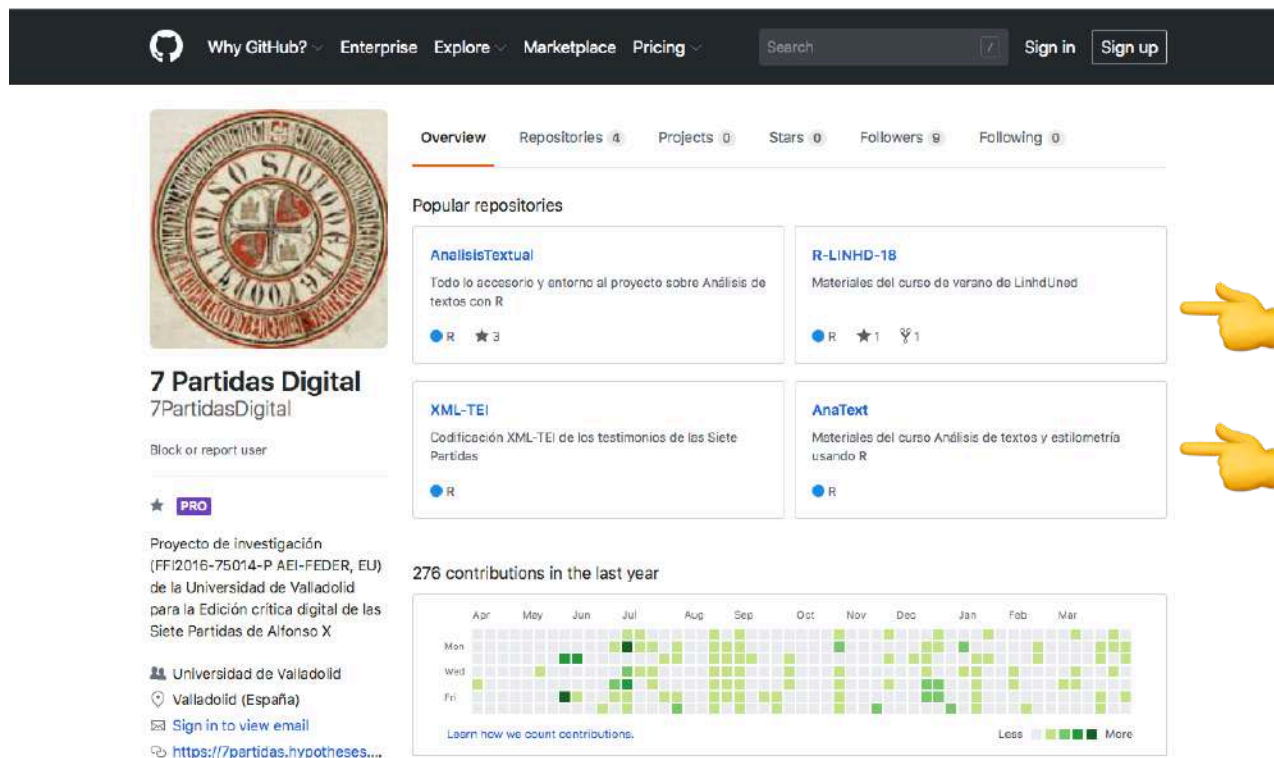


NOVELA



POPULAR

Cómo hacer estas cosas...



The screenshot shows the GitHub profile of '7 Partidas Digital'. The header includes navigation links like 'Why GitHub?', 'Enterprise', 'Explore', 'Marketplace', and 'Pricing', along with a search bar and 'Sign in'/'Sign up' buttons. The profile header features a circular logo with a cross and the text '7 PARTIDAS DIGITAL'. Below the logo, the repository name '7 Partidas Digital' is displayed, followed by a 'Block or report user' link and a 'PRO' badge. The bio states: 'Proyecto de investigación (FFI2016-75014-P AEI-FEDER, EU) de la Universidad de Valladolid para la Edición crítica digital de las Siete Partidas de Alfonso X'. The location is 'Universidad de Valladolid, Valladolid (España)', and there is a link to 'https://7partidas.hypotheses...'. The 'Overview' tab is selected, showing statistics: 4 Repositories, 0 Projects, 0 Stars, 9 Followers, and 0 Following. The 'Popular repositories' section lists four repositories: 'AnálisisTextual' (3 stars), 'R-LINHd-18' (1 star), 'XML-TEI' (0 stars), and 'AnaText' (0 stars). A '276 contributions in the last year' section shows a calendar heatmap with green squares indicating contributions. Two yellow hand emojis point to the 'R-LINHd-18' and 'AnaText' repositories.

Why GitHub? Enterprise Explore Marketplace Pricing Search Sign in Sign up

7 Partidas Digital
7PartidasDigital

Block or report user

★ PRO

Proyecto de investigación (FFI2016-75014-P AEI-FEDER, EU) de la Universidad de Valladolid para la Edición crítica digital de las Siete Partidas de Alfonso X

Universidad de Valladolid
Valladolid (España)
Sign in to view email
<https://7partidas.hypotheses...>

Overview Repositories 4 Projects 0 Stars 0 Followers 9 Following 0

Popular repositories

- AnálisisTextual**
Todo lo accesorio y entorno al proyecto sobre Análisis de textos con R
R ★ 3
- R-LINHd-18**
Materiales del curso de verano de LinhdUned
R ★ 1
- XML-TEI**
Codificación XML-TEI de los testimonios de las Siete Partidas
R
- AnaText**
Materiales del curso Análisis de textos y estilometría usando R
R

276 contributions in the last year

Apr May Jun Jul Aug Sep Oct Nov Dec Jan Feb Mar

Mon
Wed
Fri

Learn how we count contributions. Less More

github.com/7partidasdigital

Cómo hacer estas cosas...



Cuentapalabras

Estilometría, análisis de textos en R
para filólogos

Sobre el título

Guion

Cuentapalabras

José Manuel Fradejas Rueda

16 de mayo de 2019

UVa



ESTA PÁGINA Y TODO ESTE SITIO ESTÁN EN DESARROLLO

Para informar de cualquier error o sugerencia escribe a

7PartidasDigital[at]gmail[punto]com

Cuentapalabras

Estilometría, análisis de textos en R para filólogos

La *estilometría* es el análisis estadístico de textos y trata de identificar las semejanzas y diferencias que existen entre ellos para agruparlos de acuerdo con sus características lingüísticas con el objetivo de detectar señales estilísticas que puedan servir para establecer su autoría, su ubicación genérica, sus orígenes, estilo.... El término lo acuñó el filósofo polaco Wincenty Lutosławski (1863–1954) en su artículo *Principes de stylométrie* y los primeros análisis, sin conexión con Lutosławski, son los de T. C. Mendenhall (1841–1924) y su artículo *«A Mechanical solution of a literary problem»*.

El análisis de textos, también conocido como *minería de textos* es una serie de técnicas de investigación que utiliza el análisis computacional para desvelar patrones en amplios conjuntos de textos. Al hacer uso de técnicas informáticas es un ámbito que está dominado por los especialistas en informática, pero gracias a ellos disponemos de un gran número de herramientas de las

<http://www.aic.uva.es/7partidas/>

josemanuel.fradejas@uva.es
[@JMFradeRue](#)

7partidas.hypotheses.org
github.com/7partidasdigital



FFI2016-75014-P AEI-FEDER, EU

