

A modo de presentación.

El documento que ahora tiene el lector en sus manos es un intento por satisfacer, en castellano, la utilización por parte de estudiantes y profesores de un procedimiento de análisis de textos de forma automática, como es el programa de libre distribución **IRAMUTEQ**. Como se verá, en ello se ha primado el proceso de utilización y funcionamiento del software, pero introduciendo elementos conceptuales sobre el sustrato estadístico de los diversos análisis que se pueden llevar a cabo con el software que estamos introduciendo.

No se pretende que la exhaustividad sea un elemento que no se considere, pero se ha creído conveniente primar lo que se podría denominar un primer contacto con el programa y que éste fuese lo más exitoso posible, a partir de consideraciones fundamentalmente prácticas. Es por ello, que la presentación de los procedimientos a realizar con el software están expuestos y presentados a partir de las pantallas que proporciona cada análisis que contiene el programa.

Con esta visión se ha estructurado este documento de trabajo en 8 grandes apartados que son los siguientes: Un apartado que tiene que ver con los primeros conceptos a considerar al trabajar con este tipo de datos, los textos. Se incluyen aquí tanto conceptos como procedimientos estadísticos expuestos de forma conceptual con algunos ejemplos simples. El segundo y tercer apartado se adentran en explicitar la instalación del software y los parámetros de preferencias del programa. El cuarto y quinto, corresponden a la preparación del material, los archivos para procesar, y el cómo se capturan dichos archivos para ser analizados por Iramuteq. El sexto apartado, es un capítulo donde se encuentran todas las posibilidades de análisis de textos que tiene este programa. Finalmente, el séptimo y octavo se expone la selección de subtextos en función de las variables o modalidades que se deseen y cómo presenta el programa los outputs de los análisis demandados.

Índice

ALGUNOS CONCEPTOS BÁSICOS AL TRABAJAR CON IRAMUTEQ.	3
INSTALACIÓN DEL PROGRAMA.	7
PARÁMETROS DE PREFERENCIAS.	9
LA PREPARACIÓN DEL MATERIAL (FORMATO ARCHIVOS).	10
LA CAPTURA DEL ARCHIVO PARA SU ANÁLISIS.	14
ANÁLISIS DEL CORPUS.	17
Las frecuencias.	18
Análisis de especificidades.	23
Clasificación jerárquica descendente (método Reinert).	27
Análisis de Similitud.	36
Las Nubes de palabras.	39
LA SELECCIÓN DE CORPUS/SUBCORPUS.	40
LAS SALIDAS (OUTPUTS) DE LOS ANÁLISIS.	41
Terminología.	42
Referencias Bibliográficas.	43

ALGUNOS CONCEPTOS BÁSICOS AL TRABAJAR CON IRAMUTEQ.

IRAMUTEQ (Interfaz de R para el Análisis Multidimensional de los Textos y Cuestionarios) es un software libre, desarrollado por Pierre Ratinaud, en el seno del laboratorio LERASS, de la Universidad de Tolouse. El software permite realizar análisis multidimensional de textos de diferente naturaleza, como textos oficiales, páginas web, noticias, leyes preguntas de respuesta abierta de cuestionarios, etc. Especialmente útil para trabajos de investigación con un volumen de información difícil de trabajar de manera manual, este software facilita y ahorra tiempo en la interpretación de textos, mediante análisis lexicométricos.

Es un programa desarrollado con la pretensión de satisfacer las necesidades de las investigaciones sociales donde deben de analizarse materiales lingüísticos (textos).

A continuación se exponen algunos conceptos que creemos son básicos para poder trabajar con el software que nos ocupa:

.- El Corpus

El corpus es el conjunto de datos o textos científicos, literarios, informáticos, jurídicos, periodísticos, etc, que pueden servir de base a una investigación. Éste corpus lo construye el/la investigador/a, en base al objeto de estudio. Ejemplos de corpus podrían ser: el conjunto de transcripciones de entrevistas realizadas en una investigación; noticias aparecidos en diferentes diarios sobre una misma temática, o las respuestas abiertas registradas en cuestionarios sobre una misma temática, realizados a estudiantes de diferentes asignaturas.

Para que el análisis que vayamos a realizar tenga sentido, es necesario que el conjunto textual esté centrado en un temática principal o el objeto conceptual de investigación.

.- Los Textos

El corpus de texto está constituido por un conjunto de textos. La definición de cada uno de estas unidades dependerá de la naturaleza de la investigación. Retomando los ejemplos anteriores sobre el corpus de textos, encontraríamos que:

- En un estudio documental sobre noticias, cada una de ellas corresponderá a un texto.
- En un estudio con entrevistas, la transcripción de cada una de las entrevistas realizadas a diferentes sujetos de investigación correspondería a un texto. Ghiglione y Matalon (1989) recomiendan entre 20 y 30 textos, siendo 20 textos para cada grupo si se plantean estudios comparativos (Camargo, B. y Justo, A. 2013).
- En un estudio realizado a partir de respuestas en cuestionarios, el número de textos será igual. *Molina, J (2017)*.

.- Las tipologías de análisis de textos

Siguiendo el criterio de clasificación utilizado por Roy, N y Garon, R. (2013), que no es otro que el nivel necesario de participación por parte del usuario para llevar a cabo un análisis de datos textuales. De esta manera, podemos clasificar los diversos análisis realizados con textos en: **análisis automático, semiautomático y manual**.

El análisis automático es en el que el software hace la mayor parte de los análisis con un mínimo de intervención por parte del investigador. Este tipo de análisis, en muchos casos es necesario dedicar más tiempo a la preparación de los datos sin procesar e interpretar los resultados, que no el tratamiento en sí. Es un tipo de análisis de contenido que hace posible la realización de un análisis lexicométrico o estadístico textual (se buscan regularidades y correlaciones de formas gráficas específicas de un corpus). Dentro de este tipo tenemos los siguientes programas de texto: **SPAD, DTM_VIC, T-LAB, IRAMUTEQ, LEXICO3**, entre otros.

El análisis semiautomático requiere más intervenciones de los investigadores. El Software de este tipo permite un análisis lexicométrico, pero además un análisis de categorización semántica (tomando como referencia un diccionario, un tesoro o un grupo de palabras). Para ello, el investigador debe crear las relaciones entre ciertos temas y frases, generalmente escogidos en función del marco teórico-conceptual, y también en cuanto al contenido de las entrevistas. Por tanto se combina, una parte manual de agrupación o categorización con otra automática. Algunos programas que mencionan los autores son: **Sato** (Centro de análisis de textos computadora), **Semato** y **WordStat 6**.

El análisis manual, en el que el investigador tiene el control total sobre el análisis que lleva a cabo. La ventaja radica en la libertad que tiene el analista para obtener las unidades de codificación de significado, la clasificación de todos los datos obtenidos, además de ofrecer asistencia a la generación de informes de gestión y análisis (tablas, gráficos, informes relacionados con los datos, etc.). Estos tipos de programas pueden trabajar con texto basado en imágenes, sonido y/o video. Algunos de los programas del mercado son los siguientes: **ATLAS.ti 7**, **HyperRESEARCH**, **QDA Miner 4**, **MAXQDA 11**, **NVivo 10** y **Word Viewer** .

.- Los procedimientos de cálculo: Una aproximación conceptual.

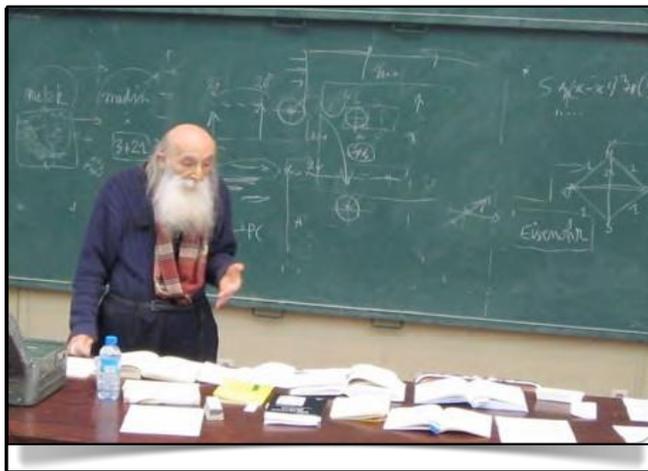
El Análisis de datos textuales es una aplicación de los métodos de análisis de datos desde la perspectiva de la escuela francesa, es decir, métodos de análisis multidimensionales exploratorios sobre datos lingüísticos. Las primeras aplicaciones fueron realizadas por Benzécri (1973), quien desarrolló el análisis factorial de correspondencias múltiples. Lebart (1994) continuó los desarrollos ante la necesidad de tratar preguntas abiertas con métodos más automáticos que la post-codificación manual que entonces se hacía y que en la mayoría de los casos aún se sigue realizando.

Actualmente nos podemos encontrar con múltiples nombres para denominar este tipo de análisis de textos. Como por ejemplo, textometría que se denomina también logometría o estadística textual y es la forma actual de la lexicometría (Lebart *et al.*, 2000). Esta área propone procedimientos de ordenamiento y de cálculos estadísticos para el estudio de un *corpus* de textos digitalizados.

Existen programas de tratamiento de datos textuales que utilizan las mencionadas técnicas, como es el caso del **PROGRAMA DTM_VIC**. Este programa es de descarga gratuita y se puede acceder a través de la siguiente dirección **WEB**: <http://www.dtmvic.com/>

Es un programa en el que el manejo no es fácil pero aporta una gama de análisis bastante amplia. En la Web se puede descargar el software, artículos y el manual en castellano.

Reinert (1998), propone el método Alceste para el análisis de textos. Que se diferencia del anterior por el algoritmo de cálculo que utiliza. Este es el tipo de algoritmo que utiliza el PROGRAMA **IRAMUTEQ**, que aquí trataremos de forma amplia.

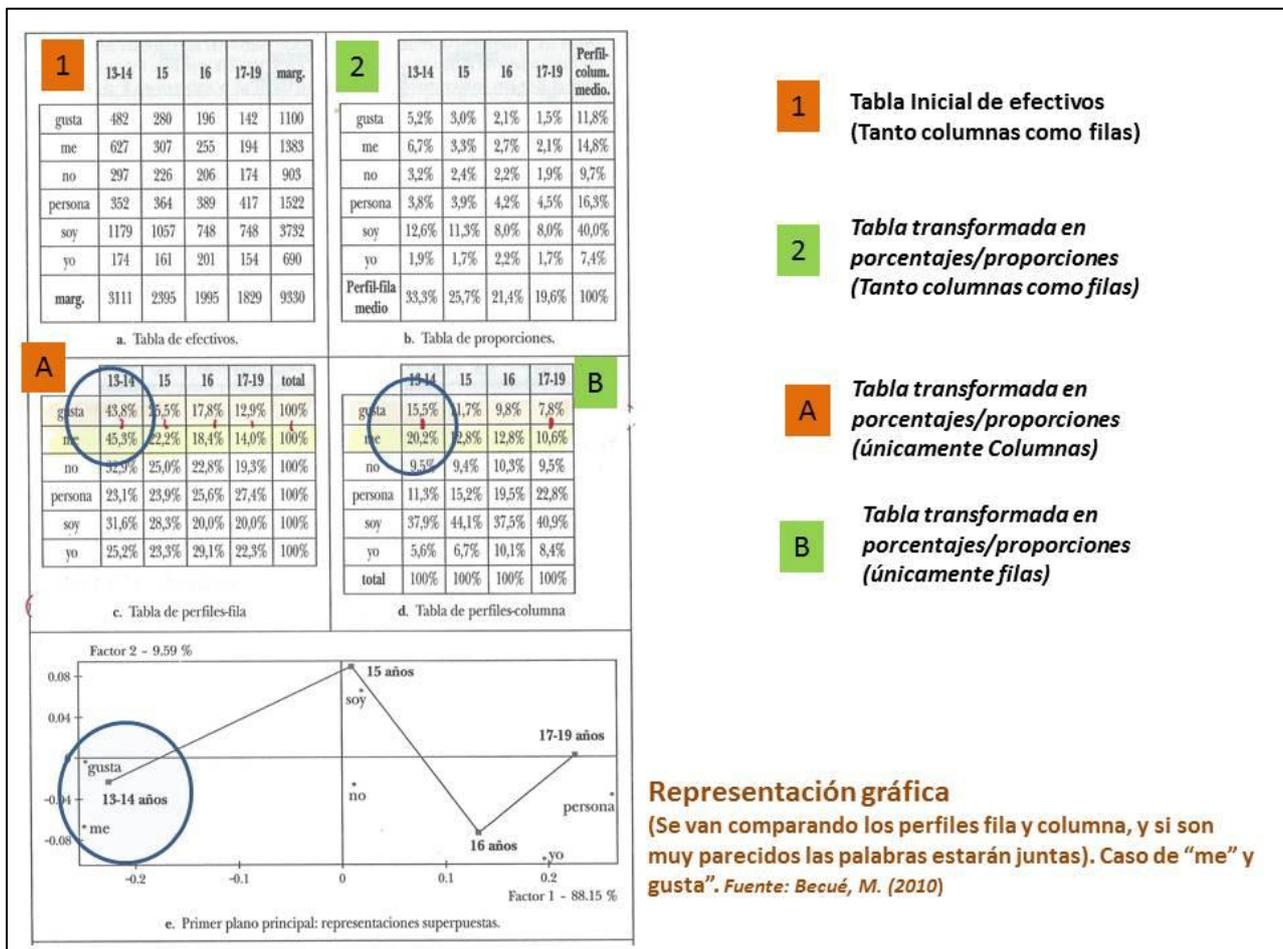


Jean-Paul Benzécri

Tanto un programa como el otro, como ya hemos indicado (se basan en cálculos multidimensionales sobre datos textuales), utilizan las técnicas de análisis de correspondencias y clasificaciones jerárquicas. Veamos a continuación como operan, a nivel conceptual, cada una de estas técnicas

El análisis de correspondencias opera mediante la comparación de los perfiles-columna y los perfiles-fila. Es decir se parte de una matriz inicial con los efectivos, para realizar transformaciones en porcentajes (proporciones), considerando la tabla original en dos que presentan las filas i las columnas, mediante el cálculo de Chi_cuadrado, nos proporciona un índice de distancia entre las diferentes categorías o variables consideradas. Así cuando menor distancia entre categorías o variables más juntas serán representadas en un espacio bidimensional (tal como muestra la imagen_1).

Imagen_1: Proceso conceptual del análisis de correspondencias múltiple



Representación gráfica
(Se van comparando los perfiles fila y columna, y si son muy parecidos las palabras estarán juntas). Caso de "me" y gusta". Fuente: Becué, M. (2010)

Recordar aquí que el análisis factorial de correspondencias múltiple es una técnica de reducción de información (al igual que el análisis factorial para variables escalares) que nos permite a partir de una tabla de frecuencias sintetizar las pautas de relaciones que se dan entre categorías/variables que serían muy difícil de ver a partir de los datos originales. Podríamos decir, tal como indica González (1991), que esta técnica nos responde a preguntas del tipo, ¿Qué categorías de una variable son más similares?, ¿Qué pares de categorías tienen perfiles diferentes?, lo cual no lleva a la creación de dimensiones o factores.

Para calcular la distancia entre las categorías/variables se utiliza la distancia de Chi-cuadrado, tal como hemos comentado, el principio seguido es el de **equivalencia distribucional**. Entendiendo por **Distribución** lo que indica el lingüista Harris, Z. (1954), que afirma que todo lo referente a la estructura de una lengua (sin recurrir al significado) puede obtenerse mediante los hechos distribucionales. Es decir, la distribución de una palabra serían todos los contextos posibles de la misma en un determinado corpus. De esta manera, el análisis de correspondencias analiza y sintetiza las características distribucionales de las palabras, **Becué (2010)**. En Ruiz (2013), se puede consultar un ejemplo de interpretación de análisis de correspondencias múltiple mediante el programa DTM_Vic.

El análisis de conglomerados o “clúster”, son una técnica de análisis que tienen como objetivo la formación de grupos con la característica de homogéneos entre los elementos que lo conforman, pero a la vez lo más diferentes a otros grupos formados. En el algoritmo de cálculo podemos encontrar dos grandes bloques de técnicas, por un lado los denominados algoritmos de clasificación automática (análisis “cluster”) y los de segmentación (análisis con variable criterio). La diferencia fundamental entre estos bloques estriba en que el análisis “cluster” en sus algoritmos no hay una diferenciación entre las variables, todas tienen un mismo estatus en el cálculo, mientras que en los de segmentación hay una variable que es la que actúa como diferenciadora de los grupos o clases (“cluster”) a partir del resto de variables que son las que la caracterizan, Gonzalez, 1991 y Lebart, I. Morineau, A. y Fénelon, J. (1985).

Dentro de la clasificación de técnicas automáticas, nos encontramos con **las jerárquicas** de las que podemos diferenciar dos tipos de algoritmos los **aglomerativos o ascendentes** y los **divisorios o descendentes**. Estas técnicas tienen la característica que cuando el objeto o variables es clasificado en una determinada etapa, no puede ser vuelto a clasificar, es por ello que existe otro tipo de procedimiento que no son jerárquicos, denominados algoritmos **dinámicos (caso del K-medias)**. Estos últimos procedimientos, permiten la reasignación en los grupos durante todo el proceso de formación de éstos.

En los programas de distribución libre que estamos comentando, en el caso del **DTM_Vic**, utiliza un procedimiento de clasificación de tipo **jerárquico ascendente**. En el caso de **IRAMUTEQ**, el procedimiento escogido es el **jerárquico descendente**. La diferencia entre estas dos maneras de hacer se sitúa en que el **jerárquico ascendente** se inicia con tantos grupos como individuos hay en la matriz de datos, por el contrario, el **jerárquico descendente**, su punto de partida es un solo grupo que incluye a todos los individuos, y en cada etapa realiza las subdivisiones en grupos/clases. Por tanto, uno funciona con un procedimiento de abajo a arriba (ascendente), y el otro de arriba abajo (descendente).

El procedimiento de cálculo implica la siguiente secuencia: Matriz de individuos por variables/categorías, Matriz de distancias entre individuos y matriz de similaridad entre los individuos. En el caso del software de tratamiento de datos textuales, se utiliza el Chi_cuadrado, como distancia y similaridad entre objetos (variables/categorías). Además se utilizan distintas maneras en la formación de los grupos, tal como muestra la Imagen_2. En la Imagen_3, se puede ver el procedimiento, mediante un ejemplo, de la formación de los grupos con una técnica de tipo descendente.

Imagen_2: Criterios formación grupos.

Formación de clases homogéneas. Clasificación o Cluster.

Herramienta de análisis multivariante exploratoria que permite agrupar los datos en clases o grupos con la característica de que dichos grupos son homogéneos **intra** pero heterogéneos entre ellos. Similitud entre los datos.

- 1.- Mide la distancia entre dos grupos como la que media entre los objetos de ambos más próximos.
- 2.- La distancia entre dos grupos es la distancia entre los objetos más distantes entre ambos grupos.
- 3.- La distancia entre dos grupos es la que media entre sus centros de gravedad.
- 4.- Para unir los grupos, se minimiza la pérdida de información ocasionada por la aglomeración (añadir al grupo), medida a través del aumento de varianza dentro de los grupos debido a la agregación.

Criterios para combinar grupos (medida de la distancia entre grupos).

1. Enlace simple:
2. Enlace completo:
3. Enlace promedio:
4. Método de Ward:

Var. Total = Var dentro + Var. entre grupos

Minimizar el aumento producido por la agregación

Fuente: González. (1991)

Imagen_3: Esquematización del proceso de formación de grupos.

Ejemplo: Consideremos 5 sujetos y dos variables. Utilizamos un algoritmo jerárquico ascendente y el criterio de salto mínimo para combinar grupo. (Adaptado de: González, 1991)

Paso_1:
Cálculo matriz de distancias (puede ser la Euclídea, Ji-Cuadrado etc.)

	1	2	3	4	5
1	0	8	10	25	1
2		0	2	5	13
3			0	9	13
4				0	34
5					0

Paso_2: Los objetos más similares (distancias más pequeñas entre ellos) son el 1 y el 5 (paso_1). Los agrupamos en un grupo, le denominamos objeto 6. Volvemos a calcular la matriz de distancias.

	2	3	4	6 (1+5)
2	0	2	5	8
3		0	9	10
4			0	25
6 (1+5)				0

Paso_3: Los objetos más similares ahora son: el 2 y el 3, ya que tienen la menor distancia (2). Al grupo (2 3) los ubicamos en un nuevo grupo, que llamamos 7. Volvemos a calcular la matriz de distancias.

	4	6 (1+5)	7 (2+3)
4	0	25	5
6 (1+5)		0	8
7 (2+3)			0

Dendograma

Distancia en la que se combinan los grupos

Se continúa el proceso hasta que quede únicamente un grupo. El gráfico que presenta las etapas de formación de los elementos en función de sus distancias correspondientes se le denomina dendograma.

.- Los procedimientos estadísticos concretos en el programa IRAMUTEQ.

En el software Iramuteq, se utiliza la siguiente secuencia de procedimientos para realizar los análisis con un corpus textual. El método principal se basa en el algoritmo ideado por Reinert (1983, 1990), éste primero fue parte fundamental en el software ALCESTE (*Analyse des Lexèmes Co-occurents dans les Enoncés Simplifiés d'un Texte*) que es, también, un software de análisis textual. Las etapas que utiliza, en el caso de IRAMUTEQ, son las siguientes:

- **Reducción de las unidades del corpus textual** (nombradas UCI: Unidades individuales de Contexto) en segmentos de texto (nombradas UCE: Unidades de Contexto Elementales). UCE representa un texto de dos o tres líneas. El tamaño de los UCE varía con arreglo al tamaño del corpus. El objetivo del análisis es proponer una clasificación de estas UCE.

- **Lematización:** Por defecto (de forma automática), las formas (las palabras) son lematizadas, es decir, que cada palabra es reemplazada por su forma canónica o su raíz (Lemaire, 2008).

- **Creación de unidades de contexto (UC):** UC es un conjunto de UCE que cuenta con un número mínimo de "formas activas " (verbos, nombres, adverbios, adjetivos) por oposición a "formas suplementarias" o "palabras-herramientas" (preposiciones, pronombres, adjetivos posesivos, ciertos verbos y adverbios frecuentes). Cada UC tiene como mínimo dos formas activas y se construyen 2 cuadros (tablas) de contingencia UC (en =filas) y las formas (en =columnas), uno para cada número mínimo).

- **Clasificación jerárquica descendente** sobre los cuadros (tablas) de contingencia creados sobre las unidades de contexto. En primer lugar, se lleva a cabo un Análisis Factorial de Correspondencias (AFC) sobre el primer cuadro (tabla) de contingencia, posteriormente las líneas se ordenan según sus aportaciones sobre el primer factor. Un segundo paso consiste en buscar a lo largo de este primer factor, la partición en 2 clases que maximiza la inercia inter-clase. Una vez realizado el paso anterior, un algoritmo de intercambio permuta cada línea de una clase a la otra y verifica la variación de inercia inter-clase. Este algoritmo es aplicado hasta que no haya más aumento de la inercia inter-clase. Finalmente, las formas específicas de una clase (definidas por el Chi2 de asociación en la clase) son retiradas de la otra clase. El análisis se realiza sobre la más grande de ambas clases y así sucesivamente hasta la obtención del número de clases definido por defecto (son 10 clases). El mismo paso es aplicado sobre otra tabla de contingencia, Delatre (2013), y así sucesivamente. En definitiva, es un procedimiento de clúster jerárquico descendente, lo cual implica partir de un solo grupo o clase, éste lo proporciona, en este caso, el análisis factorial de correspondencias, ya que incluye a todos los individuos, y en cada etapa se llevan a cabo las subdivisiones, que llegarían al final a representar a cada individuo. Este es un análisis muy interesante para poder ver estructuras textuales en un corpus de texto determinado.

Otro análisis concreto que lleva a cabo este software, es el denominado de similitud. **Es un análisis basado en la teoría de grafos (Un grafo es un conjunto de vértices -serían las palabras o formas- y aristas -las relación entre ellas-)** el propósito es el estudio de la proximidad y la relación entre los elementos de un conjunto, pero reduciendo el número de enlaces hasta llegar a "**un gráfico conectado sin ciclo**" (consiste en un camino cerrado en el que no se repite ningún vértice a excepción del primero que aparece dos veces como principio y fin del camino). Tal como se aprecia en la siguiente Imagen_4.

INSTALACIÓN DEL PROGRAMA.

Al ser un paquete informático de libre distribución asociado al programa R, su descarga es gratuita y cada uno de ellos se pueden encontrar en las siguientes direcciones web.

Programario R (Versión, 3.3.3): <https://www.r-project.org>

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To **download R**, please choose your preferred CRAN mirror.

If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

News

- **R version 3.4.0 (You Stupid Darkness) prerelease versions** will appear starting Tuesday 2017-03-21. Final release is scheduled for Friday 2017-04-21.
- **R version 3.3.3 (Another Canoe)** has been released on Monday 2017-03-06.
- **useR! 2017** (July 4 - 7 in Brussels) has opened registration and more at <http://user2017.brussels/>
- Tomas Kalibera has joined the R core team.
- The R Foundation welcomes five new ordinary members: Jennifer Bryan, Dianne Cook, Julie Josse, Tomas Kalibera, and Balasubramanian Narasimhan.
- **The R Journal Volume 8/1** is available.
- The **useR! 2017** conference will take place in Brussels, July 4 - 7, 2017.
- **R version 3.2.5 (Very, Very Secure Dishes)** has been released on 2016-04-14. This is a rebadging of the quick-fix release 3.2.4-revised.
- **Notice XQuartz users (Mac OS X)** A security issue has been detected with the Sparkle update mechanism used by XQuartz. Avoid updating over insecure channels.
- The **R Logo** is available for download in high-resolution PNG or SVG formats.
- **useR! 2016**, have taken place at Stanford University, CA, USA, June 27 - June 30, 2016.
- **The R Journal Volume 7/2** is available.
- **R version 3.2.3 (Wooden Christmas-Tree)** has been released on 2015-12-10.
- **R version 3.1.3 (Smooth Sidewalk)** has been released on 2015-03-09.

Software IRAMUTEQ: <https://sourceforge.net/projects/iramuteq/>

By using the SourceForge site, you agree to our use of cookies.

sourceforge Search [Browse](#) [Enterprise](#) [Blog](#) [Deals](#) [Help](#) [Log In](#) or [Join](#)

SOLUTION CENTERS [Resources](#) [Newsletters](#) [Cloud Storage Providers](#) [Business VoIP Providers](#) [Internet Speed Test](#) [Call Center Providers](#)

Home / Browse / Science & Engineering / Information Analysis / iramuteq

iramuteq
Brought to you by: pratinaud

Summary | Files | Reviews | Support | Wiki | Mailing Lists | Discussion

★ Add a Review
↓ 352 Downloads (This Week)
📅 Last Update: 2014-12-23

setup_iramuteq-0.7-alpha2.exe
[Browse All Files](#)

Recommended Projects

- TXM**
Unicode/XML-TEI text/corpus analysis platform
- 7-Zip**
A free file archiver for extremely high compression
- Text Analysis Markup System**

Top Searches

- iramuteq
- alceste
- iramuteq mac
- alceste (logiciel)
- evoc 2000
- mcda with gis
- .ost to .pst converter
- corpus linguistics
- batting

Description

IRAMUTEQ : Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires. Logiciel de traitement de données pour des corpus texte ou de type individus/caractères. Permet notamment de réaliser des analyses de type "ALCESTE"

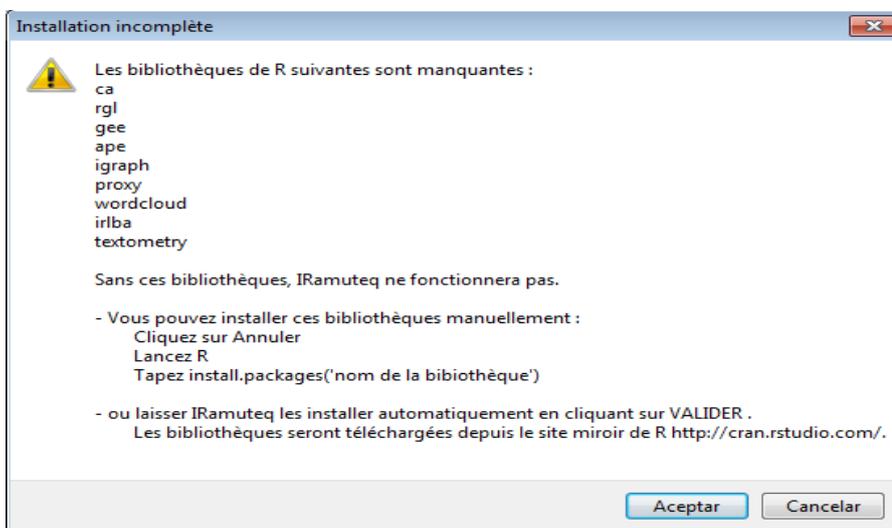
[iramuteq Web Site](#)

Categories [Information Analysis](#) [Linguistics](#) [Psychology](#) **License** [GNU General Public License version 2.0 \(GPLv2\)](#)

Las instrucciones y pautas de este documento se han verificado con la instalación de IRAMUTEQ en un entorno Windows (7 y 10)

Para poder realizar la instalación con éxito en primer lugar se ha de instalar el programario R y después el programa **IRAMUTEQ**. **- Atención es fundamental tener esta precaución a la hora de la instalación !!!!!!!!!!!.**

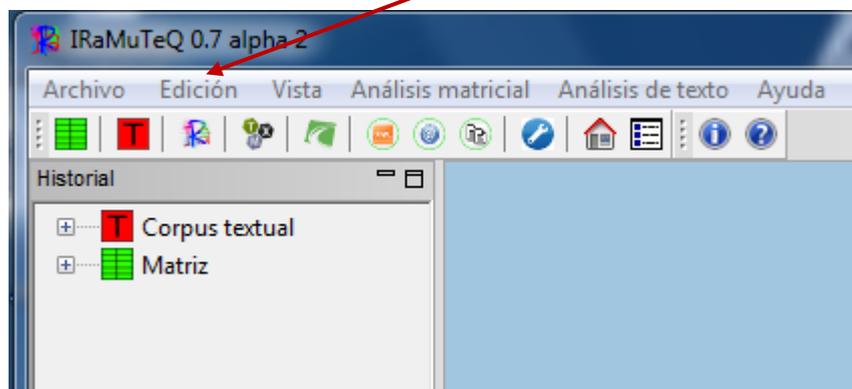
Una vez instalados los programas, la primera vez que se ejecuta **IRAMUTEQ**, nos aparecerá el siguiente mensaje, que nos indica la necesidad de instalar las librerías (les Bibliothèques), este primer mensaje está en francés (para cambiar el idioma se indica más adelante en este documento) del programa R, pulsando aceptar, se actualizarán de manera automática.



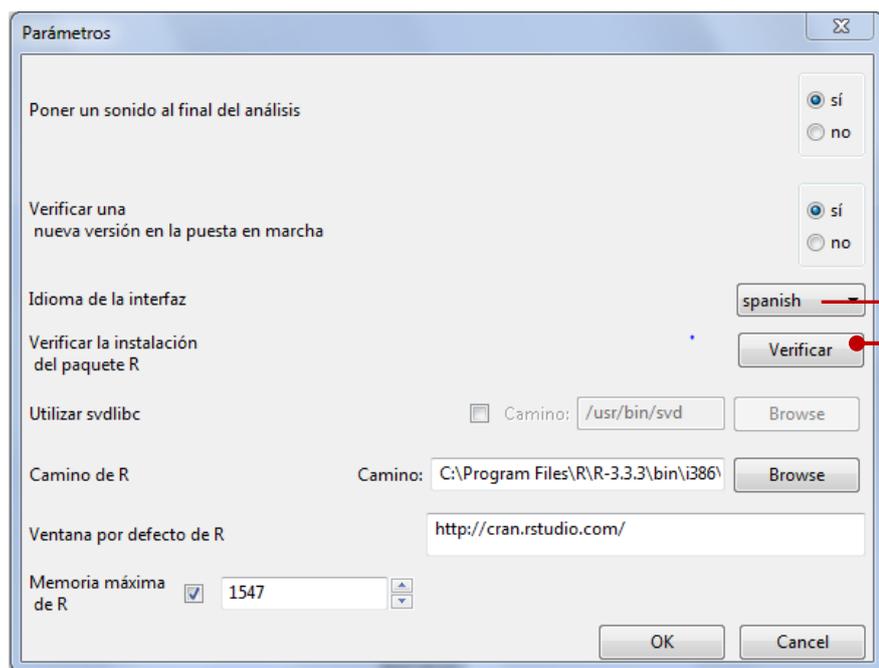
PARÁMETROS DE PREFERENCIAS.

La secuencia de activación es la siguiente:

Menú superior del programa: **Edición**



Preferencias



Cambio de idioma

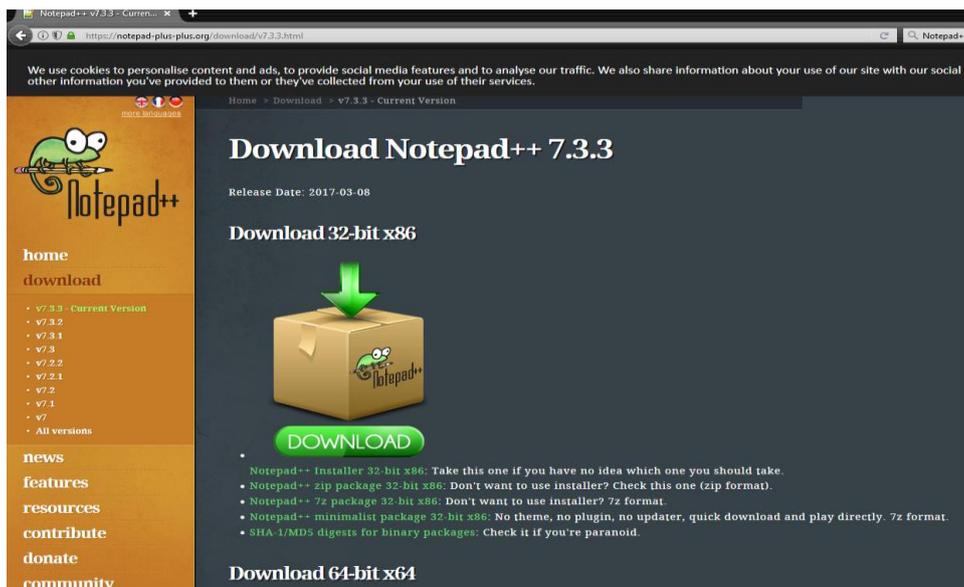
Verificación de la instalación de **R**. Se ha de verificar cuando se realiza la instalación por primera vez. Esto nos indicará si todas las librerías del programa R están instaladas para que se puedan llevar a cabo los diversos análisis que puede realizar el programa.

LA PREPARACIÓN DEL MATERIAL (FORMATO ARCHIVOS)

Como en todo programa informático se ha de preparar los datos a analizar para que puedan ser procesados por el software a utilizar. En el caso de **Iramuteq**, se trabaja con archivos **en formato UTF8** ((8-bit Unicode Transformation Format). Existen muchos programas gratuitos que nos permiten la conversión a éste formato.

Nosotros proponemos aquí el editor **Notepad+++** .

(disponible en: <https://notepad-plus-plus.org/download/v7.3.3.html>)



Respecto al formato específico en el que se ha de organizar el corpus a analizar, tenemos 3 opciones, que se pueden caracterizar de la siguiente forma, tal como muestran los 3 ejemplos:

(Recordar aquí que para el programa IRAMUTEQ, los valores de una variable los denomina “**modalidades**”). Por tanto, las tres maneras, de forma genérica, serían las siguientes:

Sin variables categoriales:

```
**** *VAR_MOD
texto texto
texto texto texto texto texto texto texto texto texto texto texto
texto texto texto texto texto texto texto texto texto
```

Una sola temática pero con variables categoriales (es como especificar el texto para cada “sujeto”:

```
**** *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 texto
```

Con variables categoriales y dos o más temáticas.

```
**** *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 texto
-*them_television
texto texto
texto texto texto texto texto texto texto texto texto texto texto
texto texto texto texto texto texto texto texto texto texto
```

Fuente: Moreno, M. y Ratinaud, P. (2015).

Ejemplos de textos analizados:**Opción_A: Cualquier texto sin variables categoriales asociadas.**

```
**** *Entrevista_Tesis
```

1.- ¿Por qué elegiste Pedagogía?, ¿Qué conocías de la Pedagogía?

Bueno, yo realmente a pedagogía llegué un poco de rebote, no fue mi primera elección de elegir, si que, yo había hecho el bachillerato científico con lo cual las primeras opciones que yo había puesto de varias universidades era biología, me llagaba la nota para bilogía de las diferentes universidades y no sé, serían cosas del destino, porque me dieron pedagogía, que además la tenía en séptima opción y la había puesto porque cuando yo pensaba en biología, pensaba en biología para explicar dentro de un centro educativo, no para ser biología en la parte práctica, en un zoo o en el ámbito de la natura, sino más pensando en la enseñanza pero de un tema concreto que me gustaba que era biología. Cuando no me lo dieron y me dieron la séptima opción, tenía la posibilidad de esperar una reasignación para ver si me daban bilogía o probar en la carrera de pedagogía para ver si se podía ajustar a lo que yo realmente quería hacer más relacionado con la docencia, si que en ese tiempo, mientras que entras en la universidad y esperas una reasignación o ver lo que haces, investigué más de pedagogía y fue lo que me hizo, finalmente, decidir que algo que había puesto en séptima opción no estaba tan alejado de lo que pretendía con la primera opción, que al final era estar dentro del ámbito educativo, aunque no tuviera que ser únicamente en el formal, y tratar con gente y poder enseñar cosas a gente, y fue lo que me ayudó a decidir que aunque fuese la séptima opción iba a probarlo y el primer año empecé a ver en las prácticas la sombra del pedagogo y todo esto que hacíamos entonces, fue cuando vi otra visión de la pedagogía que a lo mejor no era aquella que yo a lo mejor asociada más tradicional y me animó a continuar y la verdad es que luego termine cuarto y no me arrepentí en absoluto de haber escogido pedagogía.

Opción_B: Cualquier texto (con un único tema) con variables categoriales identificativas

**** *GRUP_1 *TORN_2 *EDAT_4 *ACC_3 *TREB_4 *EXPE_1

En el aula casi no habíamos hablado de innovación. No sabía qué tipo de texto esperaba el profesor. Esto en la Tasca1. La Tasca2 la encontré más motivadora y muy interesante como aprendizaje. No tuve tanta sensación de inseguridad. La Tarea 3 me hizo sentir de nuevo inseguridad. La encuentro muy interesante como reflexión y autoaprendizaje, pero creo que me faltan recursos digitales y tiempo para hacerla tan bien como quisiera.

**** *GRUP_1 *TORN_2 *EDAT_2 *ACC_2 *TREB_3 *EXPE_1

Creía que era una tarea difícil porque los artículos presentaban un lenguaje complicado. También es cierto que soy una persona que s'angustia enseguida y por tanto mostraba inseguridad, pero aun así, al final, estas tareas han sido un reto que pienso he superado positivamente.

**** *GRUP_1 *TORN_2 *EDAT_3 *ACC_2 *TREB_3 *EXPE_1

Ha sido complicado realizarlo y me he sentido angustiada y preocupada porque soy una persona que me cuesta mucho entender las lecturas y saber plasmarlo correctamente en una hoja. Cuando se me proponía me sentía insegura por si no sabía hacerlo bien.

**** *GRUP_1 *TORN_2 *EDAT_3 *ACC_1 *TREB_3 *EXPE_1

Porque tenía poco tiempo para hacer los trabajos, ya que trabajo. Aunque los conceptos trabajados en esta asignatura son importantes y está bien que los profesionales del mundo de la educación tengan una noción, creo que si la asignatura tuviera más horas podríamos asimilar mejor los conceptos que hemos aprendido. De esta manera los trabajos los hubieran realizado con más motivación y tranquilidad.

Opción_C: Cualquier texto con varios temas y variables categoriales asociadas.

**** *GRUPiNVEST_2 *TORN_2 *REDAT_4 *SEXE_2 *ACCESS_3 *TREBALL_4 *EXPSOCIOEDU_1

-*TEMA__PERQUESENTIR

En el aula casi no habíamos hablado de innovación. No sabía qué tipo de texto esperaba el profesor. Esto en la Tasca1. La Tasca2 la encontré más motivadora y muy interesante como aprendizaje. No tuve tanta sensación de inseguridad. La Tarea 3 me hizo sentir de nuevo inseguridad. La encuentro muy interesante como reflexión y autoaprendizaje, pero creo que me faltan recursos digitales y tiempo para hacerla tan bien como quisiera.

-*TEMA__DIRIESCOMPANY

Es una asignatura interesante pero compleja. Y en la que se nos pide muchísimo trabajo. Está bien para ver el proceso de un proyecto de innovación basado en la observación pero para mi gusto se profundiza en exceso. Conlleva mucho más trabajo del que debería.

-*TEMA__SUGGERENCIES

En la Tarea 1 y Tarea 2, no repetir tanto los escritos

**** *GRUPiNVEST_2 *TORN_2 *REDAT_2 *SEXE_2 *ACCESS_2 *TREBALL_3 *EXPSOCIOEDU_1

-*TEMA__PERQUESENTIR

Creía que era una tarea difícil porque los artículos presentaban un lenguaje complicado. También es cierto que soy una persona que se angustia enseguida y por tanto mostraba inseguridad, pero aun así, al final, estas tareas han sido un reto que pienso he superado positivamente.

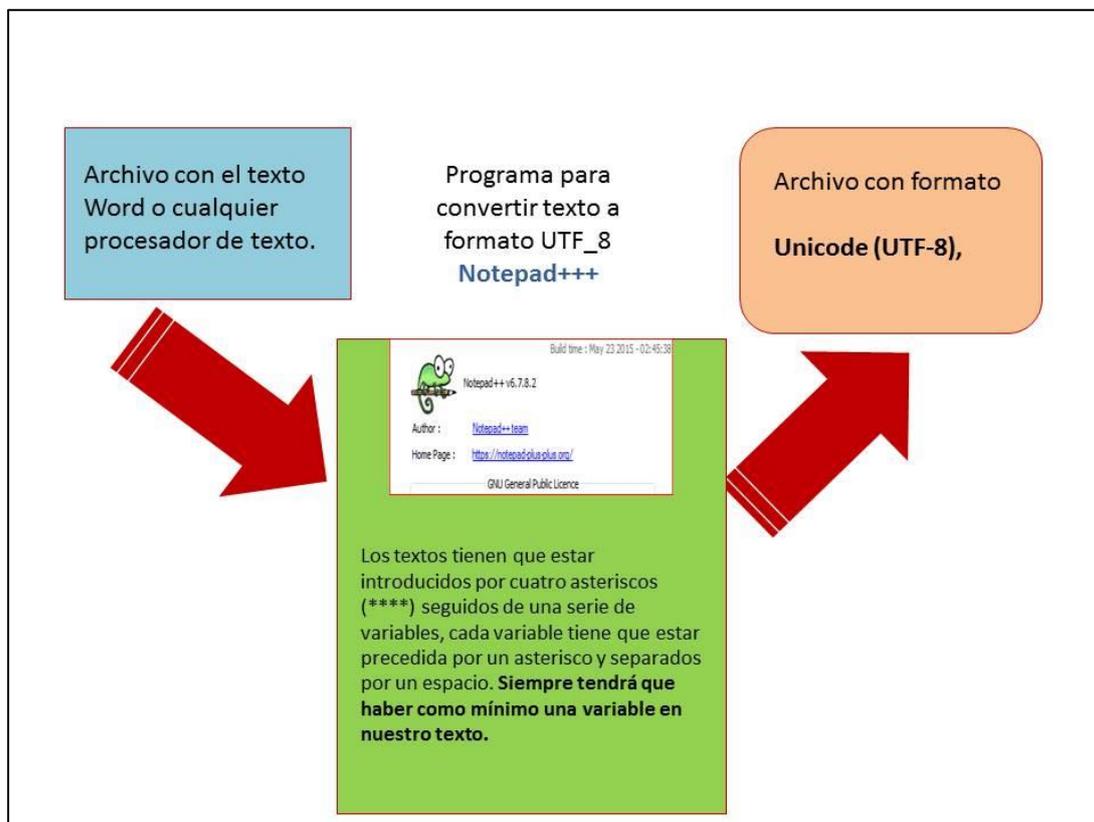
-*TEMA__DIRIESCOMPANY

Le diría que para realizar la asignatura necesita tener ganas de aprender y de trabajar, porque es una asignatura muy densa con una gran cantidad de contenidos. Además, debe ser constante en el trabajo (tanto en casa como en clase) si no quieres perderse por el camino

-*TEMA__SUGGERENCIES

Pienso que el trabajo final donde debemos realizar el artículo debería empezar desde el principio de curso e irlo haciendo a medida que el profesor explica los contenidos, porque luego te encuentras que tienes mucho trabajo y muy poco tiempo. Para el resto de tareas me ha faltado comentar los contenidos de los artículos a clase para consolidar todo.

De forma esquemática podemos ver en la siguiente gráfica el proceso para generar un archivo que pueda ser leído por el programa.



Una buena opción para realizar el proceso consiste en copiar el contenido del archivo en Notepad+++ y guardarlo con un nombre distinto al de su origen.

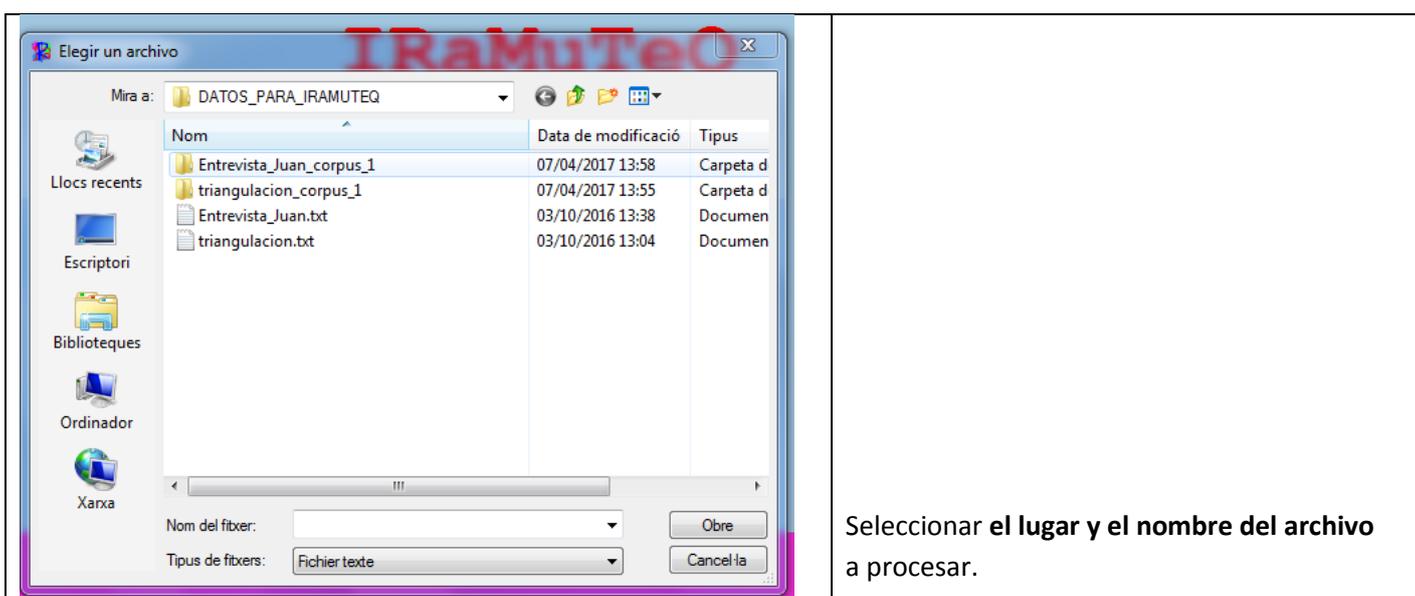
LA CAPTURA DEL ARCHIVO PARA SU ANÁLISIS.

En las opciones del menú de la parte superior de la pantalla del programa:

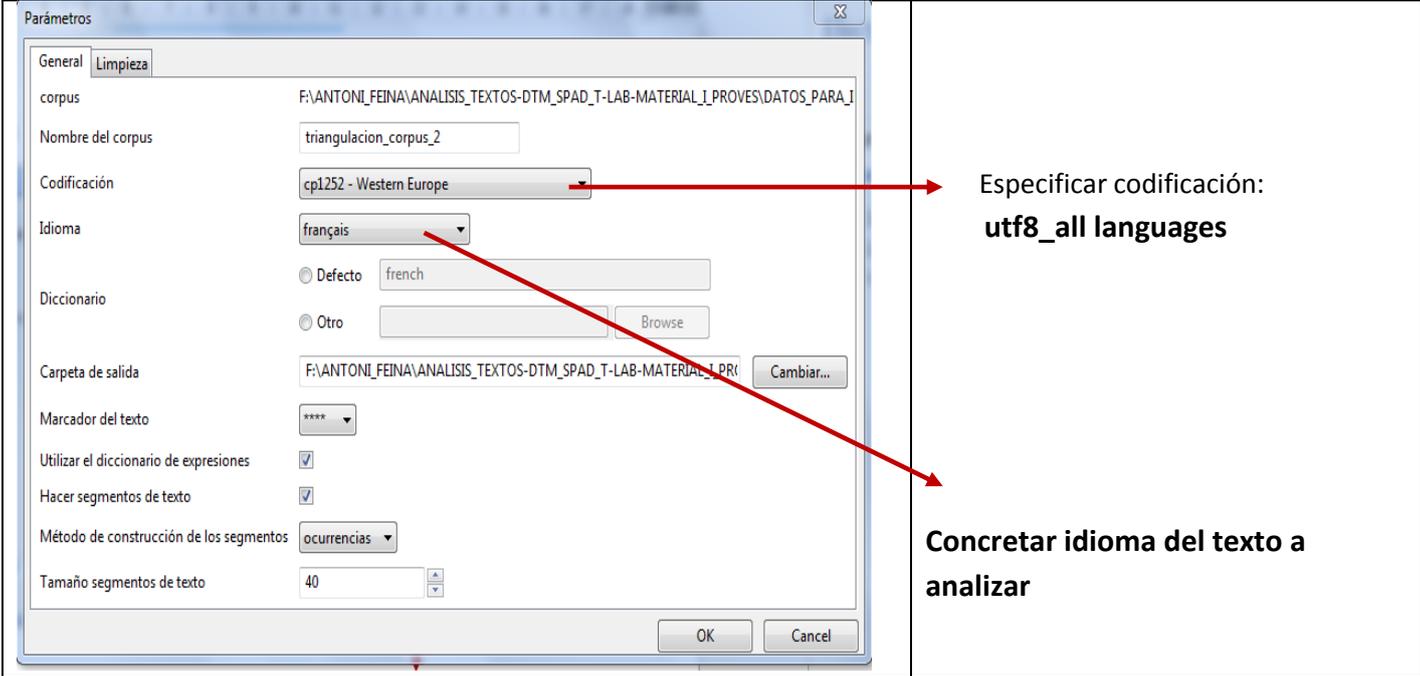
Archivo
 **1º**



 **2º**



↓ 3º Pestaña GENERAL



Parámetros

General Limpieza

corpus F:\ANTONI_FEINA\ANALISIS_TEXTOS-DTM_SPAD_T-LAB-MATERIAL_I_PROVES\DATOS_PARA_I

Nombre del corpus triangulacion_corpus_2

Codificación cp1252 - Western Europe

Idioma français

Diccionario Defecto french

Otro Browse

Carpeta de salida F:\ANTONI_FEINA\ANALISIS_TEXTOS-DTM_SPAD_T-LAB-MATERIAL_I_PROVES\DATOS_PARA_I Cambiar...

Marcador del texto ****

Utilizar el diccionario de expresiones

Hacer segmentos de texto

Método de construcción de los segmentos ocurrencias

Tamaño segmentos de texto 40

OK Cancel

Especificar codificación:
utf8_all languages

Concretar idioma del texto a analizar

Las otras características opcionales de la pestaña General:

Por regla general estas opciones se mantienen como las presenta el programa (por defecto).

Carpeta de salida: es donde el programa guardará toda la información que genere el programa. Por defecto, el programa guarda toda la información en la misma carpeta donde había sido guardado el texto que vamos a trabajar.

Marcador de texto: Hace referencia al tipo de opción que se utilizará para la separación de textos. Hay dos opciones, asteriscos o números, la que utilizaremos en este manual de uso, son los cuatro asteriscos (****). La otra opción está representada por cuatro ceros (0000).

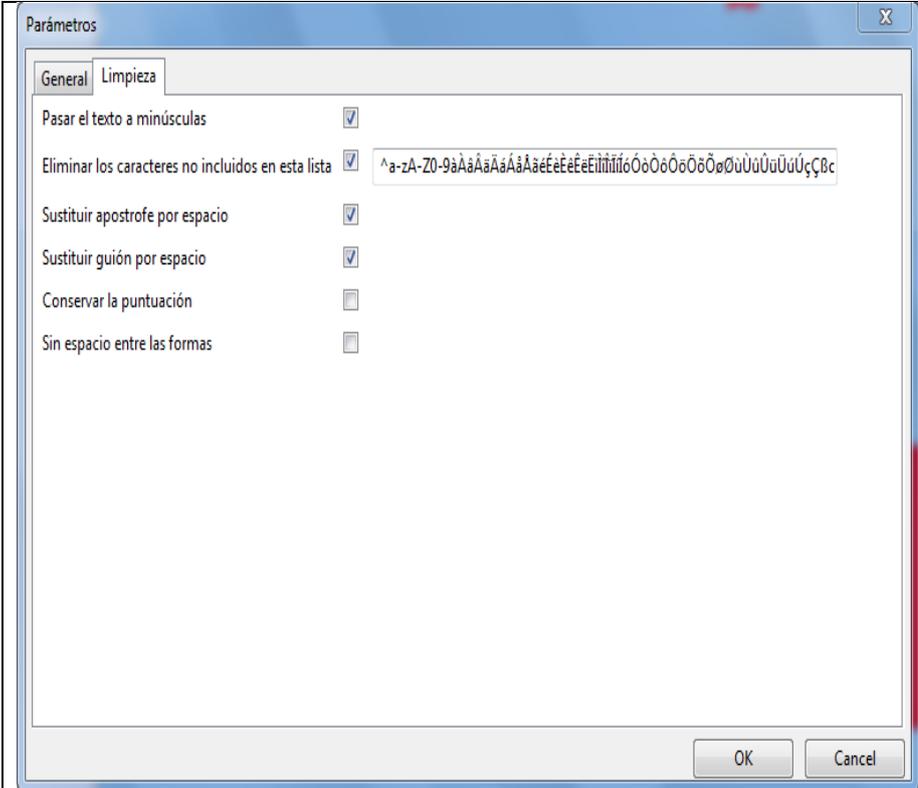
Utilizar el diccionario de expresiones: Si marcamos la casilla, todas las expresiones que hayan sido registradas en el diccionario de Iramuteq se reconocerán como una palabra única, como por ejemplo: Tasa de Natalidad. Si queremos realizar modificaciones al diccionario de una lengua, como añadir expresiones nuevas, por ejemplo, deberemos acceder a la carpeta C:\Users\"nombre de usuario" \.iramuteq\dictionnaires, escoger el que nos interesa modificar e introducir los cambios oportunos. Esto es más a nivel de programación informática.

Hacer segmentos del texto: esta casilla nos permite seleccionar si queremos dividir el texto en segmentos. Los segmentos de texto son considerados como el "ambiente de las palabras". Segmentar el texto puede resultar útil si trabajamos con textos largos, como entrevistas o leyes, pues el programa dividirá los textos en unidades más pequeñas (segmentos de texto).

Método de construcción de segmentos: aquí se elige como queremos que el programa forme el segmento. Esta división se puede basar en un número de ocurrencias o unidades lingüísticas (opción seleccionada por defecto), una serie de caracteres o párrafos. Si seleccionamos la opción de ocurrencias, el programa busca la mejor relación entre tamaño y signos de puntuación con el objetivo de tener segmentos de tamaños homogéneos, respetando el máximo posible la estructura del texto. En caso de trabajar con respuestas a cuestionarios, es preferible con la opción 'párrafos'. En caso textos más largos, se aconseja utilizar la opción que aparece en el programa por defecto.

Tamaño de segmentos de texto: aquí se define el número de unidades contenidas en los segmentos de texto

4º Pestaña LIMPIEZA



Las opciones por defecto de la limpieza del texto.

La limpieza del texto, se ha de tener en cuenta para hacer una interpretación correcta de los resultados. Además, los resultados del impacto del proceso de limpieza pueden variar mucho de unos idiomas a otros.

La elección de las otras opciones de limpieza

En el caso de elegir la opción **“conservar la puntuación”**, hemos de tener en cuenta que es necesario verificar la ausencia del punto y coma en el corpus. En caso de no hacerlo, el software no podrá abrir el corpus de texto debido a un error interno.

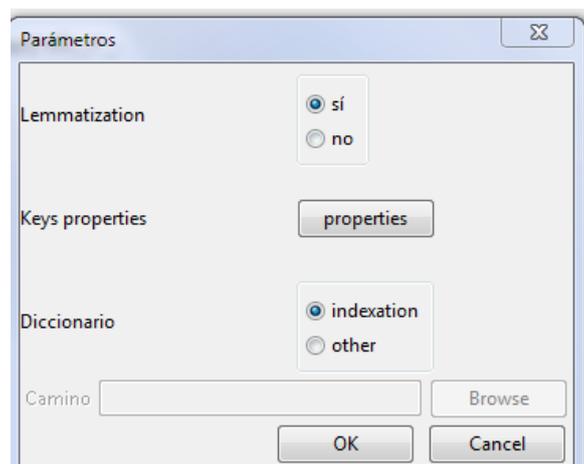
En el caso de elegir la opción **“sin espacio entre las formas”**, el software solo tendrá en cuenta el espacio como separación de las formas o unidades léxicas. Entiéndase por unidad léxica todo elemento situado entre dos espacios de la cadena escrita, o entre un espacio y un signo de puntuación. No se recomienda escoger esta opción.

Después de la especificación si damos al **OK** el programa nos cargará el texto para su análisis.

Las frecuencias.



Este análisis se obtiene pulsando el icono estadísticas, del menú de la parte superior: Una vez pulsado nos aparece la siguiente pantalla, con los concepto de Lemmatización, Keys properties (propiedades palabras) y diccionario:



Lematización:

Este proceso consiste en reducir las palabras principales a sus raíces, deviniendo así “morfemas lexicales”.

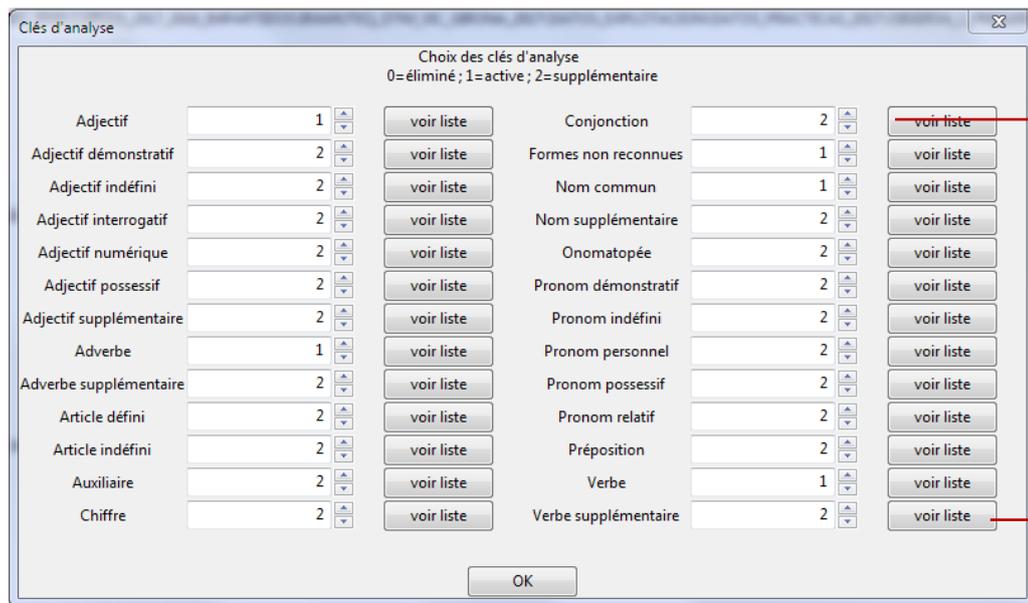
«El morfema lexical es la raíz de la palabra que remite a la familia de la cual proviene, opuesto a los morfemas derivativos que son las terminaciones que indican un matiz semántico o los morfemas flexivos que indican el género o la cantidad. Así, en la palabra “niño”, el morfema lexical es niño; sus terminaciones o, os, a, as, son los morfemas flexivos, mientras que terminaciones como ito, ita, ote, ota, corresponden a los morfemas derivativos. Alceste sólo toma en cuenta los morfemas lexicales para el análisis y elimina los morfemas flexivos o derivativos, aunque indica en un análisis suplementario todas las terminaciones de cada morfema lexical o raíz analizada.» (Alba, 2004)

Este proceso facilita la interpretación de los resultados, puesto que los verbos se presentan en infinitivo, los adjetivos en masculino y los nombres en singular. Se recomienda lematizar el corpus.

Keys properties (propiedades palabras):

Esta opción permite realizar el análisis escogiendo el tipo forma que se desea que entre a formar parte del análisis. Su pulsamos esta opción nos aparece la siguiente pantalla, dónde tenemos la oportunidad de especificar si la forma, será parte del análisis (valor 1), será eliminada (valor 0) o participará de forma suplementaria (valor 2). Por lo general, sobre todo en los primeros pasos con el programa, es preferible dejar estas opciones por defecto, las ya marcadas por el programa. También hay autores como Brigido Vizeu Camargo & Justo, 2016) se presenta, en la página 15, una sugerencia de ajustes para estudios en el ámbito de psicología. Con un mínimo de palabras activas, suplementarias y eliminadas.

También se puede comprobar las palabras que componen cada forma del corpus a analizar, para ello se ha de pulsar la tecla (voir liste). Una vez escogidos los valores de este submenú, pulsamos OK, y volvemos a la pantalla de parametros, inicial de este apartado.



Status en el análisis de la forma

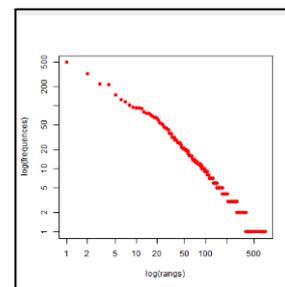
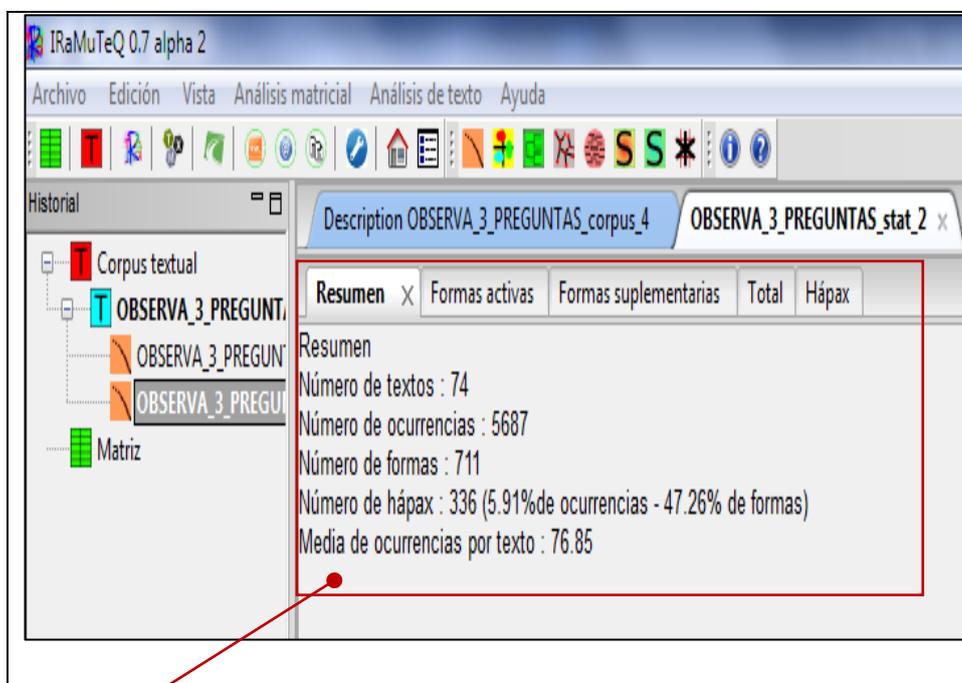
Listado de la forma

Diccionario:

Se deja por defecto la opción de indexación, tal como aparece en la primera pantalla.

***ATENCIÓN ¡!!!!!!! Este proceso que hemos explicitado aparecerá para cualquier análisis que se desee realizar con IRAMUTEQ.**

Una vez elegidos todos los parámetros, si pulsamos OK, después de unos minutos nos aparecerá el análisis de frecuencias demandado.



El gráfico presenta los logaritmos de rangos en el eje de abscisas y los logaritmos de frecuencias en el eje de ordenadas. (ley de Zipf, frecuencias previamente ordenadas en orden decreciente, presentan una relación aproximada entre el rango y la frecuencia. **Lebart, L., Salem, A. y Bécue, M. (2000).**

Como resultado del análisis se nos presentan 4 pestañas dónde se muestran distintos resultados, veamos en detalle cada una de estas pestañas:

Pestaña Resumen:

Número de textos:

Nos muestra el número de textos analizados, en este caso 74 que corresponde a 74 estudiantes que han respondido a preguntas abiertas. Si tuviésemos un texto sin variables categoriales, el número de texto sería 1.

Número de ocurrencias:

Son el total de palabras en el corpus (este número variará dependiendo si el texto está lematizado o no).

Número de formas:

Conjunto de palabras con una misma forma (verbo, adjetivo, sustantivo etc.).

Número Hápax:

Palabras que aparecen en el corpus una única vez (frecuencia = 1). También nos proporciona el porcentaje que representa respecto a las ocurrencias y las formas)

Media de ocurrencias por texto:

Es el resultado de la división entre el número de ocurrencias / número de textos.

Pestaña Formas activas:

Se pueden considerar los siguientes elementos para analizar (son las formas que entran en los análisis):

forma	Frec.	Tipos
trabajo	117	nom
más	93	adv
tarea	69	nom
hacer	66	ver
mucho	64	adv
asignatura	60	nom
tiempo	51	nom
ya	37	adv
realizar	36	ver
aprender	32	ver
decir	32	ver
final	32	nom
bueno	29	adj
saber	29	ver
sentir	29	ver
innovación	26	nom
poco	26	adv
duda	22	nom
parte	21	nom
dar	20	ver
primero	20	adj
principio	20	nom
explicar	19	ver
tema	19	nom
ir	17	ver
aula	16	nom
clase	16	nom
momento	16	nom
entender	15	ver
trabajar	15	ver
cómo	14	adv

La columna «**forma**»: presenta las formas concretas.

La columna «**frec**»: nos indica el número de veces que aparece forma en el corpus/texto.

La columna «**tipos**»: indica la categoría gramatical de la forma.

Pulsando en la parte superior (en la flecha color verde, nos ordena, por frecuencias o bien por tipos o formas en orden alfabético.

Forma	Frec.	Tipos
útil	4	adj
único	1	nom
únicamente	1	adv
último	7	adj
época	1	nom
ya	37	adv
volver	1	ver
volumen	7	nom
vocabulario	3	nom
vista	5	nom
vino	1	nom
video	3	nom
vida	1	nom
vez	12	nom
ver	7	ver
venir	5	ver
vengar venir	1	ver
valor	1	nom
valer	2	ver
vacación	1	nom
utilizando	1	nr
utilizar	2	ver

También, no podemos posicionar encima de la forma y poder ver la concordancia o las formas que se han lematizado en ella (esto se realiza pulsando el botón izquierdo del ratón, para la selección de la forma, y con el derecho la opción de ver las formas asociadas y la concordancia).

Ejemplo Formas asociadas, caso útil:

Útil: 4

Útiles: 3

Ejemplo Concordancia, caso útil:

**** *GRUP_1 *TORN_2 *EDAT_2 *ACC_2 *TREB_1 *EXPE_1
 creo que es una herramienta bastante **útil** para reflexionar sobre nuestro futuro pero que no s agobie al final le encuentras sentido a todas las experiències
 **** *GRUP_2 *TORN_1 *EDAT_1 *ACC_1 *TREB_1 *EXPE_1
 considero que la segunda parte de la asignatura es la más **útil** para realizar el artículo y debería haber más tiempo
 **** *GRUP_2 *TORN_1 *EDAT_1 *ACC_2 *TREB_1 *EXPE_1
 le diría que es una asignatura muy **útil** en cuanto a la práctica docente en un futuro ya que tendremos que innovar constantemente lo que también le diría es que se organizara bien el trabajo ya que hay un gran volumen de trabajo

Este proceso se puede repetir para cada una de las pestañas, que muestra el menú de resultados.

Pestaña Formas suplementarias:

No entran propiamente en los análisis. Están formadas mayoritariamente por conectores (artículos, conjunciones, preposiciones, etc.)

Forma	Frec.	Tipos
el	492	art_def
que	322	con
y	226	con
de	218	pre
ser	148	ver_sup
un	125	art_def
no	104	pre
a	95	pre
en	93	pre
haber	90	ver_sup
para	80	pre
yo	77	pro_per
se	76	sw
porque	72	con
tener	65	ver_sup
muy	54	adv_sup
con	52	pre
poder	45	ver_sup
por	45	pre
al	43	sw
pero	42	con
todo	36	adj_ind
bien	26	adv_sup
este	26	pro_dem
sobre	25	pre
tú	24	pro_per
como	21	con

Se pueden realizar las mismas acciones que la anterior pestaña.

Pestaña Total:

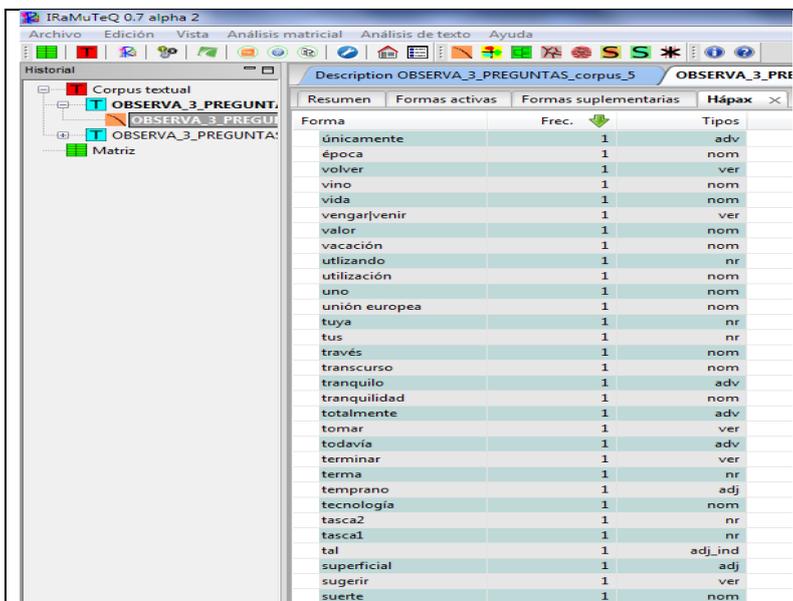
Muestra todas las formas excepto las que tienen una frecuencia igual 1.

Forma	Frec.	Tipos
el	492	art_def
que	322	con
y	226	con
de	218	pre
ser	148	ver_sup
un	125	art_def
trabajo	117	nom
no	104	pre
a	95	pre
en	93	pre
más	93	adv
haber	90	ver_sup
para	80	pre
yo	77	pro_per
se	76	sw
porque	72	con
tarea	69	nom
hacer	66	ver
tener	65	ver_sup
mucho	64	adj
asignatura	60	nom
muy	54	adv_sup
con	52	pre

Se pueden realizar las mismas acciones que la anterior pestaña.

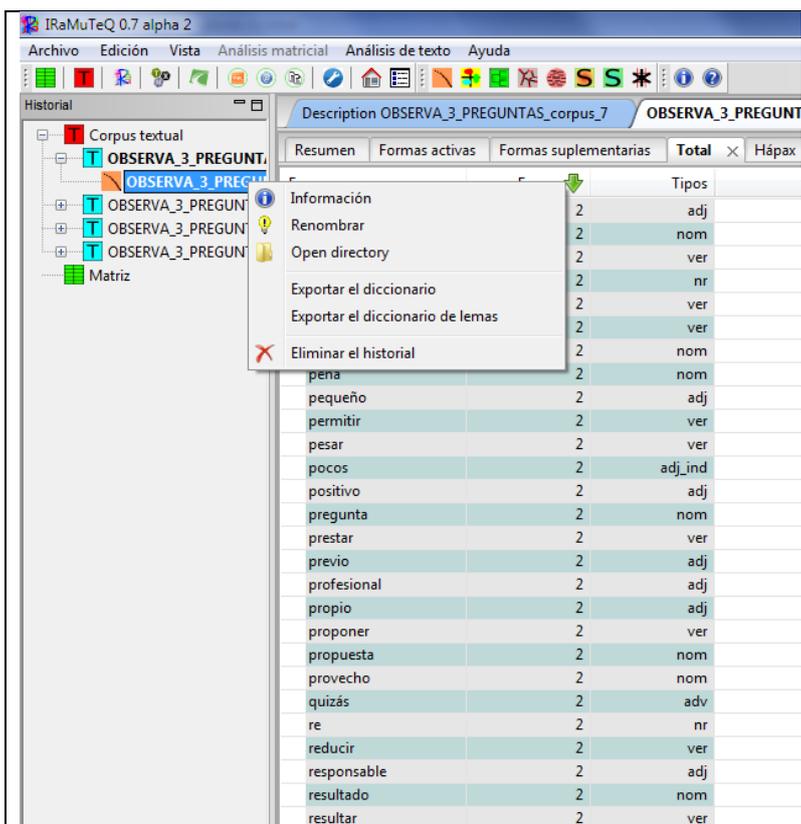
Pestaña Formas Hápax:

Son las formas con frecuencia igual a 1, pueden ser interesantes de comentar o analizar según los objetivos de la investigación.



Se pueden realizar las mismas acciones que la anterior pestaña.

Además de estos resultados se pueden llevar a cabo otras acciones que pueden resultar interesantes para las investigaciones, es el caso de seleccionar, no una palabra, sino todo el corpus analizado, para ello nos debemos situar en la parte más a la izquierda de la pantalla. Esto nos permite realizar las acciones que muestra la siguiente gráfica, pasamos a continuación a describirlas:



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.

(Extraído de: Moreno, M., & Ratinaud, P. (2015). Manual uso de Iramuteq. Versión 0.7 alpha 2. Recuperado a partir de http://iramuteq.org/documentation/fichiers/guia-ramuteq_at_download/file)

Análisis de especificidades.

Es un análisis que consiste en la especificación de las variables categoriales utilizadas y los textos. Nos proporciona la relación de formas asociadas a una variable o modalidades, además de un análisis factorial de correspondencia (en este caso cuando se tiene variables de más de dos categorías, si no es así, este análisis no lo realiza).



Para su realización debemos pulsar el icono de especificidades:

	<p>Las formas que utilizaremos en el análisis. Pueden ser, sólo las activas, activas y suplementarias o sólo suplementarias.</p> <p>Se puede seleccionar :</p> <p>a) Variable: permite seleccionar todas las modalidades de una variable.</p>
	<p>Índice: hay dos posibilidades de cálculo, la Chi2 o la ley hipergeométrica</p> <p>Umbral de frecuencia mínima utilizado para la selección de las formas.</p> <p>Selección por modalidad o modalidades</p> <p>b) Modalidad: escogeremos solo las modalidades que nos interesen dentro de esa variable.</p> <p>Es obligatorio la elección de una variable o modalidad para poder llevar a cabo el análisis.</p>

El resultado de este análisis se presenta en forma de pestañas de las que se obtienen datos referentes a las variables o modalidades seleccionadas respecto a las formas de texto.

Una vez pulsado el OK, nos aparece la siguiente pantalla, donde hay diversas pestañas que muestran la relación entre la forma o unidad léxica específica y la variable/modalidades. Esta relación se muestra con valores asociados a una forma, cuanto mayor es el valor más específica o característica es de la modalidad. Como se puede comprobar, también aparece “El signo +” significa que la forma se menciona más en ese grupo que los otros. Del mismo modo, “El signo -”, implica que esa forma la utiliza menos ese grupo que el otro.

Las pestañas de **FORMAS**, **FORMAS BANALES (SUPLEMENTARIAS)** y **TIPOS**: presentan la lista de formas, categorías gramaticales y los resultados por modalidad.

Formas	Formas banales	Tipos	Frecuencia de formas	Frecuencia de tipos	Frecuencia relativa de formas	Frecuencia relativa de tipos	AFC
formas		*TREB_1		*TREB_2	*TREB_3	*TREB_4	
pensar		1.4039	-1.4049	0.4125	-0.5931		
empezar		1.2803	0.348	-0.5548	-1.0202		
además		0.9601	0.4167	-0.4594	-0.9274		
ya		0.9595	-0.2431	-0.6374	-0.3812		
momento		0.9199	-0.5493	-0.6772	0.4317		
aula		0.9199	0.9267	-1.6984	-0.4186		
decir		0.8527	-0.8666	1.0236	-1.4059		
poder		0.838	-0.3413	-0.367	-0.4637		
cómo		0.7983	1.1814	-0.8675	-1.299		
realizar		0.7899	0.4048	-0.5783	-0.5469		
en		0.7591	-0.3844	-0.2709	-0.5684		
haber		0.7424	-0.6991	0.6499	-0.8172		
otro		0.6974	-0.6822	-0.3154	0.3329		
grupo		0.6744	0.2912	0.2605	-1.1131		
sobre		0.646	-0.7777	-1.1616	1.4361		
más		0.6213	1.3893	-1.4103	-0.5684		
no		0.5986	-0.4225	-0.3863	0.2693		
sentir		0.595	-0.269	-0.5744	0.3066		
saber		0.595	1.0722	-1.574	-0.2955		
yo		0.5739	-0.5957	-0.7353	0.8537		
nuevo		0.5673	-1.4049	-0.4214	1.0672		

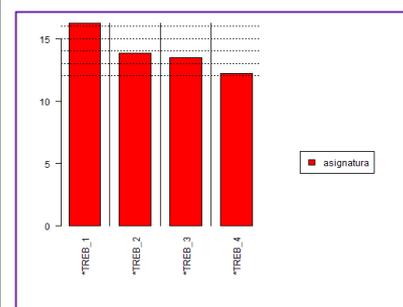
En el caso de **La pestaña de FRECUENCIAS DE FORMAS Y TIPOS**: presenta la frecuencia de cada forma y tipos gramaticales en la modalidad de la variable.

En las pestañas de **FRECUENCIAS RELATIVAS DE FORMAS Y TIPOS**: La frecuencia relativa (es la frecuencia absoluta partido por el total de la muestra) es una frecuencia que no está afectada por el tamaño de la muestra, por tanto permite la comparación. Estas frecuencias relativas están expresadas en **tantos por mil**.

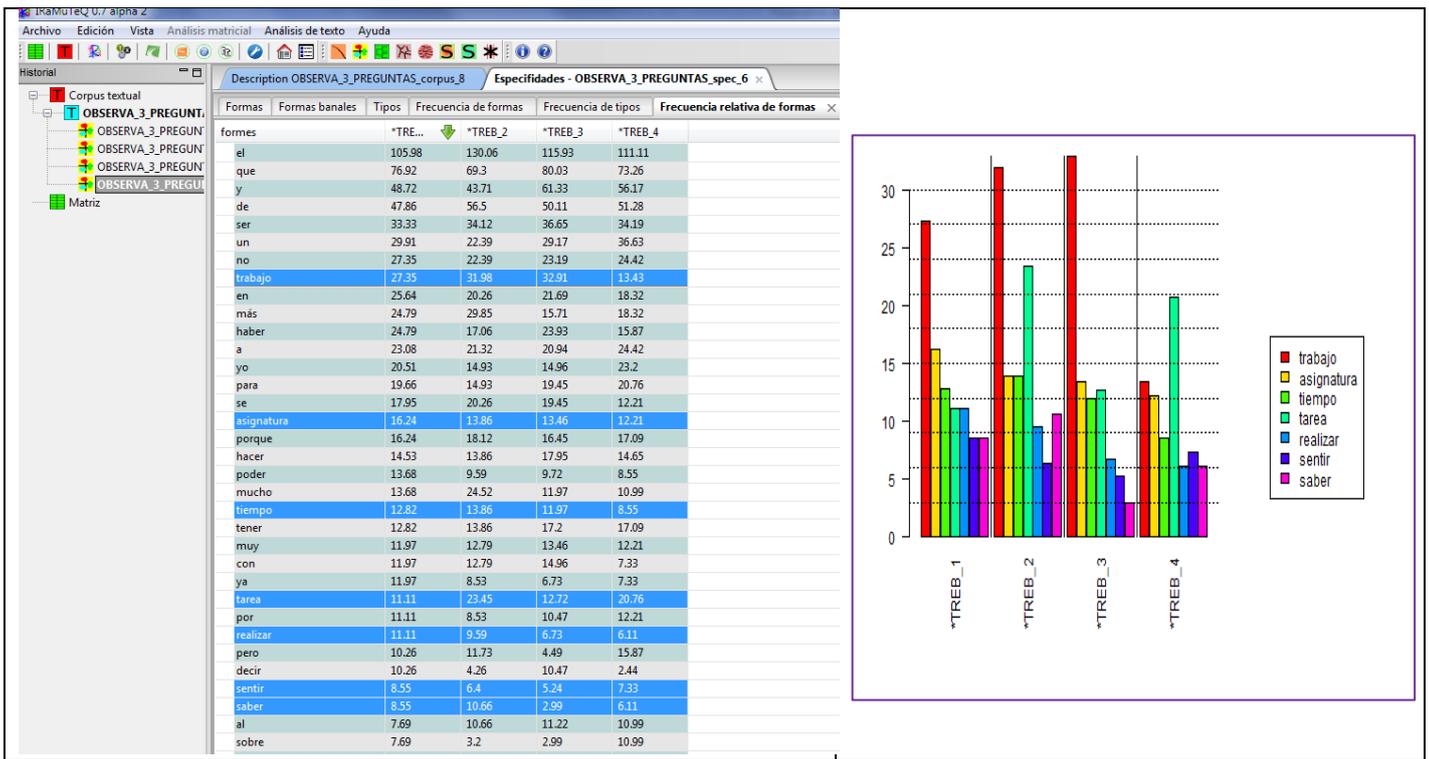
Formas	Formas banales	Tipos	Frecuencia de formas	Frecuencia de tipos	Frecuencia relativa de formas	Frecuencia relativa de tipos	AFC
formas		*TREB_1		*TREB_2	*TREB_3	*TREB_4	
el		105.98	130.06	115.93	111.11		
que		76.92	69.3	80.03	73.26		
y		48.72	43.71	61.33	56.17		
de		47.86	56.5	50.11	51.28		
ser		33.33	34.12	36.65	34.19		
un		29.91	22.39	29.17	36.63		
no		27.35	22.39	23.19	24.42		
trabajo		27.35	31.98	32.91	13.43		
en		25.64	20.26	21.69	18.32		
más		24.79	29.85	15.71	18.32		
haber		24.79	17.06	23.93	15.87		
a		23.08	21.32	20.94	24.42		
yo		20.51	14.93	14.96	23.2		
para		19.66	14.93	19.45	20.76		
se		17.95	20.26	19.45	12.21		
asignatura		16.21	13.66	13.46	12.21		
porque					17.09		
hacer					14.65		
poder					8.55		
mucho					10.99		
tiempo					8.55		
tener					17.09		
muy		11.97	12.79	13.46	12.21		
con		11.97	12.79	14.96	7.33		
ya		11.97	8.53	6.73	7.33		
tarea		11.11	23.45	12.72	20.76		
por		11.11	8.53	10.47	12.21		

Si pulsamos el ratón sobre una palabra nos aparece una ventana emergente, donde se puede obtener las formas asociadas, la concordancia o bien un gráfico sobre dichas forma diferenciando las modalidades/valores de la variable.

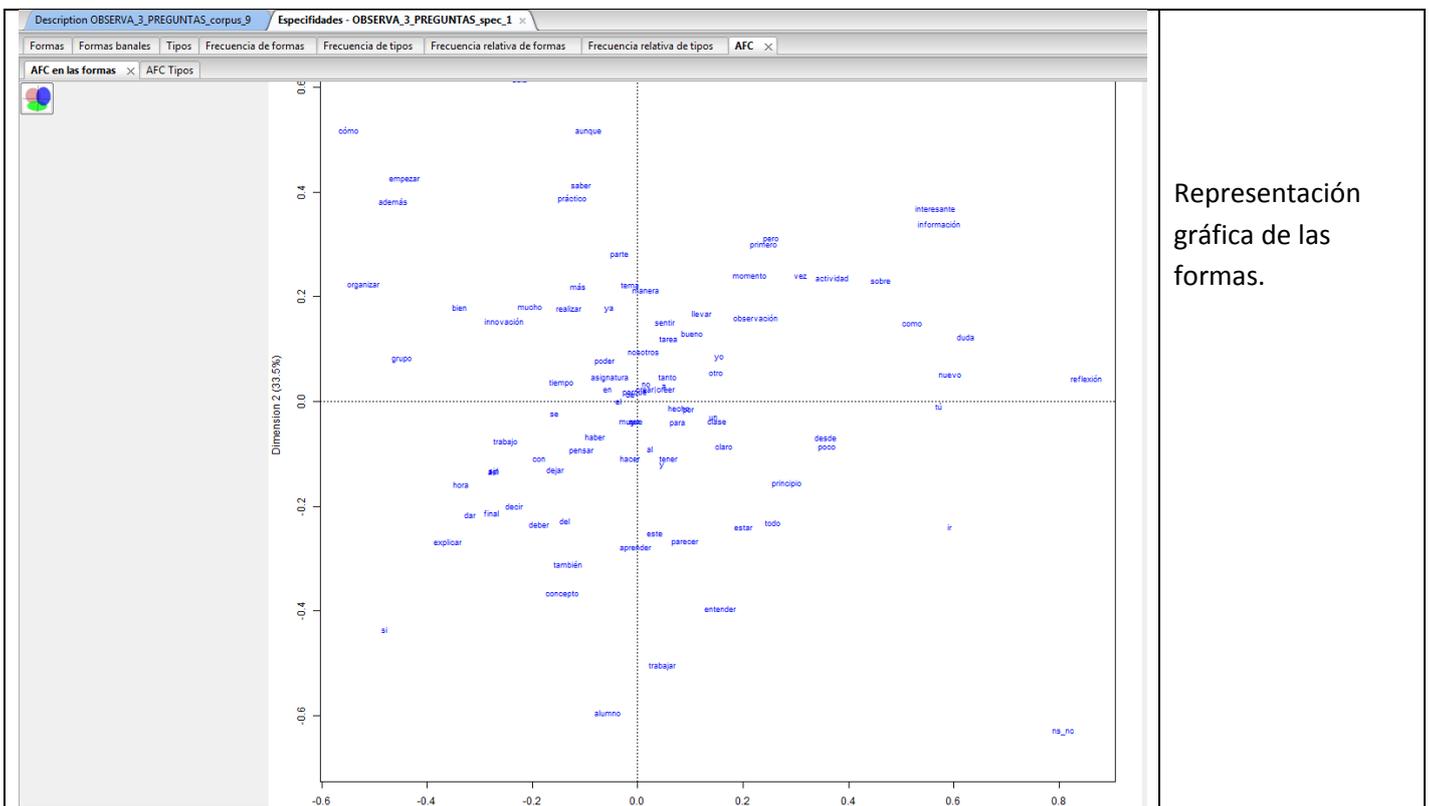
El gráfico que se obtiene es el siguiente:



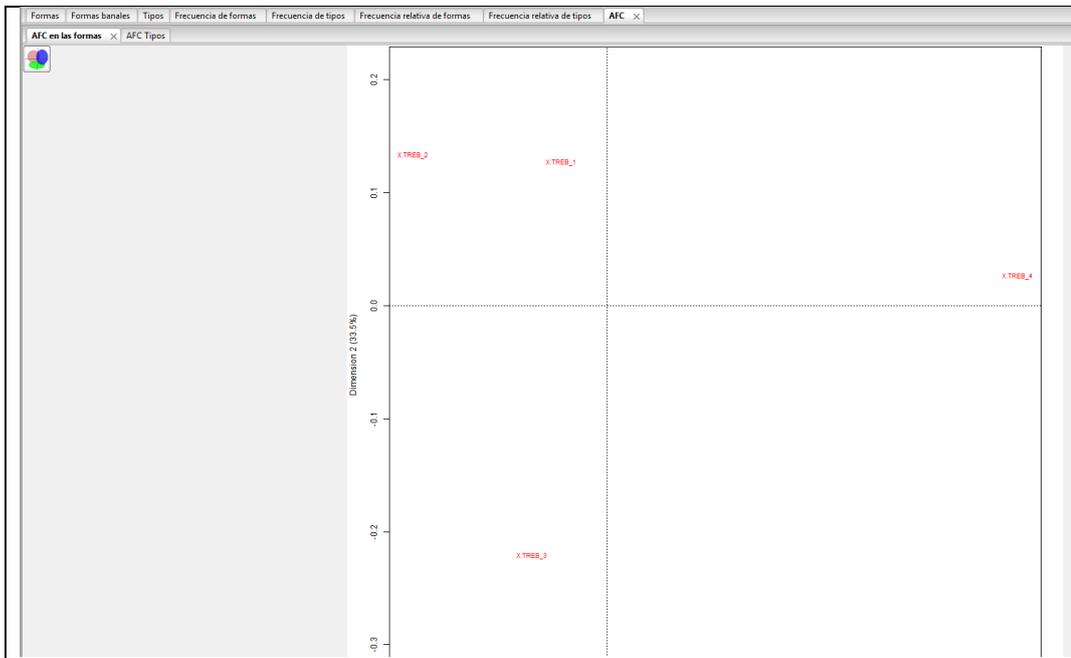
También se puede obtener una gráfica de varias formas en función de las modalidades. Para ello, sólo basta con seleccionar las modalidades y demandar el gráfico.



En la pestaña AFC (análisis factorial de correspondencias), nos proporciona unas gráficas de posición de las formas y las modalidades.



Representación gráfica de las formas.



Representación gráfica de las modalidades.

Se puede ver la correspondencia entre las formas y las modalidades, según su posición en los ejes.

Estos mismos gráficos se presentan para los tipos gramaticales.

Se puede modificar el gráfico del AFC mediante el icono 

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 gráfico (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 al 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

(Extraído de: Moreno, M., & Ratinaud, P. (2015). Manual uso de Iramuteq. Versión 0.7 alpha 2. Recuperado a partir de http://iramuteq.org/documentation/fichiers/guia-iramuteq/at_download/file)

Clasificación jerárquica descendente (método Reinert)

Este análisis muestra una clasificación jerarquizada de manera descendente (también llamado algoritmo divisivo, que opera globalmente de forma inversa a los algoritmos ascendentes: comienza con un solo grupo que incluye a todos los individuos, y en cada etapa se hace una subdivisión hasta que finalmente resultan tantos grupos como individuos) 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.

El análisis se inicia en IRAMUTEQ, pulsando el icono



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"/>
<input type="button" value="Cancel"/> <input type="button" value="Valores por defecto"/> <input type="button" value="OK"/>	

Las opciones para la realización de CHD:

Clasificación doble en RST: reagrupa los segmentos de textos más próximos. Realiza dos veces el análisis. No se recomienda su uso por el bajo aprovechamiento del corpus.

Clasificación simple sobre segmentos de texto: La clasificación se centrará en los segmentos de texto. Es el análisis estándar, recomendada para respuestas largas.

Clasificación simple en el texto: Con esta opción los textos se mantienen en su totalidad, sin dividirlos en segmentos de texto. Es la opción recomendada para las respuestas cortas.

El resto de parámetros no requieren modificación. Esto son:

Tamaño del rst1 y rst2: estas opciones sirven cuando se realiza una clasificación en RST. Permiten cambiar el tamaño de la reagrupación de segmentos de texto. Las 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 por 10, pero podemos cambiarlo en función de la heterogeneidad del corpus. Aumentar el valor, aumentaría el número de clases.

Frecuencia mínima de segmentos de texto por clase: determina el nivel mínimo bajo el que los segmentos no serán tenidos en cuenta.

Frecuencia mínima de una forma analizada: no disponible.

Número máximo de formas analizadas: hace referencia al número máximo de formas a analizar. Por defecto vienen definido para las 3000 formas activas y suplementarias más frecuentes, pero es posible modificarlo si el tamaño del corpus lo justifica. Téngase en cuenta que una forma debe de aparecer un mínimo de tres veces para ser tenida en cuenta en el análisis.

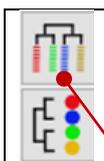
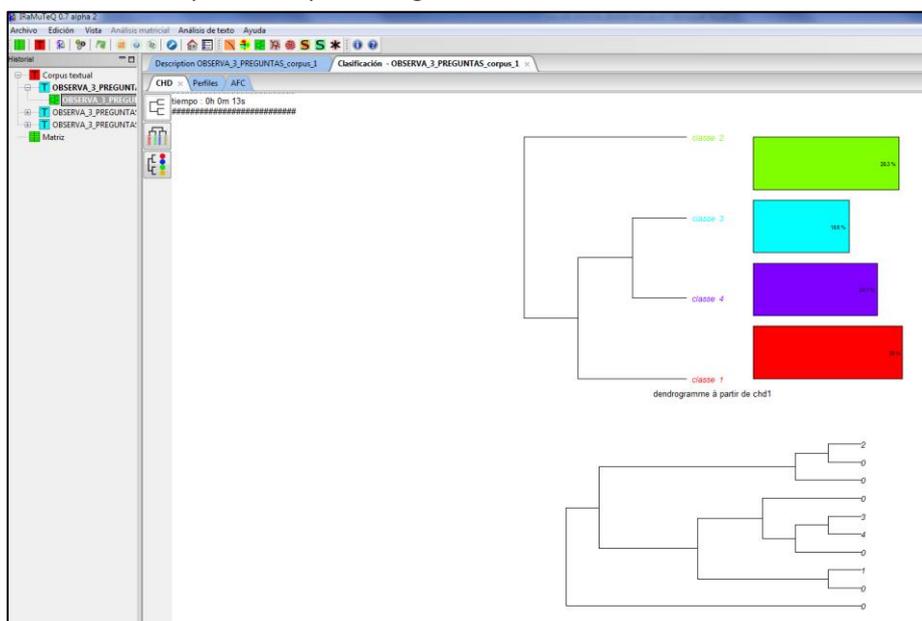
Método svd: permite la elección del algoritmo utilizado para las descomposiciones.

Método patata: si elegimos esta opción, no se realiza la segunda fase de cada partición. El análisis es más rápido, pero menos preciso.

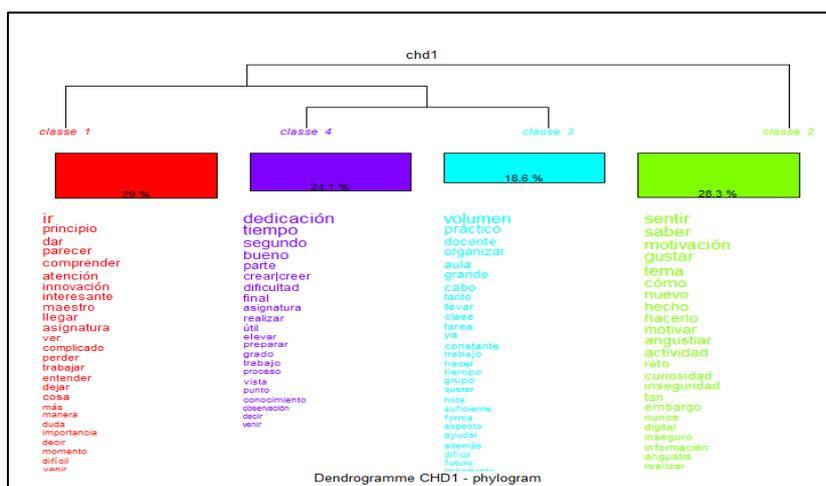
Una vez pulsado el OK, nos aparecerá una pantalla con tres pestañas que nos presentan los resultados obtenidos de la clasificación de todo el corpus analizado. Estas pestañas corresponden a:

- . **CDH:** Donde aparecen los resultados del análisis clúster.
- . **PERFILES:** Se presenta el léxico incluido en cada clase. Esto nos permitirá orientar la temática que rodea a la clase o grupo, ya que los perfiles se crean por las coocurrencias, por lo que se agrupan por la proximidad lexical entre las formas incluidas.
- . **AFC:** Muestra el análisis factorial que corresponde a la contingencia creciente de formas y lemas, que son mostrados según colores que corresponden a los de la clase a la que pertenecen.

La pestaña CDH: muestra dos grandes bloques de información, un dendograma con el historial de agrupación de las clases e información sobre las palabras que configuran cada clase.



Si pulsamos estas opciones, el dendograma se mostrará, con las formas específicas de las clases para tenerlas en consideración en la interpretación. Al mostrar estas formas características, permite la interpretación del “mundo léxico” de cada clase. Una vez interpretadas las diferencias entre las clases, se aconseja identificarlas con una etiqueta significativa del mundo representacional del que da cuenta.



En el análisis de estos dendogramas es conveniente tener en cuenta para su interpretación los siguientes aspectos:

Relaciones entre clases: Ver dónde se producen las particiones del itinerario de clasificación. En el ejemplo anterior, en un primer momento se divide el corpus en dos, formado por las clases 3 y 4 (tienen más semejanza), después se les une subcorpus de la clase 1, y finalmente la clase 2. El Análisis Jerárquico Descendiente paró, al obtener 4 clases estables, es decir, compuestas de unidades de segmentos de texto con vocabulario semejante.

Porcentaje de las clases: En los resultados también se muestra el tamaño de las clases, expresado en porcentaje respecto al total del corpus analizado.

Las formas: Las unidades léxicas sirven de orientación para hacer una primera valoración sobre el contenido lexical de cada clase. El tamaño de cada una de las formas orienta sobre la significatividad estadística de la forma dentro de ese mundo léxico. Sin embargo, para un análisis más detallado sobre esta cuestión, es necesario acudir a la segunda pestaña, denominada “perfiles”.

La pestaña PERFILES:

En esta pestaña veremos el léxico que se ha incluido en cada clase, lo cual nos orientara sobre la temática que le rodea, ya que los perfiles son agrupados con una coherencia y una proximidad lexical entre las formas incluidas en cada clase. La pantalla en el programa muestra para cada una de las clases lo siguiente:

Num: Es 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: Es el valor 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 significación asociado a Chi2

CHD		Perfiles		AFC							
1 Clase 1		2 Clase 2		3 Clase 3		4 Clase 4					
42/145		41/145		27/145		35/145					
28.97%		28.28%		18.62%		24.14%					
n...	eff. st.	eff. total	pourcentage	chi2	Type	forme	p				
12	5	6	83.33	8.99	adj	complicado	0.00271				
13	5	6	83.33	8.99	ver	perder	0.00271				
14	6	8	75.0	8.72	ver	trabajar	0.00314				
15	6	8	75.0	8.72	ver	entender	0.00314				
16	4	5	80.0	6.56	ver	dejar	0.01045				
17	4	5	80.0	6.56	nom	cosa	0.01045				
18	19	47	40.43	4.44	adv	más	0.03513				
19	4	6	66.67	4.32	nom	manera	0.03758				
20	7	13	53.85	4.3	nom	duda	0.03819				
21	3	4	75.0	4.24	nom	importancia	0.03955				
22	11	25	44.0	3.32	ver	decir	NS (0.06850)				
23	6	12	50.0	2.81	nom	momento	NS (0.09348)				
24	3	5	60.0	2.42	adj	difícil	NS (0.11948)				
25	3	5	60.0	2.42	ver	venir	NS (0.11948)				
26	8	18	44.44	2.39	adj	primero	NS (0.12186)				
27	2	3	66.67	2.12	adj	claro	NS (0.14574)				
28	2	3	66.67	2.12	nom	día	NS (0.14574)				
29	2	3	66.67	2.12	ver	profundizar	NS (0.14574)				
30	2	3	66.67	2.12	ver	necesitar	NS (0.14574)				
31	2	3	66.67	2.12	nom	educación	NS (0.14574)				
32	2	3	66.67	2.12	adj	ningún	NS (0.14574)				
33	2	3	66.67	2.12	adv	dónde	NS (0.14574)				
34	2	3	66.67	2.12	nom	cuenta	NS (0.14574)				
35	2	3	66.67	2.12	ver	contar	NS (0.14574)				
36	2	3	66.67	2.12	ver	conllevar	NS (0.14574)				
37	2	3	66.67	2.12	nom	casa	NS (0.14574)				
38	17	28	60.71	17.0	sw	al	< 0.0001				
39	17	32	53.12	11.65	con	pero	0.00064				
40	19	41	46.34	8.39	sw	se	0.00377				
41	3	3	100.0	7.51	adj_ind	nada	0.00612				
42	10	19	52.63	5.95	adj_ind	todo	0.01470				
43	21	53	39.62	4.61	pre	no	0.03176				
44	4	7	57.14	2.84	pro_per	nosotros	NS (0.09204)				
45	5	10	50.0	2.31	pre	desde	NS (0.12857)				
46	2	3	66.67	2.12	con	sino	NS (0.14574)				
47	6	13	46.15	2.05	pro_per	tú	NS (0.15215)				
48	24	64	37.5	4.06		*DAT_2	0.04402				
49	24	64	37.5	4.06		*GRUP_1	0.04402				
50	24	64	37.5	4.06		*TORN_2	0.04402				

	Formas lexicales activas asociadas a la clase
	Formas lexicales suplementarias asociadas a la clase
	VARIABLES CATEGORIALES asociadas a la clase (esto si en el texto a analizar hay valores o modalidades de una variable).

Se puede profundizar un poco más en los perfiles de las clases obtenidas mediante submenús de funciones que se pueden activar en la pantalla de salida de los perfiles mediante el posicionamiento en una fila de la forma lexical o un conjunto de formas lexicales, pulsando el botón derecho del ratón.

n...	eff. s.t.	eff. total	pourcentage	chi2	Type	forme
9	4	4	100.0	10.09	ver	llegar
10	23	51	45.1	9.95	nom	asignatura
11	5	6	83.33	8.99	ver	ver
12	5	6	83.33	8.99	adj	complicado
13	5	6	83.33	8.99	ver	perder
14	6	8	75.0	8.72	ver	trabajar
15	6	8	75.0	8.72	ver	entender
16	4	5	80.0	6.56	ver	dejar
17	4	5	80.0	6.56	nom	cosa
18	19	47	40.43	4.44	adv	más
19	4	6	66.67	4.32	nom	manera
20	7	13	53.85	4.3	nom	duda
21	3	4	75.0	4.24	nom	importancia
22	11	25	44.0	3.32	ver	decir
23	6	12	50.0	2.81	nom	momento
24	3	5	60.0	2.42	adj	difícil
25	3	5	60.0	2.42	ver	venir
26	8	18	44.44	2.39	adj	
27	2	3	66.67	2.12	adj	
28	2	3	66.67	2.12	nom	
29	2	3	66.67	2.12	ver	
30	2	3	66.67	2.12	ver	
31	2	3	66.67	2.12	nom	
32	2	3	66.67	2.12	adj	
33	2	3	66.67	2.12	adv	
34	2	3	66.67	2.12	nom	
35	2	3	66.67	2.12	ver	
36	2	3	66.67	2.12	ver	
37	2	3	66.67	2.12	nom	
38	17	28	60.71	17.0	sw	
39	17	32	53.12	11.65	con	
40	19	41	46.34	8.39	sw	
41	3	3	100.0	7.51	adj_ind	
42	10	19	52.63	5.95	adj_ind	
43	21	53	39.62	4.61	pre	

- Formas asociadas
- Chi2 por clase
- Chi2 por clases en dendograma
- Chi2 modalidades de la variable
- Gráfico de palabra
- Concordancia
- Hacer un Tgen
- Herramientas desde CNRTL (solo francés)
- Gráfico de la clase
- Segmentos repetidos
- Segmentos de texto característicos
- Nube de palabras por clase
- Exportar
- Exportar a Tropes
- Exportar a Owledge

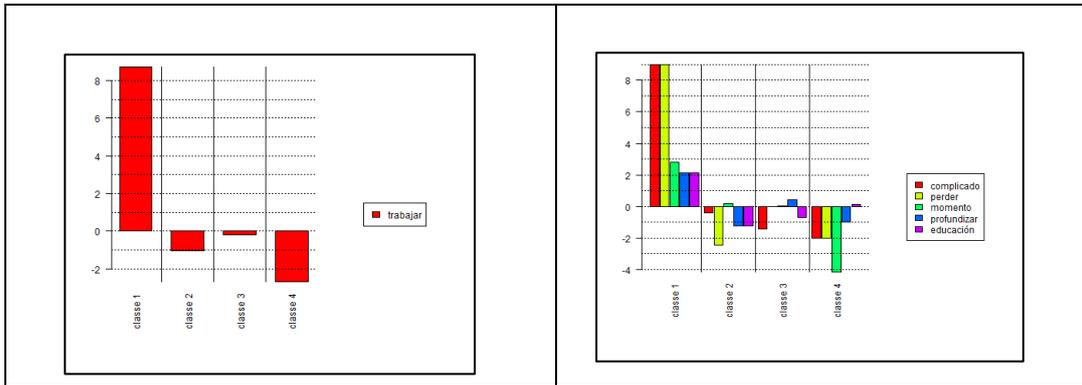
Las opciones que nos aparecen son las siguientes, tal como nos indican Moreno y Retinaud (2015):

- Formes asociadas:

Muestra los efectivos de las formas reagrupadas en un lema de la clase. Por tanto palabras lematizadas.

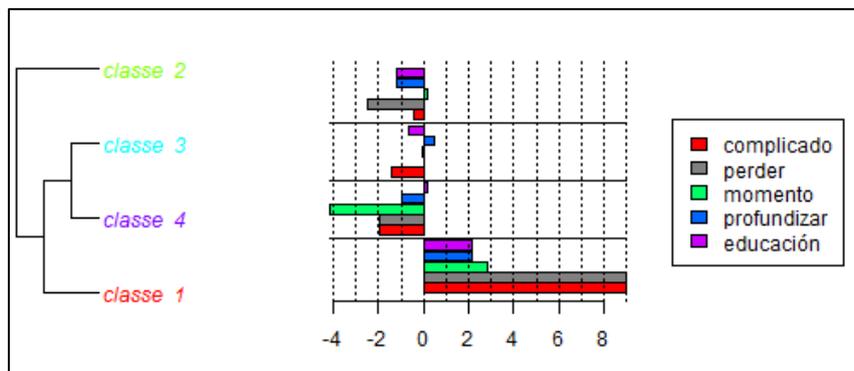
- Chi2 por clase:

Crea un gráfico que presenta la asociación de la forma a cada una de las clases. Pueden seleccionarse varias formas al mismo tiempo, mediante selección con el ratón.

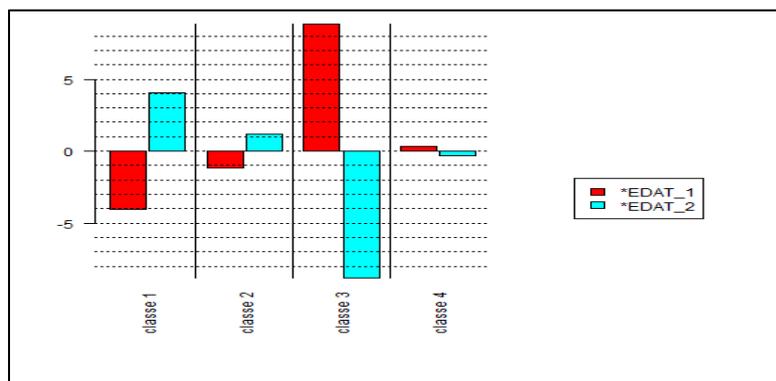


.- Chi 2 por clase en dendograma:

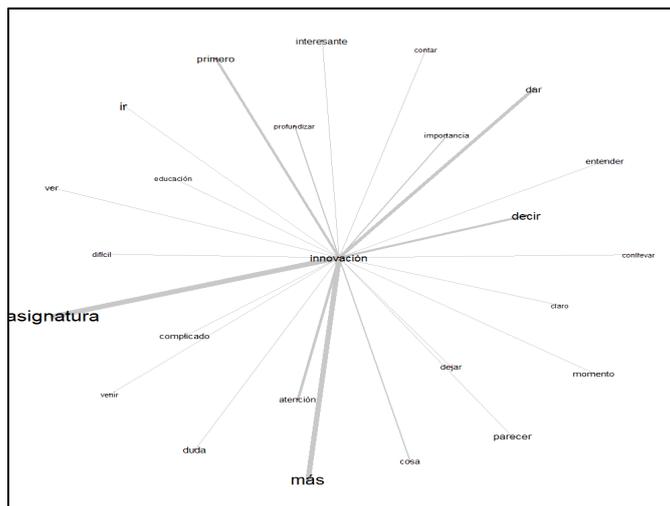
Retoma el mismo grafico que el de la opción anterior pero proyectado sobre un dendograma.



. – **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 tengamos en la clase modalidades de la variable.



. – **Gráfico de palabra:** Realizara un gráfico de similitud representando los enlaces de la forma seleccionada con las otras formas de la propia clase.

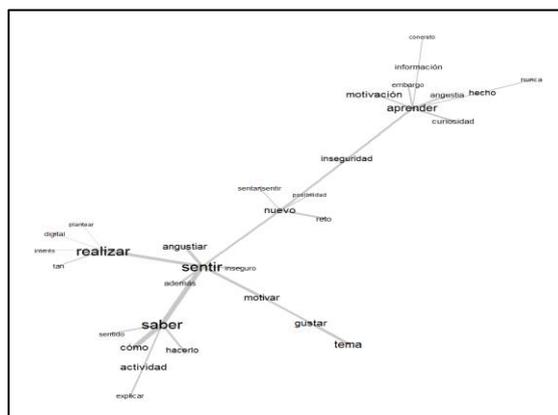


.- **Concordancia:** Nos permite obtener la concordancia de las formas y lemas seleccionados. Está disponible también para los segmentos de texto de la clase, así como para todas las clases del corpus.

. – **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.



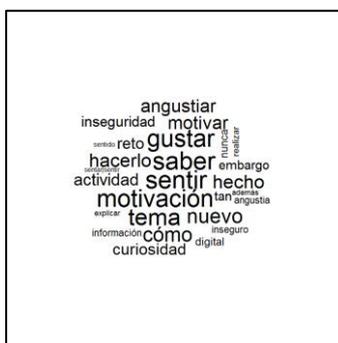
. – **Segmentos repetidos:** mostrará los efectivos y tamaños de los segmentos repetidos de la clase.

formas		
que me	11	2
no sabia	9	2
me senti	7	2
muy bien	7	2
un tema	7	2
las actividades	6	2
tema que	6	2
un poco	6	2
un tema que	6	3
a la	5	2
bien cómo	5	2
es un	5	2
sabía muy	5	2
sabía muy bien	5	3
trabajo y	5	2
la tarea	4	2
le guste	4	2
muy bien cómo	4	3
no me	4	2
no sabia muy	4	3
no sabia muy bien	4	4
que no	4	2
realizar la	4	2
sabía muy bien cómo	4	4
sin embargo	4	2
ya que	4	2
a la vez	3	3
al principio	3	2
al principio no	3	3
cómo hacerlo	3	2
de aprender	3	2
de hacer	3	2
de realizar	3	2
de trabajo	3	2
en un	3	2

. – **Segmentos de texto característicos:** hace una lista de los segmentos de textos característicos de la clase, proponiendo dos medidas: "Absoluta", 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.

```
**** *GRUP_2 *TORN_1 *EDAT_1 *ACC_1 *TREB_4 *EXPE_1
score : 20.83
que coja un tema que le guste
**** *GRUP_2 *TORN_1 *EDAT_1 *ACC_1 *TREB_1 *EXPE_2
score : 19.80
dudas porque no sabía bien cómo enfocararlo pero a la vez mejoraba mis conocimientos
**** *GRUP_2 *TORN_1 *EDAT_1 *ACC_1 *TREB_1 *EXPE_1
score : 18.77
que busque un buen tema para observar que lo motive y le guste descubrir e investigar sobre él
**** *GRUP_1 *TORN_2 *EDAT_2 *ACC_1 *TREB_3 *EXPE_1
score : 18.50
cuando hayas terminado de hacer un trabajo no sabras si lo has hecho bien o mal no obstante el profesorado te hace reflexionar y tu solo sabes perfectamente qué errores has cometido y cómo solucionarlo
```

. – **Nube de palabras por clase:** Permite realizar una nube con el léxico que seleccionemos dentro de cada clase.

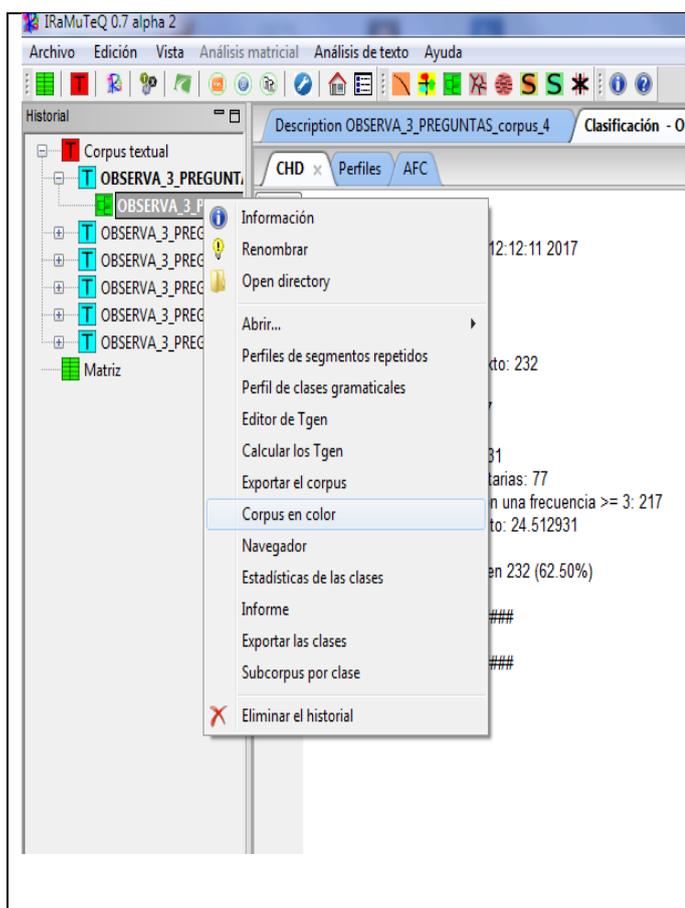


. – **Exportar:** Exporta todos los segmentos de texto presentes en la clase, puede incluso constituir un nuevo corpus.

. – **Exportar a Tropes:** Exportación al programa de análisis textual Tropes.

. – **Exportar a OwlEdge:** Exportación al programa OwlEdge.

Pero además, 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.



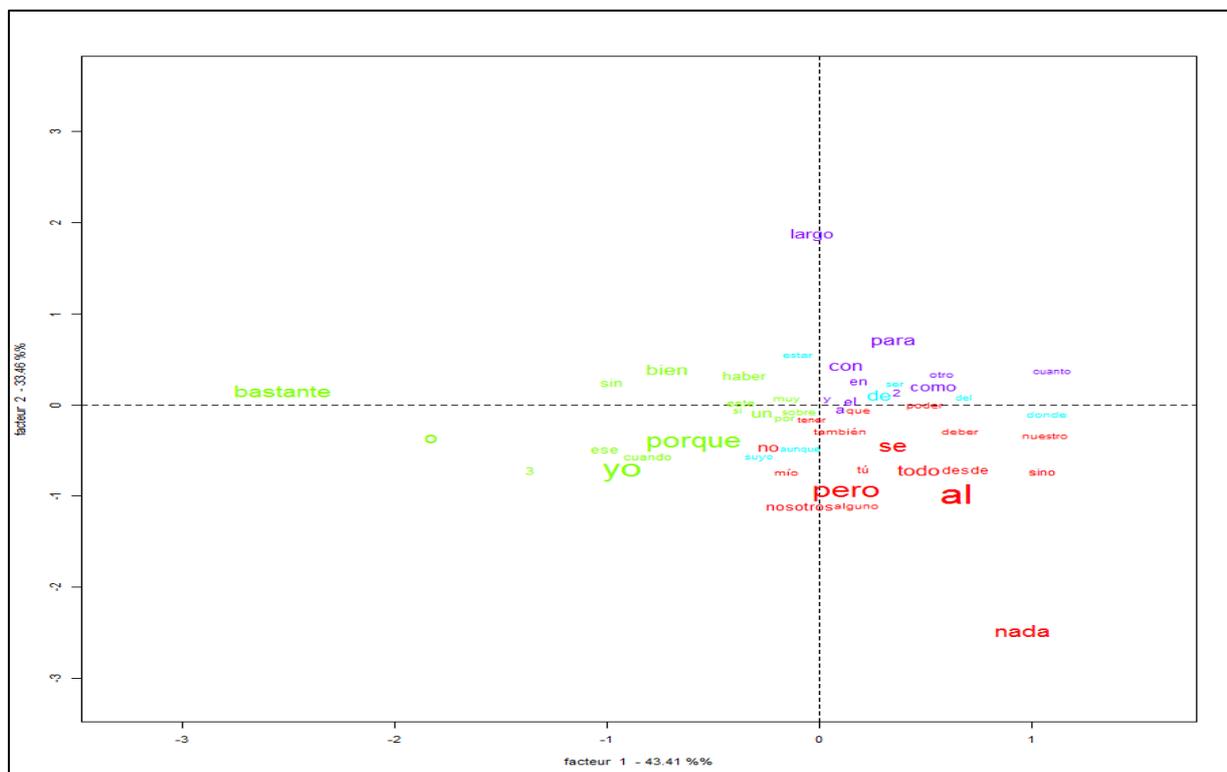
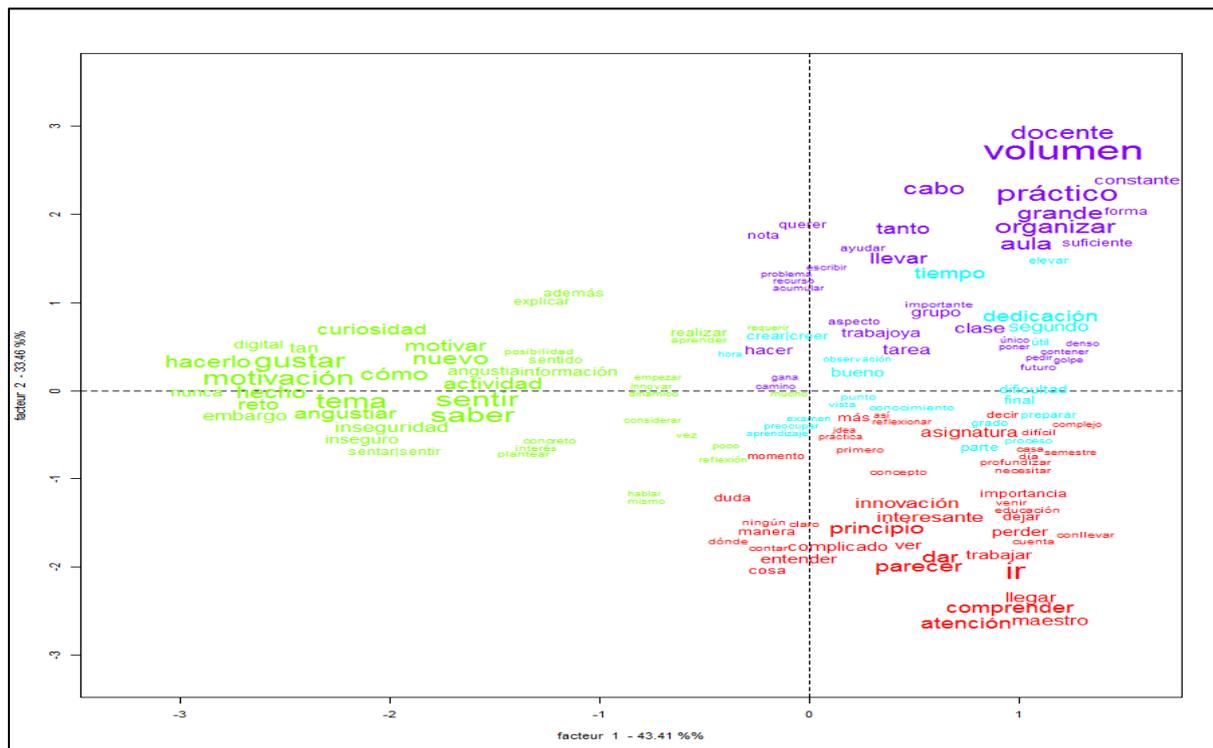
- .- **Información:** Retoma las características elegidas para la clasificación.
- .- **Abrir– anti-perfiles:** Recoge las formas significativamente ausentes en la clase.
- .- **Perfiles de segmentos repetidos:** Calcula los segmentos repetidos.
- .- **Perfiles de clases gramaticales:** Calcula los perfiles 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_classe.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.

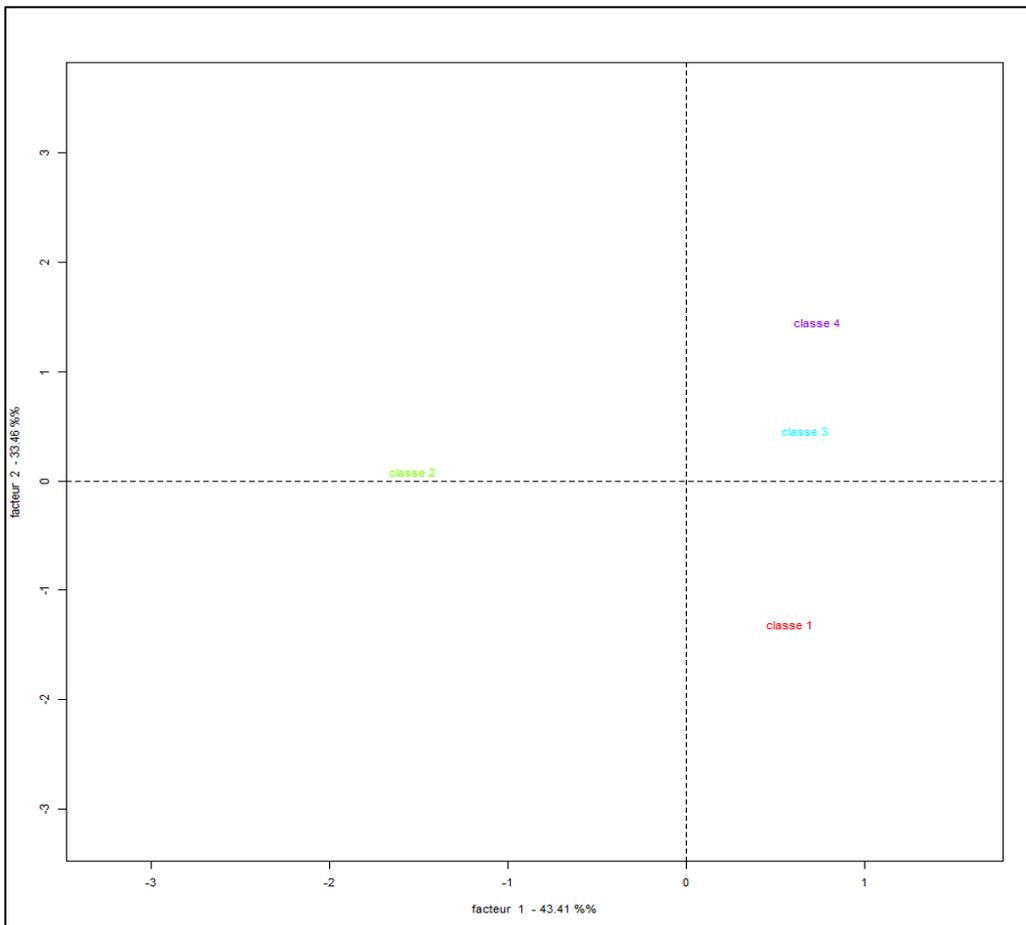
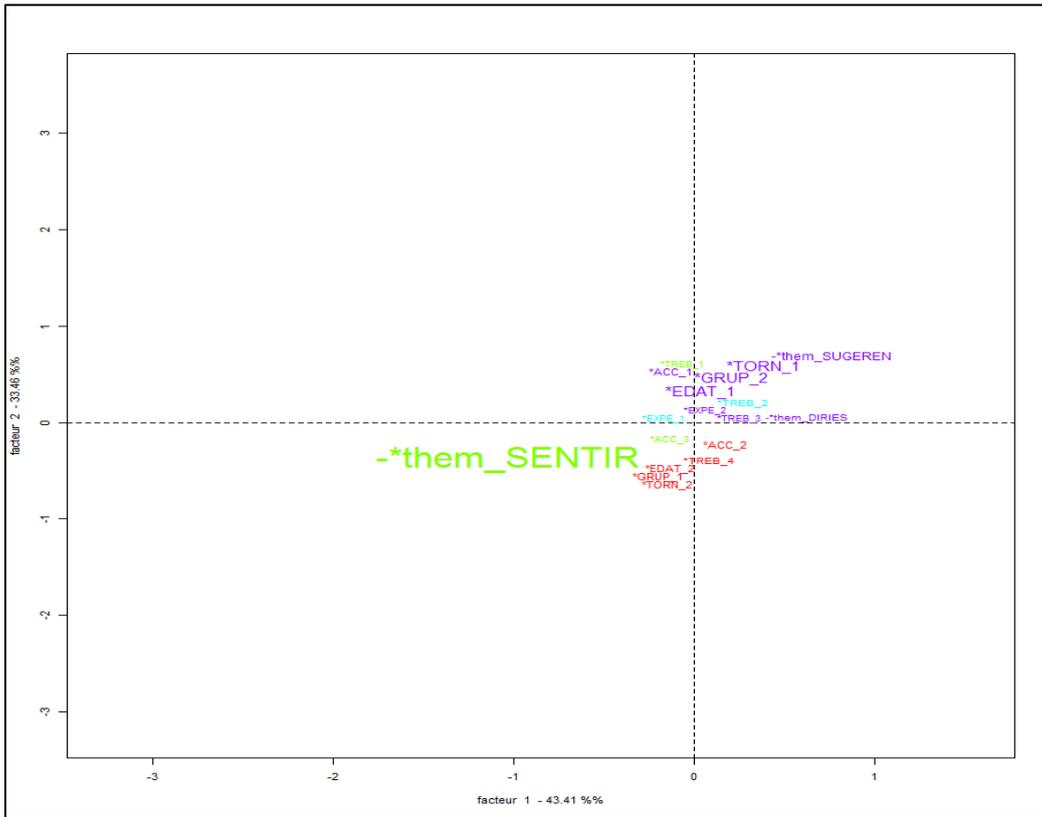
Moreno y Retinaud (2015)

La pestaña AFC:

Los resultados nos muestran el análisis factorial de correspondencias a la contingencia creciente de formas y lemas que utiliza el método de clasificación.

Los resultados son muy parecidos a los vistos en el análisis de especificidades. La diferencia estriba en que aquí se presentan las proyecciones sobre los ejes tanto de las formas activas, las formas suplementarias, las modalidades de las variables como de las clases obtenidas.



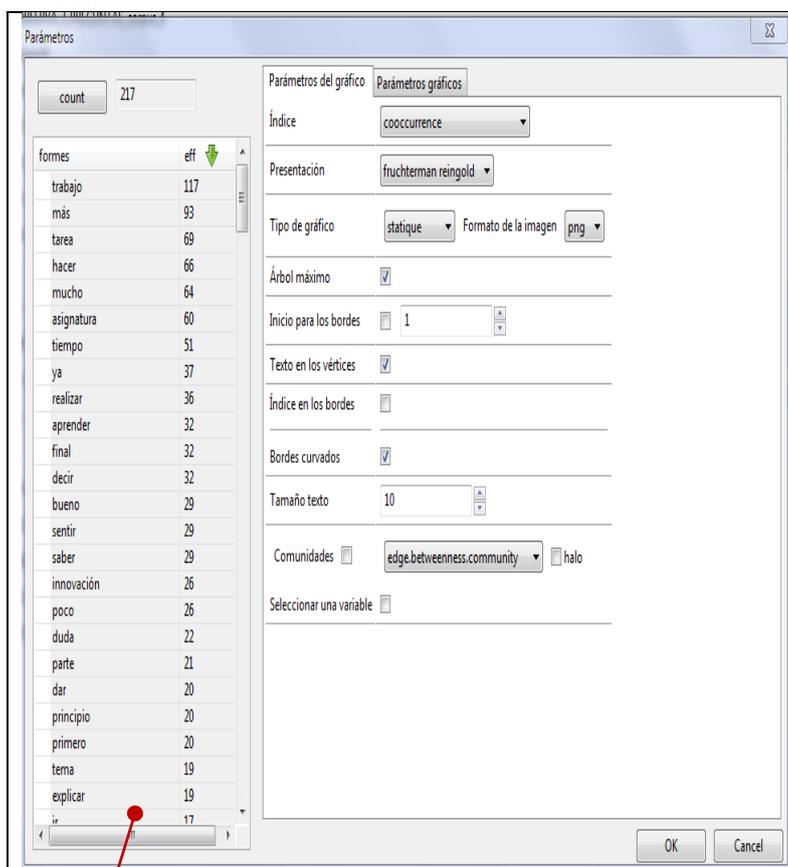


Análisis de Similitud.

Como ya se ha comentado el análisis de similitud nos proporciona las relaciones entre las diversas formas de un corpus. Es un análisis basado en la teoría de grafos. Para su obtención con IRAMUTEQ, se debe de pulsar el icono:



Nos proporciona una pantalla con tres tipos de opciones: La formas a seleccionar, Los parámetros del gráfico y parámetros gráficos. Veamos cada uno de ellos.



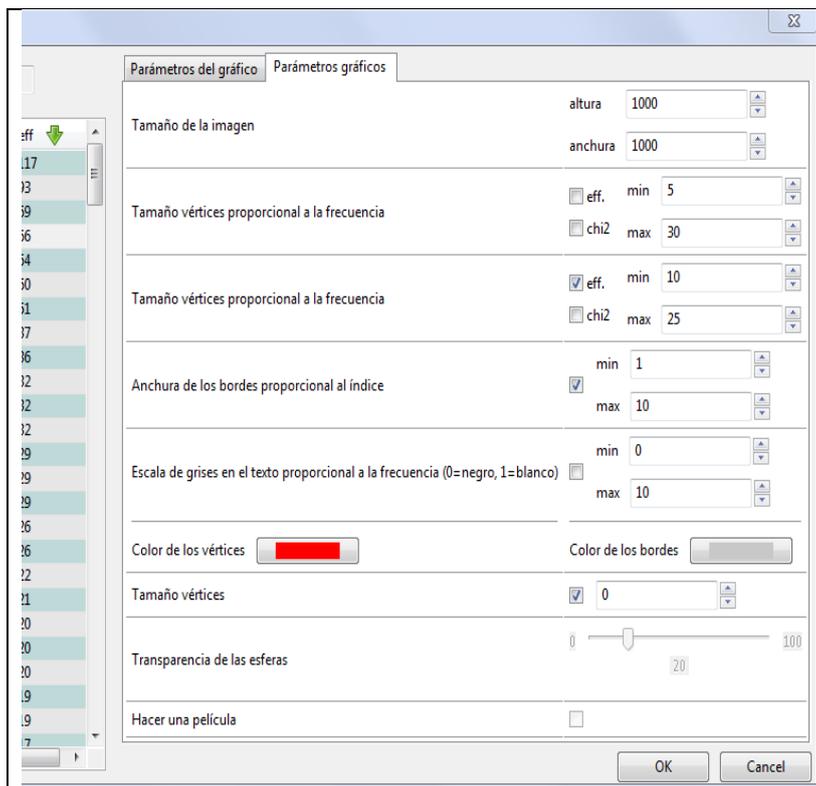
Parámetros del gráfico:

- .- **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 « igraph » 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, o bien,
 - Dynamyque: utiliza la interfaz tk de igraph. Esta interfaz permite organizar los gráficos y conservar las coordenadas para la reutilización en un gráfico estático. El 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 igraph) 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.

Moreno y Retinaud (2015)

Selección de variables

A la izquierda de la ventana es posible seleccionar las variables que se quieren incluir en el análisis. Si el corpus es muy amplio, el gráfico que generará el software será ilegible. Se recomienda en esos casos hacer una selección de las formas teniendo en cuenta sus frecuencias de aparición (eff.). Por defecto aparecen todas seleccionadas. Hágase un doble clic encima de ellas para borrar la selección. Con el Ctrl apretado se pueden hacer la selección de más de una forma a la vez.



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 χ^2 .

.- **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.

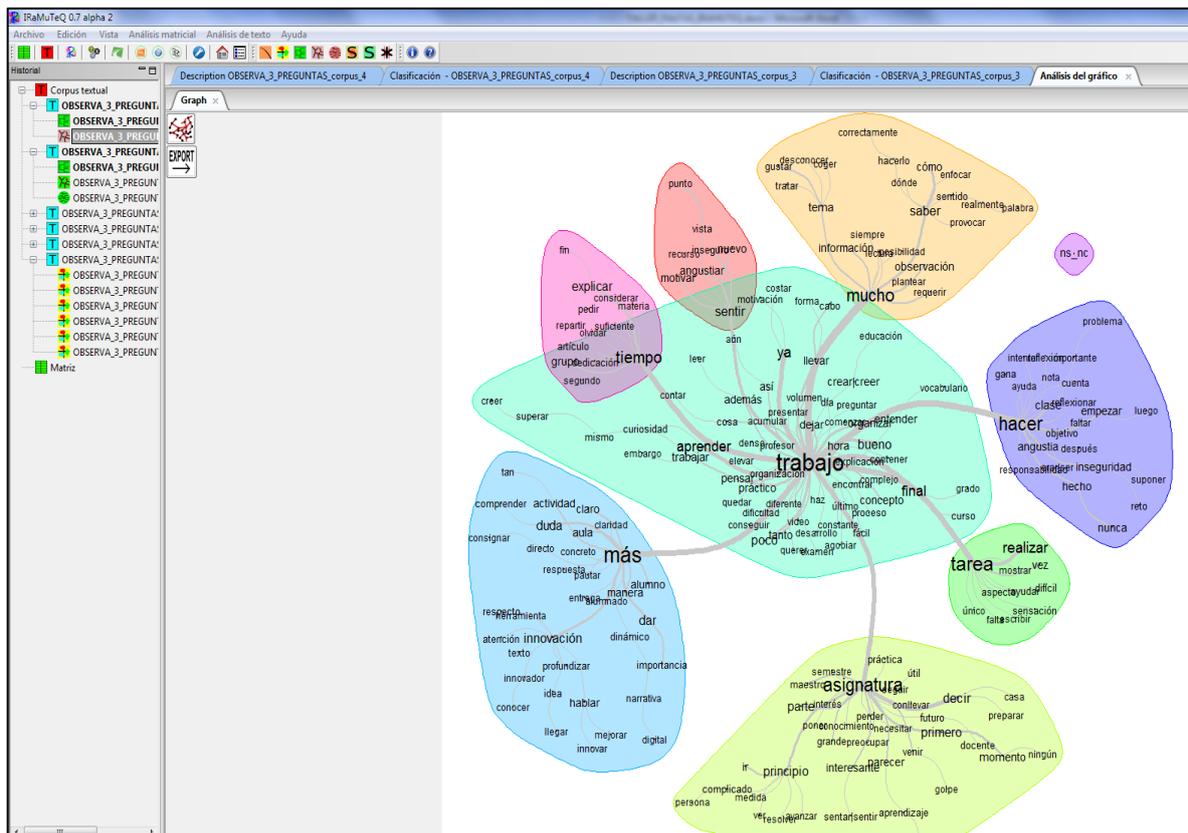
.- **Tamaños vértices:** tamaño único de los máximos del gráfico.

.- **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.

Moreno y Retinaud (2015)

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. En el ejemplo que presentamos a continuación se han escogido las opciones de parámetros del gráfico, **comunidades i halo**:



A la hora de la interpretación se ha de considerar las formas se encuentran en los nodos de la gráfica y los aristas/enlaces representan la co-ocurrencia entre ellos. A mayor frecuencia de las palabras, mayor tamaño de las mismas en el gráfico. A mayor co-ocurrencia entre palabras, más grueso se representa el enlace entre ellas. Se trata de dar cuenta de las relaciones entre las formas a partir de los objetivos de investigación.

A parte, este análisis permite, con los dos iconos que se presenta en la parte izquierda de los resultados, lo siguiente:



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



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

Las Nubes de palabras.

Es el análisis más simple ya que únicamente nos representa las palabras/formas en función de la frecuencia de aparición en el corpus.

Se activa mediante el icono:



Preferencias de nube de palabras

altura 800 anchura 800

Formato de la imagen png

Número máximo de formas 600

Formas utilizadas activas

Tamaño texto Min 5 Max 50

Color del texto Color de fondo

OK Cancel

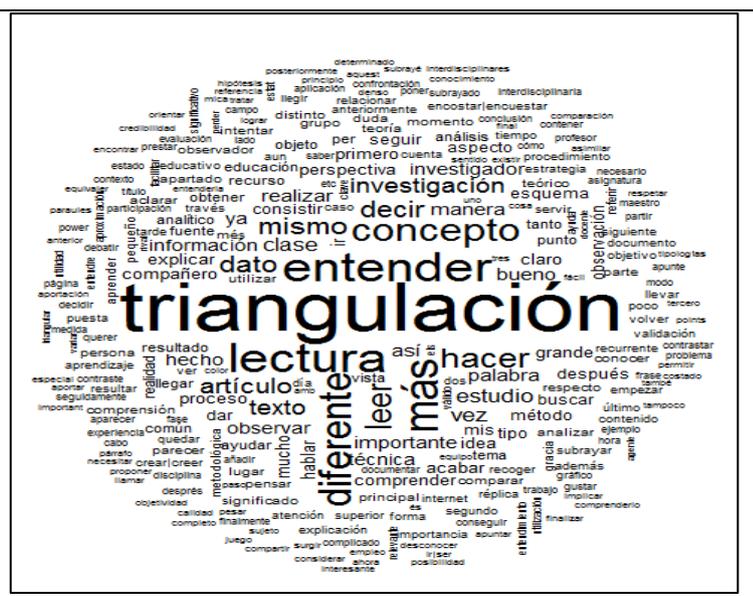
Preferencias:

- **Altura / Anchura:** fija las dimensiones de la nube.
- **Formato de la imagen:** permite guardar la imagen en .png o .svg
- **Número máximo de formas:** número máximo de formas presentes en la nube.
- **Formas utilizadas:** se puede elegir entre formas activas y/o suplementarias.
- **Tamaño texto:** establece los límites superiores e inferiores del tamaño del texto.
- **Color de Texto / Fondo:** define los colores.

Una vez dado el Ok, nos aparecerá la lista de formas que participarán en la confección de la nube de palabras/formas. Por defecto nos aparecerán todas las formas de la lista. Para hacer una selección bastara con mantener la tecla Ctrl al momento de la selección y escoger aquellas que nos interesen.

formas	eff	
triangulación	113	↓
lectura	62	
entender	62	
más	61	
diferente	52	
concepto	47	
leer	39	
hacer	37	
mismo	32	
dato	30	
decir	29	
artículo	28	
investigación	25	
texto	25	
estudio	24	
vez	24	
clase	21	
realizar	19	
así	19	
palabra	19	
manera	18	
...	10	

OK Cancel

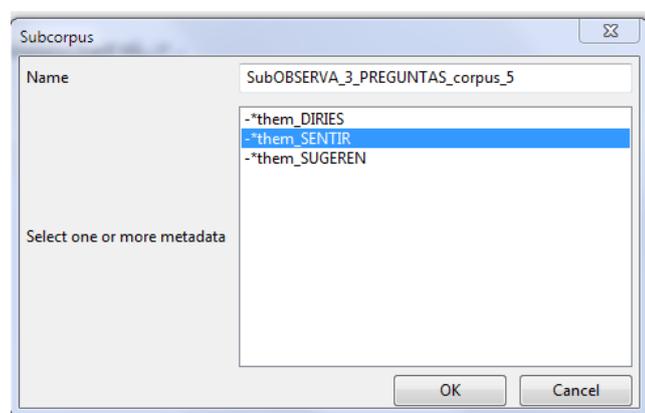


LA SELECCIÓN DE CORPUS/SUBCORPUS.

Existe una función importante en el IRAMUTEQ que nos permite trabajar con partes de un corpus. En este sentido, tenemos a nuestra disposición dos tipos de selección del corpus en subcorpus. Una que nos permite seleccionar una temática de un corpus (denominada subcorpus por temática) y la otra, la elección de una modalidad de una variable (llamada subcorpus por metadatos).

Veamos cómo se realiza cada una de estas opciones.

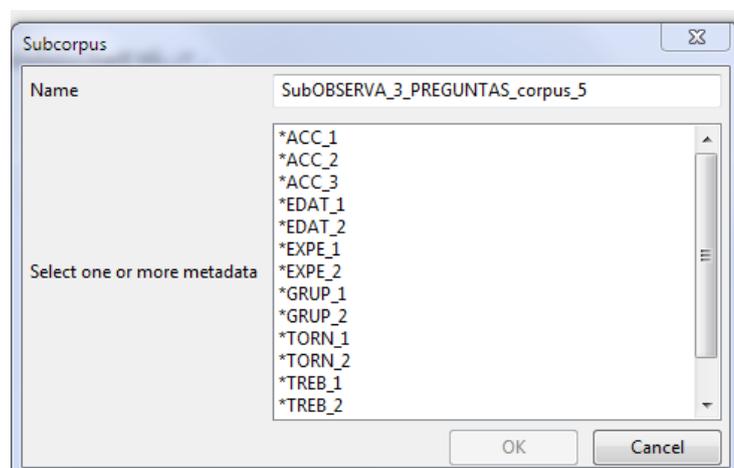
Para la creación de un subcorpus por temática, pulsar el icono:



Elegir la temática a crear como subcorpus.

Pulsar OK

Selección de un subcorpus por metadatos (modalidad variable)

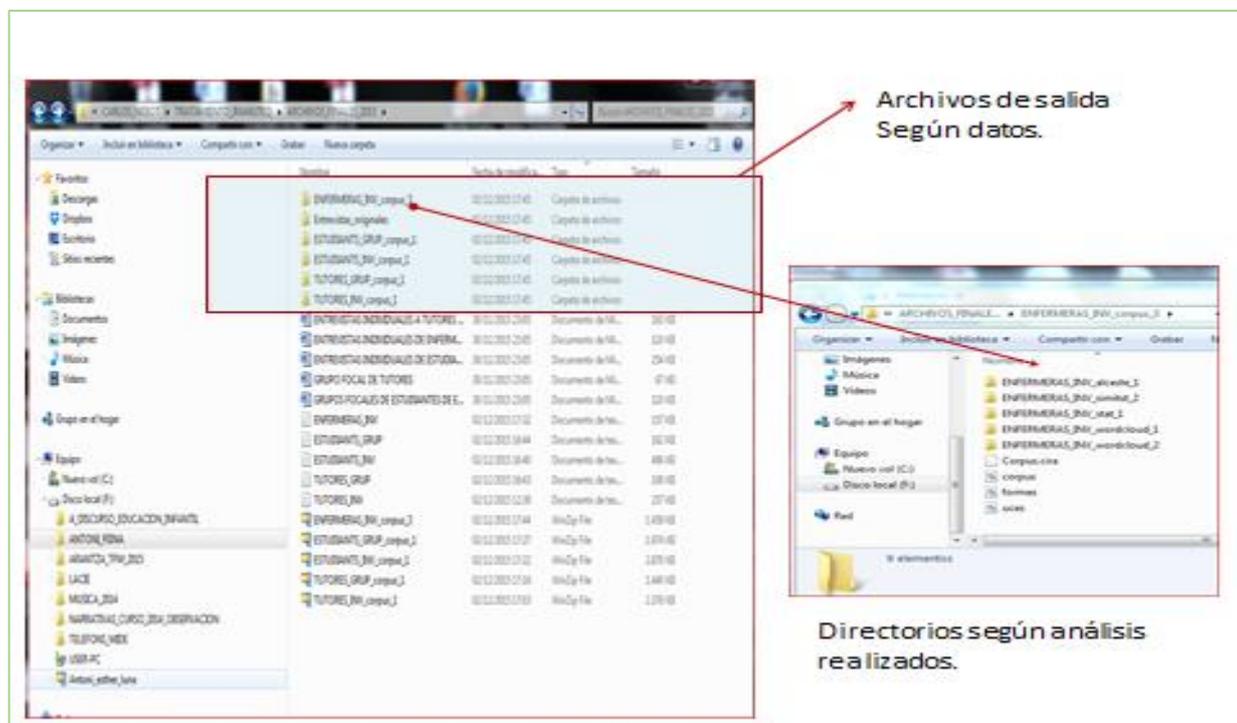


Elegir la modalidad de la variable para crear el subcorpus.

Pulsar OK

LAS SALIDAS (OUTPUTS) DE LOS ANÁLISIS.

Todos los análisis realizados con el programa se organizan sus resultados por carpetas y se ubican en el mismo lugar donde se encuentra el archivo que contiene el corpus. La carpeta corresponde a los diversos análisis que hemos ido viendo.



Terminología:

Corpus	Conjunto de textos que se han introducido al programa, siguiendo una serie de criterios y con un fin determinado.
Forma	Unidad léxica. Palabra con la que trabaja el software en caso de optar por lematizar el corpus.
Hápax	Formas que cuya frecuencia de aparición en el texto es igual a uno.
Lematizar	<p>Reducir las palabras principales a sus raíces, deviniendo así “morfemas lexicales”.</p> <p>«El morfema lexical es la raíz de la palabra que remite a la familia de la cual proviene, opuesto a los morfemas derivativos que son las terminaciones que indican un matiz semántico o los morfemas flexivos que indican el género o la cantidad. Así, en la palabra “niño”, el morfema lexical es <i>niñ</i>; sus terminaciones <i>o, os, a, as</i>, son los morfemas flexivos, mientras que terminaciones como <i>ito, ita, ote, ota</i>, corresponden a los morfemas derivativos. Alceste sólo toma en cuenta los morfemas lexicales para el análisis y elimina los morfemas flexivos o derivativos, aunque indica en un análisis suplementario todas las terminaciones de cada morfema lexical o raíz analizada.» (Alba, 2004).</p> <p>Para facilitar la interpretación de los resultados, los verbos se presentan en infinitivo, los adjetivos en masculino y los nombres en singular.</p>
Unidad de Contexto Elemental (UCE)	Lo que conocemos en Iramuteq por ‘segmento de texto’, se conoce como Unidad de Contexto Inicial en el software ALCESTE.
Unidad de Contexto Inicial (UCI)	Lo que conocemos en Iramuteq por ‘texto’, se conoce como Unidad de Contexto Inicial en el software ALCESTE.

Molina, J. (2017).

Referencias Bibliográficas:

- Abad-González, J, Blanco-Alonso, P. & García-Gallego, A. (2008). Análisis de Correspondencias y estudio de historias de vida: Una aplicación a la Encuesta de Transición Educativo-Formativa e Inserción Laboral **Universidad de León**. Facultad de Ciencias Económicas y Empresariales. Departamento de Economía y Estadística.
- Becué, M. (2010). Minería de Textos. Aplicación a preguntas abiertas en encuestas. Madrid. La Muralla.
- Benzécri J-P & et alt. (1973). *L'Analyse des Correspondances, Vol II*. Paris: Bordas.
- Benzécri J-P & col.(1981). *Pratique de l'Analyse des Données, tome III, Linguistique et Lexicologie*. Paris: Dunod.
- Camargo, B. & Justo, A. (2013). Tutorial para uso do software de análise textual IRAMUTEQ.
- Camargo, B. & Justo, A. (2016). Iramuteq Tutorial. (T. Forte, Trad.). Recuperado a partir de http://www.iramuteq.org/documentation/fichiers/IRaMuTeQ%20Tutorial%20translat%20to%20English_17.03.2016.pdf
- De Alba, M. (2004). El Método ALCESTE y su Aplicación al Estudio de las Representaciones Sociales del Espacio Urbano: El Caso de la Ciudad de México. *Papers on Social Representations*, 13, 1.1-1.20 [<http://www.psr.jku.at/>]
- Delatre, L. (2013). *Analyse des déterminants des choix de préservation des espaces agricoles et naturels dans les politiques locales d'urbanisme Apports d'une approche multi-méthodes* (Doctoral dissertation, Sciences-Po, Paris, France).
- Gambette P., Véronis J. (2010). Visualising a text with a tree cloud. In Locarek-Junge H. and Weihs C. (Eds). *Classification as a Tool for Research. Studies in Classification, Data Analysis, and Knowledge Organization. Part 3*, SpringerLink, pp. 561-569.
- Ghiglione, R. & Matalon, B. (1989). *Las encuestas sociológicas. Teorías y práctica*. México: Ed. Trillas.
- Gonzalez, B. (1991). *Análisis multivariante: Aplicación al ámbito sanitario*. Barcelona: SG editores.
- Gragé, D. & Lebart, L. (1994). *Traitements statistiques des enquêtes*. Paris: Dunod.
- Lemaire, B. 2008. . Limites de la lemmatisation pour l'extraction de significations. *Actes des 9e Journées internationales d'Analyse Statistique des Données Textuelles* :725-732.
- Lebart, I. Morineau, A. & Fénelon ; J. (1985). *Tratamiento estadístico de datos. Métodos y programas*. Barcelona: Marcombo.
- Lebart, L.(1986). Analyse statistique des réponses libres dans les enquêtes par sondage. *Revue Française du Marketing*, 109.
- Lebart, L; Salem, A. (1988). *Analyse statistique de donnes textuelles. Questions ouvertes et lexicométrie*. Paris : Bordas.

Lebart, L.; Salem, A. & Bécue, M. (2000). Análisis estadístico de textos. Milenio: Lleida.

Lebart, L. (2010). Validation techniques for textual data analysis. *Statistica Applicata - Italian Journal of Applied Statistics* Vol. 22 (1.)

Marchand, P., & Ratinaud, P. (2012). L'analyse de similitude appliquée aux corpus textuels: les primaires socialistes pour l'élection présidentielle française (septembre-octobre 2011). En *Actes des 11eme Journées internationales d'Analyse statistique des Données Textuelles*. (pp. 687–699).

Molina, J. (2017). Tutorial para el análisis de textos con el software Iramuteq. Recuperado a partir de: https://www.researchgate.net/publication/315696508_Tutorial_para_el_analisis_de_textos_con_el_software_IRAMUTEQ

Moreno, M., & Ratinaud, P. (2015). Manual de uso de Iramuteq. Versión 0.7 alpha 2. Recuperado a partir de http://iramuteq.org/documentation/fichiers/guia-iramuteq/at_download/file

Ratinaud, P. & Marchand, P. (2011). "Application de La Méthode ALCESTE Aux «gros» Corpus et Stabilité Des «mondes Lexicaux»: Analyse Du «CableGate» Avec IRAMUTEQ." 10.

Reinert, A. (1983). Une méthode de classification descendante hiérarchique: application à l'analyse lexicale par contexte. *Les cahiers de l'analyse des données*, 8(2), 187–198.

Reinert, M. (1987). Classification Descendante Hierarchique et Analyse Lexicale par Contexte- Application au Corpus des Poesies D'A. Rihbaud. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 13(1), 53–90.

Reinert, M. (1998). Quel objet pour une analyse statistique du discours? Quelques réflexions à propos de la réponse Alceste. En *JADT* (pp. 557–569). Recuperado a partir de http://w3dev.ualg.pt/~lfaisca/SMAD03/JADT_Reinert_1998.pdf

Reinert, M. (2003). Le rôle de la répétition dans la représentation du sens et son approche statistique par la méthode «ALCESTE». *Semiotica-La Haye Then Berlin-*, 147(1/4), 389–420.

Roy, N. & Garon, R. (2013). Étude comparative des logiciels d'aide à l'analyse de données qualitatives. *Recherches qualitatives – vol. 32(1)*, pp. 154-180. la reconnaissance de la recherche qualitative dans les champs scientifiques issn 1715-8702 – Recuperado de <http://www.recherche-qualitative.qc.ca/revue.html>.

Ruiz, A. (2013). El tratamiento de datos textuales con DTM_Vic 5.6. Depósito digital UB. <http://hdl.handle.net/2445/34029>.

Ruiz, A. (2016). Aproximaciones automáticas en el tratamiento de datos textuales: Programas DTM_Vic y Iramuteq. Depósito digital UB. <http://hdl.handle.net/2445/68983>

Sánchez, M. & Muñoz, A. (2001). Análisis de respuestas abiertas en un cuestionario sobre conductas relacionadas con el SIDA, *Revista de Psicología Social: International Journal of Social Psychology*, 16:3, 261-273, DOI: [10.1174/021347401317351035](https://doi.org/10.1174/021347401317351035).