

CORPUS & FRIENDS

Leticia Martín-Fuertes Moreno
Lingüista computacional en Bitext
@nimbusaeta
nimbusaeta@gmail.com



Lingwars @ UVa
9 de noviembre de 2017

AGENDA

1. **Corpus**

¿Qué es un corpus?

¿Por qué usar corpus?

Criterios para construir un corpus

Tipos de corpus

2. **Herramientas**

Consulta en línea

AntConc

3. **Proyectos de lingüística de corpus**

Friends

Otros

CORPUS

¿QUÉ ES UN CORPUS?

“[...] a text corpus is a large body of text. Many corpora are designed to contain a careful balance of material in one or more genres.”

Natural Language Processing with Python, S. Bird, E. Klein y E. Loper, 2014 ([Chapter 2](#))

«En lingüística, se llama *corpus* a un conjunto formado por miles de textos (novelas, obras de teatro, guiones de cine, noticias de prensa, ensayos, transcripciones de noticiarios radiofónicos o televisivos, transcripciones de conversaciones, discursos, etc.) y cientos de millones de formas. Son empleados habitualmente para conocer el significado y características de palabras, expresiones y construcciones a partir de los usos reales registrados. Dado el tamaño que poseen, los corpus tienen que estar en formato electrónico.»

Real Academia Española, en la [web del CORPES XXI](#)

«Esencialmente, un corpus es una carpeta llena de archivos del bloc de notas ordenados y preparados para ser analizados posteriormente.»

Carlos Loz Garrido, su blog [Filología, investigación y humanidades digitales](#)

¿QUÉ ES UN CORPUS?

¿El tamaño importa?

Corpus Text Linguistics



How big does a text have to be before it becomes a corpus?



Olfert Rahbek, Founder of wordmaps.org (2016-present)

Answered Thu



There is no formal limit. Rather, the main question would be: What does the set of data represent and is it complete? For example, a really nice corpus of essays would contain all corpuses from one or even several years, not just a sample.



John Stanley, Graduate major was Linguistics. Keep up through independent reading.

Answered Wed



As I understand the word, a corpus is not a single text; it is the total of all available texts in a category. It can relatively small: the total corpus of Pictish inscriptions would scarcely make a pamphlet. Alternatively it can be huge: it would take a shelf of books to contain the corpus of Ancient Greek literature and inscriptions.

¿POR QUÉ USAR CORPUS?



- Necesitamos grandes cantidades de datos para detectar **tendencias**, lo que es normal o típico en el uso real del lenguaje
- Los corpus también revelan **casos muy raros** o excepcionales que no encontraríamos analizando textos uno a uno
- Los humanos cometemos **errores** y somos **lentos**; los ordenadores son muchísimo más rápidos y precisos

CRITERIOS PARA CONSTRUIR UN CORPUS



- Debe ser una **gran cantidad** de texto
- Tiene que ser **representativo** de una lengua (o de un género de esa lengua)
- Tiene que estar en un **formato legible** para una máquina (p. ej., archivos de texto .txt)
- A menudo, **se anota** con información lingüística adicional (p. ej., categorías de palabras o partes del discurso)

EJEMPLO DE TEXTO SIN ANOTAR

"Arrest warrant out for Clowes' partner years before collapse."

By Daniel John

A WARRANT for the arrest of the former partner of Mr Peter Clowes was issued seven years before his Barlow Clowes investment empire collapsed, according to evidence submitted to the Parliamentary Ombudsman.

EJEMPLO DE TEXTO ANOTADO

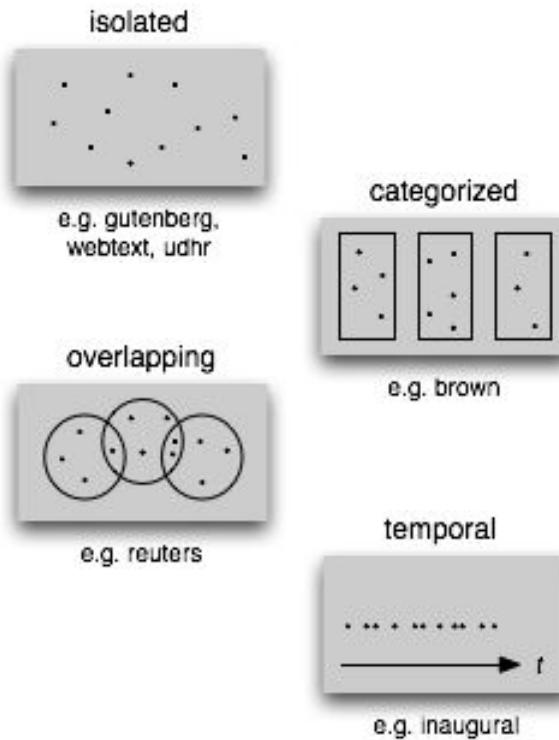
```
<head type=MAIN>
<s n=001><c PUQ>&bquo<w NN1>Arrest <w NN1>warrant <w AVP>out <w PRP>for <w
NP0>Clowes<c PUN>' <w NN1>partner <w NN2>years <w PRP>before <w NN1>collapse
<c PUN>. <c PUQ>&equo'
</head>
<head type=BYLINE>
<s n=002><w PRP>By <w NP0>Daniel <w NP0>John
</head>
<p>
<s n=003><w AT1>A <w>NN1>WARRANT <w PRP>for <w AT0>the <w NN1>arrest <w
PRF>of <w AT0>the <w DT0>former <w NN1>partner <w PRF>of <w NP0>Mr <w
NP0>Peter <w NP0>Clowes <w VBD>was <w VVN>issued <w CRD>seven <w NN2>years
<w CJS>before <w DPS>his <w NN1-NP0>Barlow <w NP0>Clowes <w NN1>investment
<w NN1>empire <w VVD>collapsed<c PUN>, <w PRP>according to <w NN1>evidence
<w VVN>submitted <w PRP>to <w AT0>the <w AJ0>Parliamentary <w
NN1>Ombudsman<c PUN>.
</p>
```

EJEMPLO DE TEXTO ANOTADO

```
<head type=MAIN>
< s n=001><c PUQ>&nbquo< w NN1>Arrest < w NN1>warrant < w AVP>out < w PRP>for < w
NP0>Clowes<c PUN>' < w NN1>partner < w NN2>years < w PRP>before < w NN1>collapse
<c PUN>. <c PUQ>&equo'
</head>
<head type=BYLINE>
< s n=002>< w PRP>By < w NP0>Daniel < w NP0>John
</head>
<p>
< s n=003>< w AT1>A < w>NN1>WARRANT < w PRP>for < w AT0>the < w NN1>arrest < w
PRF>of < w AT0>the < w DT0>former < w NN1>partner < w PRF>of < w NP0>Mr < w
NP0>Peter < w NP0>Clowes < w VBD>was < w VVN>issued < w CRD>seven < w NN2>years
< w CJS>before < w DPS>his < w NN1-NP0>Barlow < w NP0>Clowes < w NN1>investment
< w NN1>empire < w VVD>collapsed<c PUN>, < w PRP>according to < w NN1>evidence
< w VVN>submitted < w PRP>to < w AT0>the < w AJ0>Parliamentary < w
NN1>Ombudsman<c PUN>.
</p>
```

TIPOS DE CORPUS SEGÚN SU ESTRUCTURA

- El más simple: **carencia de estructura**; sencillamente, una colección de textos
- Por **categorías**, que normalmente corresponden a géneros, fuentes, autores, lenguas, etc.
- Por **categorías que se solapan**: cuando los textos se clasifican por temas, por ejemplo, puede ser interesante que un mismo texto pertenezca a dos categorías distintas.
- **Estructura temporal**: usada sobre todo en los corpus más recientes.



CORPUS BROWN

El [corpus Brown](#), creado en 1961 en la Universidad Brown (Rhode Island) fue el primer corpus electrónico en inglés en llegar al millón de palabras. Contiene textos de 500 fuentes diferentes, que fueron categorizadas por géneros, como «noticias», «editoriales», etc. Ejemplos de documentos incluidos en cada una de las categorías:

ID	File	Genre	Description
A16	ca16	news	Chicago Tribune: <i>Society Reportage</i>
B02	cb02	editorial	Christian Science Monitor: <i>Editorials</i>
C17	cc17	reviews	Time Magazine: <i>Reviews</i>
D12	cd12	religion	Underwood: <i>Probing the Ethics of Realtors</i>
E36	ce36	hobbies	Norling: <i>Renting a Car in Europe</i>
F25	cf25	lore	Boroff: <i>Jewish Teenage Culture</i>
G22	cg22	belles_lettres	Reiner: <i>Coping with Runaway Technology</i>
H15	ch15	government	US Office of Civil and Defence Mobilization: <i>The Family Fallout Shelter</i>
J17	cj19	learned	Mosteller: <i>Probability with Statistical Applications</i>
K04	ck04	fiction	W.E.B. Du Bois: <i>Worlds of Color</i>
L13	cl13	mystery	Hitchens: <i>Footsteps in the Night</i>
M01	cm01	science_fiction	Heinlein: <i>Stranger in a Strange Land</i>
N14	cn15	adventure	Field: <i>Rattlesnake Ridge</i>
P12	cp12	romance	Callaghan: <i>A Passion in Rome</i>
R06	cr06	humor	Thurber: <i>The Future, If Any, of Comedy</i>

TIPOS DE CORPUS SEGÚN SU PROPÓSITO

- **Especializado** en, por ejemplo:
 - Un género: el lenguaje periodístico.
 - Un momento: desde 2005 hasta ahora.
 - Un lugar: solo textos publicados en China.
- **General o de referencia:** sirve para obtener las características globales que presenta una lengua en un momento determinado de su historia. Por tanto, suelen ser mucho más grandes que los especializados. Por ejemplo:
 - [Corpus de Referencia del Español Actual \(CREA\)](#): 154 millones de palabras de textos escritos y orales, desde 1975 hasta 2004.
 - [British National Corpus \(BNC\)](#): contiene unos 100 millones de palabras del inglés hablado y escrito.

EL BNC

Modo	Tipo de textos y descripción		Número de palabras
Escrito (87 284 364 palabras)	Informativo (8 tipos)		70,9 millones
	1) Internacional	5) Filosofía	
	2) Ocio	6) Ciencias sociales	
	3) Arte	7) Ciencias aplicadas	
	4) Comercio y finanzas	8) Ciencias puras y naturales	
	Imaginativo (1 tipo)		
	9) Ficción		
Hablado (10 341 729 palabras)	Conversaciones informales recogidas demográficamente de la población del Reino Unido		4,2 millones
	Lenguaje dirigido a realizar una tarea, grabado en lugares específicos para eventos específicos, como reuniones de trabajo, charlas en público, etc.		6,1 millones

TIPOS DE CORPUS SEGÚN SU PROPÓSITO

- **Especializado** en, por ejemplo:
 - Un género: el lenguaje periodístico.
 - Un momento: desde 2005 hasta ahora.
 - Un lugar: solo textos publicados en China.
- **General o de referencia**: sirve para obtener las características globales que presenta una lengua en un momento determinado de su historia. Por tanto, suelen ser mucho más grandes que los especializados. Por ejemplo:
 - [Corpus de Referencia del Español Actual \(CREA\)](#): 154 millones de palabras de textos escritos y orales, desde 1975 hasta 2004.
 - [British National Corpus \(BNC\)](#): unos 100 millones de palabras del inglés hablado y escrito.
- Corpus **multilingüe**: inglés y español, o inglés americano y británico.
- Corpus **paralelo**: un mismo texto traducido a varios idiomas. Por ejemplo:
 - Corpus [CRATER](#).
 - Portal [OPUS](#) (Open Parallel corpUS).
- Corpus **histórico** o **diacrónico**: recogen textos representativos de varias épocas para poder estudiar la evolución del lenguaje. Por ejemplo:
 - Corpus [Helsinki](#): un millón y medio de palabras de textos escritos entre 700 y 1700.
 - Discursos presidenciales de EE. UU.
- Corpus **monitorizado**: está en continuo aumento. Por ejemplo, el [Bank of English](#).

HERRAMIENTAS

CONSULTA EN LÍNEA



REAL
ACADEMIA
ESPAÑOLA

Google Books

Ngram Viewer

BANCO DE DATOS DEL ESPAÑOL

Corpus Diacrónico del Español (CORDE): corpus textual de todas las épocas y lugares en que se habló español, desde los inicios del idioma hasta el año 1974.

Corpus de Referencia del Español Actual (CREA): textos escritos y orales, producidos en todos los países de habla hispana desde 1975 hasta 2004. Los textos escritos, seleccionados tanto de libros como de periódicos y revistas, abarcan más de cien materias distintas. La lengua hablada está representada por transcripciones de documentos sonoros, obtenidos, en su mayor parte, de la radio y la televisión.

Corpus	Palabras
CORDE	236709914
CREA	154212661

Origen	Palabras CORDE	Palabras CREA	Total palabras
España	196106277	85563661	281669938
Hispanoamérica	37562461	68569561	106132022
Otros	3041176	79439	3120615

Medio	Palabras CORDE	Palabras CREA	Total palabras
Libros	234056904	80677065	314733969
Prensa	2653010	70702838	73355848
Miscelánea (solo CREA)		2832758	2832758

Zonas lingüísticas de América	Palabras CORDE	Palabras CREA	Total palabras
Andina	15076129	16536515	31612644
Caribeña	3345900	9543456	12889356
Central	714341	2883979	3598320
Chilena	3983488	6494966	10478454
Mexicana	7822486	16897942	24720428
Rioplatense	6621954	16212703	22834657

BANCO DE DATOS DEL ESPAÑOL

Real Academia Española - Corpus Diacrónico del Español (CORDE)

Consulta:	lengua		
Criterios de selección:			
Autor:	Obra:		
Cronológico:	Medio:	(Todos) Libros Periódicos Revistas Miscelánea Orales	Geográfico:
Tema:	(Todos) 11.- Lírica 12.- Narrativa 121.- Breve 1211.- Relato breve tradicional 1212.- Relato breve culto		
<input type="button" value="Buscar"/> <input type="button" value="Limpiar"/>			

Resultado de la consulta al banco de datos

Cómo citar el CORPUS

Consulta:	<i>lengua, en todos los medios, en CORDE , en Narrativa , en ARGENTINA</i>
Resultado:	201 casos en 39 documentos.
<input type="button" value="Ver estadística"/>	

Filtros:	Casos
Ratio:	10

Mantener documentos (Solo para filtro sobre casos).

OBTENCIÓN DE EJEMPLOS

Recuperar	Concordancias	Normal
Clasificación:		
Agrupación:	Marcas:	

BANCO DE DATOS DEL ESPAÑOL

Consulta:	<i>lengua, en todos los medios, en CORDE , en Narrativa , en ARGENTINA</i>
Resultado:	201 casos en 39 documentos.

OBTENCIÓN DE EJEMPLOS

Recuperar	Concordancias ▾	Normal. ▾	Clasificación:	▼
Agrupación:			Marcas:	▼

Cómo citar el CORPUS

Concordancias.

Pantalla: 1 de 9. Siguiente 1 2 3 4 5 6 7 8 9 Ver párrafos

Nº	CONCORDANCIA	AÑO	AUTOR	TÍTULO	PAÍS
1	Estar con ustedes me conforta -exclamó Banyay. Su lengua, vasta, delicada y rosada, se apoyó contra los ** 1948 Biyo Casares, Adolfo La trama celeste ARGENTINA				
2	ededor de una caja que la corriente arrojara á la lengua del agua, de donde esforzados marineros no con ** 1903 Obligado, Pastor Servando Tradiciones argentinas ARGENTINA				
3	s varas de ancho por una legua de fondo, desde la lengua del agua) los primeros que adquirió allí, hast ** 1903 Obligado, Pastor Servando Tradiciones argentinas ARGENTINA				
4	oros hasta el barrio de la Ranchería, en su media lengua bozal, Síripo y Maratón, El Tartufo y la Camil ** 1903 Obligado, Pastor Servando Tradiciones argentinas ARGENTINA				
5	reñadas entre las filas de milicianos, sacando la lengua á los misters azorados, haciendoles pito catal ** 1903 Obligado, Pastor Servando Tradiciones argentinas ARGENTINA				
6	de toda la peonada, como que no tenía pelos en la lengua, declaró la niña que lo quería para su uso par ** 1903 Obligado, Pastor Servando Tradiciones argentinas ARGENTINA				
7	ques y dolamas de todo género, vino á trabarle la lengua. - ¡Qué bruto soy! - repitió Williams.- Dios ca ** 1903 Obligado, Pastor Servando Tradiciones argentinas ARGENTINA				
8	añadiendo la negrita esclava, recado en su media lengua: "Manda decir mi amita que las naranjas tambié ** 1903 Obligado, Pastor Servando Tradiciones argentinas ARGENTINA				
9	rientas varas por lengua y media de fondo, desde la lengua del agua, tras la iglesia de San Isidro, minim ** 1903 Obligado, Pastor Servando Tradiciones argentinas ARGENTINA				
10	sor con insipida galantería y melosidades á media lengua. Los representantes de Su Majestad británica p ** 1903 Obligado, Pastor Servando Tradiciones argentinas ARGENTINA				
11	de la prisión de este último. Pecolat se cortó la lengua con los dientes y la arrojó altanero al rostro ** 1844 - 1851 Alberdi, Juan Bautista Tobías o La cárcel a la vela ARGENTINA				
12	del panóptico? Poseedor cada uno de una riquísima lengua, tienen que acudir para entenderse, a las muc ** 1844 - 1851 Alberdi, Juan Bautista Tobías o La cárcel a la vela ARGENTINA				
13	ra la mordaza que ahoga la blasfemia; y atáis la lengua del desgraciado que aspira a decir palabras de ** 1844 - 1851 Alberdi, Juan Bautista Tobías o La cárcel a la vela ARGENTINA				
14	son culpables del abandono en que hoy yace. "La lengua española es una lira, que no tiene armonías en ** 1844 - 1851 Alberdi, Juan Bautista Tobías o La cárcel a la vela ARGENTINA				
15	, es decir, de su existencia política, pues en la lengua del publicista, la unidad quiere decir la patr ** 1844 - 1851 Alberdi, Juan Bautista Tobías o La cárcel a la vela ARGENTINA				
16	ía y Pernambuco. Gentes blandujas y fofas como la lengua que hablan. Pasan su vida abordo descuajados s ** 1884 Cambaceres, Eugenio Música sentimental: silbidos de un vago ARGENTINA				
17	pecto". El père Grognard, se enojó: - "Métete la lengua donde no te la vean, tú, y déjate de fastidiar ** 1884 Cambaceres, Eugenio Música sentimental: silbidos de un vago ARGENTINA				
18	ban, tenía los ojos idos, el resuello pesado y la lengua considerablemente trabada. Siguiendo una vía q ** 1884 Cambaceres, Eugenio Música sentimental: silbidos de un vago ARGENTINA				
19	el trote cabestrando al corral con una cuarta de lengua afuera y por sufrir que le hachen las púas y l ** 1884 Cambaceres, Eugenio Música sentimental: silbidos de un vago ARGENTINA				
20	Vd. agotado todo el repertorio de dícterios de la lengua, amén de los que no reza el diccionario. Vomit ** 1884 Cambaceres, Eugenio Música sentimental: silbidos de un vago ARGENTINA				
21	ás quiero que a un cristiano con quien desatar la lengua? ¿Cree que su sociedad me divierte, sin embarg ** 1884 Cambaceres, Eugenio Música sentimental: silbidos de un vago ARGENTINA				
22	ara de hueso de pato. El paladar, las encías y la lengua se hallaban invadidas también. De la faz inter ** 1884 Cambaceres, Eugenio Música sentimental: silbidos de un vago ARGENTINA				
23	hijita? repuso éste suplicando, seca la boca, la lengua empastada. ¡Ahora, ahora mismo, ven! - No, no ** 1884 Cambaceres, Eugenio Música sentimental: silbidos de un vago ARGENTINA				
24	és- a recordarles la infranqueable barrera de la lengua, cuando Tarantino pidió la palabra, para afirm ** 1962 Biyo Casares, Adolfo El gran Serafín ARGENTINA				
25	podia de los topetazos de los años, me soltó la lengua y me llevó, sin un momento de vacilación, a in ** 1962 Biyo Casares, Adolfo El gran Serafín ARGENTINA				

[Ir arriba](#) Pantalla: 1 de 9. Siguiente 1 2 3 4 5 6 7 8 9 Ver párrafos

BANCO DE DATOS DEL ESPAÑOL

Real Academia Española - Corpus Diacrónico del Español (CORDE)

Consulta:	lengua		
Criterios de selección:			
Autor:	Obra:		
Cronológico:	Medio:	(Todos) Libros Periódicos Revistas Miscelánea Orales	Geográfico: (Todos) Argentina Bolivia Chile Colombia Costa Rica
Tema:	(Todos) 11.- Lírica 12.- Narrativa 121.- Breve 1211.- Relato breve tradicional 1212.- Relato breve culto		
<input type="button" value="Buscar"/> <input type="button" value="Limpiar"/>			

Resultado de la consulta al banco de datos

Cómo citar el CORPUS

Consulta:	<i>lengua, en todos los medios, en CORDE</i>
Resultado:	36466 casos en 3869 documentos.

No se pueden ver estadísticas. Demasiados documentos.

Filtros:	<input type="button" value="▼"/>
Ratio:	10

Mantener documentos (Solo para filtro sobre casos).

No se puede filtrar. Demasiados aciertos y documentos.

OBTENCIÓN DE EJEMPLOS

No se puede Recuperar.

Demasiados documentos.

Normal.

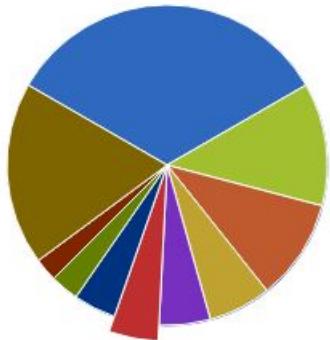
Clasificación:

Agrupación:

Marcas:

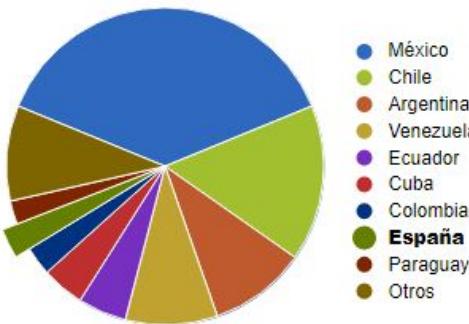
CORPES XXI

Distribución País



sostenible

Distribución País



sustentable

El Corpus del Español del Siglo XXI (Corpes XXI) se compone de textos escritos y orales procedentes de España, América, Filipinas y Guinea Ecuatorial.

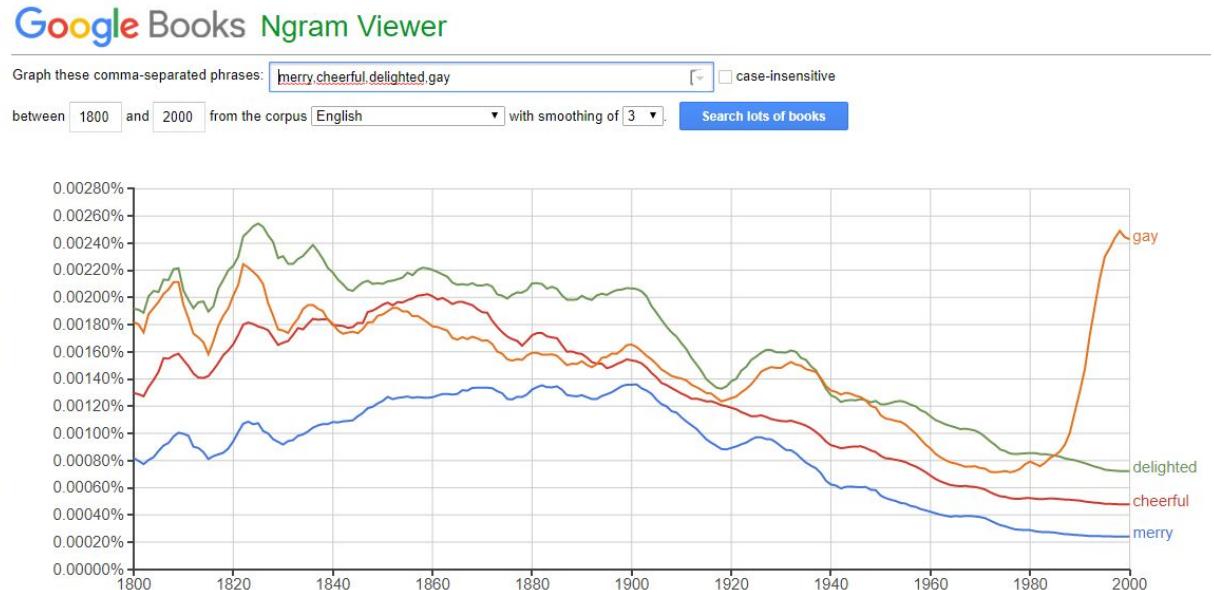
25 millones de formas por cada año entre 2001 y 2018 (en construcción).

Anotación morfosintáctica y lematización.

GOOGLE BOOKS NGRAM VIEWER

N-gramas: conjunto de n palabras.

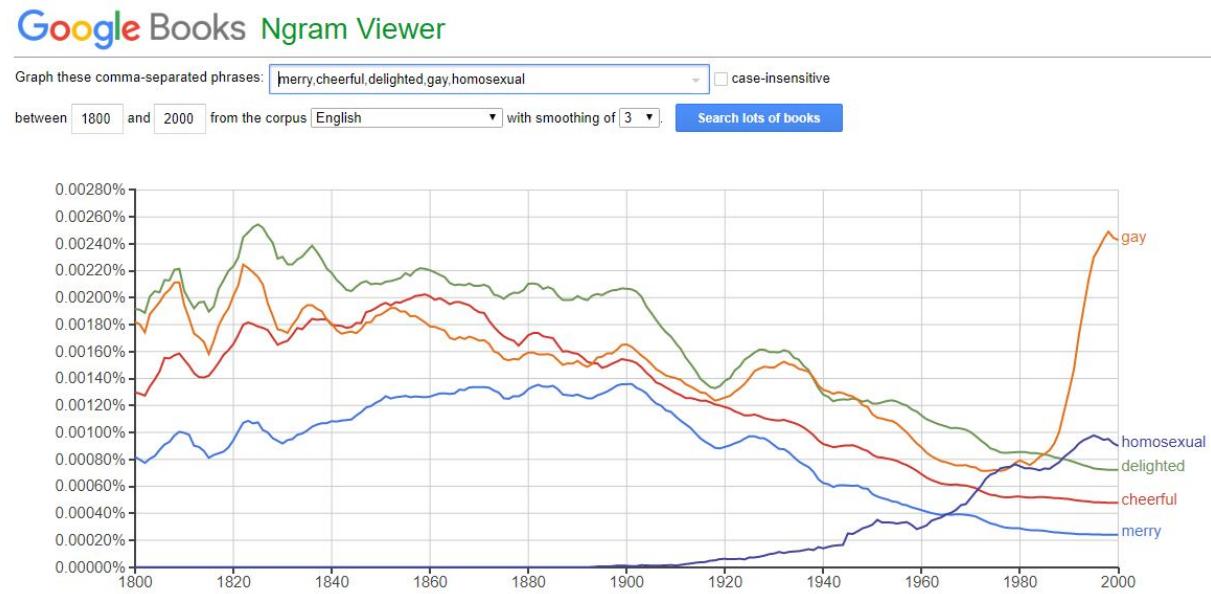
Ngram Viewer permite hacer búsquedas en Google Books, que contiene 5,2 millones de libros, publicados entre 1500 y 2008, en inglés, francés, español, alemán, chino y ruso. Monumental base de datos, gratuita, con 500 mil millones de términos obtenidos de esos libros.



GOOGLE BOOKS NGRAM VIEWER

N-gramas: conjunto de n palabras.

Ngram Viewer permite hacer búsquedas en Google Books, que contiene 5,2 millones de libros, publicados entre 1500 y 2008, en inglés, francés, español, alemán, chino y ruso. Monumental base de datos, gratuita, con 500 mil millones de términos obtenidos de esos libros.



ANTCONC

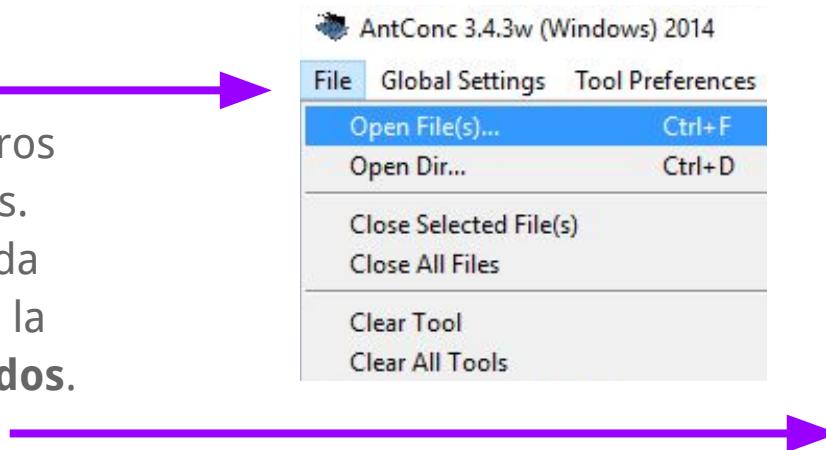


ANTCONC

- Programa desarrollado por el profesor Laurence Anthony
- Sirve para el análisis de corpus
- Ventajas:
 - Se puede [descargar gratuitamente](#)
 - Es fácil de manejar
- Desventajas:
 - Funcionalidad limitada, no puede trabajar con corpus muy grandes
 - No permite agrupaciones (lematización)

CARGAR EL CORPUS

1. En el menú de AntConc:
File -> Open File(s).
2. Seleccionamos los ficheros que componen el corpus.
3. En la parte de la izquierda de la ventana aparecerá la **lista de ficheros cargados.**
Debajo aparecerá el número de ficheros.



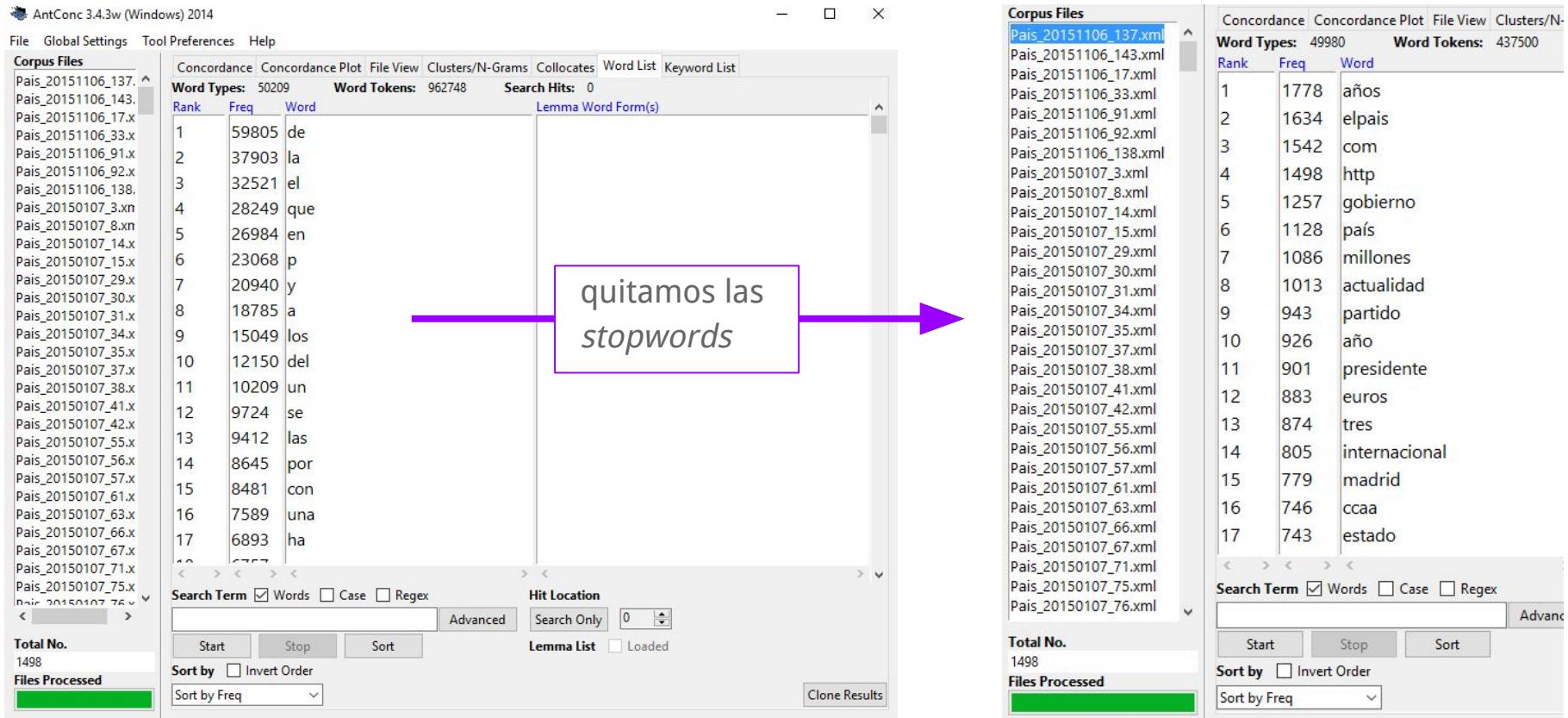
¡Esto va a ser nuestro corpus!

Artículos de *El País* de 12 ejemplares diferentes, pertenecientes a cada mes de 2015

Corpus Files
Pais_20151106_137.xml
Pais_20151106_143.xml
Pais_20151106_17.xml
Pais_20151106_33.xml
Pais_20151106_91.xml
Pais_20151106_92.xml
Pais_20151106_138.xml
Pais_20150107_3.xml
Pais_20150107_8.xml
Pais_20150107_14.xml
Pais_20150107_15.xml
Pais_20150107_29.xml
Pais_20150107_30.xml
Pais_20150107_31.xml
Pais_20150107_34.xml
Pais_20150107_35.xml
Pais_20150107_37.xml
Pais_20150107_38.xml
Pais_20150107_41.xml
Pais_20150107_42.xml
Pais_20150107_55.xml
Pais_20150107_56.xml
Pais_20150107_57.xml
Pais_20150107_61.xml
Pais_20150107_63.xml
Pais_20150107_66.xml
Pais_20150107_67.xml
Pais_20150107_71.xml
Pais_20150107_75.xml
Pais_20150107_76.xml

Total No.
1498

LISTA DE PALABRAS

AntConc 3.4.3w (Windows) 2014

File Global Settings Tool Preferences Help

Corpus Files

- Pais_20151106_137.xml
- Pais_20151106_143.xml
- Pais_20151106_17.xml
- Pais_20151106_33.xml
- Pais_20151106_91.xml
- Pais_20151106_92.xml
- Pais_20151106_138.xml
- Pais_20150107_3.xml
- Pais_20150107_8.xml
- Pais_20150107_14.xml
- Pais_20150107_15.xml
- Pais_20150107_29.xml
- Pais_20150107_30.xml
- Pais_20150107_31.xml
- Pais_20150107_34.xml
- Pais_20150107_35.xml
- Pais_20150107_37.xml
- Pais_20150107_38.xml
- Pais_20150107_41.xml
- Pais_20150107_42.xml
- Pais_20150107_55.xml
- Pais_20150107_56.xml
- Pais_20150107_57.xml
- Pais_20150107_61.xml
- Pais_20150107_63.xml
- Pais_20150107_66.xml
- Pais_20150107_67.xml
- Pais_20150107_71.xml
- Pais_20150107_75.xml
- Pais_20150107_76.xml

Word Types: 50209 **Word Tokens:** 962748 **Search Hits:** 0

Rank	Freq	Word
1	59805	de
2	37903	la
3	32521	el
4	28249	que
5	26984	en
6	23068	p
7	20940	y
8	18785	a
9	15049	los
10	12150	del
11	10209	un
12	9724	se
13	9412	las
14	8645	por
15	8481	con
16	7589	una
17	6893	ha

Corpus Files

- Pais_20151106_137.xml
- Pais_20151106_143.xml
- Pais_20151106_17.xml
- Pais_20151106_33.xml
- Pais_20151106_91.xml
- Pais_20151106_92.xml
- Pais_20151106_138.xml
- Pais_20150107_3.xml
- Pais_20150107_8.xml
- Pais_20150107_14.xml
- Pais_20150107_15.xml
- Pais_20150107_29.xml
- Pais_20150107_30.xml
- Pais_20150107_31.xml
- Pais_20150107_34.xml
- Pais_20150107_35.xml
- Pais_20150107_37.xml
- Pais_20150107_38.xml
- Pais_20150107_41.xml
- Pais_20150107_42.xml
- Pais_20150107_55.xml
- Pais_20150107_56.xml
- Pais_20150107_57.xml
- Pais_20150107_61.xml
- Pais_20150107_63.xml
- Pais_20150107_66.xml
- Pais_20150107_67.xml
- Pais_20150107_71.xml
- Pais_20150107_75.xml
- Pais_20150107_76.xml

Word Types: 49980 **Word Tokens:** 437500

Rank	Freq	Word
1	1778	años
2	1634	elpais
3	1542	com
4	1498	http
5	1257	gobierno
6	1128	país
7	1086	millones
8	1013	actualidad
9	943	partido
10	926	año
11	901	presidente
12	883	euros
13	874	tres
14	805	internacional
15	779	madrid
16	746	ccaa
17	743	estado

Search Term Words Case Regex

Hit Location Search Only 0

Lemma List Loaded

Total No. 1498
Files Processed

Start **Stop** **Sort**
Sort by Invert Order
Sort by Freq

Advanced

Search Term Words Case Regex

Hit Location Search Only 0

Lemma List Loaded

Total No. 1498
Files Processed

Start **Stop** **Sort**
Sort by Invert Order
Sort by Freq

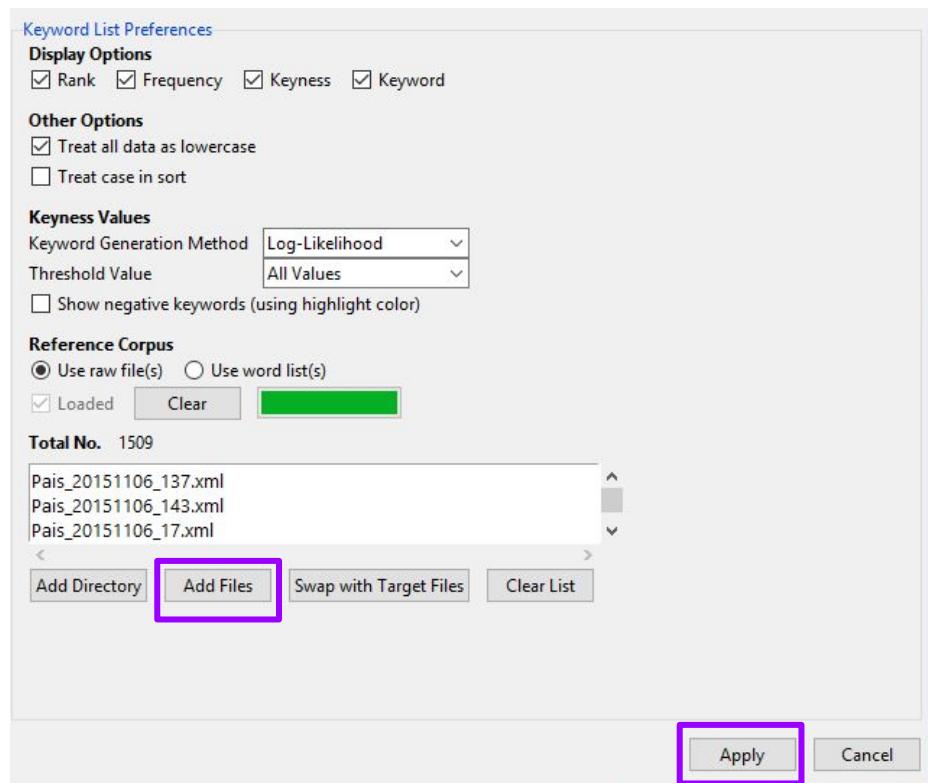
Advanced

LISTA DE PALABRAS CLAVE

Necesitaremos un corpus de referencia para que tenga con qué comparar.

Para cargar el corpus de referencia:

1. **Tool preferences** -> pestaña **Keyword List**
2. **Use raw file(s)** si vamos a usar un corpus o **Use word list(s)** si vamos a usar una lista de frecuencias
3. **Add Files** -> seleccionamos el/los archivo(s) -> **Load** (puede tardar unos minutos en cargarse) -> **Apply**



LISTA DE PALABRAS CLAVE

File Global Settings Tool Preferences Help

Corpus Files

- .DS_Store
- Pais_20150107_39.xml
- Pais_20150107_44.xml
- Pais_20150107_47.xml
- Pais_20150107_52.xml
- Pais_20150107_53.xml
- Pais_20150107_80.xml
- Pais_20150107_106.xml
- Pais_20150206_5.xml
- Pais_20150206_11.xml
- Pais_20150206_12.xml
- Pais_20150206_28.xml
- Pais_20150206_35.xml
- Pais_20150206_37.xml
- Pais_20150206_52.xml
- Pais_20150206_53.xml
- Pais_20150206_57.xml
- Pais_20150206_64.xml
- Pais_20150206_83.xml
- Pais_20150206_102.xml
- Pais_20150206_107.xml
- Pais_20150206_114.xml
- Pais_20150206_127.xml
- Pais_20150206_140.xml
- Pais_20150206_143.xml
- Pais_20150306_11.xml
- Pais_20150306_18.xml
- Pais_20150306_23.xml
- Pais_20150306_26.xml
- Pais_20150306_31.xml

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Types Before Cut: 13970 **Types After Cut:** 10289 **Search Hits:** 0

Rank	Freq	Keyness	Keyword
1	358	366.465	podemos
2	385	344.603	pp
3	248	271.242	psoe
4	245	199.246	ciudadanos
5	141	170.713	iu
6	127	151.869	rajoy
7	115	137.664	iglesias
8	90	109.344	cis
9	102	105.060	andalucía
10	134	103.432	juez
11	74	97.509	rivera
12	66	93.418	garzón
13	100	91.377	sánchez
14	83	79.667	escaños
15	78	73.822	dfaz
16	115	72.843	formación
17	266	71.857	partido

Search Term Words Case Regex **Hit Location**

wheno 0

Total No. 228 **Files Processed**

Sort by Invert Order **Reference Corpus** Loaded

Sort by Keyness

CONCORDANCIAS

File Global Settings Tool Preferences Help

Corpus Files

- Pais_20151106_137.xml
- Pais_20151106_143.xml
- Pais_20151106_17.xml
- Pais_20151106_33.xml
- Pais_20151106_91.xml
- Pais_20151106_92.xml
- Pais_20151106_138.xml
- Pais_20150107_3.xml
- Pais_20150107_8.xml
- Pais_20150107_14.xml
- Pais_20150107_15.xml
- Pais_20150107_29.xml
- Pais_20150107_30.xml
- Pais_20150107_31.xml
- Pais_20150107_34.xml
- Pais_20150107_35.xml
- Pais_20150107_37.xml
- Pais_20150107_38.xml
- Pais_20150107_41.xml
- Pais_20150107_42.xml
- Pais_20150107_55.xml
- Pais_20150107_56.xml
- Pais_20150107_57.xml
- Pais_20150107_61.xml
- Pais_20150107_63.xml
- Pais_20150107_66.xml
- Pais_20150107_67.xml
- Pais_20150107_71.xml
- Pais_20150107_75.xml
- Pais_20150107_76.xml

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Concordance Hits 1257

Hit KWIC

Line	Text	File
1	. Vicario y Serra, además, han propuesto al Gobierno de Catamarca (Argentina) la creación de u	Pais_20150107_15.xml
2	captura, según ha asegurado el delegado del Gobierno en la Comunidad Valenciana, Serafín Caste	Pais_20150107_30.xml
3	resultado de homicidio imprudente. El delegado del Gobierno ha hecho hincapié en que desde el	Pais_20150107_30.xml
4	dialogará con otros partidos para asegurar un Gobierno "fuerte". La formación invita a los dono	Pais_20150107_34.xml
5	el resto de formaciones para lograr un Gobierno municipal fuerte, amplio y plural en San	Pais_20150107_34.xml
6	presentar el mejor equipo y el mejor Gobierno cuando toque. Al respecto ha indicado que	Pais_20150107_34.xml
7	se ha agarrado a la responsabilidad de gobierno y a los datos de reducción del	Pais_20150107_42.xml
8	formación de rebajar la tensión en el Gobierno de coalición, el secretario de organizaci	Pais_20150107_55.xml
9	, Dolores López Gabarro, que dijo que el Gobierno autónomo no hace frente al desempleo. La	Pais_20150107_55.xml
10	abertzale exhibe hacia la última propuesta del Gobierno vasco sobre el desarme de ETA y	Pais_20150107_56.xml
11	daño injusto causado", además de reclamar del Gobierno español "la aplicación de una política pe	Pais_20150107_56.xml
12	roto el pacto que otorgaba estabilidad al Gobierno de CiU y amenaza con votar en	Pais_20150107_66.xml
13	sociales le han facilitado su acercamiento al Gobierno de Artur Mas y su conversión en	Pais_20150107_67.xml
14	se trata de un mero delegado: el Gobierno catalán lo convirtió desde el 23 de dicie	Pais_20150107_67.xml
15	un castigo tras varias polémicas con el Gobierno y medios conservadores españoles, uno de	Pais_20150107_67.xml
16	hay elecciones no serán para cambiar el gobierno sino para poder hacer la consulta definit	Pais_20150107_75.xml
17	epresidenta también ha lanzado críticas contra el Gobierno de Mariano Rajoy. Madrid no escucha. El	Pais_20150107_75.xml

Search Term Words Case Regex

Search Window Size 50

Gobierno

Start Stop Sort Advanced

Kwic Sort

Level 1 1R Level 2 2R Level 3 3R

Clone Results

CONCORDANCIAS

File Global Settings Tool Preferences Help

Corpus Files

- Pais_20151106_137.xml
- Pais_20151106_143.xml
- Pais_20151106_17.xml
- Pais_20151106_33.xml
- Pais_20151106_91.xml
- Pais_20151106_92.xml
- Pais_20151106_138.xml
- Pais_20150107_3.xml
- Pais_20150107_8.xml
- Pais_20150107_14.xml
- Pais_20150107_15.xml
- Pais_20150107_29.xml
- Pais_20150107_30.xml
- Pais_20150107_31.xml
- Pais_20150107_34.xml
- Pais_20150107_35.xml
- Pais_20150107_37.xml
- Pais_20150107_38.xml
- Pais_20150107_41.xml
- Pais_20150107_42.xml
- Pais_20150107_55.xml
- Pais_20150107_56.xml
- Pais_20150107_57.xml
- Pais_20150107_61.xml
- Pais_20150107_63.xml
- Pais_20150107_66.xml
- Pais_20150107_67.xml
- Pais_20150107_71.xml
- Pais_20150107_75.xml
- Pais_20150107_76.xml

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Concordance Hits FINISHED

Hit KWIC

Hit	KWIC	File
1	. Vicario y Serra, además, han propuesto al Gobierno de Catamarca (Argentina) la creación de u	Pais_20150107_15.xml
2	captura, según ha asegurado el delegado del Gobierno en la Comunidad Valenciana, Serafín Caste	Pais_20150107_30.xml
3	esultado de homicidio imprudente. El delegado del Gobierno ha hecho hincapié en que desde el	Pais_20150107_30.xml
4	dialogará con otros partidos para asegurar un Gobierno "fuerte". La formación invita a los dono	Pais_20150107_34.xml
5	el resto de formaciones para lograr un Gobierno municipal fuerte, amplio y plural en San	Pais_20150107_34.xml
6	presentar el mejor equipo y el mejor Gobierno cuando toque. Al respecto ha indicado que	Pais_20150107_34.xml
7	formación de rebajar la tensión en el Gobierno de coalición, el secretario de organizaci	Pais_20150107_55.xml
8	, Dolores López Gabarro, que dijo que el Gobierno autónomo no hace frente al desempleo. La	Pais_20150107_55.xml
9	abertzale exhibe hacia la última propuesta del Gobierno vasco sobre el desarme de ETA y	Pais_20150107_56.xml
10	daño injusto causado", además de reclamar del Gobierno español "la aplicación de una política pe	Pais_20150107_56.xml
11	roto el pacto que otorgaba estabilidad al Gobierno de CiU y amenaza con votar en	Pais_20150107_66.xml
12	sociales le han facilitado su acercamiento al Gobierno de Artur Mas y su conversión en	Pais_20150107_67.xml
13	se trata de un mero delegado: el Gobierno catalán lo convirtió desde el 23 de dicie	Pais_20150107_67.xml
14	un castigo tras varias polémicas con el Gobierno y medios conservadores españoles, uno de	Pais_20150107_67.xml
15	epresidenta también ha lanzado críticas contra el Gobierno de Mariano Rajoy. Madrid no escucha. El	Pais_20150107_75.xml
16	de Mariano Rajoy. Madrid no escucha. El Gobierno del Estado vuelve a dar la espalda	Pais_20150107_75.xml
17	Ertzaintza y el Departamento de Seguridad del Gobierno vasco. Mientras siguen las diferencias so	Pais_20150107_82.xml

Search Term Words Case Regex

Search Window Size 50

Gobierno

Kwic Sort

Level 1 1R Level 2 2R Level 3 3R

Clone Results

CONCORDANCIAS

File Global Settings Tool Preferences Help

Corpus Files

- Pais_20151106_137.xml
- Pais_20151106_143.xml
- Pais_20151106_17.xml
- Pais_20151106_33.xml
- Pais_20151106_91.xml
- Pais_20151106_92.xml
- Pais_20151106_138.xml
- Pais_20150107_3.xml
- Pais_20150107_8.xml
- Pais_20150107_14.xml
- Pais_20150107_15.xml
- Pais_20150107_29.xml
- Pais_20150107_30.xml
- Pais_20150107_31.xml
- Pais_20150107_34.xml
- Pais_20150107_35.xml
- Pais_20150107_37.xml
- Pais_20150107_38.xml
- Pais_20150107_41.xml
- Pais_20150107_42.xml
- Pais_20150107_55.xml
- Pais_20150107_56.xml
- Pais_20150107_57.xml
- Pais_20150107_61.xml
- Pais_20150107_63.xml
- Pais_20150107_66.xml
- Pais_20150107_67.xml
- Pais_20150107_71.xml
- Pais_20150107_75.xml
- Pais_20150107_76.xml

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Concordance Hits 71

Hit KWIC

1 Loyola, especialista na obra de Neruda e amigo do poeta, não exclui a possibilidade de
do poeta. Mas, conhecendo-o pessoalmente como amigo e companheiro de batalhas, acredito que hoje
2 pasar unos días en casa de un amigo en Murcia, el homicida volvió a su
3 quien aseguró que cuando pidió a su "amigo" Oriol Pujol que le facilitara contactos en
4 del Vallès, y él veía que su amigo del colegio pasaba las vacaciones en Premià
5 . Nos dice que en verano del 2014, un amigo vecino suyo y cliente de la peluquería
6 él estaba dispuesto a hacerlo, que su amigo lo proporcionaba todo, nos explica que final
7 hagan mucho daño a alguien", asegura un amigo de Antonio. El barrio de Horta regresaba
8 entonces cuando se interpuso entre ambos un amigo de la familia. El agresor, supuestamente, sa
9 y la Generalitat, que compartió con su amigo y colega, el sociólogo francés y especialist
10 imaginar me a un estudiante encontrando a un amigo en Huysmans y dedicando su vida a
11 gustado leerle. Creo que habría sido un amigo verdadero para mí. De modo que, después
12 las lágrimas de Luis recordando a su amigo. Y esa interpretación me evoca el perfil
13 tu voz y tu arte. Cohen es amigo familiar, Como es Cohen como ser humano?
14 Barcelona, jajajaja Fueras de broma, el gran amigo Leonardo, como le llamaba mi padre, no
15 carta que nos mandó al partir su amigo... Creo que esta publicada, la leyó en
16 Ford. Eso hicieron Shadee Griugius y su amigo Shervan Delban, estudiantes en la ciudad pla
17

Search Term Words Case Regex

Search Window Size 50

Kwic Sort

Level 1 1R Level 2 2R Level 3 3R

Clone Results

CONCORDANCIAS

File Global Settings Tool Preferences Help

Corpus Files

- Pais_20151106_137.xml
- Pais_20151106_143.xml
- Pais_20151106_17.xml
- Pais_20151106_33.xml
- Pais_20151106_91.xml
- Pais_20151106_92.xml
- Pais_20151106_138.xml
- Pais_20150107_3.xml
- Pais_20150107_8.xml
- Pais_20150107_14.xml
- Pais_20150107_15.xml
- Pais_20150107_29.xml
- Pais_20150107_30.xml
- Pais_20150107_31.xml
- Pais_20150107_34.xml
- Pais_20150107_35.xml
- Pais_20150107_37.xml
- Pais_20150107_38.xml
- Pais_20150107_41.xml
- Pais_20150107_42.xml
- Pais_20150107_55.xml
- Pais_20150107_56.xml
- Pais_20150107_57.xml
- Pais_20150107_61.xml
- Pais_20150107_63.xml
- Pais_20150107_66.xml
- Pais_20150107_67.xml
- Pais_20150107_71.xml
- Pais_20150107_75.xml
- Pais_20150107_76.xml

Total No. 1498
Files Processed

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Concordance Hits 150 Hit KWIC

1 Loyola, especialista na obra de Neruda e amigo do poeta, não exclui a possibilidade de
do poeta. Mas, conhecendo-o pessoalmente como amigo e companheiro de batalhas, acredito que hoje
2 con el mundo de la montaña, familiares, amigos, compañeros de trabajo y medios de comunica
3 . La obra retrata a un grupo de amigos —Jacob Torres, Martina Tresserra, Miriam To
4 la fiesta de bienvenida improvisada por los amigos se complica cuando uno de ellos decide
5 Beth. Su personaje intentó suicidarse y sus amigos no saben cómo comportarse ante ella, lo
6 encuentro, el chico lo confiesa a sus amigos y pide que le cubran. A partir
7 tendría inconveniente en reconocerme como "grandes amigos" -así se ha dicho- pero solo hace
8 pasar unos días en casa de un amigo en Murcia, el homicida volvió a su
9 acto a movilizar a sus vecinos, hermanos, amigos. Llegamos diciendo que la hacía falta que
10 quien aseguró que cuando pidió a su "amigo" Oriol Pujol que le facilitara contactos en
11 se las comió y ofrecieron a sus amigos. El escritor, mitómano, en Sicilia, escribió
12 del Vallès, y él veía que su amigo del colegio pasaba las vacaciones en Premià
13 . Nos dice que en verano del 2014, un amigo vecino suyo y cliente de la peluquería
14 él estaba dispuesto a hacerlo, que su amigo lo proporcionaba todo, nos explica que final
15 hagan mucho daño a alguien", asegura un amigo de Antonio. El barrio de Horta regresaba
16 Camino en bicicleta con un grupo de amigos desde Campo de Criptana (Ciudad Real) en
17

File

- Pais_20151106_143.xml
- Pais_20151106_143.xml
- Pais_20150107_15.xml
- Pais_20150206_6.xml
- Pais_20150206_6.xml
- Pais_20150206_6.xml
- Pais_20150206_6.xml
- Pais_20150206_147.xml
- Pais_20150306_38.xml
- Pais_20150306_71.xml
- Pais_20150407_30.xml
- Pais_20150407_121.xml
- Pais_20150407_129.xml
- Pais_20150608_88.xml
- Pais_20150608_88.xml
- Pais_20150608_119.xml
- Pais_20150707_5.xml

Search Term Words Case Regex Search Window Size
amigo Advanced 50 ▲▼

Start Stop Sort

Kwic Sort Level 1 1R ▲▼ Level 2 2R ▲▼ Level 3 3R ▲▼

Clone Results

CONCORDANCIAS

File Global Settings Tool Preferences Help

Corpus Files

- Pais_20151106_137.xml
- Pais_20151106_143.xml
- Pais_20151106_17.xml
- Pais_20151106_33.xml
- Pais_20151106_91.xml
- Pais_20151106_92.xml
- Pais_20151106_138.xml
- Pais_20150107_3.xml
- Pais_20150107_8.xml
- Pais_20150107_14.xml
- Pais_20150107_15.xml
- Pais_20150107_29.xml
- Pais_20150107_30.xml
- Pais_20150107_31.xml
- Pais_20150107_34.xml
- Pais_20150107_35.xml
- Pais_20150107_37.xml
- Pais_20150107_38.xml
- Pais_20150107_41.xml
- Pais_20150107_42.xml
- Pais_20150107_55.xml
- Pais_20150107_56.xml
- Pais_20150107_57.xml
- Pais_20150107_61.xml
- Pais_20150107_63.xml
- Pais_20150107_66.xml
- Pais_20150107_67.xml
- Pais_20150107_71.xml
- Pais_20150107_75.xml
- Pais_20150107_76.xml

Total No. 1498
Files Processed

Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Concordance Hits 174

Hit KWIC File

Hit	KWIC	File
158	Granados hace cuatro años. "Nunca fue mi amigo ni una persona de mi confianza", señaló.	Pais_20150507_23.xml
159	candidatas y candidatos. Pero ahí está su amigo Carlos Slim, ha expresado Lara, quien además	Pais_20150608_52.xml
160	pareja, mientras que el 24% lo hace con amigos. Para ver la tele, igual que para	Pais_20150707_35.xml
161	lleva nueve meses preso junto a su amigo Francisco Granados, exconsejero autonómico y	Pais_20150707_91.xml
162	primeros gais en esta institución, tenemos muchos amigos que también lo son dentro del Cuerpo,	Pais_20150907_122.xml
163	que les facilitaba el Gobierno sueco, tenían amigos y proyectos de futuro. Estaban en definitiv	Pais_20150907_135.xml
164	menudo, reciben llamadas de Suecia. Son los amigos suecos de las gemelas de 14 años, que	Pais_20150907_135.xml
165	largo de la jornada tres madres de amigas; una vecina y cinco docentes en la	Pais_20151007_101.xml
166	se los dio a tu madre?". "Una amiga en el portal de casa". Otra docente	Pais_20151007_101.xml
167	pequeña dijo en otro momento que esa amiga era "médico". El padre, aquel día, aseguró	Pais_20151007_101.xml
168	julio. Mientras tanto, la pediatra de Asunta, amiga de la familia, ha descrito a la	Pais_20151007_101.xml
169	abrazaron. Esta mañana, la progenitora de una amiga de Asunta ha contado cómo la niña,	Pais_20151007_101.xml
170	, vive como lo hacen aquí y haz amigos que sean de aquí". Javi invita a	Pais_20150507_7.xml
171	y difusa. En Facebook te comunicas con amigos y familiares. Tienes que confiar en nosotros	Pais_20150807_4.xml
172	del encuentro. El año pasado, un viejo amigo e inversor de la casa, Bono y	Pais_20150907_2.xml
173	. En lo que sí me ayudó un amigo fue en la grabación de la voz	Pais_20151106_1.xml
174	, pero estoy seguro de que hay manos amigas en todos los lados", apunta. El alemán	Pais_20151106_2.xml

Search Term Words Case Regex Search Window Size amig 50

Start Stop Sort

Kwic Sort Level 1 1R Level 2 2R Level 3 3R

Clone Results

CONCORDANCIAS

File Global Settings Tool Preferences Help

Corpus Files

- Pais_20151106_137.xml
- Pais_20151106_143.xml
- Pais_20151106_17.xml
- Pais_20151106_33.xml
- Pais_20151106_91.xml
- Pais_20151106_92.xml
- Pais_20151106_138.xml
- Pais_20150107_3.xml
- Pais_20150107_8.xml
- Pais_20150107_14.xml
- Pais_20150107_15.xml
- Pais_20150107_29.xml
- Pais_20150107_30.xml
- Pais_20150107_31.xml
- Pais_20150107_34.xml
- Pais_20150107_35.xml
- Pais_20150107_37.xml
- Pais_20150107_38.xml
- Pais_20150107_41.xml
- Pais_20150107_42.xml
- Pais_20150107_55.xml
- Pais_20150107_56.xml
- Pais_20150107_57.xml
- Pais_20150107_61.xml
- Pais_20150107_63.xml
- Pais_20150107_66.xml
- Pais_20150107_67.xml
- Pais_20150107_71.xml
- Pais_20150107_75.xml
- Pais_20150107_76.xml

Total No. 1498

Files Processed

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Concordance Hits 174

Hit KWIC

Hit	KWIC	File
158	Granados hace cuatro años. "Nunca fue mi amigo ni una persona de mi confianza", señaló.	Pais_20150507_23.xml
159	candidatas y candidatos. Pero ahí está su amigo Carlos Slim, ha expresado Lara, quien además	Pais_20150608_52.xml
160	pareja, mientras que el 24% lo hace con amigos. Para ver la tele, igual que para	Pais_20150707_35.xml
161	lleva nueve meses preso junto a su amigo Francisco Granados, exconsejero autonómico y	Pais_20150707_91.xml
162	primeros gais en esta institución, tenemos muchos amigos que también lo son dentro del Cuerpo,	Pais_20150907_122.xml
163	que les facilitaba el Gobierno sueco, tenían amigos y proyectos de futuro. Estaban en definitiv	Pais_20150907_135.xml
164	menudo, reciben llamadas de Suecia. Son los amigos suecos de las gemelas de 14 años, que	Pais_20150907_135.xml
165	largo de la jornada tres madres de amigas; una vecina y cinco docentes en la	Pais_20151007_101.xml
166	se los dio a tu madre?". "Una amiga en el portal de casa". Otra docente	Pais_20151007_101.xml
167	pequeña dijo en otro momento que esa amiga era "médico". El padre, aquel día, aseguró	Pais_20151007_101.xml
168	julio. Mientras tanto, la pediatra de Asunta, amiga de la familia, ha descrito a la	Pais_20151007_101.xml
169	abrazaron. Esta mañana, la progenitora de una amiga de Asunta ha contado cómo la niña,	Pais_20151007_101.xml
170	vive como lo hacen aquí y haz amigos que sean de aquí". Javi invita a	Pais_20150507_7.xml
171	y difusa. En Facebook te comunicas con amigos y familiares. Tienes que confiar en nosotros	Pais_20150807_4.xml
172	del encuentro. El año pasado, un viejo amigo e inversor de la casa, Bono y	Pais_20150907_2.xml
173	. En lo que sí me ayudó un amigo fue en la grabación de la voz	Pais_20151106_1.xml
174	, pero estoy seguro de que hay manos amigas en todos los lados", apunta. El alemán	Pais_20151106_2.xml

Search Term Words Case Regex Advanced 50

Start Stop Sort

Kwic Sort Level 1 1R Level 2 2R Level 3 3R

Clone Results

CONCORDANCIAS

File Global Settings Tool Preferences Help

Corpus Files

Pais_20151106_137.xml	^	
Pais_20151106_143.xml		
Pais_20151106_17.xml		
Pais_20151106_33.xml		
Pais_20151106_91.xml		
Pais_20151106_92.xml		
Pais_20151106_138.xml		
Pais_20150107_3.xml		
Pais_20150107_8.xml		
Pais_20150107_14.xml		
Pais_20150107_15.xml		
Pais_20150107_29.xml		
Pais_20150107_30.xml		
Pais_20150107_31.xml		
Pais_20150107_34.xml		
Pais_20150107_35.xml		
Pais_20150107_37.xml		
Pais_20150107_38.xml		
Pais_20150107_41.xml		
Pais_20150107_42.xml		
Pais_20150107_55.xml		
Pais_20150107_56.xml		
Pais_20150107_57.xml		
Pais_20150107_61.xml		
Pais_20150107_63.xml		
Pais_20150107_66.xml		
Pais_20150107_67.xml		
Pais_20150107_71.xml		
Pais_20150107_75.xml		
Pais_20150107_76.xml	▼	

Total No. 1498

Files Processed

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Concordance Hits 87

Hit KWIC

File

71 de Masterchef, y cuál es tu mejor amigo de los concursantes, y ya por último,
72 . Carvalho, conocido en la zona como Amigão (amigo), hacía denuncias frecuentes de casos de cor
73 . Espinosa mandó el último mensaje a un amigo a las 14,13, por lo que el robo,
74 vecino, Víctor, llamaba llorando a un arquitecto amigo para pedirle ayuda. Cada vez se escuchan
75 de corrupción y narcotráfico, entre ellos su amigo el coronel Antonio de la Guardia, y
76) a Canarias, supuestamente para visitar a una amiga íntima, aunque el presidente extremeño ha de
77 en la que navegaban volcó. Fue un amigo el que, desde un teléfono móvil desde
78 el extranjero cuenta de Granados y su amigo y socio David Marjaliza (ambos presos) que
79 Granados hace cuatro años. "Nunca fue mi amigo ni una persona de mi confianza", señaló.
80 candidatas y candidatos. Pero ahí está su amigo Carlos Slim, ha expresado Lara, quien además
81 lleva nueve meses preso junto a su amigo Francisco Granados, exconsejero autonómico y
82 se los dio a tu madre?". "Una amiga en el portal de casa". Otra docente
83 pequeña dijo en otro momento que esa amiga era "médico". El padre, aquel día, aseguró
84 julio. Mientras tanto, la pediatra de Asunta, amiga de la familia, ha descrito a la
85 abrazaron. Esta mañana, la progenitora de una amiga de Asunta ha contado cómo la niña,
86 del encuentro. El año pasado, un viejo amigo e inversor de la casa, Bono y
87 . En lo que sí me ayudó un amigo fue en la grabación de la voz

< > < >

Search Term Words Case Regex

Search Window Size 50

amig+ Advanced

Start Stop Sort

Kwic Sort

Level 1 1R Level 2 2R Level 3 3R

Clone Results

CONCORDANCIAS

File Global Settings Tool Preferences Help

Corpus Files

- Pais_20151106_137.xml
- Pais_20151106_143.xml
- Pais_20151106_17.xml
- Pais_20151106_33.xml
- Pais_20151106_91.xml
- Pais_20151106_92.xml
- Pais_20151106_138.xml
- Pais_20150107_3.xml
- Pais_20150107_8.xml
- Pais_20150107_14.xml
- Pais_20150107_15.xml
- Pais_20150107_29.xml
- Pais_20150107_30.xml
- Pais_20150107_31.xml
- Pais_20150107_34.xml
- Pais_20150107_35.xml
- Pais_20150107_37.xml
- Pais_20150107_38.xml
- Pais_20150107_41.xml
- Pais_20150107_42.xml
- Pais_20150107_55.xml
- Pais_20150107_56.xml
- Pais_20150107_57.xml
- Pais_20150107_61.xml
- Pais_20150107_63.xml
- Pais_20150107_66.xml
- Pais_20150107_67.xml
- Pais_20150107_71.xml
- Pais_20150107_75.xml
- Pais_20150107_76.xml

Total No. 1498
Files Processed

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Concordance Hits 87

Hit KWIC

Hit	KWIC	File
17	Barcelona, jajaja Fueras de broma, el gran amigo Leonardo, como le llamaba mi padre, no	Pais_20150206_86.xml
18	piloto retirado de 56 años y (supuesto) gran amigo de Ahmad Shah, a The New York	Pais_20150306_143.xml
19	el rey saliente. "Hermida era un gran amigo , una gran persona , un magnífico periodista y	Pais_20150507_2.xml
20	se parece más esta? Un saludo! Hola amiga , me encanta tu nombre. Ya verás como	Pais_20150407_58.xml
21	femenino ahí estaré yo, sea mi interlocutor amigo o enemigo ". El presidente de la Liga	Pais_20151106_135.xml
22	amigos han aportado ninguna pista. "Tu mejor amigo puede albergar los secretos más oscuros", de	Pais_20150306_143.xml
23	Masterchef, y cuál es tu mejor amigo de los concursantes, y ya por último,	Pais_20150707_108.xml
24	Granados hace cuatro años. "Nunca fue mi amigo ni una persona de mi confianza ", señaló.	Pais_20150507_23.xml
25	leal, su brújula y su guía . Nuestra amiga del alma —de mi mujer, Chelo, y	Pais_20150407_115.xml
26	contra su régimen y añadía: Mi querido amigo Sarkozy sufre un desorden mental. Fui yo	Pais_20150306_27.xml
27	lombianos, acaba de colaborar con Alejandro Sanz, amigo de Shakira (ambos cantaron juntos La tortura	Pais_20150608_70.xml
28	todo, un gran español ". ¿Sobre todo? Ser amigo , buena persona , excelente periodista, está b	Pais_20150507_2.xml
29	baires y decir chorraditas. Si acabé siendo amigo de skinheads en mi adolescencia era porque	Pais_20150608_6.xml
30	empresa pública de América Latina a su amiga y antigua colaboradora Graça Foster. La inte	Pais_20150206_125.xml
31	las lágrimas de Luis recordando a su amigo . Y esa interpretación me evoca el perfil	Pais_20150107_54.xml
32	I lleva nueve meses preso junto a su amigo Francisco Granados, exconsejero autonómico y	Pais_20150707_91.xml
33	inédita, que Christian Boltanski dedica a su amigo Mario Merz, fallecido en 2003, uno de los	Pais_20151106_117.xml

Search Term Words Case Regex

Search Window Size

amig+ Advanced 50

Start Stop Sort

Kwic Sort

Level 1 1L Level 2 2L Level 3 2R

Clone Results

GRÁFICO DE CONCORDANCIAS

File Global Settings Tool Preferences Help

Corpus Files

- Pais_20151106_137.xml
- Pais_20151106_143.xml
- Pais_20151106_17.xml
- Pais_20151106_33.xml
- Pais_20151106_91.xml
- Pais_20151106_92.xml
- Pais_20151106_138.xml
- Pais_20150107_3.xml
- Pais_20150107_8.xml
- Pais_20150107_14.xml
- Pais_20150107_15.xml
- Pais_20150107_29.xml
- Pais_20150107_30.xml
- Pais_20150107_31.xml
- Pais_20150107_34.xml
- Pais_20150107_35.xml
- Pais_20150107_37.xml
- Pais_20150107_38.xml
- Pais_20150107_41.xml
- Pais_20150107_42.xml
- Pais_20150107_55.xml
- Pais_20150107_56.xml
- Pais_20150107_57.xml
- Pais_20150107_61.xml
- Pais_20150107_63.xml
- Pais_20150107_66.xml
- Pais_20150107_67.xml
- Pais_20150107_71.xml
- Pais_20150107_75.xml
- Pais_20150107_76.xml

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Concordance Hits 14 Total Plots 14

HIT FILE: 39 FILE: Pais_20150206_15.xml

--	--

No. of Hits = 1
File Length (in chars) = 3670

HIT FILE: 44 FILE: Pais_20150206_31.xml

--	--

No. of Hits = 1
File Length (in chars) = 5752

HIT FILE: 74 FILE: Pais_20150206_90.xml

--	--

No. of Hits = 1
File Length (in chars) = 4088

HIT FILE: 98 FILE: Pais_20150306_12.xml

--	--

No. of Hits = 1
File Length (in chars) = 8858

HIT FILE: 121 FILE: Pais_20150306_82.xml

--	--

No. of Hits = 1
File Length (in chars) = 5192

HIT FILE: 138 FILE: Pais_20150306_125.xml

--	--

No. of Hits = 1
File Length (in chars) = 2979

HIT FILE: 389 FILE: Pais_20150206_96.xml

--	--

No. of Hits = 1
File Length (in chars) = 3188

Total No. 1498
Files Processed

Search Term Words Case Regex

por último

COAPARICIONES

Palabras que aparecen junto al término buscado.

Podemos decidir qué colocaciones nos interesan: solo las de la izquierda, solo las que están a una distancia máxima de 2, etc.

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword					
Total No. of Collocate Types: 185			Total No. of Collocate Tokens: 392		
Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate
1	28	28	0	12.07359	angela
2	27	7	20	0	y
3	16	6	10	0	que
4	16	14	2	0	de
5	12	6	6	0	la
6	12	12	0	9.92782	canciller
7	11	4	7	9.49618	hollande
8	11	0	11	0	el
9	9	2	7	0	en
10	8	8	0	9.53925	alemana
11	6	1	5	0	ha
12	6	4	2	0	a
13	5	4	1	0	con
14	5	2	3	6.34811	ayer
15	4	2	2	0	un
16	4	0	4	0	también
17	4	0	4	0	se
18	4	2	2	0	pero

Search Term Words Case Regex
 Advanced Start Stop Sort
Min. Collocate Frequency

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword					
Total No. of Collocate Types: 63			Total No. of Collocate Tokens: 178		
Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate
1	69	69	0	11.10493	mariano
2	20	20	0	0	de
3	12	12	0	0	a
4	6	6	0	0	que
5	4	4	0	0	con
6	3	3	0	0	y
7	3	3	0	9.38872	casqueiro
8	2	2	0	2.44781	presidente
9	2	2	0	7.56275	prepara
10	2	2	0	2.12364	país
11	2	2	0	2.86231	ni
12	2	2	0	8.94126	cué
13	1	1	0	3.35630	última
14	1	1	0	6.94126	zapatero
15	1	1	0	10.26319	wozniak
16	1	1	0	10.26319	viúdez
17	1	1	0	0	uno
18	1	1	0	5.03437	triunfo

Search Term Words Case Regex
 Advanced Start Stop Sort
Min. Collocate Frequency

"Fuerza" de la colocación

Número de palabras a la derecha y a la izquierda

COAPARICIONES

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword					
Total No. of Collocate Types: 51			Total No. of Collocate Tokens: 98		
Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate
1	1	1	0	12.12421	terció
2	1	1	0	12.12421	merkelstreichelt
3	1	1	0	12.12421	malentendido
4	1	1	0	12.12421	cerrarlas
5	28	28	0	12.07359	angela
6	1	1	0	11.12421	tubella
7	1	1	0	11.12421	trucados
8	1	1	0	11.12421	csu
9	1	1	0	10.53925	elmau
10	1	1	0	9.80229	contrapartida
11	1	1	0	9.80229	aseverado
12	1	1	0	9.53925	anti
13	1	1	0	8.53925	discusiones
14	1	1	0	8.42377	fotógrafos
15	1	1	0	8.31686	doncel
16	2	2	0	8.26623	subrayó
17	1	1	0	8.21732	müller
18	1	1	0	8.21732	amplia

Se puede ordenar por frecuencia o por “fuerza” de la colocación

Search Term Words Case Regex

Window Span Same
From... 1L To... 0

Min. Collocate Frequency

Sort by Invert Order

Sort by Stat

Start Stop Sort

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword					
Total No. of Collocate Types: 63			Total No. of Collocate Tokens: 178		
Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate
1	1	1	0	11.26319	sovint
2	1	1	0	11.26319	epicentros
3	1	1	0	11.26319	demoscópica
4	1	1	0	11.26319	almuñécar
5	69	69	0	11.10493	mariano
6	1	1	0	10.26319	wozniak
7	1	1	0	10.26319	viúdez
8	1	1	0	9.67823	báñez
9	3	3	0	9.38872	casqueiro
10	2	2	0	8.94126	cué
11	1	1	0	7.67823	manresa
12	2	2	0	7.56275	prepara
13	1	1	0	7.56275	catalá
14	1	1	0	6.94126	zapatero
15	1	1	0	6.73963	mente
16	1	1	0	6.40521	moncloa
17	1	1	0	5.97779	escena
18	1	1	0	5.67823	catalanes

Search Term Words Case Regex

Window Span Same
From... 1L To... 0

Min. Collocate Frequency

Sort by Invert Order

Sort by Stat

Start Stop Sort

LISTA DE PALABRAS

Rank	Freq	Word	Rank	Freq	Word	Rank	Freq	Word	Rank	Freq	Word
1	800	xd	19	200	va	37	122	vaya	55	102	madre
2	760	uk	20	187	gracias	38	118	quiero	56	101	jo
3	477	pues	21	185	asi	39	117	cosas	57	97	xa
4	456	jajaja	22	185	ni	40	116	vale	58	95	da
5	404	bueno	23	184	xo	41	115	viene	59	94	año
6	393	he	24	182	jajaj	42	114	otro	60	92	asiq
7	382	voy	25	169	jajajaja	43	113	jajajaj	61	92	espero
8	347	tengo	26	157	hora	44	113	sitio	62	91	digo
9	328	ver	27	153	hacer	45	112	finde	63	91	quedamos
10	298	estoy	28	152	mañana	46	111	dia	64	90	pa
11	293	jeje	29	147	jaja	47	110	curro	65	90	viernes
12	279	creo	30	141	era	48	110	entonces	66	88	salgo
13	270	algo	31	141	mas	49	109	vas	67	87	felicidades
14	261	ir	32	140	guay	50	106	aqui	68	87	gente
15	251	ok	33	137	quedar	51	106	tiempo	69	86	hecho
16	246	casa	34	127	alguien	52	105	has	70	85	cn
17	226	vamos	35	126	xq	53	103	domingo	71	85	trabajo
18	209	puedo	36	125	semana	54	102	iqual	72	84	madrid

LISTA DE PALABRAS

Rank	Freq	Word	Rank	Freq	Word	Rank	Freq	Word	Rank	Freq	Word
1	800	xd	19	200	va	37	122	vaya	55	102	madre
2	760	uk	20	187	gracias	38	118	quiero	56	101	jo
3	477	pues	21	185	asi	39	117	cosas	57	97	xa
4	456	jajaja	22	185	ni	40	116	vale	58	95	da
5	404	bueno	23	184	xo	41	115	viene	59	94	año
6	393	he	24	182	jajaj	42	114	otro	60	92	asiq
7	382	voy	25	169	jajajaja	43	113	jajajaj	61	92	espero
8	347	tengo	26	157	hora	44	113	sitio	62	91	digo
9	328	ver	27	153	hacer	45	112	finde	63	91	quedamos
10	298	estoy	28	152	mañana	46	111	dia	64	90	pa
11	293	jeje	29	147	jaja	47	110	curro	65	90	viernes
12	279	creo	30	141	era	48	110	entonces	66	88	salgo
13	270	algo	31	141	mas	49	109	vas	67	87	felicidades
14	261	ir	32	140	guay	50	106	aqui	68	87	gente
15	251	ok	33	137	quedar	51	106	tiempo	69	86	hecho
16	246	casa	34	127	alguien	52	105	has	70	85	cn
17	226	vamos	35	126	xq	53	103	domingo	71	85	trabajo
18	209	puedo	36	125	semana	54	102	iqual	72	84	madrid

Ortografía SMS,
abreviaturas,
acortamientos

LISTA DE PALABRAS

Rank	Freq	Word	Rank	Freq	Word	Rank	Freq	Word	Rank	Freq	Word
1	800	xd	19	200	va	37	122	vaya	55	102	madre
2	760	uk	20	187	gracias	38	118	quiero	56	101	jo
3	477	pues	21	185	asi	39	117	cosas	57	97	xa
4	456	jajaja	22	185	ni	40	116	vale	58	95	da
5	404	bueno	23	184	xo	41	115	viene	59	94	año
6	393	he	24	182	jajaj	42	114	otro	60	92	asiq
7	382	voy	25	169	jajajaja	43	113	jajajaj	61	92	espero
8	347	tengo	26	157	hora	44	113	sitio	62	91	digo
9	328	ver	27	153	hacer	45	112	finde	63	91	quedamos
10	298	estoy	28	152	mañana	46	111	dia	64	90	pa
11	293	jeje	29	147	jaja	47	110	curro	65	90	viernes
12	279	creo	30	141	era	48	110	entonces	66	88	salgo
13	270	algo	31	141	mas	49	109	vas	67	87	felicidades
14	261	ir	32	140	guay	50	106	aqui	68	87	gente
15	251	ok	33	137	quedar	51	106	tiempo	69	86	hecho
16	246	casa	34	127	alguien	52	105	has	70	85	cn
17	226	vamos	35	126	xq	53	103	domingo	71	85	trabajo
18	209	puedo	36	125	semana	54	102	iqual	72	84	madrid

Ortografía SMS,
abreviaturas,
acortamientos

Risas

LISTA DE PALABRAS

Rank	Freq	Word	Rank	Freq	Word	Rank	Freq	Word	Rank	Freq	Word
1	800	xd	19	200	va	37	122	vaya	55	102	madre
2	760	uk	20	187	gracias	38	118	quiero	56	101	jo
3	477	pues	21	185	asi	39	117	cosas	57	97	xa
4	456	jajaja	22	185	ni	40	116	vale	58	95	da
5	404	bueno	23	184	xo	41	115	viene	59	94	año
6	393	he	24	182	jajaj	42	114	otro	60	92	asiq
7	382	voy	25	169	jajajaja	43	113	jajajaj	61	92	espero
8	347	tengo	26	157	hora	44	113	sitio	62	91	digo
9	328	ver	27	153	hacer	45	112	finde	63	91	quedamos
10	298	estoy	28	152	mañana	46	111	dia	64	90	pa
11	293	jeje	29	147	jaja	47	110	curro	65	90	viernes
12	279	creo	30	141	era	48	110	entonces	66	88	salgo
13	270	algo	31	141	mas	49	109	vas	67	87	felicidades
14	261	ir	32	140	guay	50	106	aqui	68	87	gente
15	251	ok	33	137	quedar	51	106	tiempo	69	86	hecho
16	246	casa	34	127	alguien	52	105	has	70	85	cn
17	226	vamos	35	126	xq	53	103	domingo	71	85	trabajo
18	209	puedo	36	125	semana	54	102	iqual	72	84	madrid

Ortografía SMS,
abreviaturas,
acortamientos

Risas

Muletillas

LISTA DE PALABRAS

Rank	Freq	Word	Rank	Freq	Word	Rank	Freq	Word	Rank	Freq	Word
1	800	xd	19	200	va	37	122	vaya	55	102	madre
2	760	uk	20	187	gracias	38	118	quiero	56	101	jo
3	477	pues	21	185	asi	39	117	cosas	57	97	xa
4	456	jajaja	22	185	ni	40	116	vale	58	95	da
5	404	bueno	23	184	xo	41	115	viene	59	94	año
6	393	he	24	182	jajaj	42	114	otro	60	92	asiq
7	382	voy	25	169	jajajaja	43	113	jajajaj	61	92	espero
8	347	tengo	26	157	hora	44	113	sitio	62	91	digo
9	328	ver	27	153	hacer	45	112	finde	63	91	quedamos
10	298	estoy	28	152	mañana	46	111	dia	64	90	pa
11	293	jeje	29	147	jaja	47	110	curro	65	90	viernes
12	279	creo	30	141	era	48	110	entonces	66	88	salgo
13	270	algo	31	141	mas	49	109	vas	67	87	felicidades
14	261	ir	32	140	guay	50	106	aqui	68	87	gente
15	251	ok	33	137	quedar	51	106	tiempo	69	86	hecho
16	246	casa	34	127	alguien	52	105	has	70	85	cn
17	226	vamos	35	126	xq	53	103	domingo	71	85	trabajo
18	209	puedo	36	125	semana	54	102	iqual	72	84	madrid

Ortografía SMS,
abreviaturas,
acortamientos

Risas

Muletillas

Verbos en 1^a
persona (sobre
todo en sing.)

LISTA DE PALABRAS

Rank	Freq	Word	Rank	Freq	Word	Rank	Freq	Word	Rank	Freq	Word
1	800	xd	19	200	va	37	122	vaya	55	102	madre
2	760	uk	20	187	gracias	38	118	quiero	56	101	jo
3	477	pues	21	185	asi	39	117	cosas	57	97	xa
4	456	jajaja	22	185	ni	40	116	vale	58	95	da
5	404	bueno	23	184	xo	41	115	viene	59	94	año
6	393	he	24	182	jajaj	42	114	otro	60	92	asiq
7	382	voy	25	169	jajajaja	43	113	jajajaj	61	92	espero
8	347	tengo	26	157	hora	44	113	sitio	62	91	digo
9	328	ver	27	153	hacer	45	112	finde	63	91	quedamos
10	298	estoy	28	152	mañana	46	111	dia	64	90	pa
11	293	jeje	29	147	jaja	47	110	curro	65	90	viernes
12	279	creo	30	141	era	48	110	entonces	66	88	salgo
13	270	algo	31	141	mas	49	109	vas	67	87	felicidades
14	261	ir	32	140	guay	50	106	aqui	68	87	gente
15	251	ok	33	137	quedar	51	106	tiempo	69	86	hecho
16	246	casa	34	127	alguien	52	105	has	70	85	cn
17	226	vamos	35	126	xq	53	103	domingo	71	85	trabajo
18	209	puedo	36	125	semana	54	102	iqual	72	84	madrid

Ortografía SMS,
abreviaturas,
acortamientos

Risas

Muletillas

Verbos en 1ª
persona (sobre
todo en sing.)

Adverbios y
nombres de
tiempo y lugar

GRÁFICO DE CONCORDANCIAS

Concordance Hits 2606 Total Plots 1

HIT FILE: 1 FILE: LOS DE SIEMPRE.txt



No. of Hits = 2606
File Length (in chars) = 885659

Concordance Hits 1389 Total Plots 1

HIT FILE: 1 FILE: LOS DE SIEMPRE.txt



No. of Hits = 1389
File Length (in chars) = 885659

Concordance Hits 2643 Total Plots 1

HIT FILE: 1 FILE: LOS DE SIEMPRE.txt



No. of Hits = 2643
File Length (in chars) = 885659

Concordance Hits 1806 Total Plots 1

HIT FILE: 1 FILE: LOS DE SIEMPRE.txt



No. of Hits = 1806
File Length (in chars) = 885659

Concordance Hits 504 Total Plots 1

HIT FILE: 1 FILE: LOS DE SIEMPRE.txt



No. of Hits = 504
File Length (in chars) = 885659

Concordance Hits 4428 Total Plots 1

HIT FILE: 1 FILE: LOS DE SIEMPRE.txt



No. of Hits = 4428
File Length (in chars) = 885659

Concordance Hits 3013 Total Plots 1

HIT FILE: 1 FILE: LOS DE SIEMPRE.txt



No. of Hits = 3013
File Length (in chars) = 885659

Concordance Hits 9 Total Plots 1

HIT FILE: 1 FILE: LOS DE SIEMPRE.txt



No. of Hits = 9
File Length (in chars) = 885659

Concordance Hits 42 Total Plots 1

HIT FILE: 1 FILE: LOS DE SIEMPRE.txt



No. of Hits = 42
File Length (in chars) = 885659

Concordance Hits 14 Total Plots 1

HIT FILE: 1 FILE: LOS DE SIEMPRE.txt



No. of Hits = 14
File Length (in chars) = 885659

Concordance Hits 11 Total Plots 1

HIT FILE: 1 FILE: LOS DE SIEMPRE.txt



No. of Hits = 11
File Length (in chars) = 885659

Concordance Hits 37 Total Plots 1

HIT FILE: 1 FILE: LOS DE SIEMPRE.txt



No. of Hits = 37
File Length (in chars) = 885659

Miembros
del grupo

Otros amigos,
parejas, exparejas...

PROYECTOS DE LINGÜÍSTICA DE CORPUS

FRIENDS



Teoría de las «formas de las historias» (shapes of stories) de Kurt Vonnegut



SYUZHET

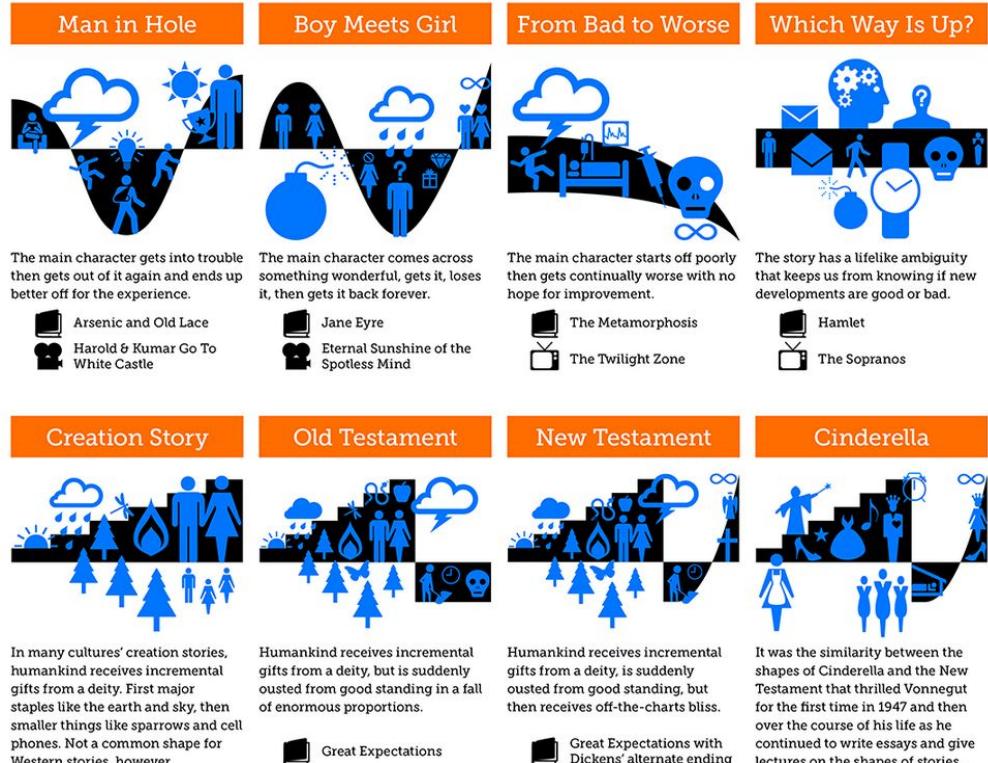


Imagen sacada de: Gizmodo

SYUZHET

syuzhet vs. fabula

La *fabula* sigue el orden cronológico, real (dentro de la ficción) de los acontecimientos

Star Wars: I, II, III, III y medio (Rogue One), IV, V, VI, VII, VIII, IX

El *syuzhet* es el orden narrativo de los acontecimientos, el orden en que se nos presentan

Star Wars: IV, V, VI, I, II, III, VII, III y medio (Rogue One), VIII? IX?

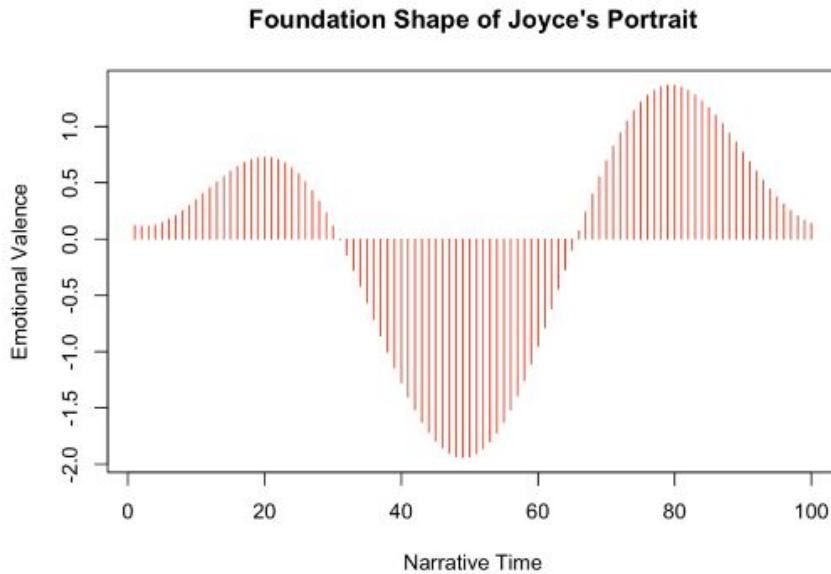
SYUZHET

4 lexicones en inglés de análisis de opinión o *sentiment analysis*:

- Syuzhet lexicon ([Matthew L. Jockers](#))
- AFINN Word Database ([Finn Arup Nielsen](#))
- Opinion Lexicon ([Minqing Hu and Bing Liu](#))
- **NRC Emotion Lexicon** ([Saif M. Mohammad y Peter D. Turney](#))

Para cada palabra devuelve una lista de 10 valores numéricos:

- 8 emociones (ira, ilusión, asco, miedo, alegría, tristeza, sorpresa, confianza)
- 2 sentimientos (negativo o positivo)

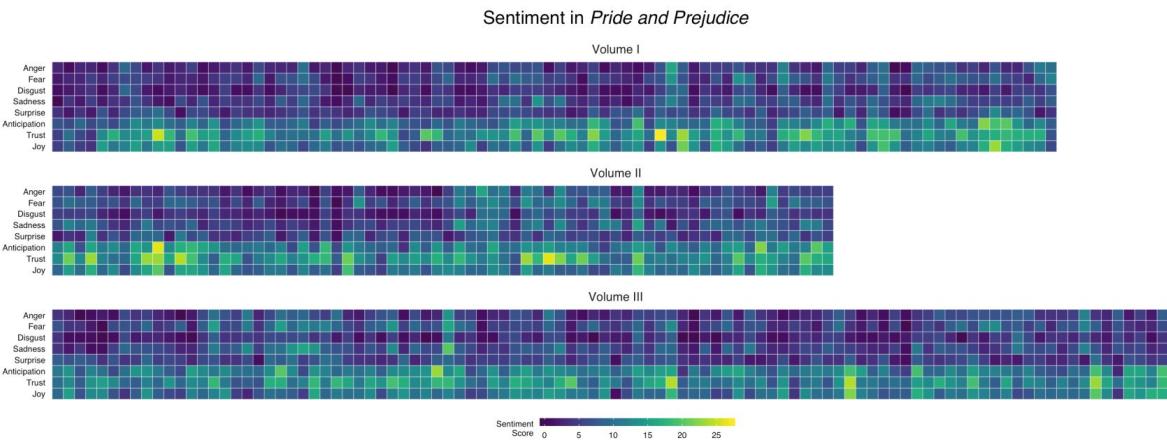
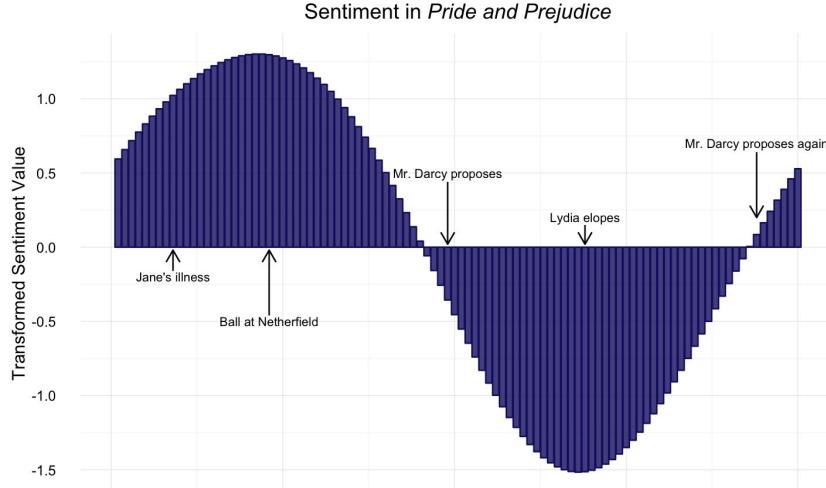


LA IDEA



data science ish

A Blog by Julia Silge



1. Joy to the World, and Also Anticipation, Disgust, Surprise...
(sobre el paquete syuzhet)
2. You Must Allow Me to Tell you How Ardently I Admire and Love NLP
(syuzhet aplicado a las novelas de Jane Austen)
3. If I Loved NLP Less, I Might Be Able To Talk About It More (más sobre syuzhet)

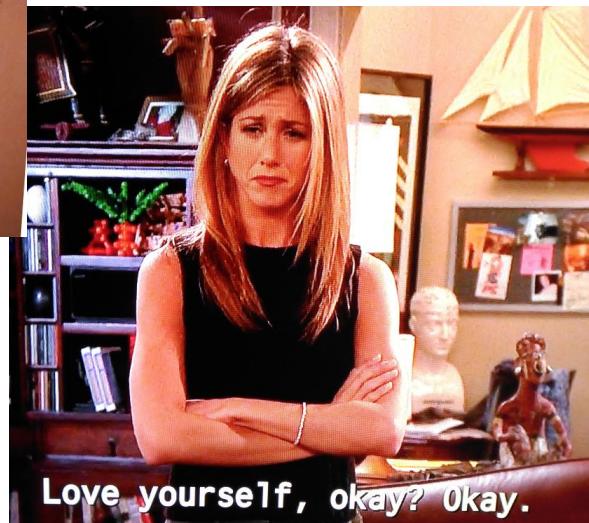
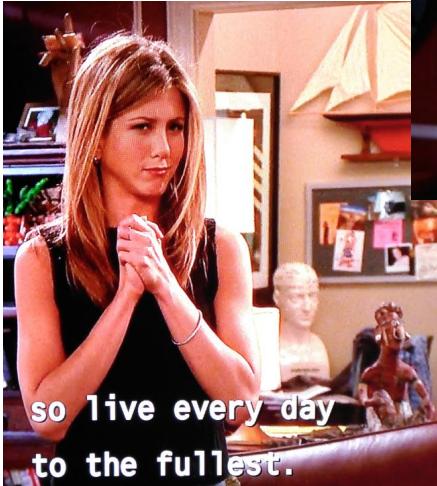
EJEMPLOS

Una de las frases con más ira de la temporada 3



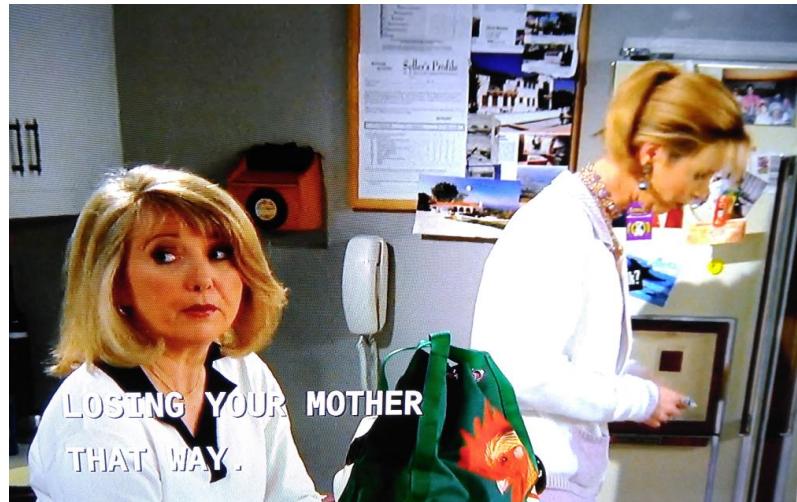
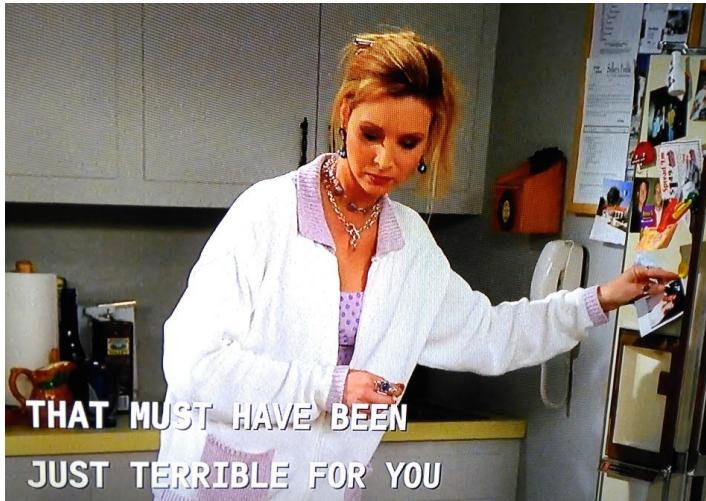
EJEMPLOS

Una de las frases más alegres de la temporada 5



EJEMPLOS

Una de las frases más tristes de la temporada 3



PROBLEMAS DEL MÉTODO

1. Que funciona palabra por palabra, y no es lo ideal.
 - Las expresiones u oraciones no se analizan como expresiones u oraciones, sino como la suma de sus palabras:

very happy ≠ very + happy

```
library(syuzhet)
> get_nrc_sentiment("very")
  anger anticipation disgust fear joy sadness surprise trust negative positive
1      0              0      0    0      0          0      0      0      0      0
> get_nrc_sentiment("happy")
  anger anticipation disgust fear joy sadness surprise trust negative positive
1      0              1      0    0    1          0      0      1      0      1
> get_nrc_sentiment("very happy")
  anger anticipation disgust fear joy sadness surprise trust negative positive
1      0              1      0    0    1          0      0      1      0      1
```

I don't want to be with you ≠ I + don't + want + to + be + with + you

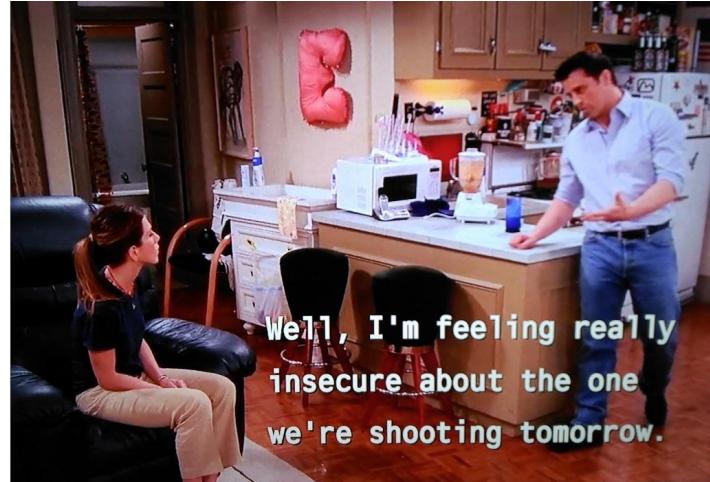
PROBLEMAS DEL MÉTODO

1. Que funciona palabra por palabra, y no es lo ideal.

- Todo es literal: se pierden las metáforas...

```
> get_nrc_sentiment("Joey: Well, I'm feeling really insecure about the one we are shooting tomorrow")
```

anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive	
1	3	2	1	3	1	2	1	1	3	1



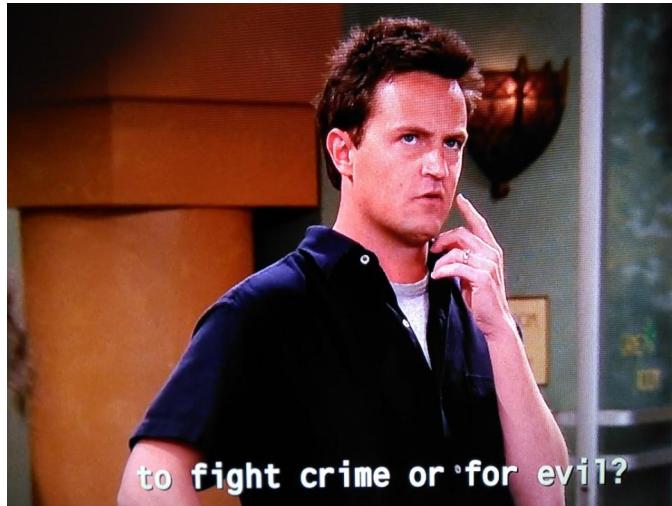
PROBLEMAS DEL MÉTODO

1. Que funciona palabra por palabra, y no es lo ideal.

- Todo es literal: y, claro, se pierden las bromas, la exageración, el sarcasmo...

```
> get_nrc_sentiment("Chandler: Should I use my invisibility to fight crime or for  
evil?")
```

	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive
1	3	0	1	2	0	1	0	0	3	0



PROBLEMAS DEL MÉTODO

1. Que funciona palabra por palabra, y no es lo ideal.
 - Todo es literal: y, claro, se pierden las bromas, la exageración, el sarcasmo...

```
> get_nrc_sentiment("Ross: I hate Chandler, the bastard ruined my life")
anger anticipation disgust fear joy sadness surprise trust negative positive
1           2           0           3           2           0           3           0           0           0           3           0
```



PROBLEMAS DEL MÉTODO

1. Que funciona palabra por palabra, y no es lo ideal.
 - Todo es literal: y, claro, se pierden las bromas, la exageración, el sarcasmo...

```
> get_nrc_sentiment("Rachel: Ohh! Lucky me! Oh my God! That is good news, Ross! I think that's")
```

	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive
1	0	2	0	1	3	0	2	2	0	3



PROBLEMAS DEL MÉTODO

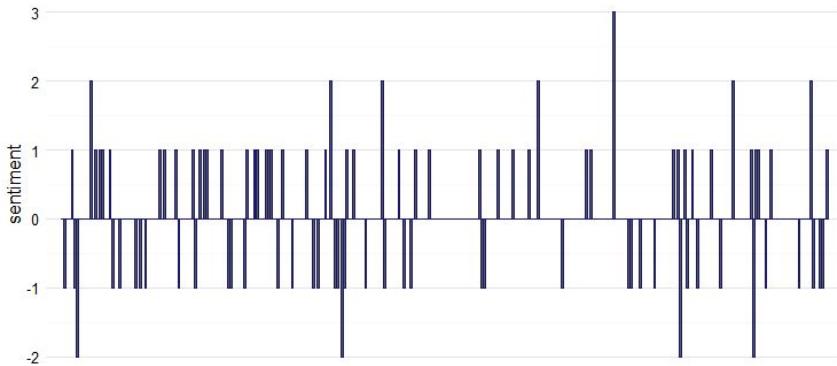
2. Que pronunciar una palabra no implica que la escena transmita el sentimiento asociado a ella.

```
> joy_items <- which(s10nrc$joy > 3)
> s10texto[joy_items]
[1] "Joey: Oh food. No, sex. Food! Sex! Food! Se-I don't know! Good God, I don't"
[2] "Mike: Phoebe you're so beautiful. You're so kind, you're so generous. You're
so wonderfully"
```

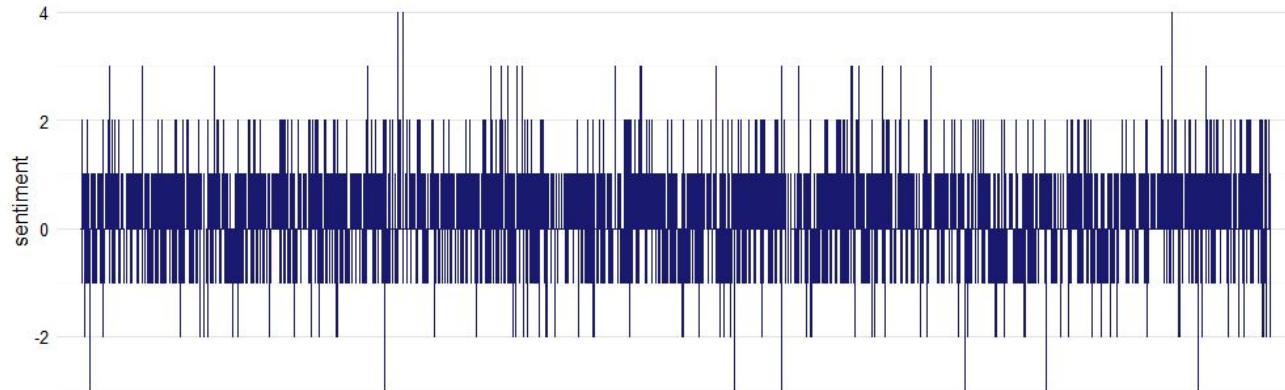


¡A PINTAR!

s01e13



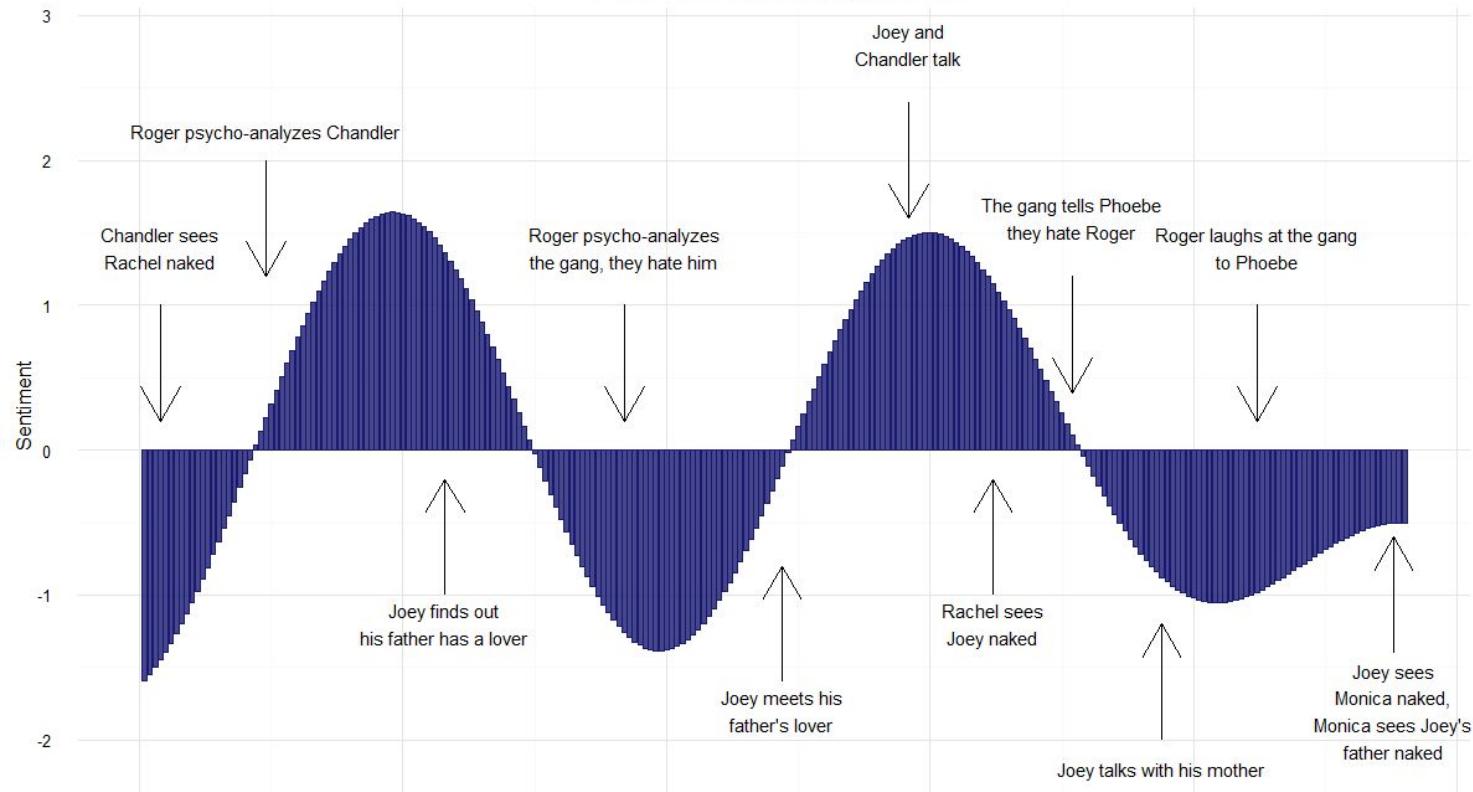
s01



transformada
de Fourier

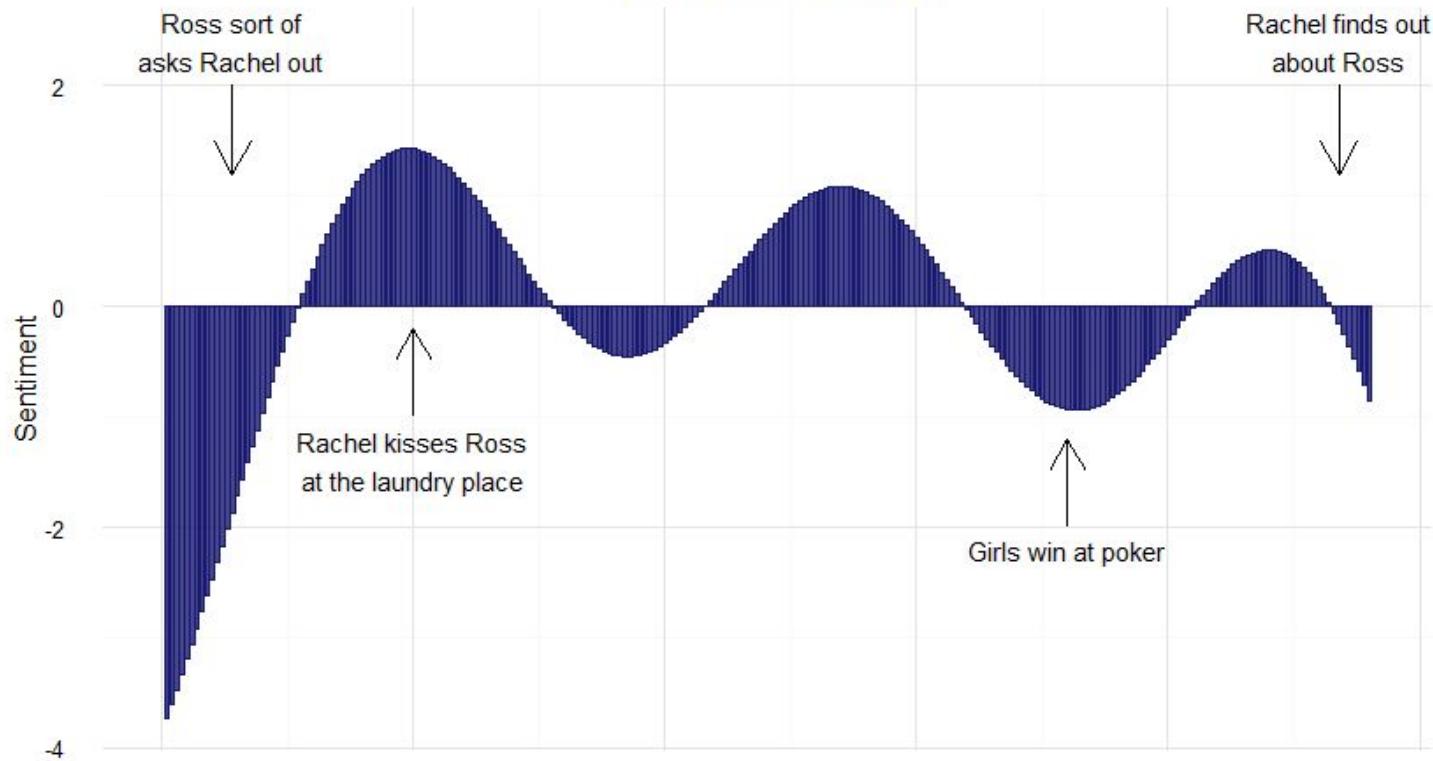
UN EPISODIO

Sentiment in Season 1 Episode 13



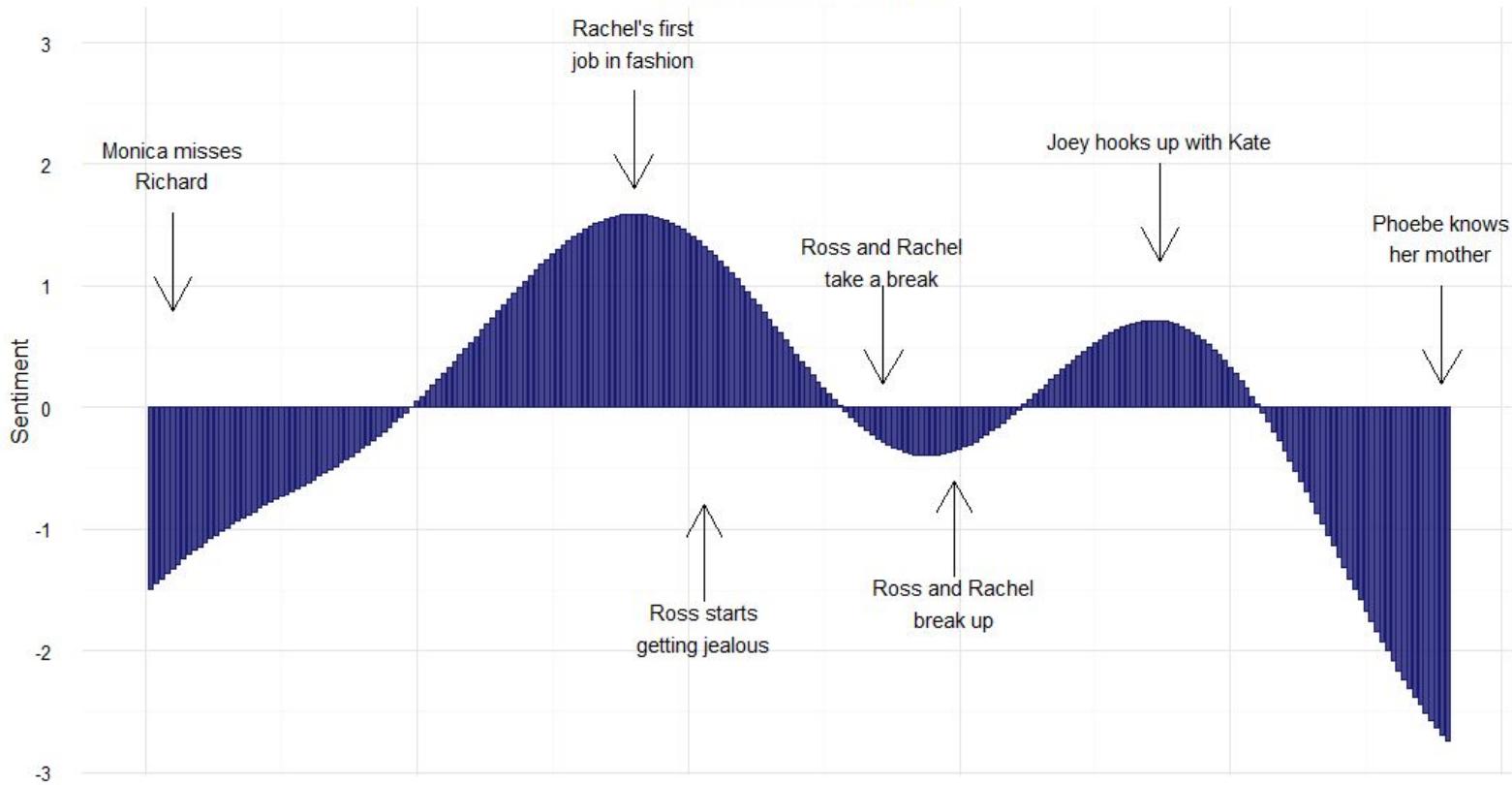
UNA TEMPORADA

Sentiment in Season 1

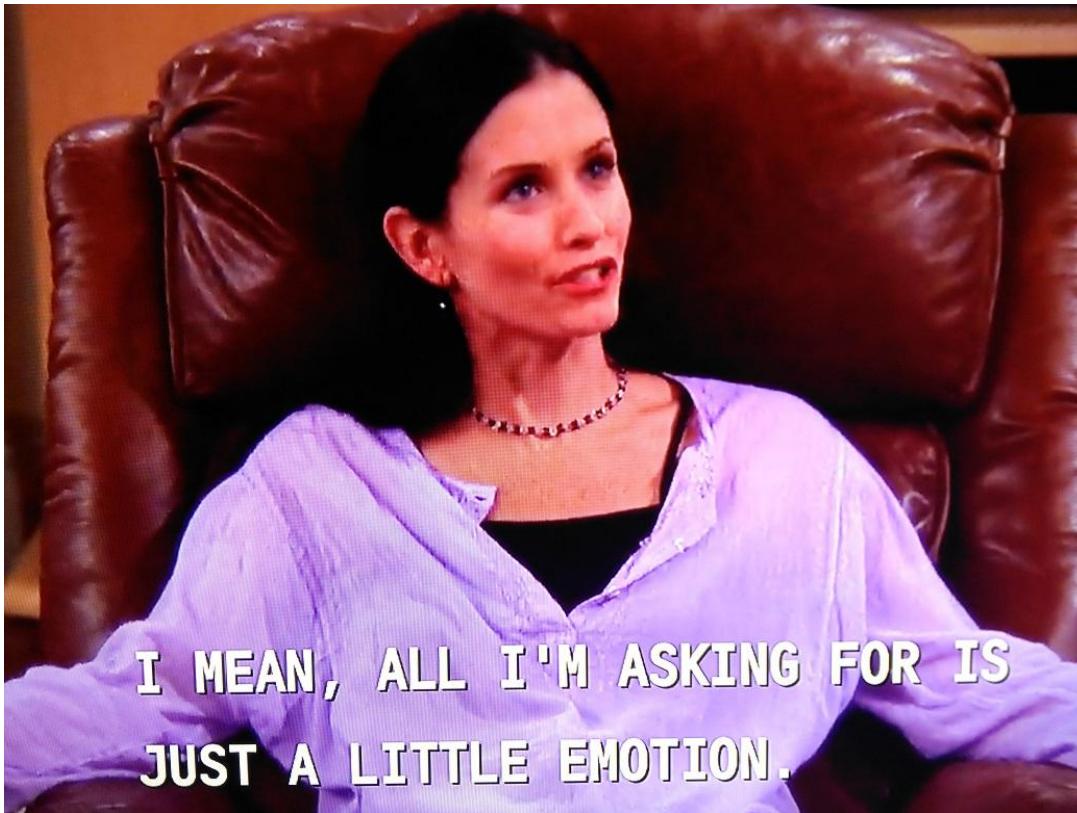


UNA TEMPORADA

Sentiment in Season 3



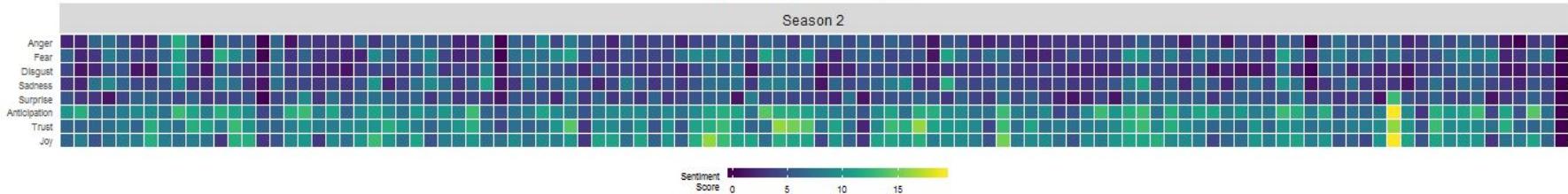
EMOCIONES



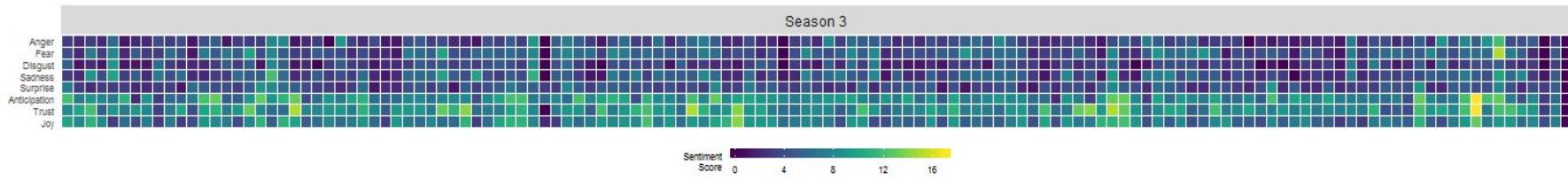
I MEAN, ALL I'M ASKING FOR IS
JUST A LITTLE EMOTION.

EMOCIONES POR TEMPORADA

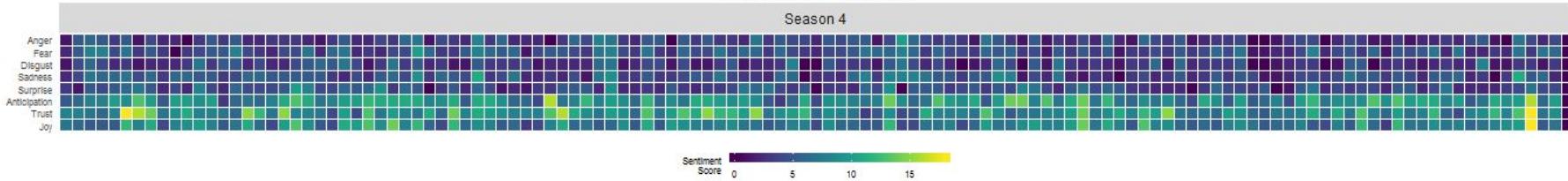
Sentiment in Season 2



Sentiment in Season 3

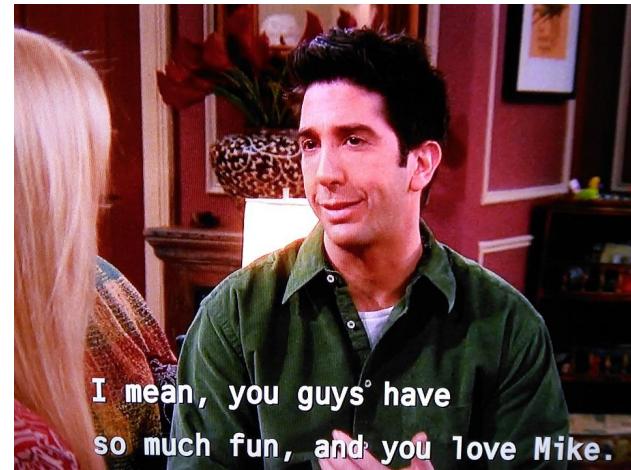
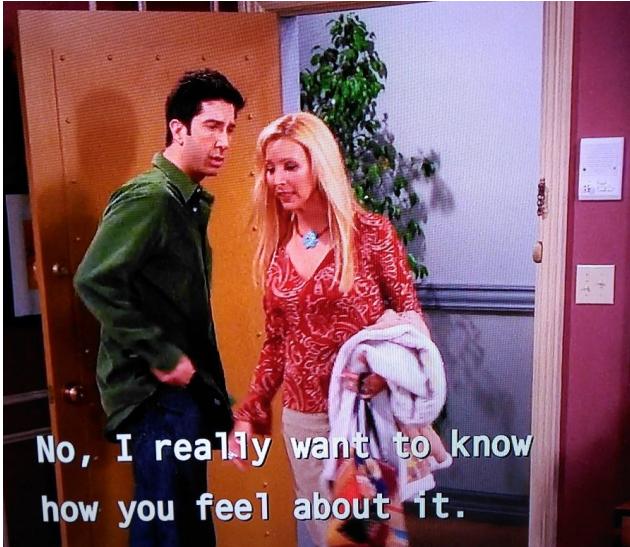
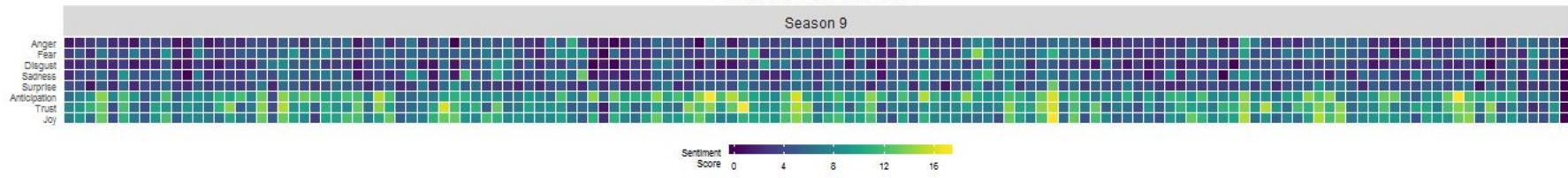


Sentiment in Season 4



EMOCIONES POR TEMPORADA

Sentiment in Season 9



CONCLUSIONES

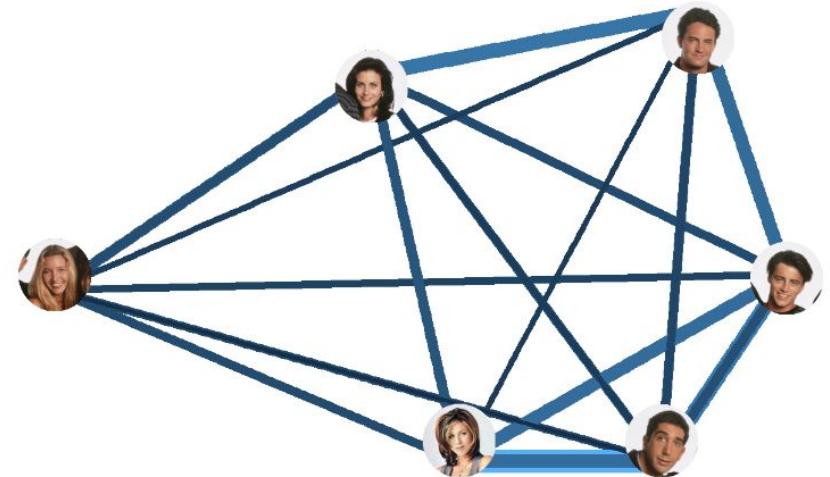
- No es lo mismo texto novelado que texto dialogado. Ni una novela con un protagonista que una historia coral. Sería interesante comparar estos resultados con obras de teatro, con otras series...
- Estaría bien tener un paquete o una API para acceder al corpus de *Friends* bien corregido y formateado
- Hay que avanzar hacia un análisis de opinión que vaya más allá de la suma de palabras
- Hacen falta más recursos para análisis de opinión en español

MÁS FRIENDS

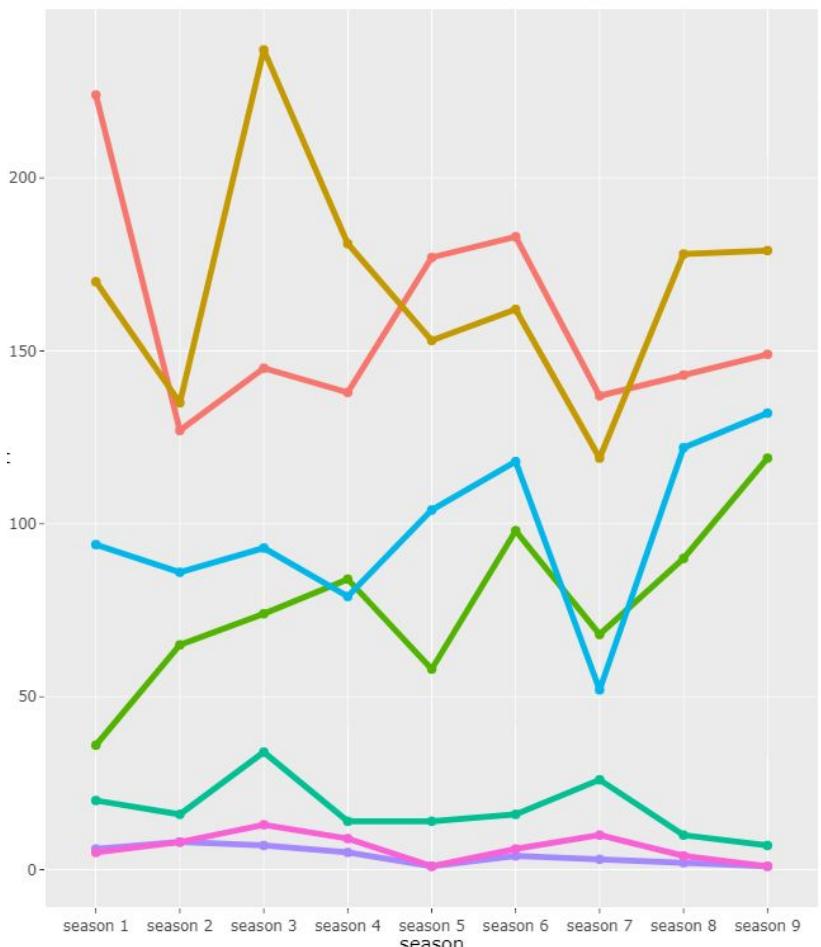
Friends: No. of Lines per Character



Friends: Graph by Characters #Co-Mentions in Script



The One With Friends (Giora Simchoni)

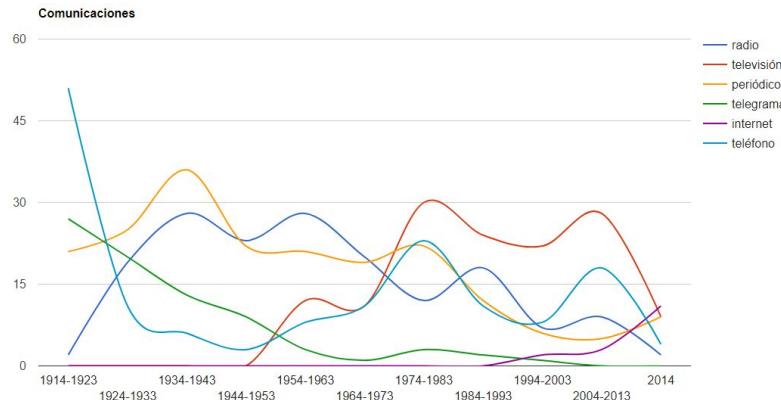
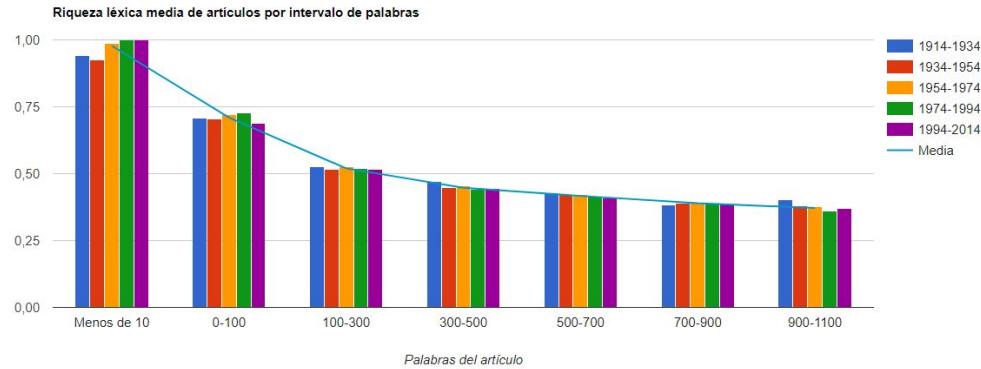


Taller de minería de textos con R (Claudia Guirao)

ARACNE

Aracne es un estudio sobre la variación de la riqueza lingüística en la prensa española desde 1914 hasta 2014.

«¿Hablamos peor que antes,
tenemos un vocabulario menos
rico?»



SUGERENCIAS PARA INVESTIGAR EN LINGÜÍSTICA DE CORPUS

- Distribución léxica
 - *después de + participio (después de pronunciado su discurso)*
 - *en + gerundio (en viendo...»)*
 - *aparte + nombre / aparte de + nombre (aparte Finlandia)*
 - *anteanoche / anoche*
 - *en + año / en el + año*
 - *es fuerza*
 - *encarnado / rojo*
 - *grande (sin apocopar) + nombre*
 - *mejor comprender / comprender mejor*
 - *en caso tal / en tal caso*
 - *eufemismos (inútil, discapacitado, lisiado, cojo)*
- Distribución ortográfica
 - *Beijing / Pekín (¿2008?)*
 - *Yugoeslavia / Yugoslavia, Checoeslovaquia / Checoslovaquia*
 - *subs- / sus-*
- Longitud de las oraciones a lo largo del tiempo
- Media de palabras por oración

BIBLIOGRAFÍA

- [Fun with Google Ngrams](#)
- [Guía rápida de análisis de corpus con AntConc](#) (Borja Navarro, Universidad de Alicante)
- [Canal de YouTube de Laurence Anthony](#): tutoriales sobre AntConc
- [Corpus Analysis with AntConc](#)
- ['Star Wars' en clase de lengua: cómo hacer un corpus de subtítulos](#)
- [Sabina en corpus presente \(I\): creación](#)
- [Creación del corpus de Aracne](#)



¡GRACIAS POR VUESTRA ATENCIÓN!



Lingwars