

# MANUAL

# USO DE



IRaMUTEQ

Versión 0.7 alpha 2

Mariola MORENO y Pierre RATINAUD

11/03/2015

# INDICE

Qué es Iramuteq .....	3
Instalación.....	3
Formato texto de trabajo .....	3
Comenzar a trabajar con el programa .....	5
Opciones de análisis.....	8
1. Estadística Textual.....	9
2. Especificidades y Análisis Factorial.....	11
3. Clasificación Método Reinert .....	14
4. Análisis de Similitud .....	22
5. Nubes de Palabras .....	25
6. Subcorpus Metadatos.....	25
7. Subcorpus por Temática .....	26

## ¿Qué es IRAMUTEQ?

Iramuteq es un programa informático que nos permite realizar un análisis multidimensional de textos y cuestionarios, el cual ha sido desarrollado por el profesor **Pierre Ratinaud** en el seno del laboratorio de investigación **LERASS**, de la universidad de Toulouse.

Es compatible con todos los tipos de sistemas operativos, totalmente gratuito y se puede descargar directamente desde su página web: <http://www.iramuteq.org/telechargement>

Su utilidad es muy amplia, como por ejemplo en los trabajos de investigación, donde tenemos un volumen de información bastante grande y difícil de abarcar manualmente, Iramuteq nos facilitará y ahorrará mucho tiempo en su lectura y su posterior análisis.

## Instalación de Iramuteq...

Al tratarse de una interfaz del software R, es indispensable descargarse éste para que el programa pueda funcionar correctamente, por lo tanto:

**Paso 1:** Descargar e instalar el software R (<http://www.r-project.org>). La última versión 0.7 alpha 2 ha sido probada con **R 3.1.2** en Windows XP / Windows 7 y en Mac OS X y la versión **R 2.15.1** en Ubuntu 14.10.

**¡OJO!** Visitar la [pagina web de iramuteq](http://sourceforge.net/projects/iramuteq/) (<http://sourceforge.net/projects/iramuteq/>) y comprobar qué versión de R ha sido testada para la versión de iramuteq que va a instalar.

**Paso 2:** Descargar e instalar Iramuteq (<http://sourceforge.net/projects/iramuteq/>)

En el primer lanzamiento de Iramuteq éste instalara las librerías de R que necesita para su funcionamiento (ca, rgl, ape, proxy, gee, igraph, wordcloud, irlba) y sus dependencias. Sobre GNU / Linux este proceso puede tardar unos minutos.

**Paso 3:** Para verificar que la instalación se haya completado, abriremos Iramuteq y nos dirigiremos a la pestaña de « Edición » -> Preferencias (Iramuteq -> preferencias en Mas Os X) y pulse sobre el botón de "Verificar la instalación".

Seguidos estos pasos, tendría que funcionar correctamente, no obstante para cualquier duda o problema pueden dirigirse a: [iramuteq-users@lists.sourceforge.net](mailto:iramuteq-users@lists.sourceforge.net).

## Iramuteq instalado, damos formato al texto de trabajo...

Una vez que tenemos ya nuestro programa instalado podemos comenzar con la puesta en forma de la información con la que queremos trabajar. Para ello es importante que el texto esté volcado sobre **Openoffice** o **Libreoffice** y codificado en UTF-8 (esta opción se realiza al

guardar nuestro texto, primeramente encontramos la pestaña en la que introduciremos el « nombre que le queremos dar al archivo » y debajo se encuentra la que nos permite desglosar el « tipo » de formato en el que queremos guardar nuestro archivo, ahí tenemos que escoger la opción de “texto codificado”).



**¡OJO!** Nunca utilizar los editores de texto: ~~Bloc de notas, Wordpad y Microsoft Office~~ para trabajar con Iramuteq.

Los textos tienen que estar introducidos por cuatro asteriscos (\*\*\*\*) seguidos de una serie de variables, cada variable tiene que estar precedida por un asterisco y separados por un espacio.  **Siempre tendrá que haber como mínimo una variable en nuestro texto.**

Se pueden utilizar los caracteres comprendidos de la A-Z o numéricos de 0-9 pero nunca caracteres especiales como apóstrofes, signos de puntuación, a excepción de la barra baja, que será el método de separación entre variable y modalidad.

\*\*\*\* \*VAR\_MOD

La **variable** es la parte general y la **modalidad** la parte específica. Ejemplo:

\*\*\*\* \*año\_2012 \*sex\_h

texto  
texto texto texto texto texto texto texto texto texto texto texto texto  
texto texto texto texto texto texto texto texto texto

\*\*\*\* \*año\_2013 \*sex\_f

texto  
texto texto texto texto texto texto texto texto texto texto texto texto  
texto texto texto texto texto texto texto texto texto

También se pueden introducir **temáticas** dentro de las variables, para ello utilizaremos un guion bajo y un asterisco antes de denominar a la temática, es decir:

\*\*\*\* \*VAR\_MOD

-\*Temática

Ejemplo:

\*\*\*\* \*año\_2012 \*sex\_h

-\*them\_crisiseconmica

texto  
texto texto texto texto texto texto texto texto texto texto texto texto  
texto texto texto texto texto texto texto texto texto

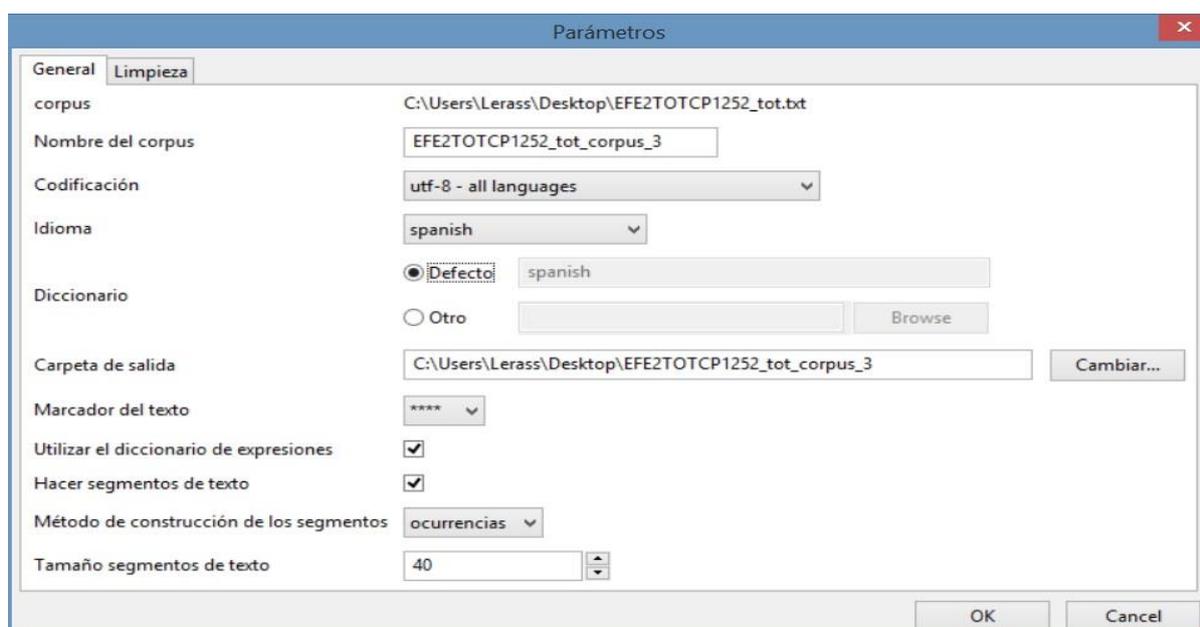


- ❖ **Marcador de texto:** \*\*\*\*. No modificaremos esta configuración que viene dada por defecto con los cuatro asteriscos.
- ❖ **Utilizar el diccionario de expresiones:** hay determinadas expresiones que son compuestas y están unidas por apóstrofes o guiones, al seleccionar esta opción todas las expresiones que hayan sido registradas en el diccionario de Iramuteq serán reconocidas como una única palabra, por ejemplo: Producto Interior Bruto.



Se pueden aportar modificaciones en las diferentes lenguas disponibles de Iramuteq, para ello iremos al perfil de usuario de nuestro ordenador, buscaremos « *iramuteq/dictionaires* » escogeremos el diccionario de la lengua que queremos modificar e introduciremos los cambios oportunos.

- ❖ **Hacer segmentos de texto:** esta opción permite dividir el texto en segmentos.
- ❖ **Método de construcción de los segmentos:** aquí elegiremos donde queremos que se formen los segmentos, las opciones posibles son :
  - Caracteres: se tomara como criterio de segmentación el número de caracteres que contienen.
  - Ocurrencias: en este caso el criterio será el número de ocurrencias.  
\*Ocurrencia es: una unidad lingüística.
  - Párrafos: el cambio de línea marca el fin de un segmento.
- ❖ **Tamaño de segmentos de texto:** define el número de unidades contenidas en el segmento de texto.





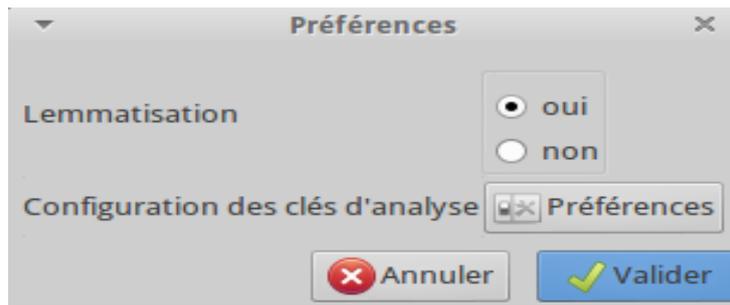


Figura 2 Lematización

- Con la opción de « **lematización** » diremos SI o NO queremos realizarla.



¿Qué significa lematización? Quiere decir que los verbos se pasan al infinitivo, los nombres al singular y los adjetivos al masculino, en definitiva, quitar el género, número y conjugación a las palabras (los diccionarios de los diferentes idiomas están disponibles en el repertorio de *iramuteq/dictionnaires*)

- **Configuration des clés d'analyse:** en esta opción podemos determinar las formas activas y las suplementarias, es decir, las formas relevantes para nuestro análisis (activas) y las que no aportan ningún valor (suplementarias). Para realizar cualquier modificación iremos a « preferencias » y podremos realizar las modificaciones que consideremos pertinentes (estos valores están registrados en el fichero *key.cfg* accesible en el *repertoire.iramuteq*)

Una vez que tenemos estos primeros detalles definidos le daremos a « VALIDAR » y automáticamente comenzará a producirse la indexación de nuestro corpus, este proceso llevara unos minutos, dependiendo del tamaño del corpus.



Llamaremos **corpus** a ese conjunto de textos que hemos introducido en nuestro programa siguiendo una serie de criterios y con un fin determinado.

***Han pasado unos minutos y el proceso ha finalizado...***

Veremos que en la parte de izquierda de Iramuteq aparece un nuevo archivo que se llamará del mismo nombre del documento que hemos introducido para trabajar, es decir que si mi archivo se llamaba Crisis, en Iramuteq será « *Crisis\_Corpus1* » y si lo seleccionamos y pulsamos el botón derecho de nuestro ratón, veremos que se abren distintas posibilidades de análisis, o bien podemos hacerlo a través de la barra de tareas del mismo programa, realícelo como le resulte más cómodo.



**Opciones de Análisis...**

# 1. ESTADÍSTICAS



Esta opción reproduce el resultado de todas nuestras formas: activas y suplementarias y la lista de « hápax ».

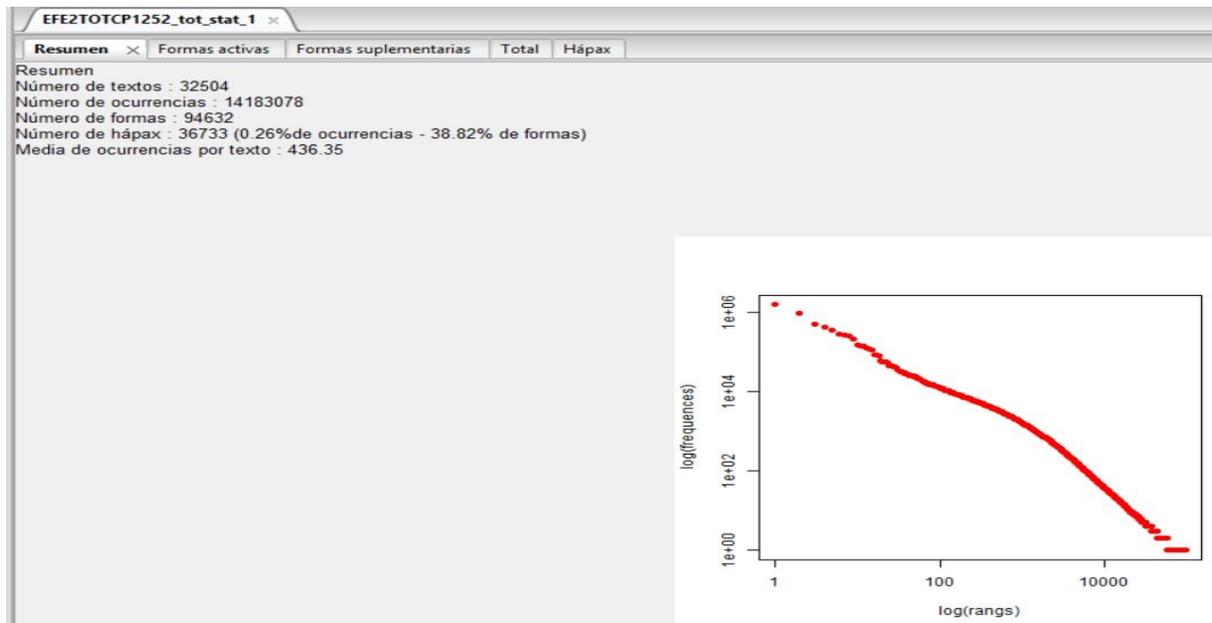


Figura 3 Resultados análisis textual

En estos primeros resultados podremos ver las siguientes informaciones:

El Gráfico, presenta en el eje de abscisas los logaritmos de los rangos y en el eje de ordenadas se reflejan los logaritmos de frecuencia de los textos.

- *Número de textos*: es el número de textos que tiene nuestro corpus.
- *Número de ocurrencias*: son el total de palabras (este número variara dependiendo si el texto esta lematizado o no)
- *Número de formas*: total número de formas del corpus.
- *Número de Hápax*: en esta lista se encuentran las palabras que aparecen solo una vez en todo el corpus.
- *Media de ocurrencias por texto*: es el resultado de la división entre el número de ocurrencias / número de textos.

Forma	Frec. ↓	Tipos
crisis	57658	nom
económico	56723	adj
gobierno	44141	nom
país	41359	nom
millón	39608	nom
presidente	32378	nom
españa	32158	nom
ciento	27442	adj
político	25849	adj
euro	25610	nom
nuevo	25192	adj
europeo	24776	adj
español	24363	adj
primero	24235	adj
estado	22266	nom
madrid	21162	nom

Figura 4 Tabla de indexación

Si nos posicionamos sobre la ventana de formas activas, suplementarias, global o total, veremos que hay tres columnas.

- ✓ La columna de «**forma**»: están presentes las formas.
- ✓ La columna de «**frec**»: nos indica el número de veces de aparición de la forma a la que acompaña.
- ✓ La columna «**tipos**»: señala la categoría gramatical de la forma.

Además de estos datos, dentro de esta pestaña de « Estadística » hay más opciones de análisis. Podremos visualizar una información complementaria al seleccionar una forma y pulsar el botón derecho de nuestro ratón, donde se nos abrirá un menú que tendrá las opciones de:

- **Formas asociadas**: aquí podremos visualizar todas las formas que se han lematizado respecto a la forma que hemos señalado, es decir, si tenemos la forma niño, en sus formas asociadas podrán aparecer niño, niña, niños, niñas.
- **Concordancia**: en este caso podremos ver los segmentos de texto donde aparece la forma elegida. Es decir, si yo selecciono la palabra « **casa** », y los segmentos donde aparece son: Mi **casa** es muy bonita.

El problema que encuentran los compradores es que la **casa** cuesta 100 millones, por ese motivo dice el agente inmobiliario que la **casa** aún no se ha vendido, todos los segmentos donde aparezca “casa” me aparecerán.

Si en vez de seleccionar una palabra del corpus lo hacemos sobre la totalidad del mismo, se desplegará la siguiente ventana:

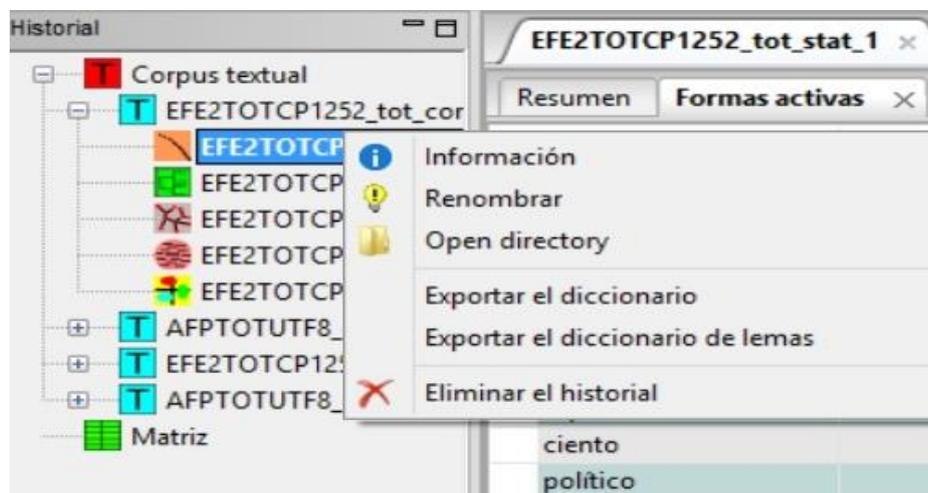


Figura 5 Opciones suplementarias

- **Información:** permite recuperar las informaciones del análisis.
- **Renombrar:** esta opción es para cambiarle el nombre al fichero de estadística textual que hemos generado.
- **Open directory:** nos dirige a la carpeta donde se encuentran todos los archivos que hemos creado al realizar este análisis de estadística (habrá un Excel de las palabras activas, otro de las suplementarias, etc.). Para cada análisis que realizamos, Iramuteq crea un nuevo repertorio en la misma carpeta en la que está guardado nuestro corpus, éste se llama « NomDuCorpus-Stat-x ».
- **Exportar el diccionario:** genera un fichero « *dictionary.csv* » que recoge las ocurrencias → Utilizar la tabulación como separador. Este diccionario está en el propio formato que utiliza Iramuteq, por lo tanto está listo para ser utilizado directamente para el análisis.
- **Exportar el diccionario de lemas:** genera un fichero « *lemme.csv* » que recoge los lemas, y también las variantes y sus efectivos. (Igualmente, hay que utilizar la tabulación como separador)
- **Eliminar el historial:** borrar el análisis del historial.

## 2. ESPECIFICIDADES Y AFC

Esta opción nos permite realizar un análisis factorial, cruzando las formas activas con las variables.

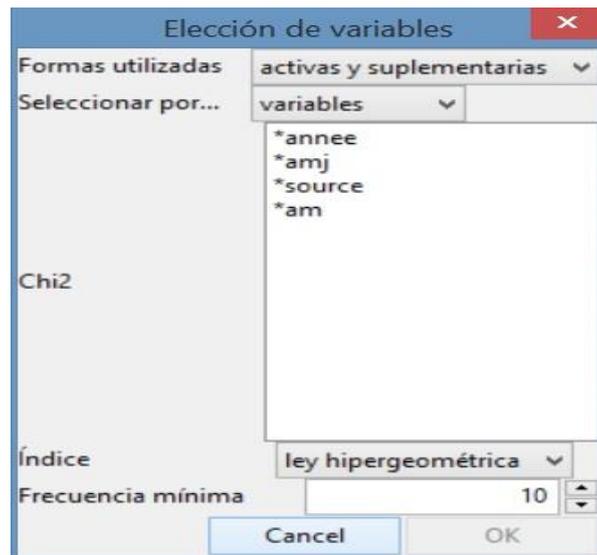


Figura 6 Elección variables AFC

El menú de « *elección de variables* » nos ofrece las siguientes opciones:

- **Formas utilizadas:** escogeremos las formas que queremos incluir para nuestro análisis.
    - Activas y suplementarias
    - Activas
    - Suplementarias
  - **Seleccionar por :**
    - a) Variable: permite seleccionar todas las modalidades de una variable.
    - b) Modalidad: escogeremos solo las modalidades que nos interesen dentro de esa variable.
- Es obligatorio seleccionar una variable o modalidad.
- **Índice:** hay dos posibilidades de cálculo, la Chi2 o la ley hypergeométrica.
  - **Frecuencia mínima:** Sólo se mostraran las formas en las que el número de ocurrencias que estén por encima del umbral que marquemos en todo el corpus.



**¡OJO!** Si seleccionas varias variables, solo se tendrá una en cuenta.

Una vez que lanzamos nuestro análisis factorial, obtendremos el número de efectivos de formas y tipos.

Formas	Formas banales	Tipos	Frecuencia de formas	Frecuencia de tipos	Frecuencia relativa de formas	Frecuencia relativa de tipos	AFC
formes		*annee_2009	*annee_2010	*annee_2011	*annee_2012	*annee_2013	*annee_2014
ciento	inf	49.8945	-2.8008	-inf	-inf	-41.0457	
dólar	inf	-6.1089	6.2684	-183.2634	-141.9882	-11.9039	
mundial	267.169	-0.9383	0.4239	-93.015	-96.3475	-11.0478	
millón	170.0573	3.3672	32.563	-131.9956	-81.8366	-4.7662	
crisis	168.2447	2.0492	-0.3558	-58.0817	-67.848	-5.7678	
reunión	105.7526	1.1289	-22.8007	-17.2054	-24.8753	-13.6957	
presidente	62.6729	0.8185	-59.6874	-4.4421	-8.1168	-0.9694	
unir	62.6378	-3.4813	-2.4306	-16.3381	-8.9612	-1.7318	
internacional	50.7694	-2.6159	0.5832	-24.7143	-5.6469	-1.4609	
precio	47.7432	4.5571	45.489	-74.8068	-45.7153	-0.4386	

Figura 7 Análisis de Especificidades

## 2.1 Reedición de gráficos de AFC



Pulsando este botón que tenemos a la izquierda del análisis de dendogramas podremos acceder a la siguiente ventana para realizar una reedición del análisis de dendograma y realizar ciertos retoques en nuestro análisis factorial.

The dialog box 'Reedición AFC' contains the following settings:

- Tipo de gráfico: 2D
- Formato de la imagen: png
- Representación: datos
- Variables: activas
- anchura: 800
- altura: 800
- Tamaño texto: 9
- Tomar los x primeros puntos:  50
- Tomar los x primeros puntos por clase:  30
- Limitar los puntos por la Chi2 que enlaza las clases:  4
- Evitar recuperaciones:
- Tamaño del texto proporcional a la frecuencia:  min: 5 max: 40
- Tamaño texto proporcional al chi2:  min: 5 max: 40
- Factor x: 1
- Factor y: 2
- Factor z: 3
- Transparencia de las esferas: 10 (slider from 1 to 100)
- Hacer una película:

Figura 8 Reedición AFC

- **Tipo de gráfico:** a elegir entre 2D y 3D.
- **Formato de la imagen:** escoger el formato en el que queremos guardar la imagen (png o svg).
- **Representación:** seleccionar datos o correlaciones para su presentación.
- **Variable:** Elección de formas a analizar (activas, suplementarias, clases, estrellas).
- **Anchura / Altura:** definir las dimensiones del grafico (píxeles).
- **Tamaño del texto:** determinar el tamaño del texto.
- **Tomar los primeros puntos:** tomar los x primeros puntos que tienen los chi2 más fuertes en el conjunto de la clasificación.
- **Tomar los primeros puntos por clase:** tomar los x primeros puntos chi2 más fuertes por clase.
- **Limitar los puntos por la chi2 que enlaza la clase:** umbral mínimo de enlace a la clase para aparecer.
- **Evitar recuperaciones:** esta opción evita la recuperación de las etiquetas de los puntos producidos, pero permite una mejor legibilidad.
- **Tamaño del texto proporcional a la frecuencia:** establece el tamaño del texto en función del efectivo de la forma en el corpus.
- **Tamaño texto proporcional al Chi2:** estructura el tamaño del texto en función del chi2 relacionado a la clase.
- **Factor X/ factor Y/ factor Z:** hace referencia respectivamente eal factor en abscisa (X), ordenadas (Y) o el factor en nivel (Z).
- **Transparencia de las esferas:** permite ajustar la transparencia de las esferas en los gráficos 3D.
- **Hacer una película:** disponible únicamente para los gráficos en 3D, produce un formato .gif animado que representa una rotación de 360 grados del gráfico.

### 3. CLASIFICACIÓN MÉTODO REINERT



Se trata de realizar una clasificación jerarquizada de manera descendente según el método descrito por Reinert (1983, 1986, 1991), definido por clases lexicales, donde cada una de ellas representa una temática y pueden ser descritas según el vocabulario que las define.

Figura 9 Análisis Método Reinert

Esta es la ventana de opciones que se nos despliega cuando vamos a realizar un análisis de clasificación del Método Reinert:

- **Clasificación:**
  - a) **Doble en RST:** reagrupa los segmentos de texto más próximos y realiza dos veces este análisis.
  - b) **Simple sobre segmento de textos:** la clasificación llevara consigo los segmentos de texto.
  - c) **Simple en el texto:** mantiene la integridad de los textos e intenta reagrupar los más próximos.
  
- **Tamaño de rst1 y rst2:** estas opciones se seleccionarán para la clasificación « *doble en rst* » y nos permiten cambiar el tamaño de la reagrupación de segmentos de texto. Estas unidades representan el número de formas activas reagrupadas por línea.
  
- **Número de clases terminales de la fase 1:** determina el número de clases de la primera parte de la clasificación, por defecto viene definida sobre 10, pero la ajustaremos en función del tamaño de nuestro corpus.
  
- **Frecuencia mínima de segmentos de texto por clase:** determina un nivel mínimo bajo el cual, los segmentos de texto no serán tenidos en cuenta.
  
- **Frecuencia mínima de una forma analizada:** esta opción no está disponible.
  
- **Número máximo de formas analizadas:** se refiere al número máximo de formas a analizar, por defecto viene definido para 3000 formas activas y suplementarias más frecuentes. Pero este número puede ser modificado según el tamaño de tu corpus. → Una forma debe aparecer un mínimo de 3 veces para ser retenida.
  
- **Método pour SVD:** nos permite elegir el algoritmo utilizado para las descomposiciones, en valores particulares.

- **Modo patata:** esta opción suprime la segunda fase de cada partición. Realiza un análisis más rápido pero un poco menos preciso. Es aconsejable seleccionar esta opción en primer lugar para obtener una visión de los resultados de manera más rápida, y ya después realizar otro análisis sin seleccionar esta opción para realizarlo de forma más exhaustiva.

Ya hemos realizado las selecciones correspondientes a nuestras necesidades, así que es la hora de lanzarlo y ver qué resultados obtendremos...Pasados unos minutos, horas...(dependiendo del tamaño de nuestro corpus), se nos abrirán diferentes pestañas :

- **CDH:** aquí veremos los resultados. Y se distinguirán tres bloques :
  1. Constituye las estadísticas textuales del corpus, donde se podrá ver la media de ocurrencias por segmento y el número de segmentos de clases sobre la totalidad del corpus.
  2. En ésta se representa la partición, el tamaño y porcentaje de las clases.
  3. Representa las clases no retenidas.

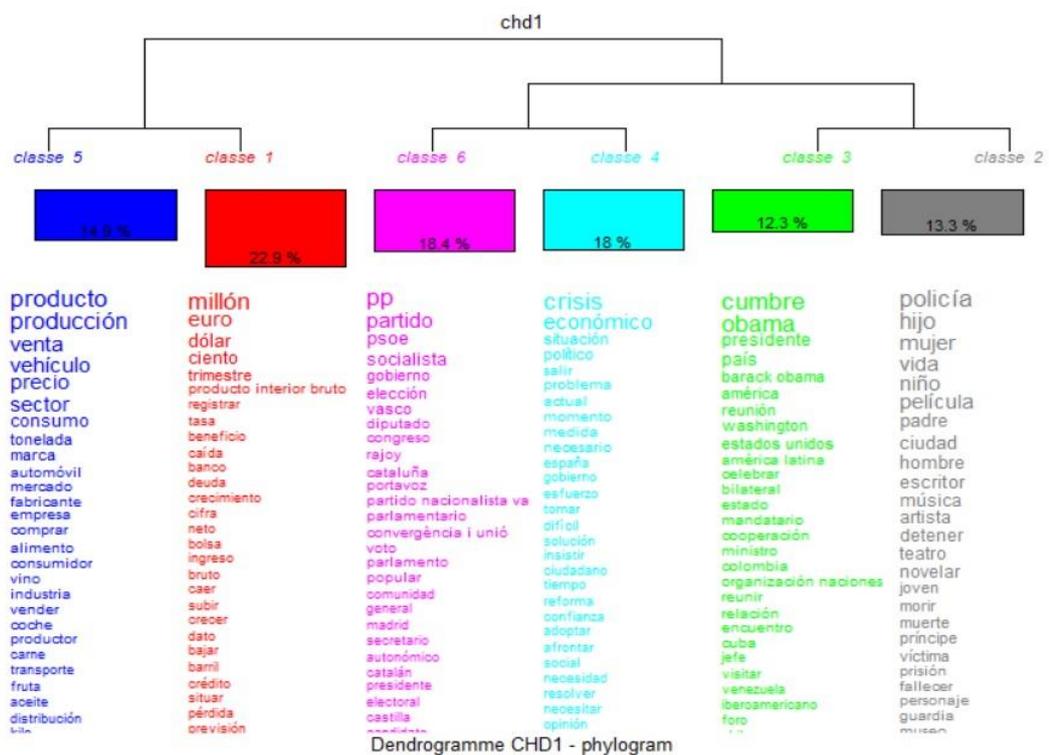


Figura 10 Dendograma CHD

- **PERFILES:** aquí veremos el léxico que se ha incluido en cada clase, así como nos orientara sobre la temática que le rodea, ya que los perfiles son agrupados con una cordura y una proximidad lexical entre las formas incluidas.

CHD		Perfiles x		AFC							
<b>1 Clase 1</b>		<b>2 Clase 2</b>		<b>3 Clase 3</b>		<b>4 Clase 4</b>		<b>5 Clase 5</b>		<b>6 Clase 6</b>	
76170/332316		44314/332316		41034/332316		59872/332316		49624/332316		61302/332316	
22.92%		13.33%		12.35%		18.02%		14.93%		18.45%	
n...	↑	eff. s.t.	eff. total	pourcentage	chi2	Type	forme				
0		20270	28326	71.56	41464.16	nom	millón				
1		14358	20892	68.72	26473.74	nom	euro				
2		7924	8945	88.59	22435.07	nom	dólar				
3		12902	20439	63.12	19924.41	adj	ciento				
4		4329	4867	88.95	12187.56	nom	trimestre				
5		3696	4155	88.95	10384.29	nom	producto interior bruto				
6		3859	5623	68.63	6763.83	ver	registrar				
7		3001	3853	77.89	6666.37	nom	tasa				
8		2810	3669	76.59	6047.97	nom	beneficio				
9		3930	6267	62.71	5723.66	nom	caída				
10		4636	8047	57.61	5617.38	nom	banco				

Figura 11 Perfiles de clasificación Método Reinert

**Num:** el número de la línea, nos permite encontrar la clasificación inicial.

**Eff. S.T:** abreviación de efectivos de segmentos de texto (debe contener mínimo una forma)

**Eff.total:** son los efectivos totales de segmentos de texto contenidos en todo el corpus.

**Pourcentage:** es el resultado de dividir eff. ST y eff Total.

**Chi2:** experimenta la fuerza del enlace entre la forma y la clase.

**Type:** categoría gramatical de la forma.

**Forme:** es la forma o el lema en cuestión.

**P:** es el umbral de significado asociado a Chi2



Si pulsamos el comando CTRL+F podremos buscar una forma en el perfil que tenemos abierto en ese momento y pulsaremos CTRL + Shift + F si la forma que queremos buscar es de todos los perfiles.

- **AFC:** nos mostrara un análisis factorial correspondiente a la contingencia creciente de formas y lemas.

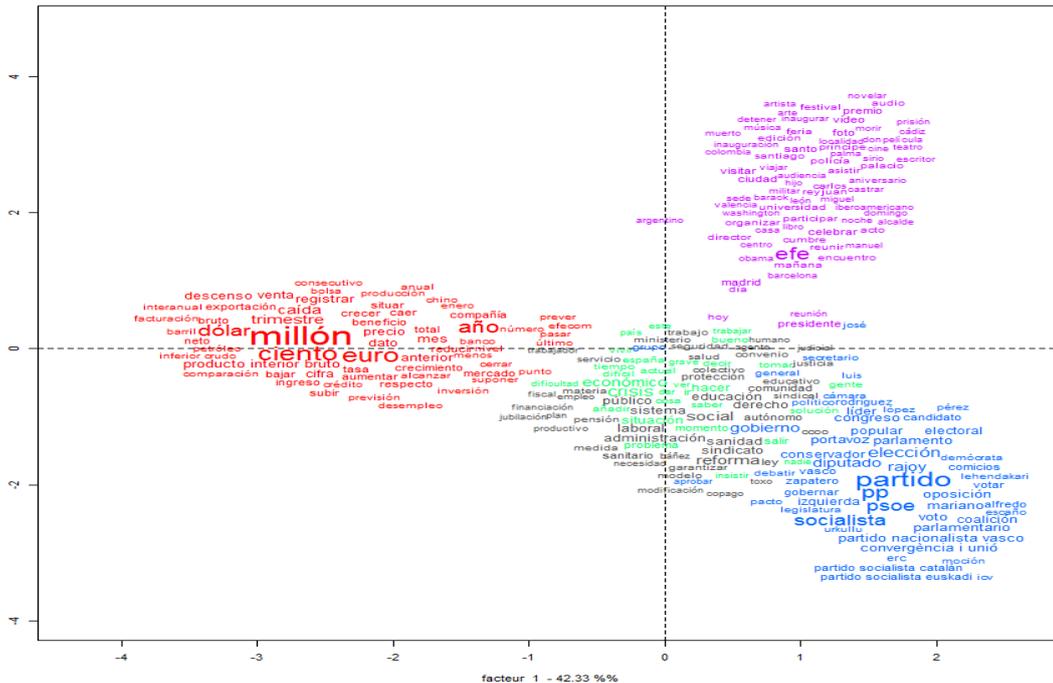


Figura 32 Ejemplo AFC

- **FACTOR:** presenta los valores y los porcentajes de cada clase.
- **3D GRAFICO:** produce un gráfico de distancias mínimas sobre las formas en 3 dimensiones.

Para profundizar en nuestro estudio de perfiles, Iramuteq propone un menú suplementario de funciones, que lo obtendremos al pulsar el botón derecho del ratón.

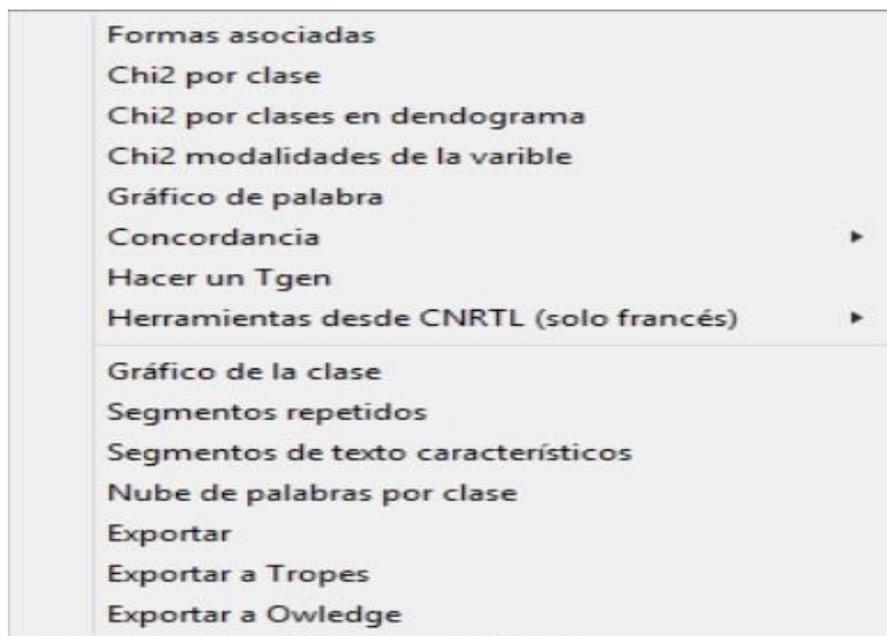


Figura 13 Menú suplementario perfiles

- **Formes asociadas:** muestra los efectivos de las formas reagrupadas en un lema de la clase.
- **Chi2 por clase:** crea un gráfico que presenta la asociación de la forma a cada una de las clases. Pueden seleccionarse varias formas al mismo tiempo, teniendo presionado el botón de CTRL.

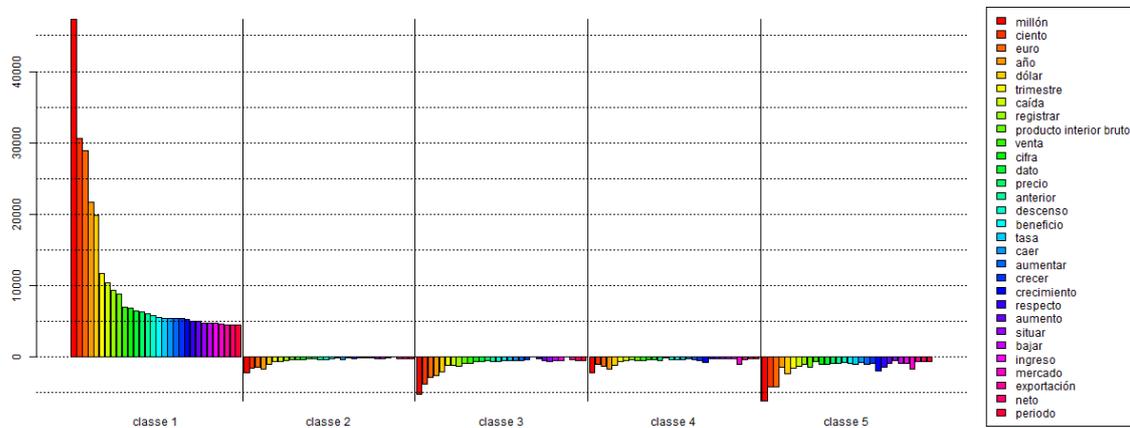


Figure 14 Chi2 por clase

- **Chi 2 por clase en dendograma :** retoma el mismo gráfico que el de la opción anterior pero proyectado sobre un dendograma.
- **Chi2 modalidades de la variable:** esta opción crea un gráfico que representa el chi2 de asociación de las modalidades de la variable seleccionada a cada una de las clases.
- ➔ Necesita que le demos el formato de « *variable\_modalidad* ».
- **Gráfico de palabra:** realizara un gráfico de similitud representando los enlaces de la forma seleccionada con las otras formas de la clase.

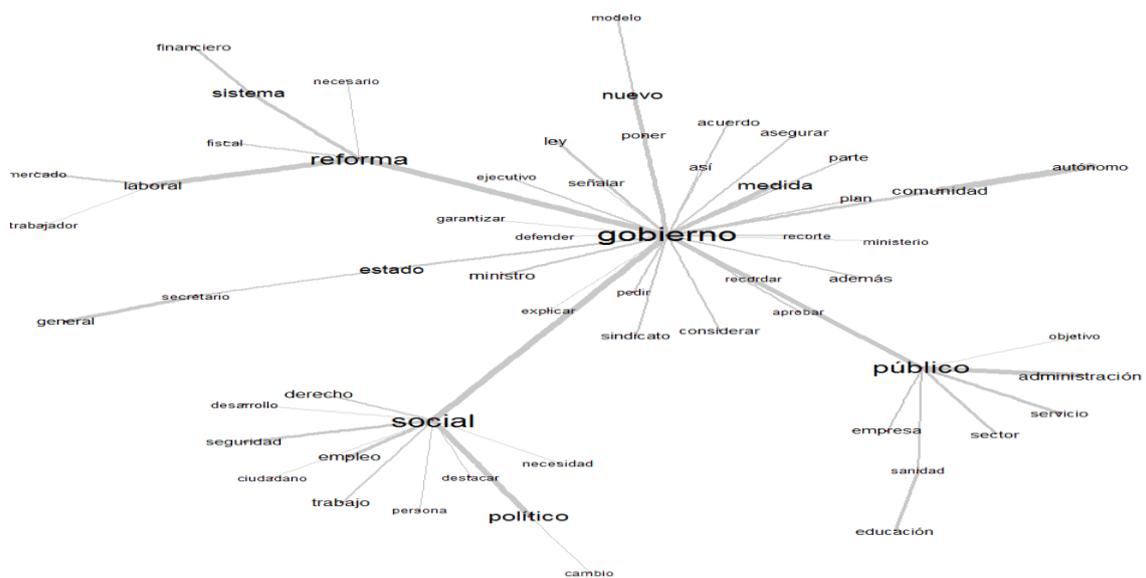


Figura 15 Gráfico de palabras

- **Concordancia:** propone hacer la concordancia de las formas y lemas seleccionados. Está disponible también para los segmentos de texto de la clase, así como para todas las clases del corpus.
- **Hacer un Tgen:** un TGEN es una reagrupación de formas o lemas que son tratados conjuntamente como un todo.
- **Herramientas desde CNRTL:** analiza la base de datos del Centro Nacional de Recursos Textuales y Léxicos (<http://www.cnrtl.fr/>) a partir de la forma o lema seleccionado, para realizar este análisis necesita estar conectado a Internet y solo funciona en francés.
- **Gráfico de la clase:** se trata de un análisis de similitud llevado a una tabla de ausencia y presencia (0/1) que cruza las unidades elegidas en la línea y las formas activas de la clase en una columna. La matriz de similitud está construida en columnas.
- **Segmentos repetidos:** mostrará los efectivos y tamaños de los segmentos repetidos de la clase.
- **Segmentos de texto característicos:** hace una lista de los segmentos de textos característicos de la clase, proponiendo dos medidas: "*Absoluto*", en la que los segmentos de textos son clasificados en función de la suma de los Chi2 y la relación a la clase de las formas activas que contiene y "*Relativo*", en la que los segmentos de texto son clasificados en función de la media de los Chi2 y el enlace a la clase de formas activas que contiene.
- **Nube de palabras por clase:** permite realizar una nube con el léxico que seleccionemos dentro de cada clase.
- **Exportar:** exporta todos los segmentos de texto presentes en la clase, puede incluso constituir un nuevo corpus.
- **Exportar a Tropes:** exportación al programa de análisis textual Tropes.
- **Exportar a Owledge:** exportación al programa OwlEdge.

El método Reinert no acaba aquí, si seleccionamos el análisis y pulsamos el botón derecho del ratón sobre el archivo que ha generado para realizar este análisis, podremos obtener más opciones.

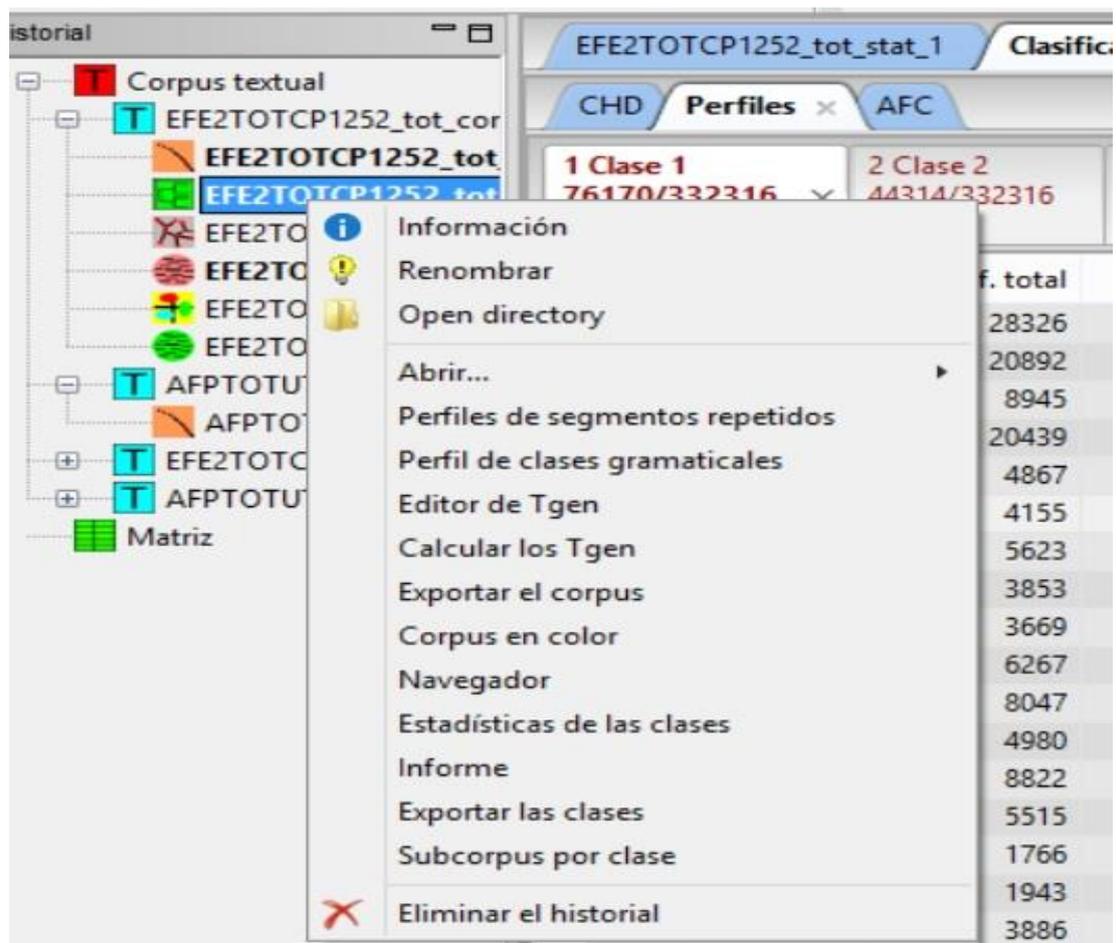


Figura 16 Opciones clasificación Reinert

- **Información:** retoma las características elegidas para la clasificación.
- **Abrir – anti-perfiles:** recoge las formas significativamente ausentes en la clase.
- **Perfiles de segmentos repetidos:** calcula los segmentos repetidos.
- **Perfiles de clases gramaticales:** calcula los perfiles de los tipos gramaticales.
- **Editor de Tgen**
- **Calcular los Tgen**
- **Exportar el corpus:** esta función permite exportar el corpus separado en segmentos de texto y añadiendo una variable con asterisco a la clase del segmento de texto. Los segmentos de texto aislados de la clasificación aparecerán con la variable \*clase\_0.
- **Corpus en color:** esta función crea un fichero .html exportando el corpus entero y cada segmento aparecerá del color de su clase. Los segmentos de texto no clasificados son escritos en negro y el fichero creado es accesible en el repertorio del análisis.
- **Navegador:** matriz que retoma todas las formas tomando su chi2 como enlace de cada clase.
- **Estadísticas de la clase:** esta función crea un fichero .csv en el repertorio del análisis (*stat\_par\_clase.csv*) incluyendo los siguientes registros en cada una de las clases : el

número total de ocurrencias, formas diferentes, hapax, segmentos de texto y el ratio de número de hapax/número de formas.

- **Informe:** crea un fichero rapport.txt que tiene las estadísticas generales de la clasificación y los perfiles.
- **Exportar las clases:** nos permite llevar a otro lugar las clases obtenidas.
- **Subcorpus de las clases:** crea un corpus a partir de la clase.
- **Eliminar el historial:** elimina este análisis del historial de Iramuteq.

## 4. ANALISIS DE SIMILITUD



La mayoría de los índices propuestos en este análisis tienen lugar en la librería « proxy » de R.

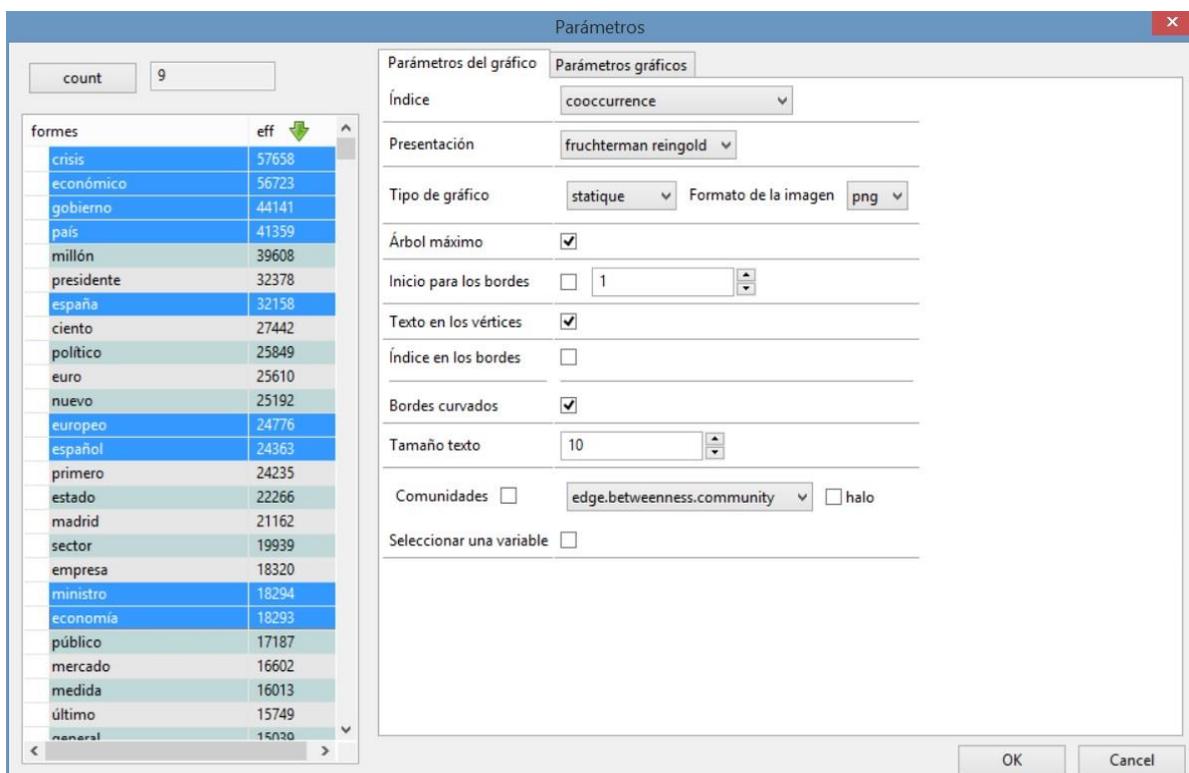


Figura 17 Parámetros análisis de similitud

Este menú se compone de tres partes:

1. **La lista de formas:** en el cuadro de la izquierda, están presentadas las formas y su frecuencia, por defecto aparecen todas seleccionadas para deseccionarlas basta con un doble clic de nuestro ratón, para hacer nuestra selección mantendremos CTRL seleccionado y clicando con nuestro ratón sobre las palabras deseadas estas se incluirán en nuestra lista de selección.
2. **Parámetros del gráfico:** esta pestaña se activa automáticamente cuando abrimos este menú y permite elegir las características del gráfico :
  - **Índice:** tipo de índice de similitud que queremos utilizar.

- Presentación: modo de presentación de los datos. Las opciones son aleatorias, para ello se recomienda ver la librería « igrph » para tener más detalles.
- Tipo de gráfico: en esta opción podremos elegir el método de presentación del gráfico, que podrá ser :
  - *Statique*: produce una imagen png o svg.
  - *Dynamyque*: utiliza la interfaz tk de igrph. Esta interfaz permite organizar los gráficos y conservar las coordenadas para la reutilización en un gráfico estático.
  - *3D*: produce un gráfico en tres dimensiones que se abre en una ventana rgl.
- + Formato de la imagen: permite guardar la imagen en .png o .svg
- + Árbol máximo: calcula el eje máximo.
- + Inicio para los bordes: con esta opción podemos eliminar los bordes en los que el valor del índice es inferior o igual al umbral elegido.
- + Texto en los vértices: fija el texto sobre los puntos álgidos de los gráficos.
- + Índice en los bordes: mete el texto sobre las bordes del gráfico.
- + Bordes curvados
- + Tamaño texto: indica el tamaño del texto sobre el grafico.
- + Comunidades: realiza un cálculo global (si queréis tener más detalles sobre ello id a la documentación de la *librería igrph*) y aproxima los conjuntos léxicos más próximos en cada bloque.
- + Halo: representa las comunidades por una agrupación de color.
- + Seleccionar una variable: se abre después de la validación de un nuevo menú, permitiendo elegir una variable o varias modalidades. Las formas características de modalidades se representan en color.

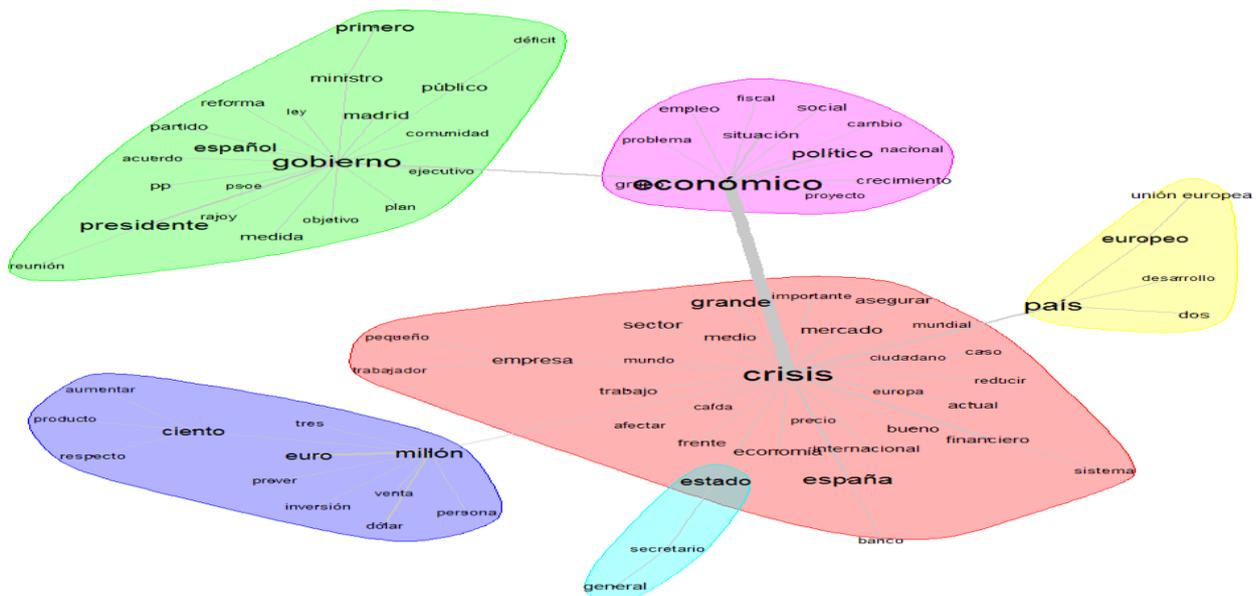


Figura 4 Ejemplo Análisis de Similitud con opción Halo seleccionada

La tercera pestaña que se abre en este apartado de análisis de similitud es la de **« Parámetros gráficos »**

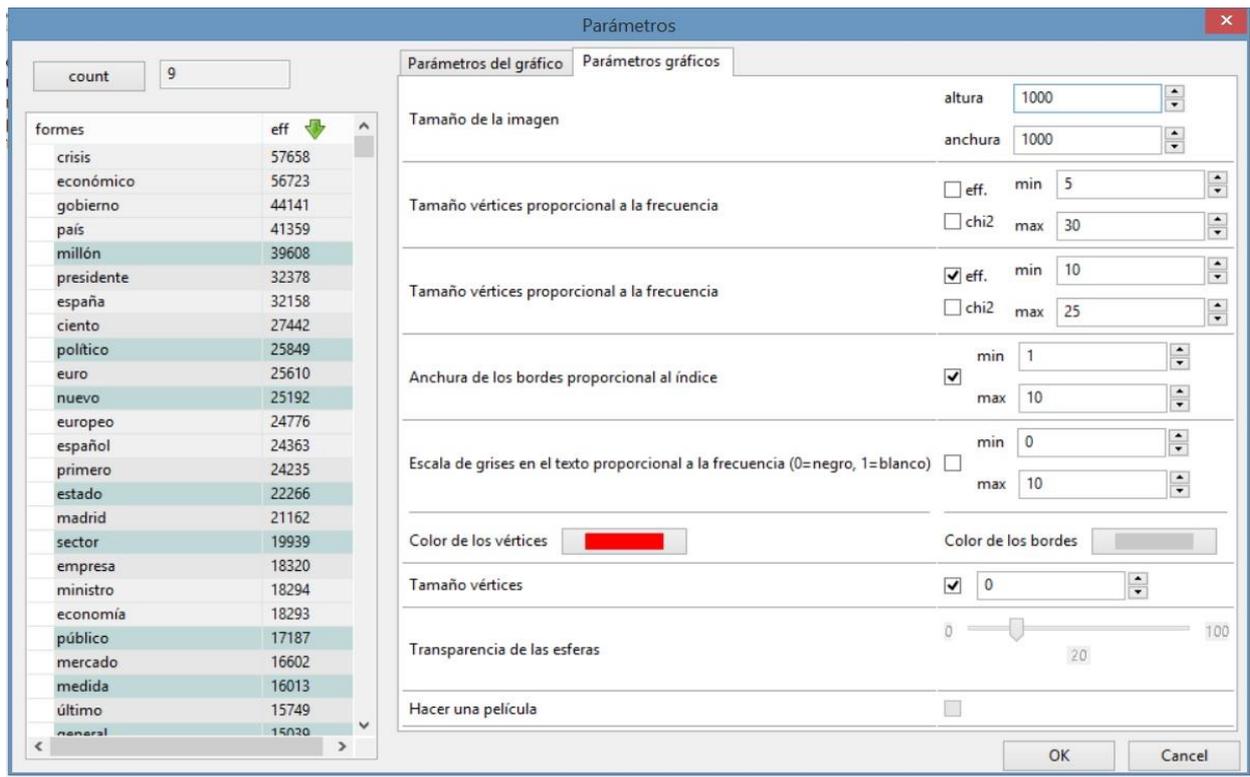


Figura 19 Análisis de Similitud “Parámetros Gráficos”

- **Tamaño de la imagen:** el tamaño de la imagen producida en píxeles. Seleccionaremos la altura y anchura de la misma.
- **Tamaño vértices proporcional a la frecuencia:** asigna un tamaño máximo proporcional al efectivo de la forma o a su chi2.
- **Anchura de los vértices proporcional al índice:** establece una anchura de los vértices proporcionales al valor del índice.
- **Escala de grises en el texto proporcional a la frecuencia:** ajusta el color del texto de los valores máximos en función del efectivo de la forma. Siendo 1 el valor para blanco y 0 para el negro.
- **Color de los vértices/ Color de los bordes:** selección de color de vértices y bordes.
- **Tamaño vértices:** tamaño único de los máximos del gráfico.
- **Transparencia de las esferas:** ajusta la transparencia de las esferas para los gráficos en 3D.
- **Hacer una película:** permite hacer una película representando un tour completo de un gráfico en 3D.

Y como siempre, una vez que hemos incluido los parámetros que queremos realizar a nuestro análisis, lo lanzaremos y obtendremos los resultados, que se abrirán en una nueva pestaña a la izquierda, presentados en dos iconos.



Este icono permite abrir de nuevo el menú de los parámetros conservando las últimas selecciones hechas.



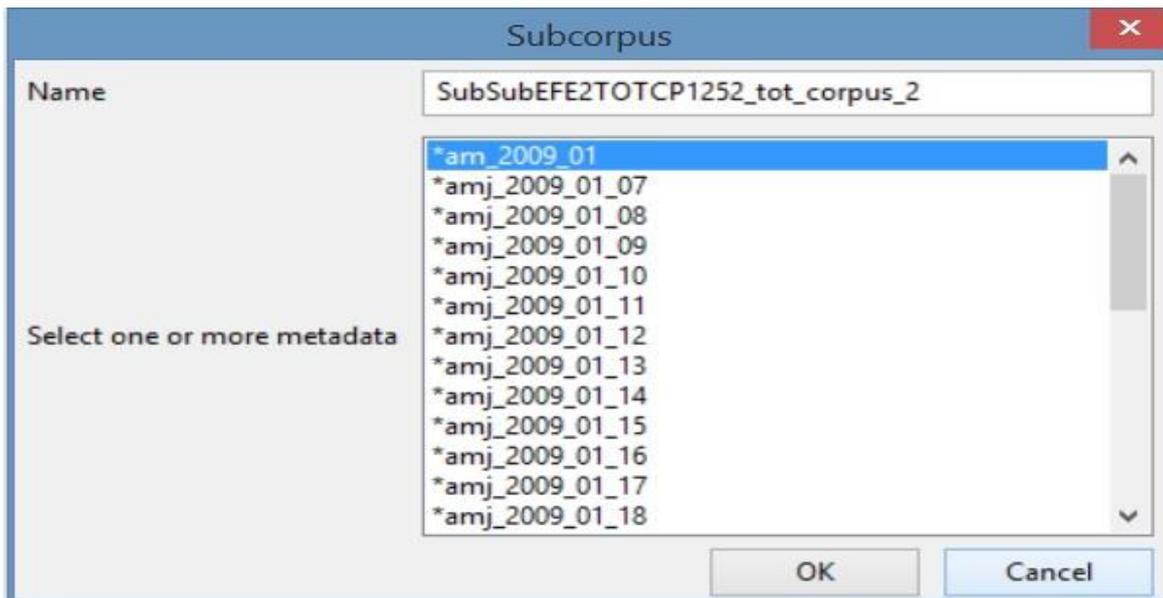


Figura 21 Subcorpus Metadatos

## 7. SUBCORPUS POR TEMÁTICA



En el caso de los corpus con temática, nos permite realizar un subcorpus de la temática que seleccionemos.



**INFORMACIÓN** Se está desarrollando una nueva versión de Iramuteq, por ello os recomendamos ver las informaciones que iremos actualizando en nuestra página:

<http://www.iramuteq.org/news>

**¡¡DISFRUTAR DE VUESTROS ANÁLISIS CON IRAMUTEQ!!**



## ANEXOS : INDICE IMÁGENES

Figura 1 Configuración previa al lanzamiento de un corpus.....	7
Figura 2 Lematización.....	8
Figura 3 Resultados análisis textual .....	9
Figura 4 Tabla de indexación.....	10
Figura 5 Opciones suplementarias .....	11
Figura 6 Elección variables AFC .....	12
Figura 7 Analisis de especificidades.....	13
Figura 8 Reedición AFC.....	13
Figura 9 Análisis Método Reinert .....	15
Figura 10 Dendograma CHD .....	16
Figura 11 Perfiles clasificación Método Reinert.....	17
Figura 12 Ejemplo AFC.....	18
Figura 13 Menu suplementario perfiles .....	18
Figure 14 Chi2 por clase .....	19
Figura 15 Gráfico por palabra.....	19
Figura 16 Opciones clasificación Reinert.....	21
Figura 17 Parámetros análisis de similitud.....	22
Figura 18 Ejemplo Análisis de Similitud con opción Halo seleccionada.....	23
Figura 19 ADS “Parámetros gráficos” .....	24
Figura 20 Preferencias de nube de palabras y ejemplo .....	25
Figura 21 Subcorpus Metadatos .....	26