

# ¿Qué se escribe respecto al marxismo en redes sociales? Análisis de patrones de texto a través de *Twitter* por medio de *Data Mining*.

Alfredo Olguín

Círculo de Estudios Economía Marxista  
para entender el Siglo XXI

“... se afirma un fatalismo histórico, una eliminación del hombre y de su práctica social, una acción de la técnica como *fuera natural* de la sociedad, como *ley natural social*” Lukács, G. (2009)

## Resumen

La tecnología que se ha venido gestando desde el siglo XX ha permitido un desarrollo acelerado de las capacidades para obtener información, especialmente en el ámbito de las tecnologías de la información. Este estudio trata de analizar por medio de patrones de texto las opiniones públicas en Twitter sobre el pensamiento crítico. El objetivo es obtener algunas nociones que permitan entender las expresiones e ideas que se gestan en esta red social. Para ello, se desarrolló un programa en *R* por medio de minería de datos para obtener información y analizar cerca de 2,000 mensajes y 18,000 palabras. Se desarrolla un programa completo para la limpieza de datos y se crean estructuras de documentos *corpus* para el análisis de los mismos. Los resultados apuntan a que la opinión pública no favorece totalmente a la teoría crítica. A manera de conclusión se observa que existe una opinión variada pero que es necesario quizás seguir desarrollando e innovando el pensamiento crítico.

## 1. Introducción

Las relaciones sociales de producción, en su esplendor actual, permiten la interacción humana por medio de dispositivos digitales que les evita la fatiga o dificultad de la comunicación personal. El desarrollo tecnológico ha brindado la oportunidad de dar acceso casi universal a redes comunicación por medio computacional. Son conocidos los amplios beneficios tecnológicos al desarrollo capitalista que esto conlleva; precarización de trabajo por medio de *outsourcing*, reducción de costos e incrementos inigualables de la productividad del trabajo, ende incremento del trabajo excedente, la explotación y contribuir a la caída tendencial de la tasa de ganancia<sup>1</sup>.

Para el marxismo contemporáneo es necesario incorporarse a esta nueva estructura de reproducción característica de tener la tecnología a disposición de la disciplina e investigación, los beneficios los han experimentado las ciencias biológicas con el desarrollo de la bioestadística por ejemplo. El pensamiento crítico ha dado interesantes propuestas en este campo, haciendo una revisión rápida nos encontramos con famosos y concurridos sitios web dedicados a la ideología marxiana y el pensamiento crítico con una perspectiva actual<sup>2</sup>.

Sin contar también con las importantes colaboraciones en revistas académicas que promueven el estudio de la ideología de Marx<sup>3</sup>.

El objetivo de este estudio será aplicar una de las técnicas más utilizadas en tecnologías de la información (IT por sus siglas en inglés) generalmente en el ámbito empresarial llamada *data mining* y aplicar un análisis en el comportamiento social en *Twitter* respecto al marxismo para tratar de comprender un poco del ambiente social que el tema genera. El análisis se conduce por medio de R 3.2.2 "Fire Safety". R es un lenguaje de programación de código abierto especializado en estadística y entendimiento gráfico-computacional disponible en el sitio web de la [organización](#). Este trabajo desarrolla los pasos lógicos para la obtención de los datos, análisis completo de frecuencias y un análisis de relación de texto. El código de programa está disponible para su reproducción y se muestra conforme el desarrollo de la investigación.

---

<sup>1</sup>Valle y Mendieta (2012) muestran un estudio empírico amplio para el caso de EE.UU. donde se presentan múltiples estimaciones de la tasa de ganancia.

<sup>2</sup>Ver *Marxismo Crítico*, *Marxist Internet Archive*, *Econo Marx21*, *Marxism 21*, *Marxismo21*, *Tiempos Críticos*.

<sup>3</sup>Ver *Review of Radical Political Economics*

### 1.1. *Data mining* y algunas de sus aplicaciones

*Data mining* es un concepto muy simple que descansa en la idea de la obtención de datos por medio de algún lenguaje de programación, el uso no está limitado a un programa específico<sup>4</sup>. También llamada minería de datos, es conjunto de técnicas provee una aproximación al entendimiento de un fenómeno por medio del análisis estadístico y/o el uso de aprendizaje computacional llamado *machine learning*.

Hay dentro de las ciencias de la computación un muy amplio campo de desarrollo en este ámbito, inclusive se ha tratado de catalogar la minería de datos una nueva forma de acercarse al conocimiento desligada a un método científico positivista convencional. Por ejemplo, Korb (2004) discute que quizás hay una potencial nueva filosofía de las ciencias en el *data mining* con resultados poco favorables. La realidad es que las técnicas de minería de datos no ascienden al grado de ciencia, más bien son una herramienta útil de aproximación al entendimiento de los fenómenos sin llegar a explicarlos.

Las técnicas de *data mining* tienen una infinidad de aplicaciones prácticas desde el análisis de datos, búsqueda de patrones, modelos de agrupamiento o *clustering* (Ver Figura 1), desarrollo de algoritmos predictivos, estadística descriptiva e incluso análisis de comportamiento. Lawrence, Kudyba y Klimberg (2008) sugieren que principalmente se utilizan para convertir información en una clase de conocimiento usualmente conocido como *business intelligence*.

Apoyado fuertemente en la minería de datos la inteligencia empresarial o *business intelligence* es un campo de análisis que surge a finales del siglo XX pero aún con problemas en nuestros días a pesar de su desarrollo, por ejemplo Thamir y Poulis (2015) critican la forma en que las empresas tratan de resolver sus problemas de falta de información con distintas metodologías, incluyendo *data mining* sin lograr ningún avance en sus objetivos de obtener resultados relevantes, poniendo en contraste que se pueden tener y analizar cantidades enormes de información sin resultar útil<sup>5</sup>.

Tomando en cuenta los problemas discutidos que la minería de datos puede conllevar, este artículo propone realizar un análisis por medio de las técnicas mencionadas y llegar a conclusiones pertinentes teniendo como objetivos: primero, tomar en cuenta la opinión pública para la teoría marxiana

---

<sup>4</sup>Ver Gallardo et al (2014) donde desarrollan varias aplicaciones para minería de datos en los lenguajes *Perl* y *Python*.

<sup>5</sup>Los trabajos de Osborne (2013) son muy claros al respecto

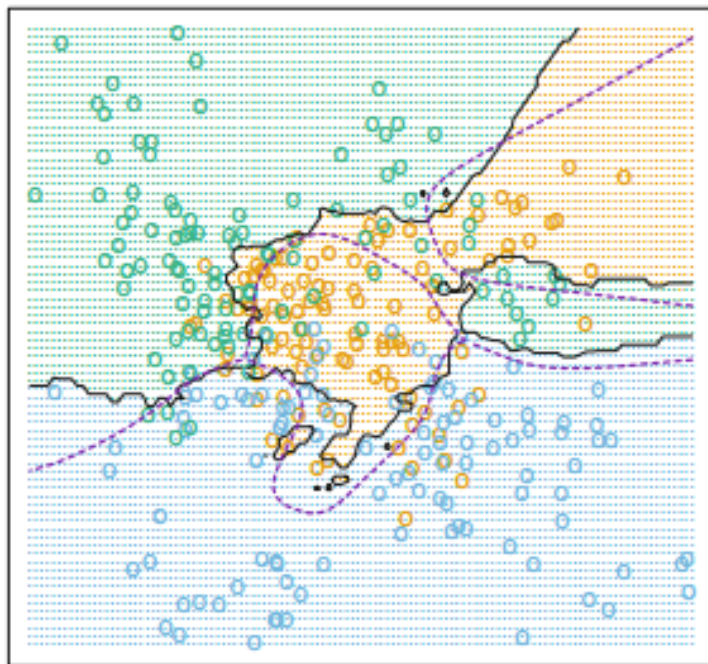


Figura 1: Ejemplo de Hastie, Tibshirani y Friedman (2009)[pp. 464] donde realizan un agrupamiento por método de Nearest Neighbors. El modelo consiste en hacer un modelo de agrupación datos dadas sus cualidades, en ese caso colores.

y someterlo al análisis de datos. Segundo, conocer cuáles son los conceptos e ideas más cercanas a la gente por medio de búsqueda en patrones de texto de la sociedad que hace uso del *microblogging Twitter*. Tercero, contribuir llevando técnicas nuevas al desarrollo del pensamiento crítico y abrir la puerta a la discusión para un análisis más profundo.

## 1.2. La importancia de *Twitter* para el pensamiento crítico

Carlson (2011) muestra una perspectiva interesante de la historia de *Twitter*. Para las aplicaciones prácticas de este estudio resulta interesante tomar en cuenta que desde sus inicios en 2006 el crecimiento del uso de las redes sociales ha sido exponencial y sigue con un fuerte desarrollo. Las limitaciones tecnológicas de la pequeña compañía fundadora de *Twitter* llamada Odeo

(dónde trabajaban 14 personas corría sus servicios en una computadora case-  
ra "IBM Thinkpad") se ha convertido en proveedora de una interesante gran  
base de datos, al día se opera con cerca de 500 millones de mensajes diarios<sup>6</sup>.  
y la posibilidad de utilizar la información disponible de manera gratuita.

Más allá de las innovaciones que este servicio en línea trajo a la produc-  
ción capitalista, resulta ser una fuente casi ilimitada de información que los  
usuarios arrojan. El tener al momento cantidades masivas de información  
que se comprenden como pensamientos de los usuarios nos da acceso casi  
inmediato e ilimitado a observar expresiones u ideas desde la fuente origi-  
nal, la sociedad misma. Esto sin importar su ideología u concepción frente al  
marxismo *ex-ante* del análisis.

Para el pensamiento crítico es necesario entender cuáles son las opiniones  
de la población general, entender cuáles son los juicios que se han formado  
hacia el pensamiento marxiano ayudaría a entender problemas fundamentales  
que el marxismo tiene en su difusión como ideología y herramienta teórica. Si  
bien el marxismo no es quizás la ideología más aceptada por la academia dada  
su naturaleza crítica al sistema capitalista<sup>7</sup> podremos entender las ideas más  
generales si se utilizan los avances tecnológicos disponibles a nuestros días.  
Más que formar un juicio de los mensajes de los usuarios de *Twitter*, es una  
aproximación a conocer las ideas que a un concepto o término rodean.

La metodología presentada tiene la capacidad de resolver preguntas im-  
portantes como ¿existe una aceptación o rechazo social al marxismo?, ¿qué  
autores están en boga?, ¿existen zonas geográficas dónde el marxismo está  
siento mencionado?, ¿que clase de acercamientos se tienen en los diferentes  
idiomas? etc.

## 2. Metodología

El trabajo se desarrolló en *Rstudio* versión 0.98.1103 de código abierto,  
siendo un ambiente de programación integrado para el lenguaje *R*. Las carac-  
terísticas del hardware son limitadas pero suficientes para el análisis: Linux  
Mint 17.2 *Rafaela-Cinamon* procesador *Intel* de cuatro núcleos i3 con 4Gb  
de memoria RAM.

<sup>6</sup>Ver estadísticas en tiempo real <http://www.internetlivestats.com/twitter-statistics/>

<sup>7</sup>Ver Astarita, R. "Fundamentos metodológicos en economía neoclásica y economía  
política: Nota de clase para Desarrollo Económico", Facultad de Ciencias Económicas,  
UBA)

## 2.1. Prerrequisitos y librerías

Los paquetes que se utilizaron se muestran del código a continuación.

```
> library(twitteR)
> library(RCurl)
> library(httr)
> library(tm)
> library(wordcloud)
> library(ggplot2)
```

Cada uno de ellos desempeña funciones específicas desde la extracción, interacción con el servidor hasta el despliegue gráfico de los datos, entre ellos *twitteR* es el paquete más importante para el análisis. El paquete *twitteR* es una interfaz de programación (API por sus siglas en inglés) web creada por Jeff Gentry (2015) que realiza la interacción con el servidor personal del usuario en • bajo el dominio de desarrollador ya previamente autenticado y obtiene los mensajes dados los criterios del programador. El paquete da acceso y soporte completo al desarrollo de la aplicación cuyo fin es la obtención de información de los servicios de •. A continuación se muestra la forma de autenticación que se utilizó para este trabajo.

```
> consumer_key    = ("My_consumer_key")
> consumer_secret = ("My_consumer_secret")
> access_token    = ("My_access_token")
> access_secret   = ("My_access_secret")
> setup_twitter_oauth(consumer_key,
+                     consumer_secret,
+                     access_token,
+                     access_secret)
```

Por cuestiones de seguridad de información personal las claves originales no se muestran en el código anterior.

## 2.2. Obtención, preprocesamiento y limpieza de los datos

La obtención de los datos se observan en el siguiente código. Dentro de los criterios de selección se eligió buscar 2,000 mensajes o *tweets* que incluyeran

la palabra "marxismo". Las condiciones de elección a continuación son que los mensajes sean en español y que se realice la búsqueda de manera aleatoria como opción predeterminada de la función. El objeto "my\_mining" creado es una lista de caracteres que no es apta para el análisis en su forma bruta.

```
> my_mining <- searchTwitter("marxismo",
+                             n=2000,
+                             lang = "es")
```

Una vez realizada la descarga de la lista de caracteres exitosamente se prosigue a hacer un manejo de los mismos. Es comprensible que al descargar la información exista información irrelevante, por ejemplo, se encuentran números, signos de puntuación, vínculos web, texto sin sentido o referencias a otros usuarios (@usuario por ejemplo) que no son analizables. Dentro de los 2,000 mensajes descargados se encontraron patrones de texto que entraban en conflicto con la manipulación de los datos<sup>8</sup>, el código siguiente transforma los datos y elimina los mensajes que se identificaron como indeseables.

```
> length(my_mining)
>
> my_mining_text <- sapply(my_mining[-c(871,
+                               1246,
+                               1711,
+                               1763,
+                               1817)],
+                           function(x) x$getText())
```

Posteriormente, el elemento lista se transforma a objeto de caracteres en el código anterior y a continuación se transforma a una lista de documentos *corpus* gracias al paquete *tm* cargado en la primera línea del código. Se generan 1,995 documentos de texto distintos; en cada uno se encuentra organizada un mensaje respectivamente.

Una vez transformados los elementos de lista se realizó una limpieza de los mismos. Al texto se le remueven todos los números, se transforman todas las palabras a minúsculas, se remueven los signos de puntuación y particularidades del lenguaje español (Artículos, pronombres, preposiciones etc.) que carecen de información.

<sup>8</sup>Se encontraron 5 mensajes con un patrón de código Utf-8 no manipulable, se consideraron como errores de origen y se eliminaron del análisis.

```
> my_mining_tcorpus <- tm_map(my_mining_tcorpus,
                             removeNumbers)
>
> my_mining_tcorpus <- tm_map(my_mining_tcorpus,
                             content_transformer(tolower))
>
> my_mining_tcorpus <- tm_map(my_mining_tcorpus,
                             removePunctuation)
>
> my_mining_tcorpus <- tm_map(my_mining_tcorpus,
                             function(x)removeWords(x,
                             stopwords("spanish")))
>
> my_mining_tcorpus <- tm_map(my_mining_tcorpus,
                             removeWords, c("marxismo"))
```

Por último se remueve la palabra "marxismo" del texto limpio para evitar la sobre estimación de la palabra en el análisis. Siendo el patrón original de texto buscado, es evidente que estará dentro de todos los mensajes.

### 3. Resultados

A continuación se crea en base a la colección de documentos (*corpus*) el análisis del patrón de texto. Primero se procede a hacer una exploración del contenido en los 1,995 documentos. Para este fin primero se realizó una nube de palabras o *Word Cloud* en la figura 2. El código siguiente genera la figura consecuente.

En la nube de palabras es posible apreciar los patrones más comunes, los términos más utilizadas en texto relacionado al marxismo a modo de hacer un análisis exploratorio. Se encontraron y procesaron 18,259 palabras en todos los mensajes, 5,163 palabras distintas. La lectura de la figura nos muestra aquellos vocablos más comunes en un tamaño mayor y en color rojo, las palabras más pequeñas y grises son aquellas que se repiten al menos 20 veces (están al menos en 1% de los *tweets*).

```
> wordcloud(my_mining_tcorpus,
+          min.freq = 20,
+          random.order = F,
```





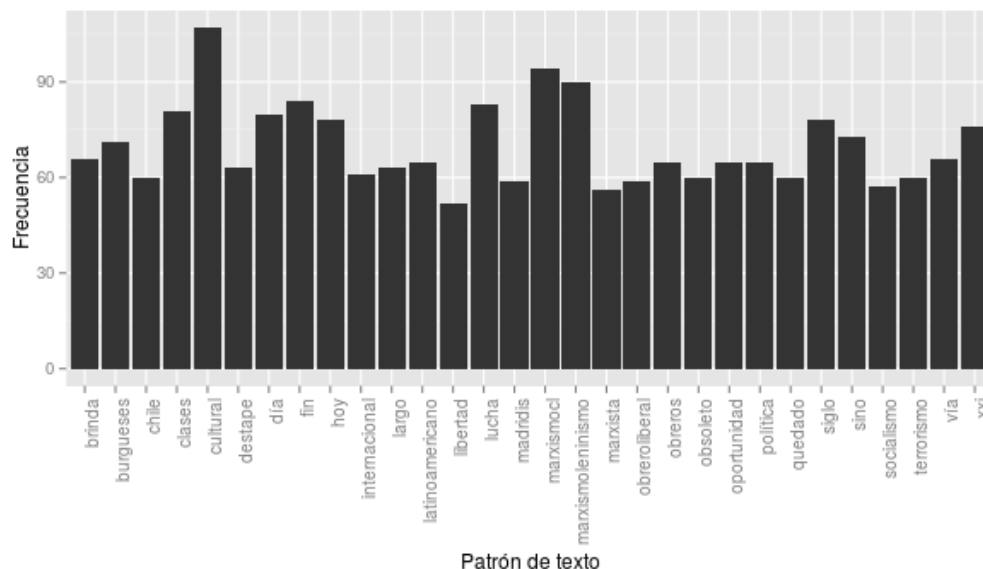


Figura 3: *Términos con las frecuencias más altas*

mostrados entre comillas y la relación de los demás términos con esta palabras, un criterio de selección de asociaciones menores "0.5" son descartadas. El estimador tiene un rango de 0 a 1, siendo 0 entendido como el término no se relaciona y 1 siempre se relaciona.

Interesantes observaciones del análisis recaen en la relación de: cultura y decadente (0.85), cultura y barrera (0.85), lucha y obsoleto(0.85), lucha y obreros (0.82), clases y burgueses (0.78), clases y obsoleto (0.86), política y narcoterrorismo (0.70), política y marxismolenismo(0.56) entre otros. Debemos recordar que a todas estas relaciones debemos siempre tener en cuenta la palabra "marxismo" que fue eliminada previamente para evitar sobreestimación en la manipulación de los datos pero está presente en todos los mensajes. Las tablas de todos términos que se analizaron están disponibles en el Anexo.

Haciendo el análisis para "Marx" obtenemos las siguientes relaciones.

"Marx"

<i>sacristán</i>	<i>manuel</i>	<i>acaso</i>	<i>aniversario</i>	<i>esperaba</i>
0.52	0.49	0.44	0.44	0.44
<i>después</i>	<i>muriese</i>	<i>antoniom</i>	<i>murió</i>	<i>burguesía</i>

0.44	0.43	0.41	0.41	0.39
chavismo	eddy			
0.39	0.33			

El análisis de relaciones denota que existe una influencia notable de Manuel Sacristán<sup>9</sup> y menciones sobre el aniversario de la muerte de Karl Marx. También una alta relación al chavismo<sup>10</sup>.

## 4. Conclusiones

Este trabajo es una aproximación a entender los patrones de texto, mensajes u ideas que se relacionan en torno al marxismo en las redes sociales. Una aproximación por medio del análisis de 2,000 mensajes que incluyeran la palabra "marxismo" arrojó interesantes nociones respecto a lo que se discute en *Twitter*. Se encontraron 5,163 términos distintos y 18,259 vocablos como total de los textos. Dentro de los resultados más visibles encontramos repetidas menciones a autores conocidos como Manuel Sacristán y constantes menciones a países como Chile, regiones como Latinoamérica e incluso ciudades específicas como Madrid. Se encontraron dentro de los mensajes expresiones poco favorables pero que deben ser tomadas en cuenta para seguir desatollando el pensamiento crítico por ejemplo "obsoleto", "dictadura" y "desigualdad". Inclusive se encontraron mensajes que relacionan el término marxismo y conocidos políticos en la nube de palabras.

Comprendiendo de manera superflua comentarios aparentemente aislados y unidos por una misma palabra llegamos a observar que a la opinión de los usuarios de *Twitter* sobre la teoría marxista quizás no tiene una opinión muy favorable. Aunque se observan también relaciones que nos muestran malestar social en el sistema capitalista como "cultura" y "decadente", inclusive "terrorismo" y "estado".

Dentro de todo, hay que recordar que es la opinión general de los usuarios lo que se analiza y en ningún momento es un estudio determinista sino una investigación que pretende comprender la dinámica social que puede tornarse

<sup>9</sup>Filósofo y traductor destacado de obras marxistas, entre ellos autores como Gramsci (2013) y Lukács (1982, 2009). Pero sin limitarse a la corriente marxiana también tradujo a Schumpeter (2012) por ejemplo.

<sup>10</sup>Notar que hay una tendencia generalizada en el la opinión común de relacionar al chavismo, no es de sorprender que se relacione con Marx en algunos mensajes de los usuarios de *Twitter* dada la propaganda de socialismo que se le adjudica.

distinta; cada día se producen cerca de 500 millones de mensajes en *Twitter* con un crecimiento sostenido. Es menester de los constructores de la teoría avanzar para llevar la crítica a los actores del capitalismo voraz en que vivimos y más importante aún, proveer y desmitificar la obsolescencia de la teoría marxiana para que el trabajador convencional la comprenda y tenga presente. Una prueba de ello es este estudio mismo, aplicar tecnologías avanzadas para tratar de construir el pensamiento crítico, seguir trayendo al siglo XXI los aportes que la teoría crítica del sistema de producción actual puede proveer.

## Referencias

1. Clarson, Nicholas, 2011, 'The Real History Of Twitter', Business Insider, Available on: <http://www.businessinsider.com/how-twitter-was-founded-2011-4>.
2. Gallardo, Ó., Ovelleiro, D., Gay, M., Carrascal, M., and Abian, J. (2014). A collection of open source applications for mass spectrometry data mining. *Proteomics*, 14(20), 2275-2279. doi:10.1002/pmic.201400124
3. Gentry, J., (2015). Package 'twitteR' R Based Twitter Client version 1.1.9, Cran R.
4. Gramsci, A, y Sacristán Luzón, M 2013, *Antología*, n.p.: Madrid, España : Ediciones Akal, 2013.
5. Hastie, T, Friedman, J, and Tibshirani, R 2009, *The Elements Of Statistical Learning : Data Mining, Inference, And Prediction*, n.p.: New York, New York : Springer Verlag, 2009.
6. Korb, K 2004, 'Introduction: Machine Learning as Philosophy of Science', *Minds and Machines*, 14, 4, pp. 433-440.
7. Lawrence, K, Kudyba, S, and Klimberg, R 2008, *Data Mining Methods And Applications*, n.p.: Boca Raton Florida : Auerbach, c2008.
8. Lukács, G, Sartelli, E, y Sacristán Luzón, M 2009, *Historia Y Conciencia De Clase : Estudios De Dialectica Marxista*, n.p.: Buenos Aires, Argentina : Ediciones RyR, c2009,

9. Lukács, G, y Sacristán Luzón, M 1982, Estetica, n.p.: Barcelona : Grijalbo, 1982-9999,
10. Osborne, JW 2013, Best Practices In Data Cleaning : A Complete Guide To Everything You Need To Do Before And After Collecting Your Data, n.p.: Thousand Oaks, California : SAGE, c2013.
11. Schumpeter, J, Estapé, F, Schumpeter, E, Sacristán Luzón, M, García-Durán de Lara, J, Serra, N, y Schumpeter, J 2012, Historia Del Análisis Económico, n.p.: Barcelona : Editorial Planeta, 2012.
12. Thamir, A, and Poulis, E 2015, 'Business Intelligence Capabilities and Implementation Strategies', International Journal Of Global Business, 8, 1, pp. 34-45.
13. Valle Baeza, A, and Mendieta Munoz, I 2012, 'What Is the Relationship between the Rates of Interest and Profit? An Empirical Note for the U.S. Economy, 1869-2009', Investigacion Economica, 71, 280, pp. 163-183,

## 5. Anexo

Código 1, las asociaciones de texto. Se despliega el código que genera las relaciones documento-término o término-documento bajo la estructura del *corpus* limpio. El número debajo de cada palabra representa el porcentaje de asociación.

```
> assoc<-findAssocs(dtm, c("cultura",
                           "lucha",
                           "clases",
                           "xxi",
                           "oportunidad",
                           "política"), corlimit=0.5)

"Cultura"

      barrera    decadente    derrumbar    impacta    martillazo
      0.85        0.85        0.85        0.85        0.85
rafaelvalera    primer      punto      siempre
      0.80        0.76        0.76        0.51
```

"Lucha"

obsoleto	quedado	madridis obreroliberal	
0.85	0.85	0.84	0.84
sino	hoy	día	burgueses
0.76	0.74	0.73	0.77
obreros			
0.82			

"Clases"

obsoleto	quedado	madridis obreroliberal	obreros
0.86	0.86	0.85	0.85 0.83
burgueses	sino	hoy	día
0.78	0.77	0.74	0.73

"xxi"

siglo	largo	brinda latinoamericano
0.95	0.91	0.89 0.89
fin	internacional	amé destape
0.78	0.77	0.75 0.86

"Oportunidad"

brinda	destape	largo latinoamericano
0.99	0.98	0.97 0.95
internacional	fin	siglo amé
0.89	0.88	0.87 0.82

"Política"

cuestione	extraña	moralmente
0.82	0.82	0.82
jcpastrana	narcot	castro
0.81	0.81	0.80
terrorismo marxismoleninismo		
0.70	0.56	