

MANUAL PARA EL USUARIO



Mariola MORENO / Pierre RATINAUD

Diciembre 2022

ÍNDICE

▪ ¿Qué es IRaMuTeQ y para qué sirve?	3
▪ Instalación de IRaMuTeQ.....	4
▪ Codificación del corpus.....	5
▪ Primeros conceptos	7
▪ Insertar el corpus	7
▪ Análisis: Estadística Textual	11
▪ Análisis: Nube de palabras	14
▪ Análisis: Especificidades y AFC.....	14
▪ Análisis de Similitud.....	17
▪ Análisis: Clasificación Descendente Jerarquizada (CDH).....	20
▪ Subcorpus por metadatos o temática.....	29
▪ Miscelánea	29

▪ ¿Qué es IRaMuTeQ y para qué sirve?

IRaMuTeQ (Interfaz de R para el Análisis Multidimensional de los Textos y Cuestionarios) es un software desarrollado por Pierre Ratinaud (2009) en el seno del Laboratorio de investigación aplicada a las ciencias sociales de la Universidad de Toulouse (LERASS), y que es de gran utilidad para realizar investigaciones interdisciplinarias ya que permite realizar un análisis multidimensional de textos y cuestionarios, cualquiera que sea la temática de éstos.

Se trata de un software de libre acceso (distribuido bajo la licencia GNU GPL), la lógica de éste es una interfaz en wxPython con código en Python que desde el punto del análisis lexical se ocupará del tratamiento textual y de la generación de tablas clonadas y que va también a escribir scripts para el software R que va estar al origen de todos los análisis estadísticos y la producción de todos los gráficos que aparecerán en la interfaz, por lo que hay una interacción directa entre IRaMuTeQ y R.

Asimismo, este software es compatible con todos los sistemas operativos, lo cual permite que todos los investigadores puedan hacer uso de él sin problema, ya que se puede descargar directamente desde su página web: <http://www.iramuteq.org/telechargement>, en esta página el usuario podrá encontrar de manera detallada todos los pasos que se deben llevar a cabo para su instalación, y así pues, al ser un programa que funciona con el software R verá que será necesaria la instalación de este para que IRaMuTeQ pueda funcionar correctamente. Además, recomendamos al usuario visitar esta página web con cierta periodicidad ya que en ella se incluirán actualizaciones sobre el propio programa o informaciones susceptibles de ser interesantes para los usuarios.

A medida que nos adentremos más sobre los diferentes análisis que nos ofrece este programa, podremos comprobar que su utilidad es muy amplia y especialmente en trabajos de investigación con corpus de gran tamaño, donde tenemos un volumen de información bastante grande y difícil de abarcar manualmente, IRaMuTeQ nos ahorrará bastante tiempo y nos ayudará con el análisis de resultados.

De igual modo, es interesante tener en cuenta que se trata de una metodología que propone un doble enfoque, de una parte, cuantitativo ya que realiza análisis estadísticos, cálculos comparativos y en algunos análisis se toma como referencia la ausencia o presencia para trazar los resultados, pero por otra parte, es esencial el enfoque cualitativo ya que será labor del investigador la interpretación y contextualización de los datos para darle sentido a los resultados obtenidos.

En línea con lo expuesto en las líneas anteriores, podríamos decir que IRaMuTeQ nos permitiría:

- Ayuda a la interpretación de los datos sin alejarse del contexto del corpus, mostrándonos por coexistencias léxicas o palabras recurrentes y sobre-representadas.
- Detectar estadísticas descriptivas, obteniendo por ejemplo formas características, así como las temáticas presentes en el corpus.
- Reparar en la estructura, interacciones, distancia entre las formas del corpus, gracias a lo cual podríamos vislumbrar los campos lexicales de nuestro corpus, la cronología del discurso...
- Permite trabajar con corpus de diferentes idiomas (pero no de manera simultánea), lo que hace que la lengua del corpus no sea una barrera para poder trabajar con IRaMuTeQ. Además, permite que el propio usuario pueda modificar el diccionario, puede ser útil ya sea en lenguas que se encuentran insertadas en el programa de manera experimental o para enriquecer el diccionario existente de lenguas ya consolidadas.

▪ **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 2020 ha sido probada con R 4.0.3 en Windows y en Mac OS X.

Importante: Visitar la página web de IRaMuTeQ (<http://sourceforge.net/projects/iramuteq/>) y comprobar qué versión de R ha sido testada para la versión de IRaMuTeQ que va a instalar y así cerciorarnos de que todo vaya a funcionar correctamente.

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 y que el camino de R ha sido dirigido correctamente, abriremos IRaMuTeQ y nos dirigiremos a la pestaña de «*Edición*» -> *Preferencias* y pulsamos sobre el botón de "*Verificar la instalación*", en el caso de que alguna librería no se haya vinculado correctamente, esta verificación nos ayudará a marcarle el camino y que pueda instalarse correctamente.

Seguidos estos pasos, tendría que funcionar correctamente, no obstante, para cualquier duda o problema pueden lanzar su cuestión detallando el problema al fórum: iramuteq-users@lists.sourceforge.net.

▪ Codificación del corpus

Una vez que tenemos ya nuestro programa instalado también es importante tener el corpus con el que queremos trabajar correctamente codificado para que IRaMuTeQ pueda leerlo, y para ello es importante tener en cuenta una serie de cuestiones en cuenta:

- Es aconsejable utilizar editores de texto como: Openoffice, Libreoffice, Notepad++ y evitar el uso de Microsoft Office o bloc de notas para trabar en IRaMuTeQ.
- Tenemos que guardar nuestro documeto en txt, texto bruto, para ello es recomendable guardarlo 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 una pestaña que nos permite desglosar el « tipo » de formato en el que queremos guardar nuestro archivo, ahí tenemos que escoger la opción de “texto codificado”. Es importante tenerlo en cuenta, ya que más adelante veremos que a la hora de introducir nuestro corpus, una de las primeras preguntas que nos lance el programa será el formato en el que está registrado el documento. Es importante renombrar nuestro corpus con un nombre corto, ya que a la hora de realizar nuestros análisis IRaMuTeQ utilizará ese nombre para renombrar los diferentes ficheros y subficheros que se vayan generando, lo que provocará que si le otorgamos un nombre demasiado largo al corpus cuando hayamos realizado múltiples análisis y el programa tenga que generar sub-ficheros puede generar problemas para guardarlo.
- El inicio de cada uno de los textos que contenga nuestro corpus tiene que estar precedido por introducidos por cuatro asteriscos (****) seguidos por una o más variables, cada variable tiene que estar precedida por un asterisco un guión bajo y su modalidad.

Importante: para la codificación de los metadatos 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 del guión bajo, que será el método de separación entre variable y modalidad.

Véamos un ejemplo de cómo debería codificarse el corpus:

```
**** *VAR1_MOD1 *VAR2_MOD2
```


Hyperbase², en los cuales podremos trabajar directamente con el mismo corpus sin cambiar la codificación de éste.

▪ **Primeros conceptos**

Antes de adentrarnos en los diferentes análisis que nos ofrece IRaMuTeQ, resulta interesante detenernos en ciertos conceptos que veremos de manera recurrente en muchos de nuestros resultados:

- **Ocurrencia:** secuencia de caracteres delimitados por dos caracteres (sea un espacio o un carácter de puntuación): Nos dará el tamaño de nuestro corpus.
- **Forma:** secuencia de dos ocurrencias idénticas: será el índice de nuestro corpus y veremos que se dividirán en Activas – Suplementarias.
- **Lematización:** es una opción del programa, lo cual significa quitar el género, número y conjugación a las formas de nuestro corpus.
- **Segmento de texto:** será una unidad tenida en cuenta en algunos análisis de IRaMuTeQ, aproximadamente están formado por 40 formas, el software tratará de guardar lo máximo posible el lenguaje natural, por lo que tendrá en cuenta signos de puntuación fuertes como el punto o signos de interrogación/exclamación para cortar el texto. Los segmentos nos permitirán reparar en los co-textos del corpus, es decir, el léxico que gira alrededor de cada forma.

▪ **Insertar el corpus**

Una vez que ya hemos codificado nuestro texto teniendo en cuenta los aspectos detallados en el apartado anterior, podemos insertar el corpus en IRaMuTeQ. Para ello, iremos a la pestaña de “Abrir un Corpus”, seguiremos los siguientes pasos:

² <http://hyperbase.unice.fr/hyperbase/>

1. Seleccionamos la opción **Abrir un Corpus**: buscamos nuestro archivo en la carpeta que lo tengamos guardado y hacemos clic en « abrir ». Inmediatamente se nos abrirá una pestaña para elegir nuestras preferencias, la primera ventana será la de “General”:

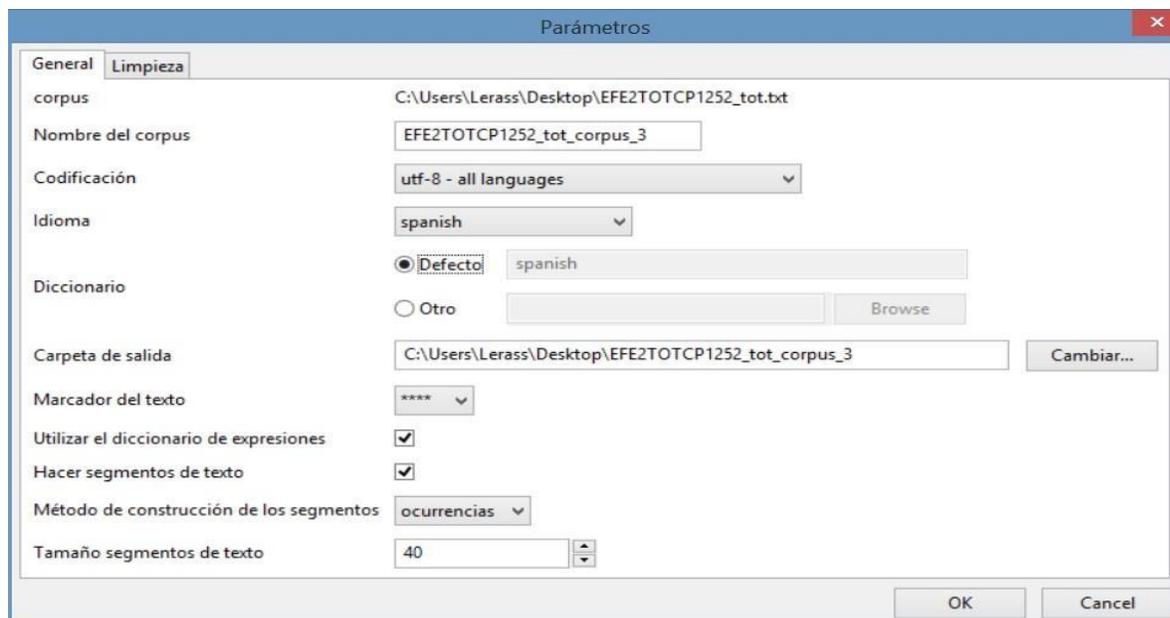


Ilustración 1: Parámetros generales al insertar un corpus

2. **Codificación**: si hemos registrado nuestro corpus en UTF8 (es lo más aconsejable), debemos seleccionar la opción de « UTF8 all languages ».
3. **Idioma**: tenemos que indicar la lengua en la que está escrito el texto de nuestro corpus.

Importante: IRaMuTeQ no puede analizar simultáneamente varias lenguas, para ello introduciremos un corpus para cada uno de los idiomas que queramos analizar. Es interesante tener en cuenta que el programa permite trabajar en todas las lenguas, el único inconveniente es que si el idioma en el que queremos trabajar no está en la lista que se despliega en IRaMuTeQ no tendremos la opción de lematizar el corpus ni podremos ver la categoría gramatical de las formas.

4. **Diccionario**: se podrá elegir entre el diccionario que nos marca por defecto el programa o también se podrá seleccionar “otro” para buscar el archivo del diccionario con el que queramos trabajar.
5. **Carpeta de salida**: indica la carpeta en la que se almacenarán todos los archivos de los análisis que realicemos en IRaMuTeQ.
6. **Marcador de texto**: ****. No modificaremos esta configuración que viene dada por defecto con los cuatro asteriscos.

En esta segunda pestaña nos encontraremos con las opciones siguientes, las cuales de manera general no haría falta modificar:

- ***Pasar el texto a minúsculas:*** al activar esta opción dejaríamos todo nuestro texto en minúsculas.
- ***Eliminar los caracteres no incluidos en esta lista:*** en esta lista aparecen todos los caracteres alfanuméricos existentes, no obstante, si hay alguno de su lengua que no encuentra presente lo puede añadir, al igual que puede eliminar todos aquellos que no considere necesarios. En el caso de una lengua que no tenga caracteres alfanuméricos debería modificar obligatoriamente esta lista para poder realizar su análisis bien.
- ***Sustituir apóstrofe por espacio:*** sustituir apóstrofes por espacios.
- ***Sustituir guión por espacio:*** sustituir guiones por espacios.
- ***Conservar la puntuación:*** en el caso de seleccionar esta opción tiene que verificar la ausencia del punto y coma en su corpus.
- ***Sin espacio entre las formas:*** si lo selecciona, IRaMuTeQ solo tendrá en cuenta el espacio como separación de las formas.

Tras haber realizado ya la selección en esta primera pestaña podemos es el momento de dar OK para lanzar la creación de nuestro corpus. Solo habrá una pequeña ventana que se abrirá antes y es otra de « ***preferencias*** » sobre la lematización y configuración de las claves del análisis.

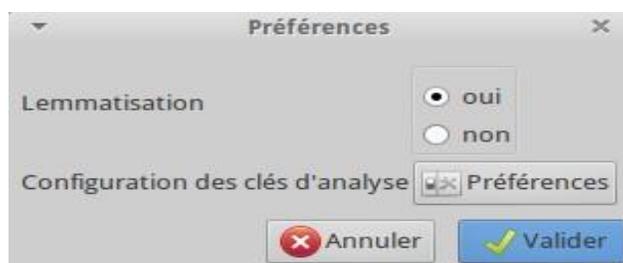


Ilustración 3: Opción de lematización del corpus

Recordemos que lematización significa que los verbos se pasan al infinitivo, los nombres al singular y los adjetivos al masculino, en definitiva, consiste en quitar el género, número y conjugación a las palabras (los diccionarios de los diferentes idiomas están disponibles en el repertorio de iramuteq/dictionnaires). Será elección del investigador el seleccionar o no esta opción, dependiendo de las preferencias y objetivos de su análisis.

Además de la opción de lematizar también tenemos otra opción que se llama “*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 tanto 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 repertorie.iramuteq).

Una vez que tenemos estos primeros detalles definidos le daremos a « *Valider* » y automáticamente comenzará a producirse la indexación de nuestro corpus, este proceso llevará unos minutos, dependiendo del tamaño del corpus. Veremos que en la parte de izquierda de IRaMuTeQ aparece un nuevo archivo que se llamará del mismo nombre que le hemos dado a nuestro corpus, 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.

▪ **Análisis: Estadística Textual**

Uno de los primeros análisis que podemos lanzar es el de estadística textual, ya que nos dará un informe cuantitativo sobre nuestro corpus.

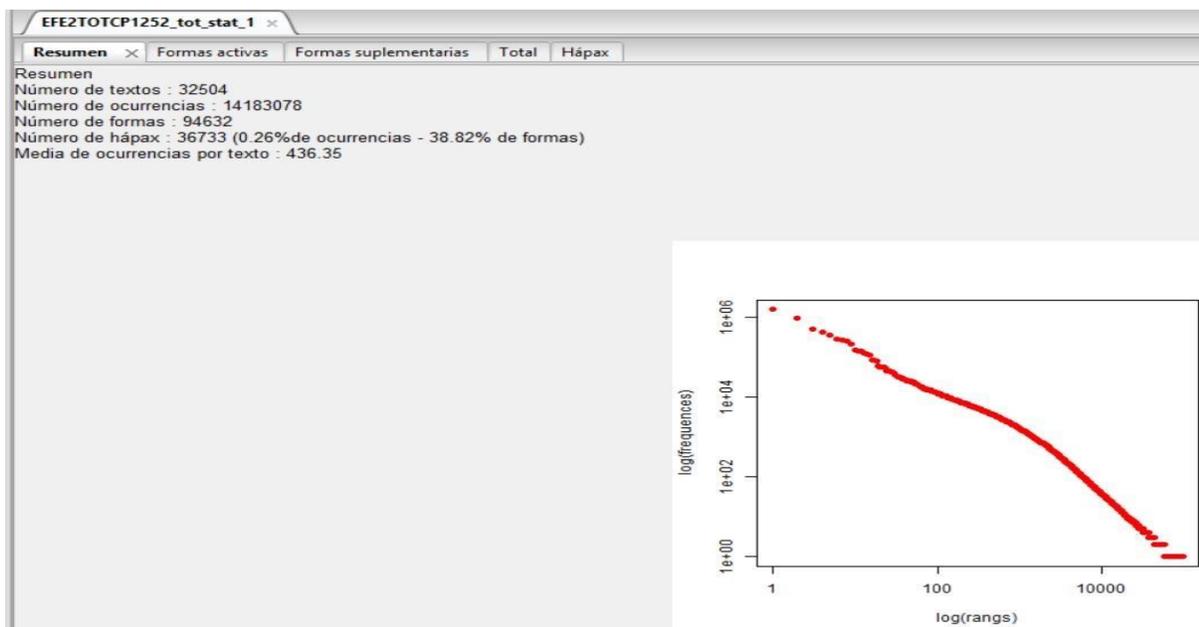


Ilustración 4: Análisis Estadística textual - resumen

En estos primeros resultados podemos ver que se despliega una ventana que contiene diferentes pestañas.

En la primera, llamada “*resumen*”, se indica:

- Número de textos: es el número de textos que tiene nuestro corpus.
- Número de ocurrencias: son el total de ocurrencias del corpus.
- 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.

Y a la derecha vemos un gráfico que nos muestra que el contenido del corpus respecta la distribución de frecuencia de la *ley de Zipf*, válido para todas las lenguas, 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.

En la segunda pestaña de esta ventana vemos “formas activas”:

Resumen	Formas activas	Formas suplementarias	Total	Hápax
Forma	Frec. ↓			
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
européo	24776			adj
español	24363			adj
primero	24235			adj
estado	22266			nom
madrid	21162			nom

Ilustración 5: Análisis Estadística Textual-formas activas

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.

Recordemos que de manera general las formas activas son los adverbios, adjetivos, sustantivos y verbos, mientras que las formas suplementarias son las preposiciones, las conjunciones, los verbos y las formas hápax son aquellas que aparecen una única vez en todo el corpus.

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:

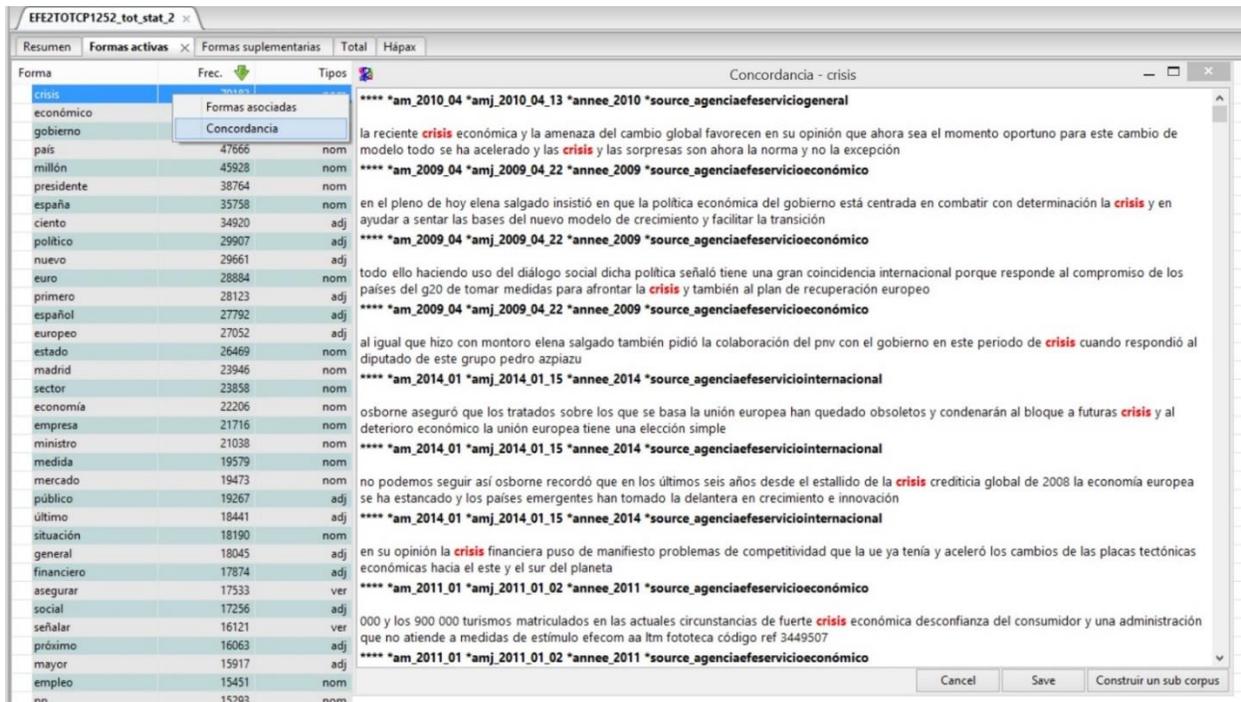


Ilustración 6: Análisis Estadística Textual - Formas asociadas / Concordancia

- **Formas asociadas:** aquí podremos visualizar todas las formas que se han lematizado respecto a la forma que hemos señalado. Por ejemplo, económico podría estar asociado, económica, económicas, económicos.
- **Concordancia:** en este caso podremos ver los segmentos de texto donde aparece la forma elegida en nuestro corpus.

serán elegidas por el usuario. Verá si una forma está representada o no desde un punto de vista estadístico comparando una parte con la integralidad de las otras partes del corpus.

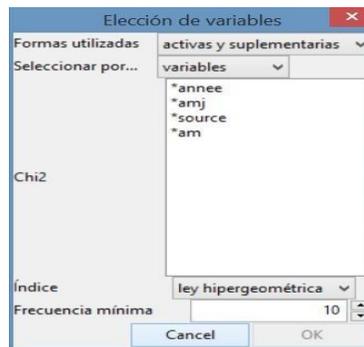


Ilustración 8: Análisis Especificidades y AFC-Elección variable

Al seleccionar el análisis de especificidades se despliega el menú de «elección de variables» que nos ofrece las siguientes opciones:

- Formas utilizadas: escogeremos las formas que queremos incluir para nuestro análisis: Activas y suplementarias / Activas / Suplementarias.
- Seleccionar por: Variable: permite seleccionar todas las modalidades de una variable / Modalidad: escogeremos solo las modalidades que nos interesen dentro de esa variable.

Importante: Es obligatorio seleccionar una variable o modalidad. Debe ser únicamente una.

- Índice: hay dos posibilidades de cálculo, la Chi2 o la ley hypergeométrica.
- Frecuencia mínima: Sólo se mostrarán las formas en las que el número de ocurrencias que estén por encima del umbral que marquemos en todo el corpus.

Una vez que lanzamos nuestro análisis factorial, obtendremos el número de efectivos de formas y tipos, como ilustra la imagen siguiente:

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	-141.9802	-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	

Ilustración 9: Análisis Especificidades y AFC – Resultados



Reedición de gráficos 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.

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: 1 10 100
Hacer una película:
OK Cancel

Ilustración 10: Análisis Especificiades y AFC - Reedición de gráficos AFC

En esta ventana vemos que aparecen diferentes opciones para poder editar nuestro gráfico factorial.

- 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 el 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.

▪ Análisis de Similitud

Otra de las opciones que nos ofrece IRaMuTeQ es el llamado análisis de similitud. Este análisis a diferencia de la nube de palabras, sí que representa las correlaciones existentes entre las formas, de modo que el lugar que ocupan y el lazo que vemos entre las formas en la imagen arbórea que obtendremos en nuestro análisis es importante. Es por ello que es muy útil para visualizar los espacios léxicos de nuestro corpus.

Al seleccionar este análisis veremos que se despliega el siguiente menú en la primera pestaña:

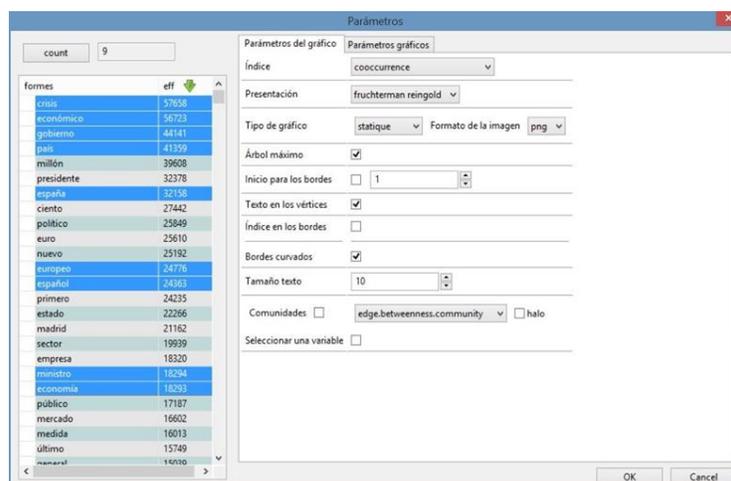


Ilustración 11: Análisis de Similitud - Menú

En el lado izquierdo se encuentran todas 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 que queramos incluir en nuestro análisis.

- Índice: tipo de índice de similitud que queremos utilizar, veremos que hay diferentes opciones.
- Presentación: modo de presentación de los datos, nos ofrecerá diferentes diseños arbóreos de presentar nuestro análisis.
- 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: incluye el texto sobre las bordes del gráfico.
- Bordes curvados: nos da la opción de seleccionarlo o no.
- Tamaño texto: indica el tamaño en el que queremos que aparezca el texto.
- Comunidades: realiza un cálculo global y aproxima los conjuntos léxicos más próximos en cada bloque, a modo de globos.
- Halo: representa las comunidades por una agrupación de color.
- Seleccionar una variable: permite elegir una variable o varias modalidades. Las formas características cada modalidad se representan en el color que se le otorga a cada una en la leyenda del gráfico y las palabras que aparecen en negro es porque no son características de ninguna de las modalidades.

En la segunda pestaña del menú de parámetros del gráfico vemos:

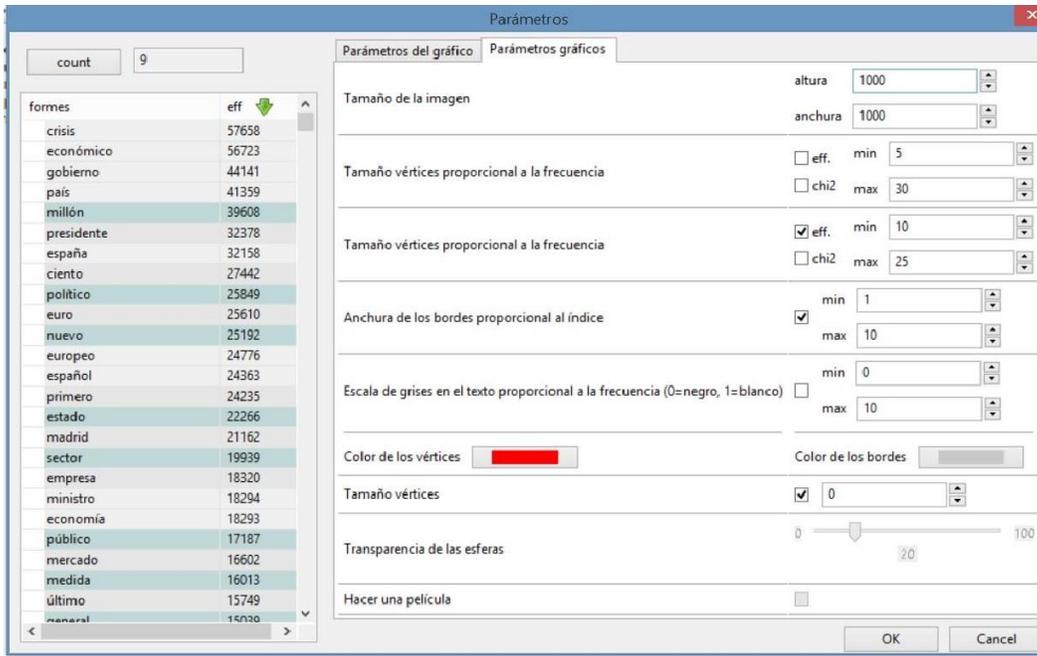


Ilustración 12: Análisis de Similitud - Menú (2)

- 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.

El icono de la izquierda permite abrir de nuevo el menú de los parámetros conservando las últimas selecciones hechas.

El botón de *Export* nos exportara el último grafico en formato graphml. Este gráfico puede ser abierto con los programas Gephi y Visone por ejemplo.

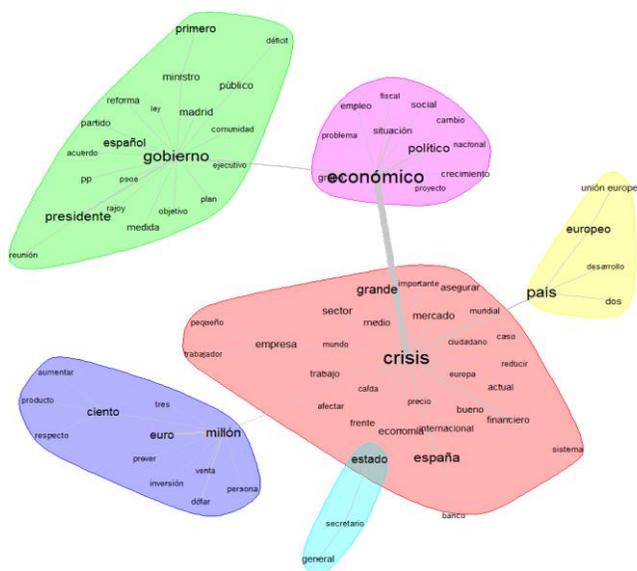


Ilustración 13: Análisis de Similitud - Ejemplo

▪ Análisis: Clasificación Descendente Jerarquizada (CDH)

Otro de los análisis que nos ofrece IRaMuTeQ es la clasificación jerarquizada de manera descendente método propuesto por Reinert (1983) e integrado en el software. El algoritmo de Clasificación Descendente Jerarquizada (CDH) permite definir las clases lexicales, partiendo de un corpus global para descender progresivamente hacia las clases terminales. El análisis parte de una matriz de presencia/ausencia que cruza las formas plenas retenidas y los segmentos de texto, cada línea es un segmento de texto y cada columna es una forma activa y en la matriz se codificará 1 si la forma está presente en el segmento de texto y 0 si está ausente, el objetivo de este análisis es reagrupar en forma de clase los segmentos de texto que contienen las mismas formas porque va a postular que si contienen las mismas palabras es porque hablan de lo mismo. Las clases terminales obtenidas son los conjuntos de segmentos de texto que tienen tendencia a compartir el mismo léxico. Es posible describir estas

clases a partir de su perfil lexical, es decir, a partir de las palabras que son estadísticamente sobrerrepresentadas en los segmentos que componen las clases.

Cuando seleccionamos este análisis se son despliega el siguiente menú de parámetros:

Parámetros	
Clasificación	<input type="radio"/> doble en RST <input checked="" type="radio"/> simple sobre segmentos de texto <input type="radio"/> simple en el texto
Tamaño de rst1	12
Tamaño de rst2	14
Número de clases terminales de la fase 1	10
Frecuencia mínima de segmentos de texto por clase	0
Frecuencia mínima de una forma analizada (2=automático)	2
Número máximo de formas analizadas	3000
método svd	irlba
Modo Patata (menos preciso, más rápido)	<input type="checkbox"/>
Cancel Valores por defecto OK	

Ilustración 14: Análisis CDH - Menú

- Clasificación: Doble en RST: reagrupa los segmentos de texto más próximos y realiza dos veces este análisis / Simple sobre segmento de textos: la clasificación llevara consigo los segmentos de texto / Simple en el texto: mantiene la integralidad 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. Cuando se trata de grandes corpus, este valor se puede aumentar para que el algoritmo pueda encontrar la heterogeneidad en el interior del corpus, o cuando es muy pequeño se puede disminuir.
- Frecuencia mínima de segmentos de texto por clase: determina un nivel mínimo bajo el cual, los segmentos de texto no serán tenidos en cuenta. El valor del tamaño mínimo de una clase es calculado dividiendo el número de segmentos de texto entre el número de clases que nosotros le demandamos.
- 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 más frecuentes. Pero este número puede ser modificado según el tamaño de tu corpus.

Importante: Una forma debe aparecer un mínimo de 3 veces para participar en los análisis.

- Método para 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: en el que veremos información sobre el análisis realizado en la clasificación. 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. En ésta se representa la partición, el tamaño y porcentaje de las clases

Vemos también el esquema de un dendograma compuesto por el número de clases que hemos solicitado, cuando vemos 0 es porque esa clase no ha sido retenida por ser demasiado pequeña. Y también vemos otro dendograma con las clases que se han determinado en nuestro análisis, en este caso, el tamaño de cada barra es proporcional al número de segmentos de texto que contiene. Clickando sobre el icono de la izquierda obtendremos un dendograma donde debajo de cada clase se le añade el vocabulario estadísticamente sobrerrepresentado en cada una de las clases (por χ^2 decreciente).

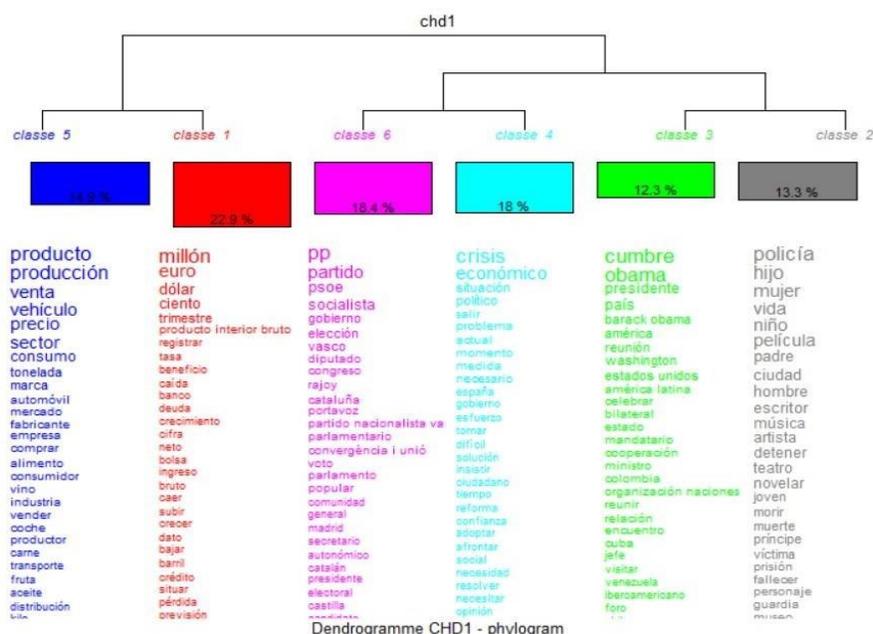


Ilustración 15: Análisis CDH - Dendograma

Otra de las pestañas que vemos en este análisis de la clasificación lexical es el de **perfiles**, aquí veremos el léxico que se ha incluido en cada clase, y veremos las formas sobrerrepresentadas en cada clase ordenadas por Chi2 decreciente, 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. Cada perfil contiene primero las formas activas en gris, le siguen las formas suplementarias en azul y por último los metadatos en naranja. Asimismo, vemos los siguientes datos en cada uno de los perfiles de las diferentes clases obtenidas:

CHD		Perfiles x		AFC		
1 Clase 1	2 Clase 2	3 Clase 3	4 Clase 4	5 Clase 5	6 Clase 6	
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

Ilustración 16: Análisis CDH – Perfiles

- Num: el número en el que está posicionada la forma.
- Eff. S.T: abreviación de efectivos de segmentos de texto que contiene la clase.
- Eff.total: son los efectivos totales de segmentos de texto contenidos en la clasificación final.
- Pourcentage: es el resultado de dividir eff. ST y eff Total.
- Chi2: algoritmo que expresa 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.

Astucia: 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.

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

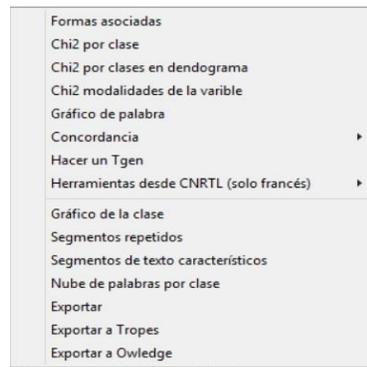


Ilustración 17: Análisis CDH – Menú suplementario perfiles

- Formas 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.

Astucia: Pueden seleccionarse varias formas al mismo tiempo, teniendo presionado el botón de CTRL.

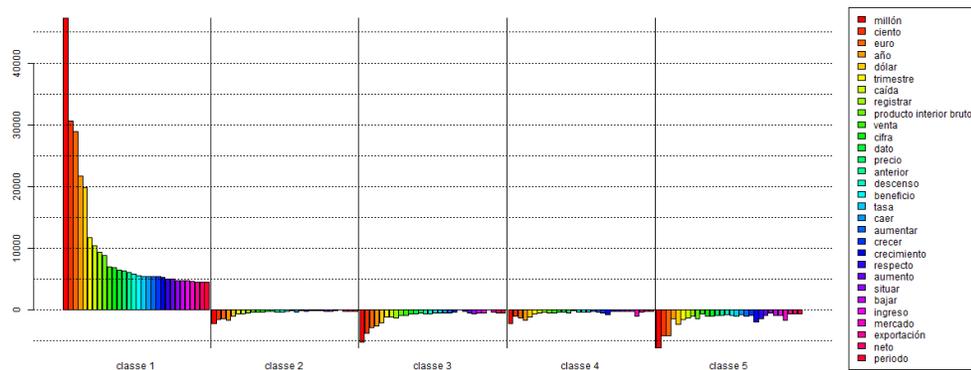


Ilustración 18: Análisis CDH - Perfiles: Chi2 por clase

- 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.

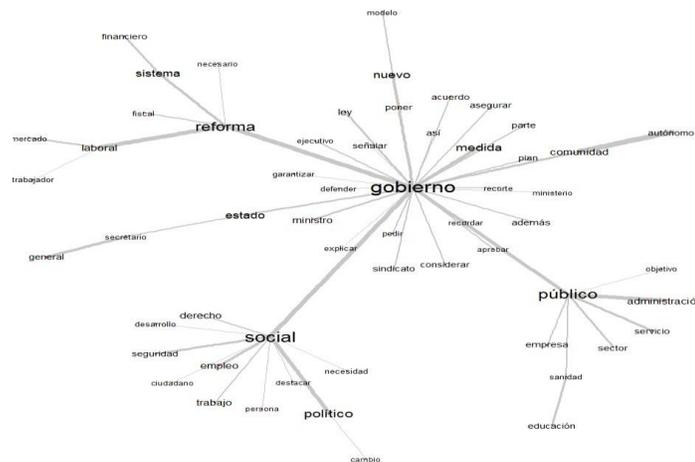


Ilustración 19: Análisis CDH - Perfiles: Gráfico de palabra

- 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. Da otra lectura de la composición de una clase. Consiste en realizar un ADS sobre los segmentos de texto de la clase.
- 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 más 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.

La tercera pestaña que vemos desplegada en nuestra clasificación lexical es la **AFC**: nos mostrara un análisis factorial correspondiente a la contingencia creciente de formas y lemas.

- Información: retoma las características elegidas para la clasificación.
- Abrir – anti-perfiles: recoge las formas significativamente ausentes en la clase. Es muy útil para hacer una lectura completa de las temáticas de nuestro corpus, ya que habrá formas completamente ausentes en algunas de las clases y ello nos dará indicios para interpretar y contextualizar mejor la clase.
- Perfiles de segmentos repetidos: calcula los segmentos repetidos. Es muy interesante ya que puede servir para que las formas pierdan ambigüedad y podamos verificar si el significado que le hemos otorgado a dicha forma corresponde con el que vemos en el corpus.
- 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 *classe_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 χ^2 como enlace de cada clase.
- Estadísticas de la clase: esta función crea un fichero .csv en el repertorio del análisis (stat_par_classe.csv) incluyendo los siguientes registros en cada una de las clases: el número total de ocurrencias, formas diferentes, hápax, segmentos de texto y el ratio de número de hapax/número de formas.
- Informe: creará 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. Es muy interesante en corpus de grandes dimensiones, ya que nos permitirá realizar todos los análisis disponibles en IRaMuTeQ pero con el contenido de la clase extraída.
- Eliminar el historial: elimina este análisis del historial de IRaMuTeQ.

▪ Subcorpus por metadatos o temática

Nos permite realizar un corpus a partir de una de nuestras variables o temática, y dispondremos de todas las opciones anteriores para realizar su análisis.

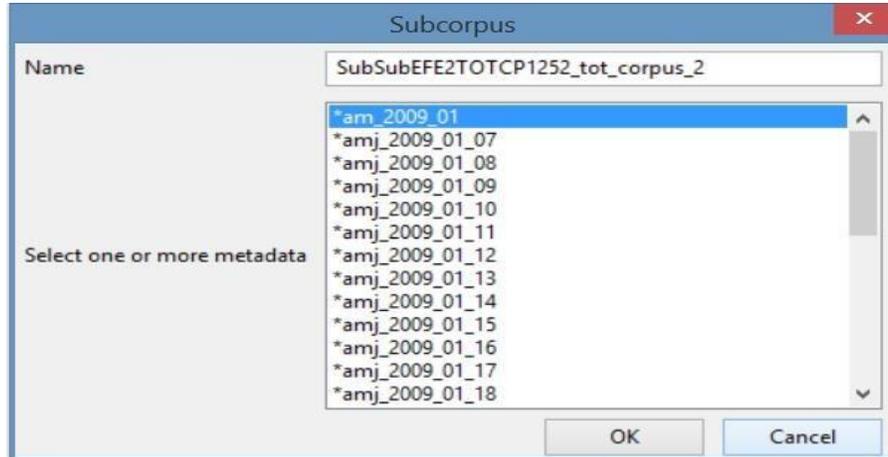


Ilustración 22: Análisis Subcorpus por metadatos – Menú

▪ Miscelánea

Es importante tener en cuenta que todos los archivos que se generen al realizar los análisis que realicemos en IRaMuTeQ se quedarán registrados en nuestro disco duro (la carpeta donde hemos señalado al insertar nuestro corpus).

Además, si seleccionamos alguno de los análisis realizados y clicamos el botón derecho del ratón veremos que aparece el siguiente menú:

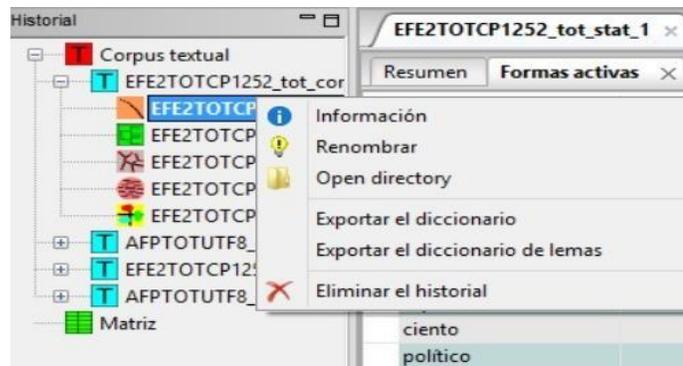


Ilustración 23: Menú análisis

- Información: permite recuperar las informaciones del análisis.
- Renombrar: esta opción es para cambiarle el nombre al análisis que hemos generado.
- Open directory: nos dirige a la carpeta donde se encuentran todos los archivos que hemos creado al realizar el análisis sobre el que estamos, en este caso por ejemplo sería el de estadística textual, por lo tanto en la carpeta encontraremos un archivo 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.
- Exportar el diccionario: genera un fichero « dictionary.csv » que recoge todas las formas. Para abrir el fichero correctamente, seleccionar la tabulación como separador.
- 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).

Importante: 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>

Algunos ejemplos de publicaciones en las que se ha usado IRaMuTeQ en la metodología de investigación:

- Beukes, E. W., Chundu, S., Ratinaud, P., Andersson, G., & Manchaiah, V. (2022). Experiential Characteristics among Individuals with Tinnitus Seeking Online Psychological Interventions: A Cluster Analysis. *Brain Sciences*, 12(9), 1221. <https://doi.org/10.3390/brainsci12091221>
- Figeac J., Smyrnaiois N., Salord T., Cabanac G., Fraissier O., Ratinaud P., Seffusatti, F. (2020). Information-sharing practices on Facebook during the 2017 French presidential campaign: an "unreliable information bubble" within the extreme right. *Communications: The European Journal of Communication Research*, 45(1), p. 648–670.
- Marchand P., Baroche C. (2021). Dynamiques émotionnelles dans la négociation policière de crise. *Le Travail Humain*, 84(2), 113-138
- Marchand, P., & Ratinaud, P. (2021). Discours politiques et expressions citoyennes : le poids des mots. *Explorateur*. <https://explorateur.univ-toulouse.fr/discours-politiques-et-expressions-citoyennes-le-poids-des-mots>
- Moreno M. (2016). Los mundos lexicales de la crisis en las agencias de prensa: una comparación Francia-España. In Pietrini D., Wenz K. dir., *Dire la crise: mots, textes, discours / Dire la crisi: parole, testi, discorsi / Decir la crisis: palabras, textos, discursos*. Frankfurt : Peter Lang, p.187-199.
- Moreno M. (2021). 40 Aniversario de la Constitución: así se habló en la prensa digital. En, Eiroa San Francisco, M. dir. *La Transición en directo: narrativas digitales de una historia reciente*. Madrid: Síntesis. p. 127-145
- Moreno, M., Ratinaud, P., Marchand, P. (2017). Le traitement de la crise économique par les agences de presse : une comparaison France/Espagne. *Mots. Les langages du politique*, 115, 51-72.
- Ratinaud P., Marchand P. (2015). Des mondes lexicaux aux représentations sociales. Une première approche des thématiques dans les débats à l'Assemblée nationale (1998-2014). *Mots. Les langages du politique*, 108, 57-77.
- Sebbah B., Sire G., Smyrnaiois N. (2020). Journalisme et plateformes : de la symbiose à la dépendance. *Sur le journalisme*, Vol. 9,1, 6-11.
- Smyrnaiois N., Tsimboukis P., Loubère L. (2021). La controverse autour de Didier Raoult et de sa proposition thérapeutique contre le Covid-19 sur Twitter : analyse de réseaux et de discours. *Communiquer. Revue de communication sociale et publique*, 32, 63-81.
- Smyrnaiois, N., & Ratinaud, P. (2015). Twitter : ce que les graphes nous disent de #CharlieHebdo. *Inaglobal*. <http://www.inaglobal.fr/sciences-sociales/article/twitter-ce-que-les-graphes-nous-disent-de-charliehebdo-8099>
- Souillard N., Sebbah B., Loubère L., Thiong-Kay L., Smyrnaiois N. (2020). Les Gilets jaunes, étude d'un mouvement social au prisme de ses arènes médiatiques. *Terminal*, 127.

Tabla de figuras

Ilustración 1: Parámetros generales al insertar un corpus.....	8
Ilustración 2: Parámetros de limpieza al insertar el corpus.....	9
Ilustración 3: Opción de lematización del corpus.....	10
Ilustración 4: Análisis Estadística textual - resumen.....	11
Ilustración 5: Análisis Estadística Textual-formas activas	12
Ilustración 6: Análisis Estadística Textual - Formas asociadas / Concordancia	13
Ilustración 7: Análisis nube de palabras - menú.....	14
Ilustración 8: Análisis Especificidades y AFC-Elección variable.....	15
Ilustración 9: Análisis Especificidades y AFC – Resultados.....	15
Ilustración 10: Análisis Especificidades y AFC - Reedición de gráficos AFC	16
Ilustración 11: Análisis de Similitud - Menú	17
Ilustración 12: Análisis de Similitud - Menú (2)	19
Ilustración 13: Análisis de Similitud - Ejemplo	20
Ilustración 14: Análisis CDH - Menú	21
Ilustración 15: Análisis CDH - Dendograma	23
Ilustración 16: Análisis CDH – Perfiles.....	23
Ilustración 17: Análisis CDH – Menú suplementario perfiles	24
Ilustración 18: Análisis CDH - Perfiles: Chi2 por clase	25
Ilustración 19: Análisis CDH - Perfiles: Gráfico de palabra.....	25
Ilustración 20: Análisis CDH – AFC	27
Ilustración 21: Análisis CDH - Opciones suplementarias de análisis	27
Ilustración 22: Análisis Subcorpus por metadatos – Menú	29
Ilustración 23: Menú análisis	29