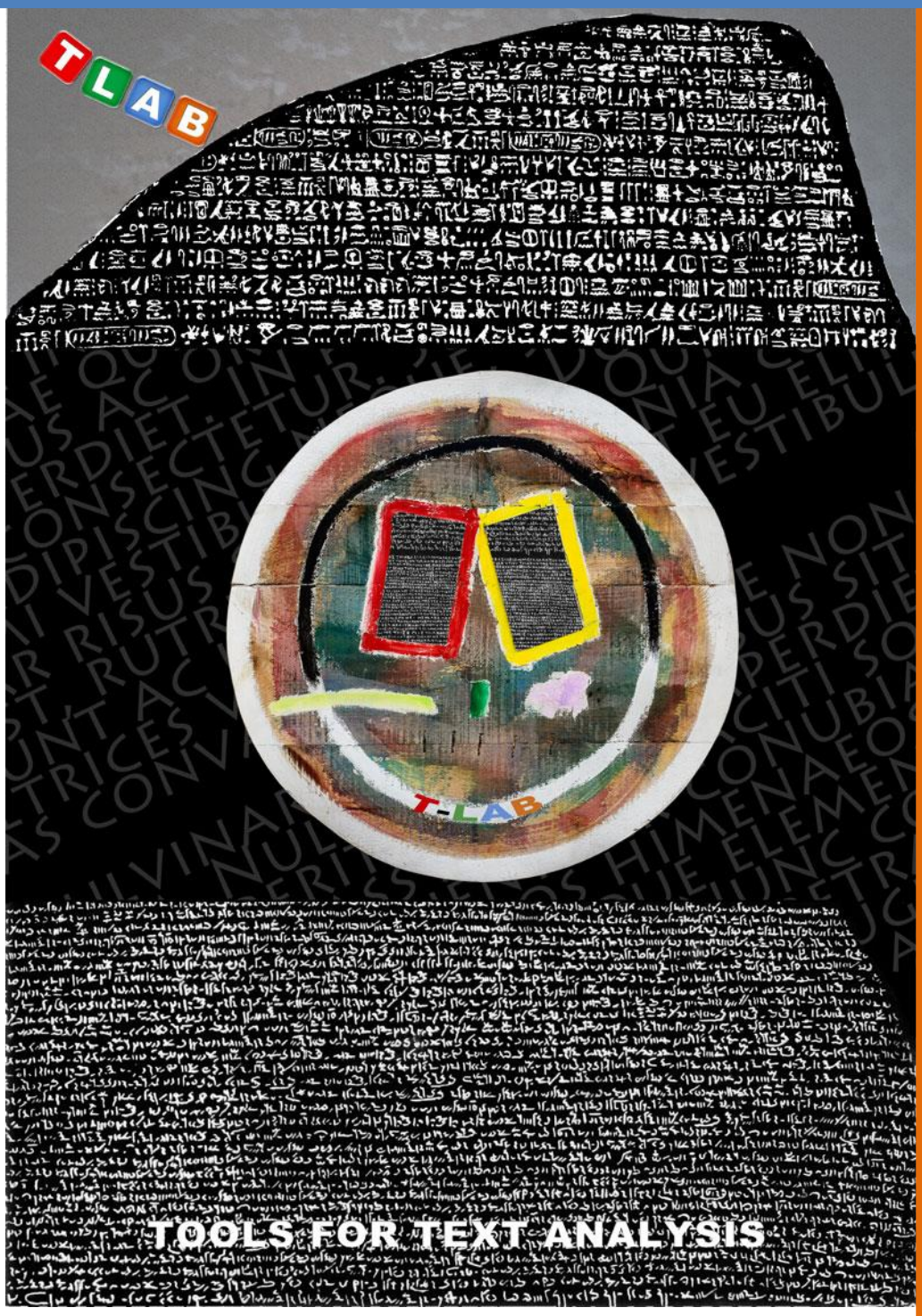
T-LAB 10

Guía de Inicio Rápido



The above artwork has been realized for T-LAB
by Claudio Marini (http://www.claudiomarini.it/)
in collaboration with Andrea D'Andrea.

Herramientas para el Análisis de Textos

Website: https://www.tlab.it/
E-mail: info@tlab.it

T-LAB: qué hace y qué permite hacer
(Extracto del Manual del Usuario)

T-LAB es un software compuesto por un conjunto de **herramientas lingüísticas, estadísticas y gráficas para el análisis de los textos.** Dichas herramientas se pueden emplear en las siguientes prácticas de investigación: Análisis del Contenido, Sentiment Analisis, Análisis semántico, Análisis Temático, Minería de Textos, Mapas Perceptuales, Análisis del Discurso y Network Text Analysis.

En efecto, gracias a las herramientas de **T-LAB**, los investigadores pueden gestionar ágilmente actividades de análisis como las siguientes:

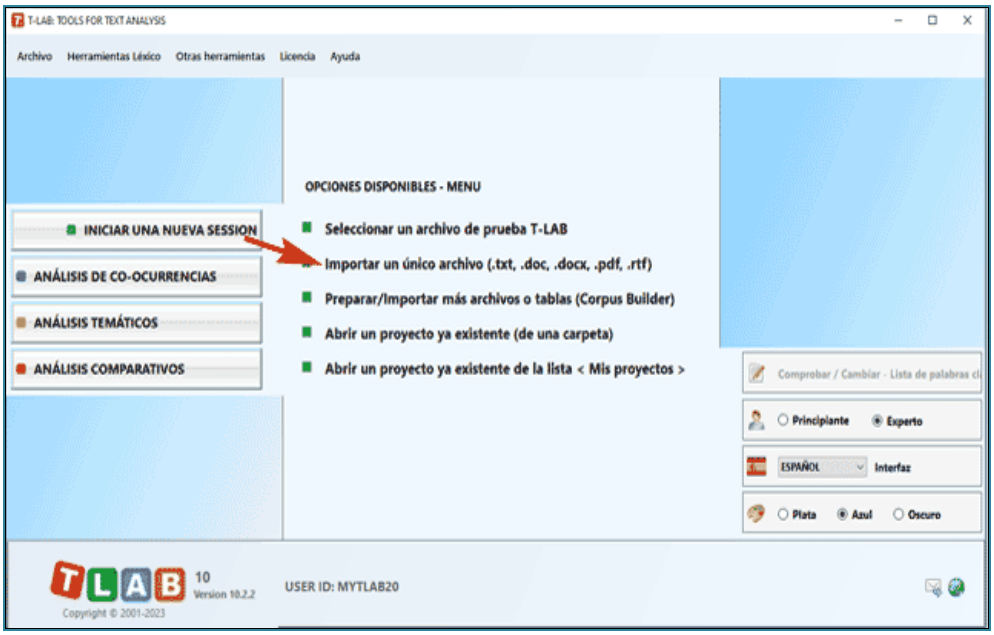
- Explorar, medir y mapear las **relaciones de co-ocurrencia entre palabras-clave**;
- Hacer una **clasificación automática** de las unidades de contexto y de los documentos, bien a través de una **metodología top-down** (es decir, mediante **categorías predefinidas**) o bien utilizando una **metodología bottom-up** (es decir, mediante el análisis de **temas emergentes**);
- Comprobar qué **unidades lexicales** (es decir, palabras o lemas), qué **unidades de contexto** (es decir, frases o párrafos) y qué **temas** son 'típicos' de subconjuntos específicos de determinados textos (p. ej.: discursos de líderes políticos, entrevistas con categorías específicas de personas, etc.);
- Aplicar categorías para la **sentiment analysis**;
- Ejecutar diferentes tipos de **análisis de las correspondencias** y de **análisis de los clústeres**;
- Generar **mapas semánticos** que representen **aspectos dinámicos del discurso** (es decir, relaciones secuenciales entre palabras o temas);
- Representar y analizar cualquier texto como si fuera una **red de relaciones**;
- Obtener medidas y representaciones gráficas sobre **textos y discursos tratados como sistemas dinámicos**;
- Personalizar y aplicar, tanto al análisis lexical como al análisis de contenido, **diferentes tipos de diccionarios**;
- Analizar todo el **corpus** o sólo algunos de sus **subconjuntos** (p. ej.: grupos de documentos) utilizando diferentes listas de palabras-clave
- Verificar los contextos de ocurrencia (p. ej.: **concordancias**) de palabras y lemas;
- Crear, explorar y exportar diferentes tipos de **tablas de contingencia** y **matrices de co-ocurrencias**.

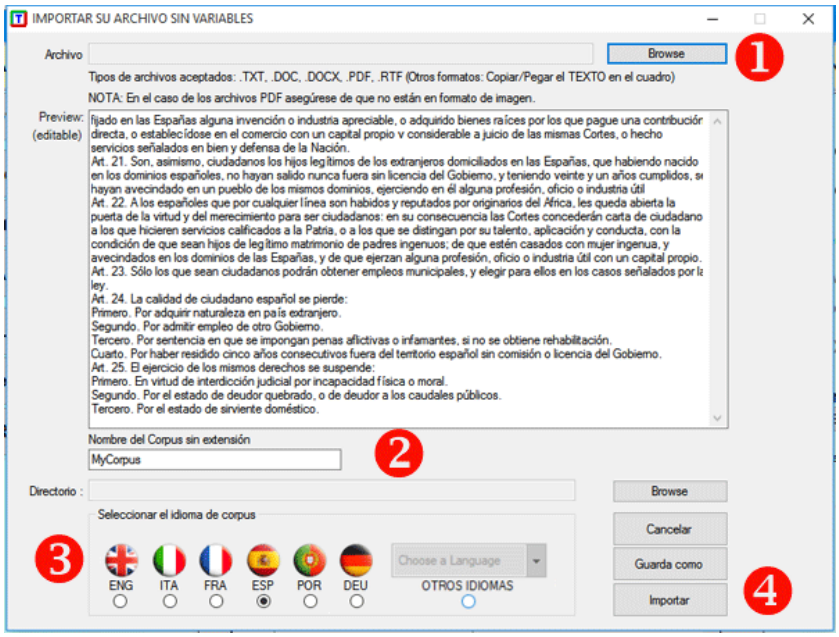
La interfaz de **T-LAB** es muy fácil de utilizar y los textos a analizar pueden ser de varios tipos:

- un único texto (ej. una entrevista, un libro, etc.);
- un conjunto de textos (ej. más entrevistas, páginas web, artículos de periódicos, respuestas a preguntas abiertas, mensajes Twitter, etc.).

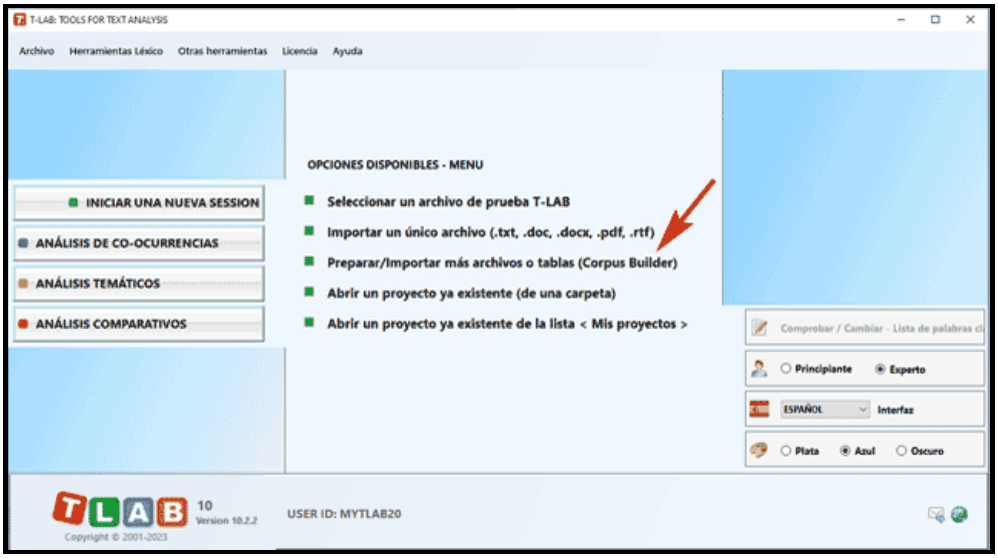
Todos los textos a analizar pueden ser codificados con **variables** categoriales y pueden incluir un identificativo (**Unique Identifier**) que corresponde a unidades de contexto o casos (ej. respuestas a preguntas abiertas).

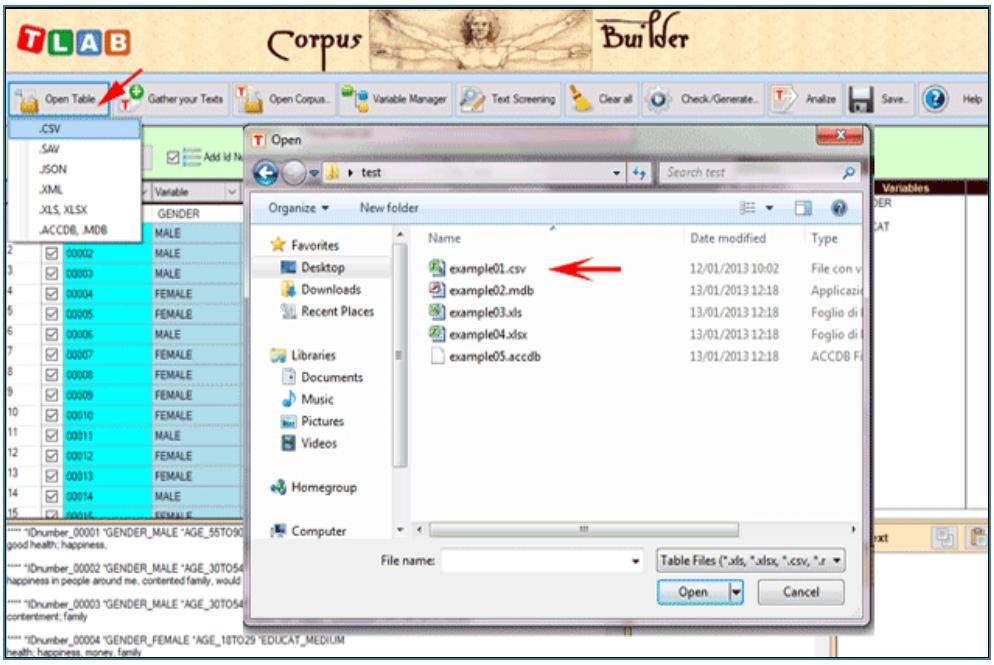
En el caso de un **único documento** (o corpus considerado como único texto), **T-LAB** no necesita nada más: es suficiente seleccionar la opción 'Importar un único archivo...'' (véase abajo).

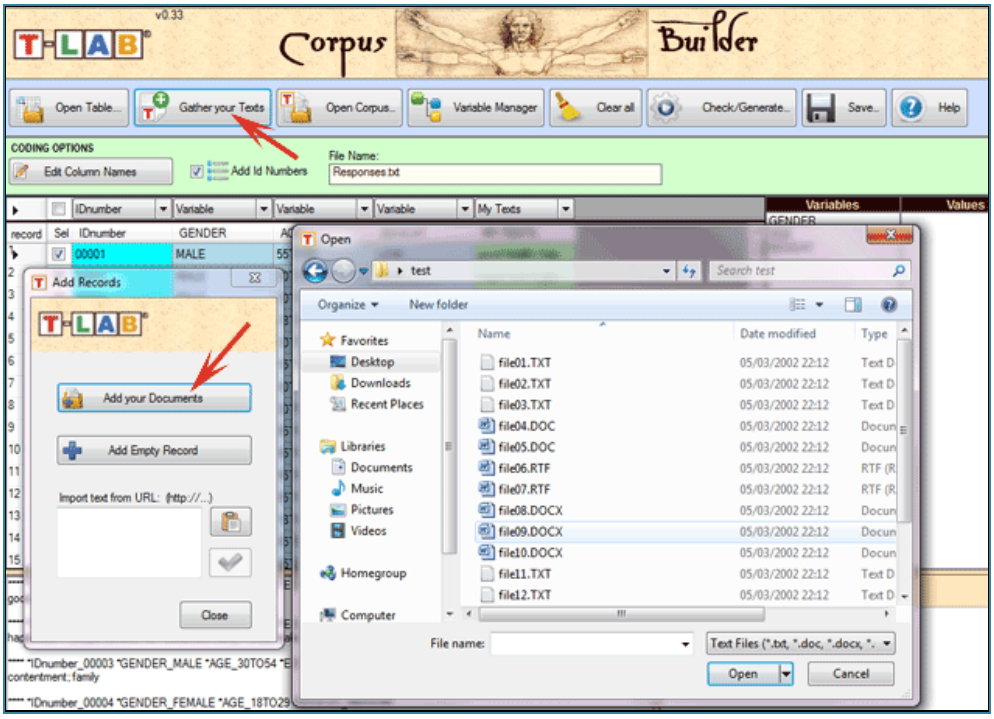




Cuando, en cambio, el corpus está compuesto por **más textos** y/o cuando se utilizan codificaciones que remiten al uso de alguna **variable**, la preparación del corpus requiere el uso del módulo **Corpus Builder** (véase abajo) que permite transformar los textos a analizar en un corpus codificado y listo para la importación.



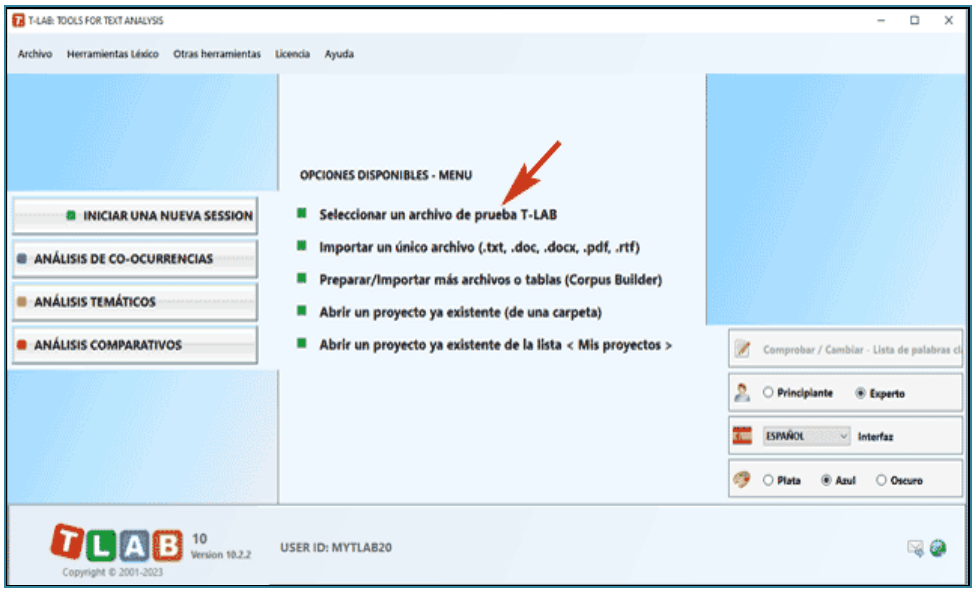




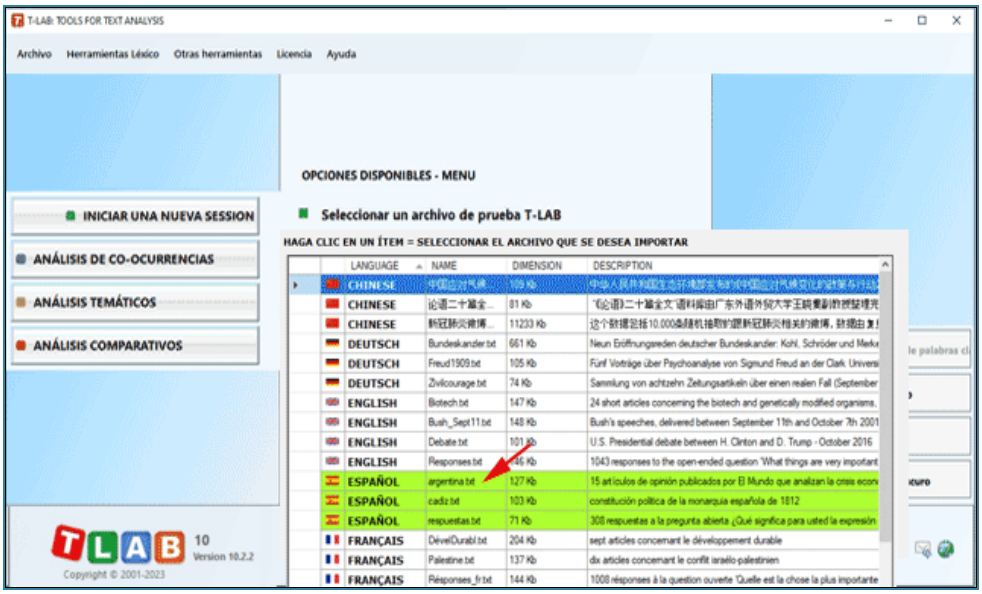
Nota: En el estado actual, **T-LAB** puede analizar archivos/corpus de hasta 90 MB de tamaño (es decir, aprox. 55.000 páginas en formato texto), garantizando el uso integrado de las distintas herramientas. Para más informaciones, véase la sección 'Requisitos y Prestaciones' del Help/Manual.

Para verificar rápidamente las funciones del software son suficientes seis pasos:

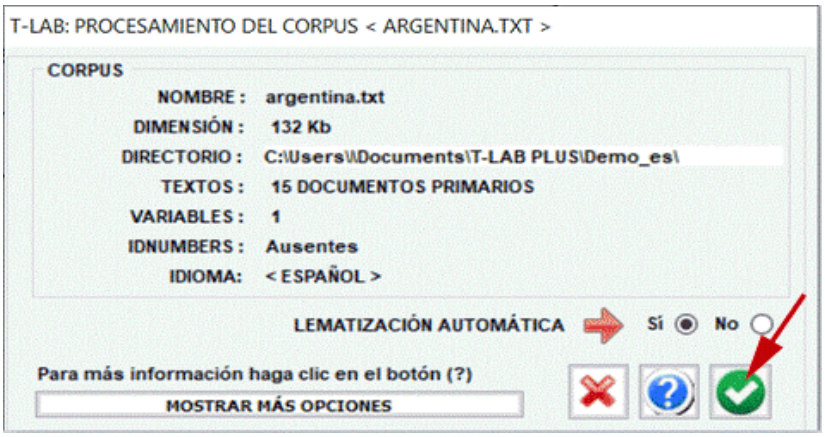
1 - Pulsar la opción 'Seleccionar un archivo de prueba T-LAB'



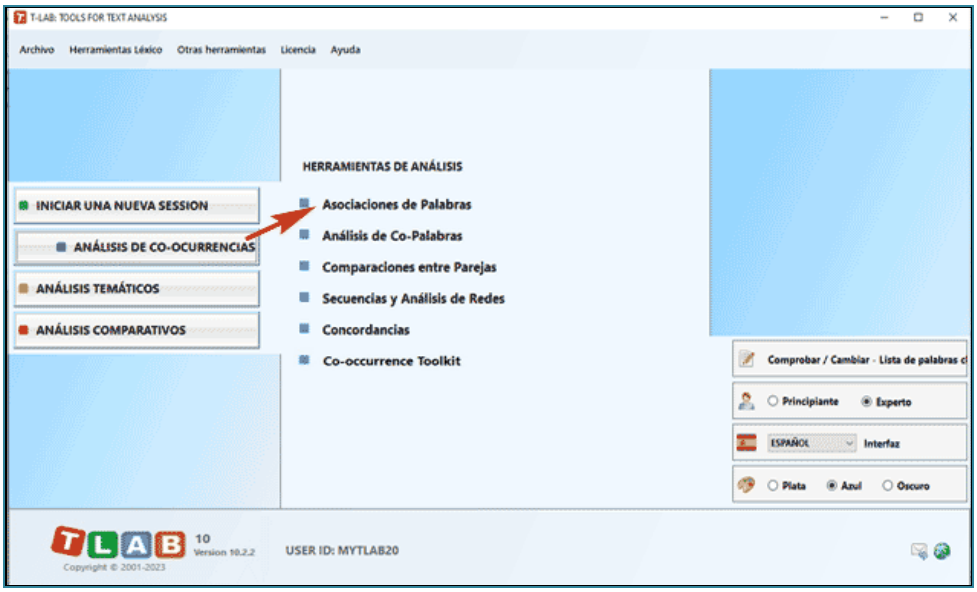
2 - Seleccionar un corpus a analizar



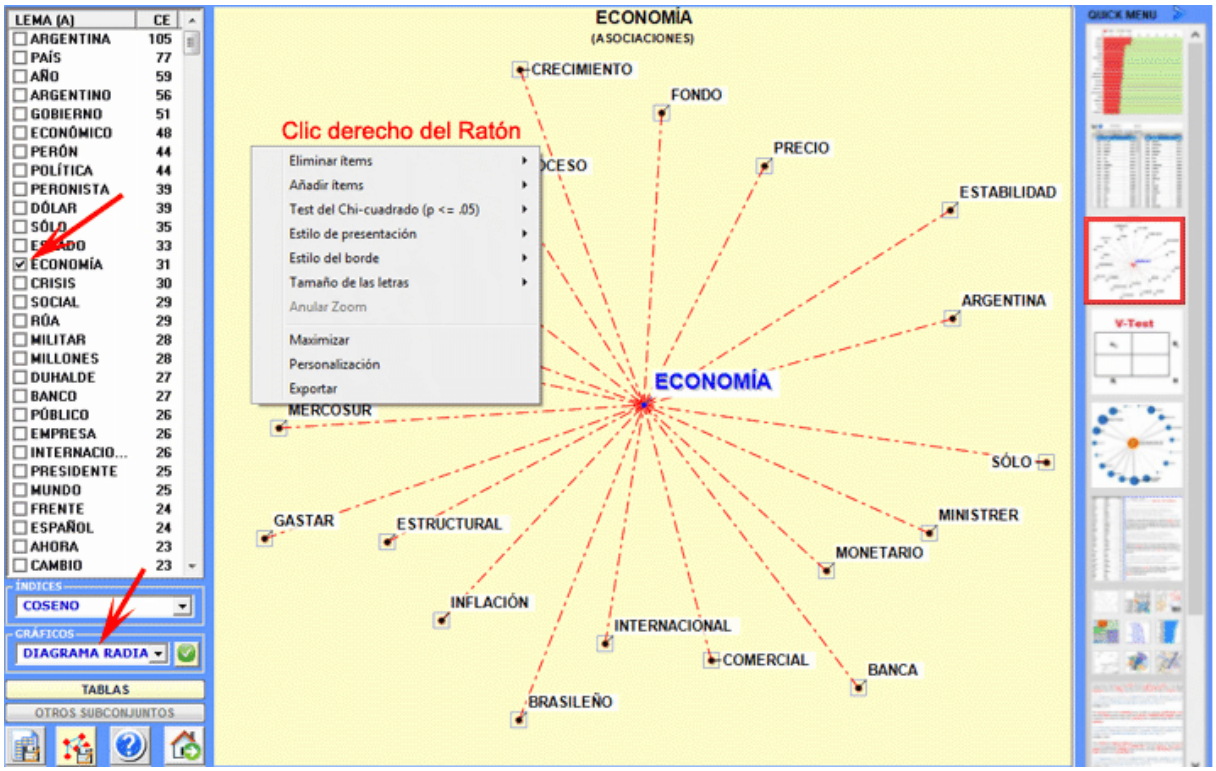
3 - Pulsar "ok" en la ventana de Configuración

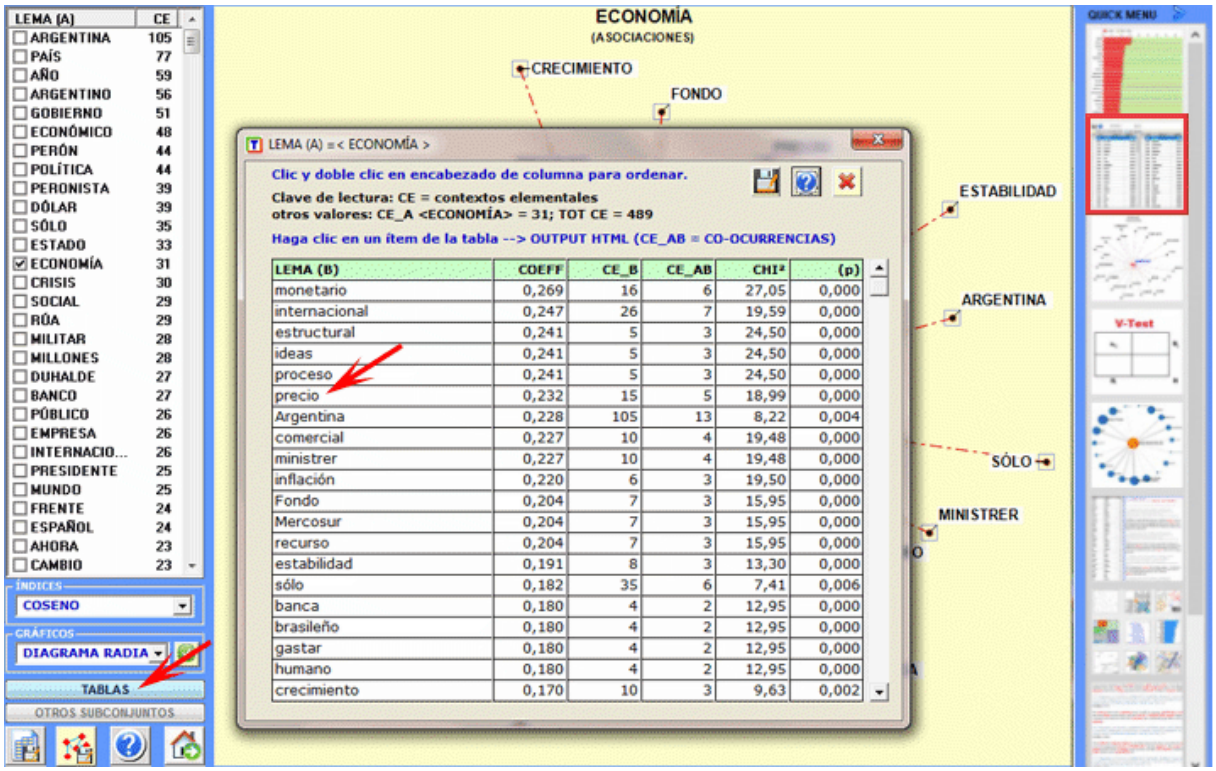


4 - Seleccionar una herramienta en uno de los submenús de "Análisis"

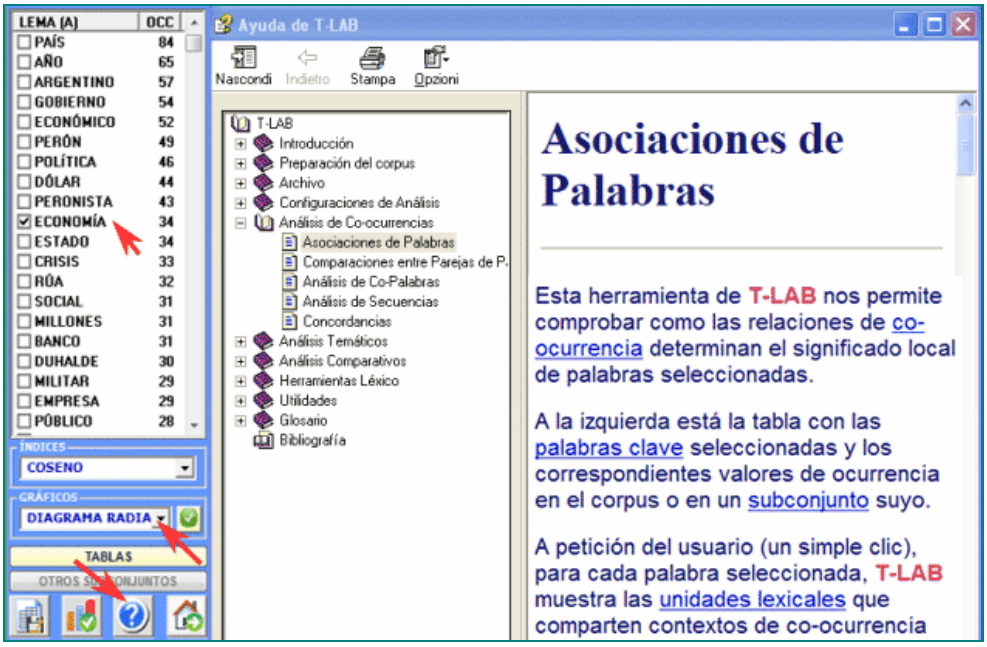


5 - Verificar los resultados





6 - Utilizar la ayuda contextual para interpretar gráficos y tablas.



Esta sección introductoria proporciona las informaciones básicas para entender cómo funciona **T-LAB** y cómo se puede utilizar.

Desde el punto de vista externo, el uso del software está organizado por la **interfaz**, es decir por el **menú principal**, los submenús y las **funciones** que lo componen.

Desde el punto de vista lógico, además de la interfaz, el sistema **T-LAB** está organizado por dos componentes principales:

- la **base de datos**, es decir el lugar informático en el que el **corpus** en input (es decir el texto o el conjunto de textos a analizar) está representado como un conjunto de **tablas** en las que se registran las **unidades de análisis**, sus características y sus relaciones recíprocas;

- los **algoritmos**, o sea subconjuntos de **instrucciones** que permiten usar la interfaz del usuario, consultar y modificar la base de datos, crear ulteriores tablas con datos en ésta contenidos, efectuar **cálculos estadísticos** y producir **outputs** que representen las relaciones entre los datos analizados.

Para entender cómo funciona **T-LAB** y cómo puede usarse es muy importante tener claro qué unidades de análisis se archivan en su base de datos y cuáles algoritmos estadísticos se usan en los distintos análisis. En efecto, las tablas de datos analizadas están siempre constituidas por filas y columnas cuyos membretes corresponden a las unidades de análisis archivadas en la base de datos, mientras que los algoritmos regulan los procesos que permiten descubrir relaciones significativas entre los datos y extraer informaciones útiles.

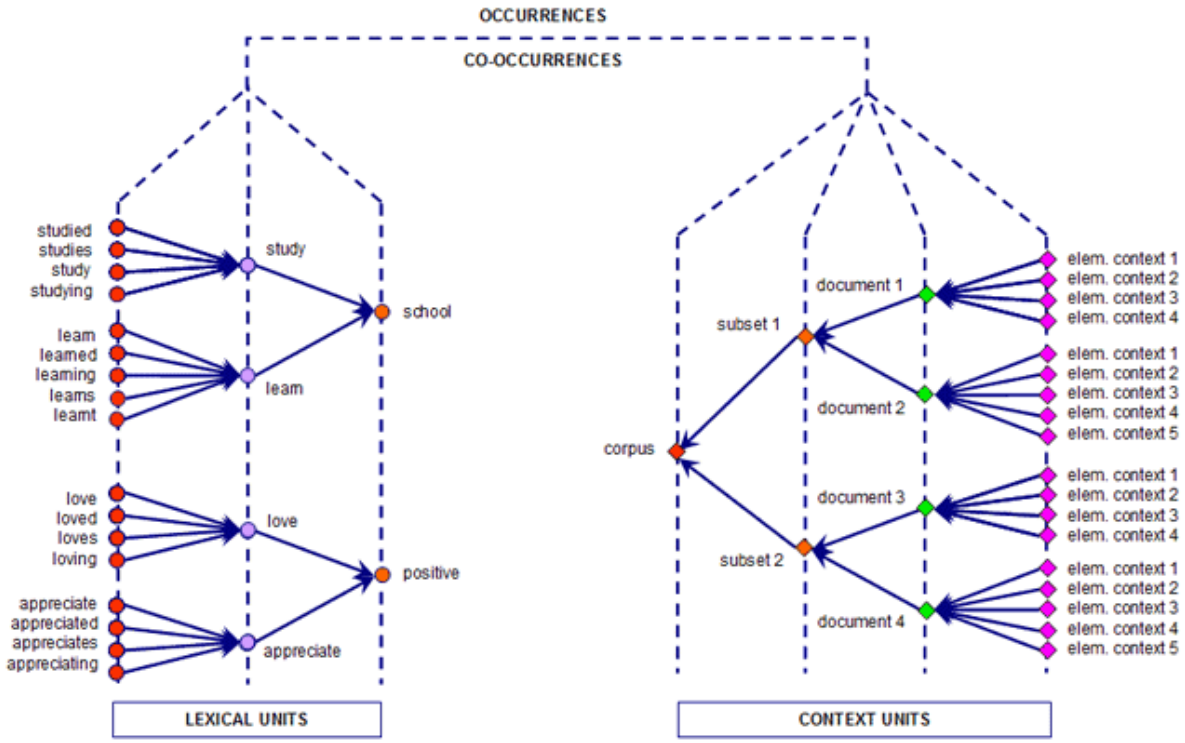
Las **unidades de análisis** de **T-LAB** son de dos tipos: **unidades lexicales** y **unidades de contexto**.

A - las **UNIDADES LEXICALES** son palabras, simples o múltiples, archivadas y clasificadas en base a un cierto criterio. En particular, en la base de datos **T-LAB**, cada unidad lexical constituye un registro clasificado con dos campos: **palabra** y **lema**. En el primer campo (**palabra**) se enumeran las palabras así como aparecen en el corpus, mientras que en el segundo (**lema**), se enumeran las etiquetas atribuidas a grupos de unidades lexicales clasificadas según criterios lingüísticos (ej. **lematización**) o a través de diccionarios y plantillas semánticas definidas por el usuario.

B - las **UNIDADES DE CONTEXTO** son porciones de texto en las que se puede dividir el corpus. En particular, en la lógica **T-LAB**, las unidades de contexto pueden ser de tres tipos:

B.1 documentos primarios correspondientes a la subdivisión "natural" del corpus (ej. entrevistas, artículos, respuestas a preguntas abiertas, etc.), o sea a los **contextos iniciales** definidos por el usuario;
B.2 contextos elementales, correspondientes a las unidades sintagmáticas (ej. fragmentos de texto, frases, párrafos) en las cuales cada documento primario puede ser subdividido;
B.3 subconjuntos del corpus que corresponden a grupos de documentos primarios atribuibles a la misma categoría (es. entrevistas de "hombres" o de "mujeres", artículos de un determinado año o de un determinado periódico, y así sucesivamente) o a clústers temáticos conseguidos a través de específicos instrumentos **T-LAB**.

El siguiente diagrama muestra las posibles relaciones que **T-LAB** nos permite analizar entre unidades lexicales y unidades de contexto.

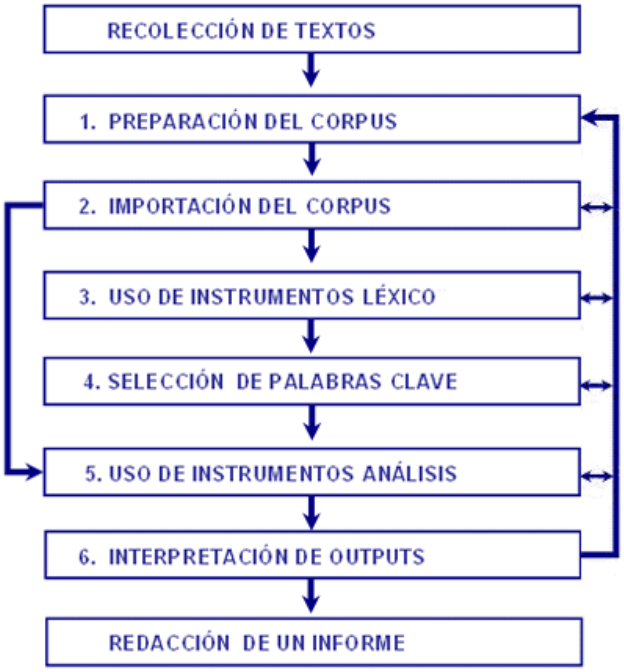


A partir de esta organización de la base de datos, **T-LAB** permite - automáticamente - explorar y analizar las relaciones entre las unidades de análisis de **todo el corpus** o de sus **subconjuntos**.

En **T-LAB**, la elección de cualquier instrumento de análisis (clic del ratón) activa siempre un proceso semi-automático que, con pocas o simples operaciones, genera algunas tablas input, aplica algún algoritmo de tipo estadístico y crea algunos outputs (ver diagrama siguiente).

Hipotéticamente, cada **proyecto** de trabajo en el que se usa **T-LAB** está constituido por el conjunto de actividades analíticas (operaciones) que tienen por objeto el mismo **corpus** y está organizado por una **estrategia** y por un **plan** del usuario. Por lo tanto, inicia con la **recolección de textos** a analizar y termina con la **redacción de un informe**.

La sucesión de las distintas fases está ilustrada en el siguiente diagrama:



NOTA:
- Las seis fases numeradas, desde la preparación del corpus a la interpretación de los output, tienen el soporte de los instrumentos **T-LAB** y son siempre reversibles;
- Por medio de las **configuraciones automáticas T-LAB** se pueden evitar dos fases (3-4); sin embargo, a los fines de la **calidad** de los resultados se recomienda la ejecución de las mismas.

Intentamos ahora comentar, una tras otra, las distintas fases.

1 - La **PREPARACIÓN DEL CORPUS** consiste en la transformación de los textos a analizar en un archivo (**corpus**) que puede ser elaborado por el software.

En el caso de un único texto (o corpus considerado como único texto), **T-LAB** no necesita nada más.

Cuando, en cambio, el corpus está formado por varios textos y hay **códigos** que hacen referencia a algunas **variables** se hace necesario utilizar el modulo **Corpus Builder,** que permite transformar los textos a analizar en un corpus codificado y listo para la importación.

NOTA:

- Al término de la fase de preparación se recomienda crear una nueva carpeta de trabajo en cuyo interior sólo se encuentre el archivo corpus a importar;

- Se recomienda, durante los análisis, mantener el archivo del corpus y la carpeta de trabajo relativa en un disco duro del ordenador donde se ha instalado **T-LAB**. De no ser así, podría ralentizarse la ejecución de los diferentes procedimientos y el programa podría llegar a mostrar errores.

2 - **LA IMPORTACIÓN DEL CORPUS** consiste en una serie de procesos automáticos que transforman el corpus en un conjunto de tablas integradas en la **base de datos T-LAB**.

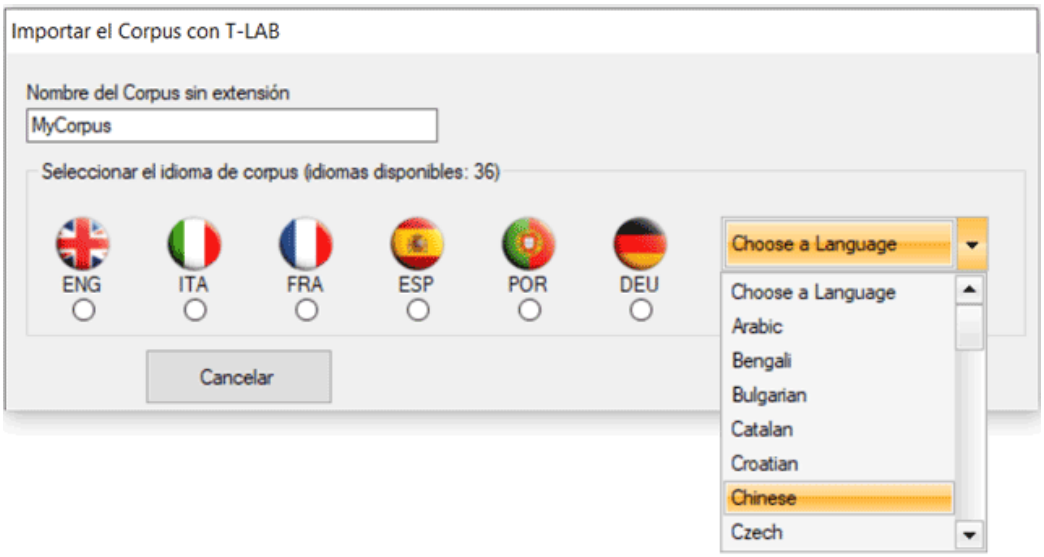
Durante el proceso de importación, **T-LAB** realiza los tratamientos siguientes: **normalización** del corpus; detección de **multi-palabras** y **palabras vacías**; segmentación en **contextos elementales**; **lematización** automática o **stemming**; construcción del **vocabulario**; selección de las **palabras clave**.

A continuación, se presenta el listado completo de los 30 idiomas para los cuales **T-LAB** prevé la posibilidad de implementar procesos de lematización automática y de stemming.

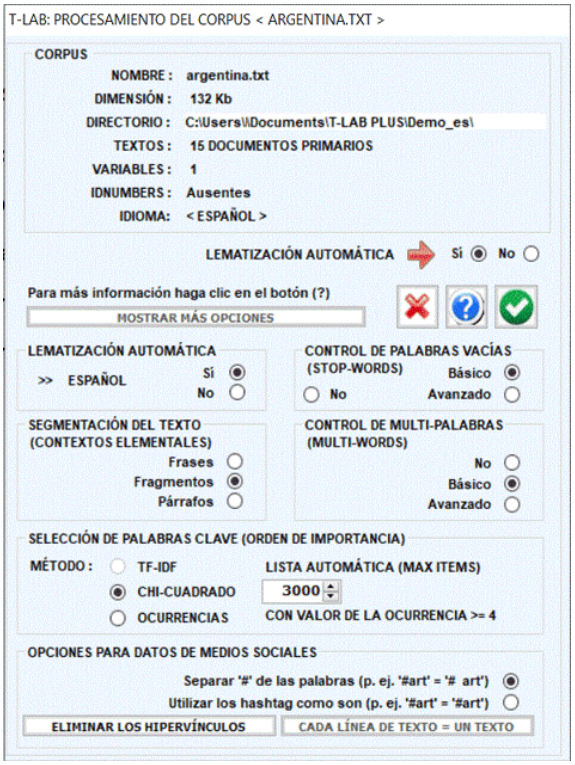
LEMATIZACIÓN: alemán, catalán, croata, eslovaco, español, francés, inglés, italiano, latín, polaco, portugués, rumano, ruso, serbo, sueco y ucraniano.

STEMMING: árabe, bengalí, búlgaro, checo, danés, finlandés, griego, hindi, húngaro, indonesio, marathi, noruego, persa y turco.

En cualquier caso, sin lematización automática y / o mediante diccionarios personalizados, el usuario puede analizar textos en **todos los idiomas**. Lo importante es que las palabras estén separadas por espacios y/o signos de puntuación.



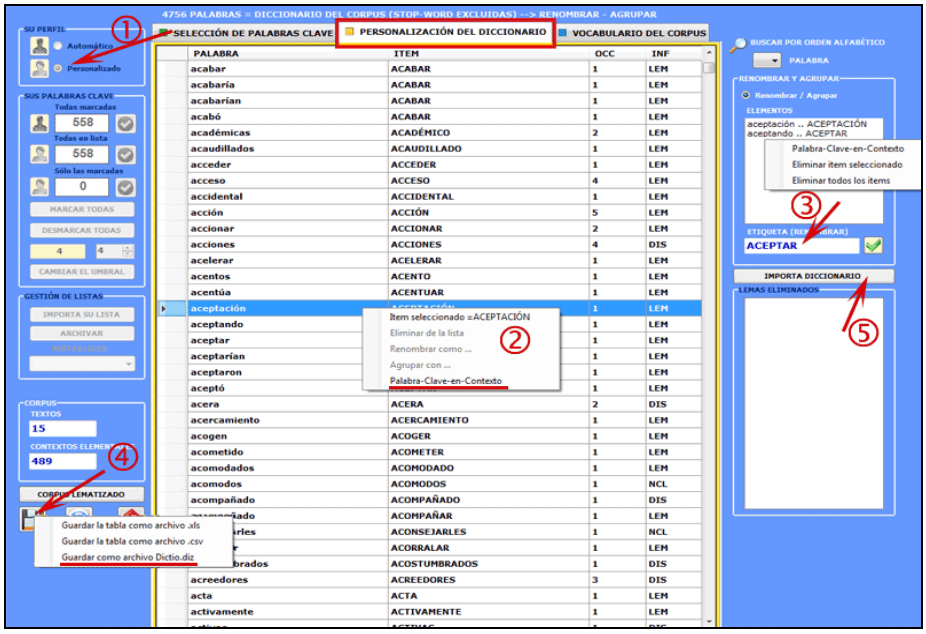
Una vez seleccionado el idioma, la intervención del usuario será necesaria para definir las opciones indicadas en la ventana siguiente:



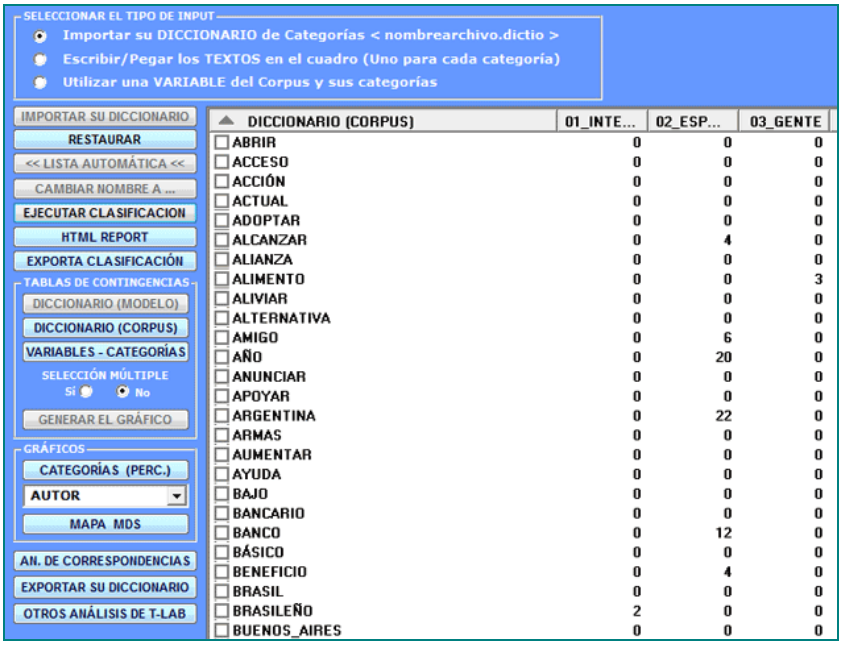
NOTA: Puesto que las diferentes opciones determinan el tipo y la cantidad de unidades de análisis (es decir las unidades de contexto y las unidades lexicales), diversas opciones determinan diversos resultados de análisis. Por esta razón, todos los outputs de **T-LAB** (es decir gráficos y tablas) utilizados en el manual del usuario y en la ayuda en red son solo indicativos.

3 - EL USO DE LAS HERRAMIENTAS LÉXICO está destinado a verificar el correcto **reconocimiento** de las unidades lexicales y a personalizar su **clasificación**, es decir a verificar y modificar las selecciones automáticas hechas por **T-LAB**.

Las modalidades de las diversas intervenciones están descritas en las correspondientes voces de la ayuda (y del manual).
En particular se redirecciona a la correspondiente voz de la ayuda (y del manual) para una descripción detallada del proceso **Personalización del Diccionario**. De hecho, cualquier modificación a las voces del diccionario (p. ej.: agrupación de dos o más ítems) incide tanto en el cálculo de las ocurrencias como en el cálculo de las co-ocurrencias)

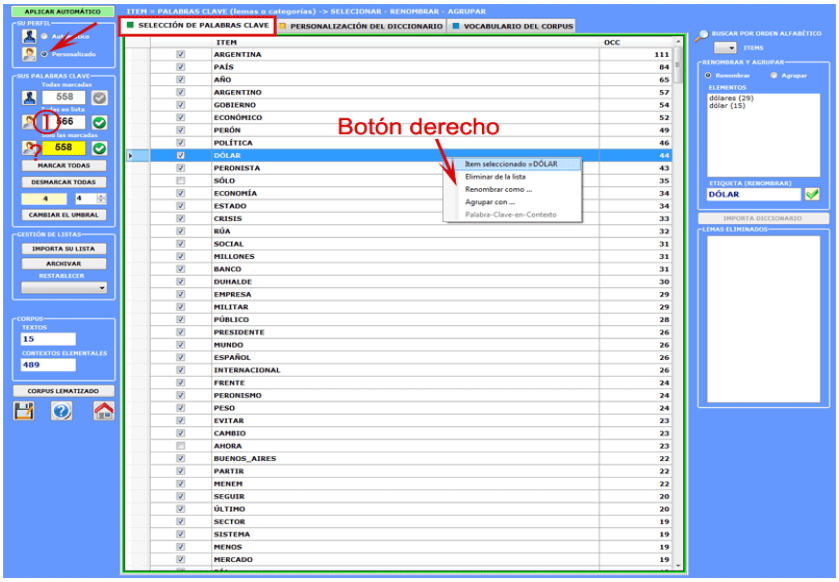


Nota: Cuando el usuario quiere aplicar esquemas de codificación que agrupen diferentes palabras o lemas en unas pocas categorías (de 2 a 50), y sin perder ninguna información lexical, se aconseja utilizar la herramienta **Clasificación basada en los Diccionarios** incluido en el submenú **Análisis Temáticos** (véase abajo).



4 - LA SELECCIÓN DE LAS PALABRAS-CLAVE consiste en la predisposición de una o más listas de unidades lexicales (palabras, lemas o categorías) a utilizar para crear las tablas de datos a analizar.

La opción **configuración automática** pone a disposición listas de **palabras clave** seleccionadas por **T-LAB**; sin embargo, dado que la elección de las unidades de análisis es muy relevante en relación a las sucesivas elaboraciones, se aconseja vivamente el uso de la **configuración personalizada**. De este modo el usuario podrá elegir la modificación de la lista sugerida por **T-LAB** y/o crear listas que correspondan mejor con sus objetivos de investigación.



En la creación de estas listas, son válidos los siguientes criterios:

- verificar la **relevancia** cuantitativa (total de las ocurrencias) y cualitativa (no banalidad del significado) de los distintos términos;
- verificar las **limitaciones** de los instrumentos analíticos que se desean utilizar;
- verificar si el conjunto de los términos es compatible con la propia **estrategia** de investigación (ver punto siguiente: 5).

5 - **EL USO DE LOS INSTRUMENTOS DE ANÁLISIS** está destinado a la producción de outputs (tablas y gráficos) que representan **relaciones significativas** entre las unidades de análisis y que permiten hacer **inferencias**.

Actualmente **T-LAB** incluye quince diversas herramientas de análisis y cada una de ellas tiene su propia lógica; es decir, cada herramienta utiliza algoritmos específicos y produce output específicos.

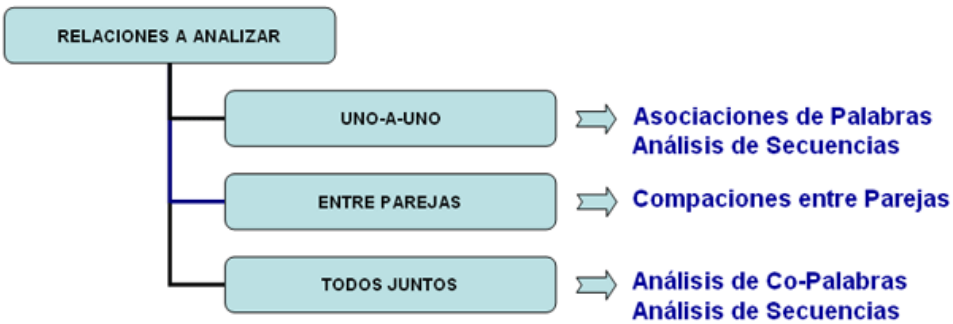
Consecuentemente, dependiendo de la tipología de textos que quiera analizar y de los objetivos que quiera alcanzar, el usuario debe decidir, cada vez que implemente una, qué instrumentos son más apropiados para su estrategia de análisis.

Para este propósito, además de la distinción entre instrumentos para **análisis de co-ocurrencias**, **análisis comparativos** y **análisis temáticos**, puede ser útil tomar en cuenta que algunos de estos nos permiten obtener nuevas **unidades del análisis** que se pueden incluir en otros procesos.

Sin embargo, teniendo en cuenta que el uso de las herramientas **T-LAB** puede ser circular y reversible, podríamos escoger tres puntos de inicio (start points) que corresponden a los tres sub-menús de ANÁLISIS:

A : INSTRUMENTOS PARA ANÁLISIS DE CO-OCCURRENCIAS

Estos instrumentos permiten analizar varios tipos de relaciones entre las palabras clave.

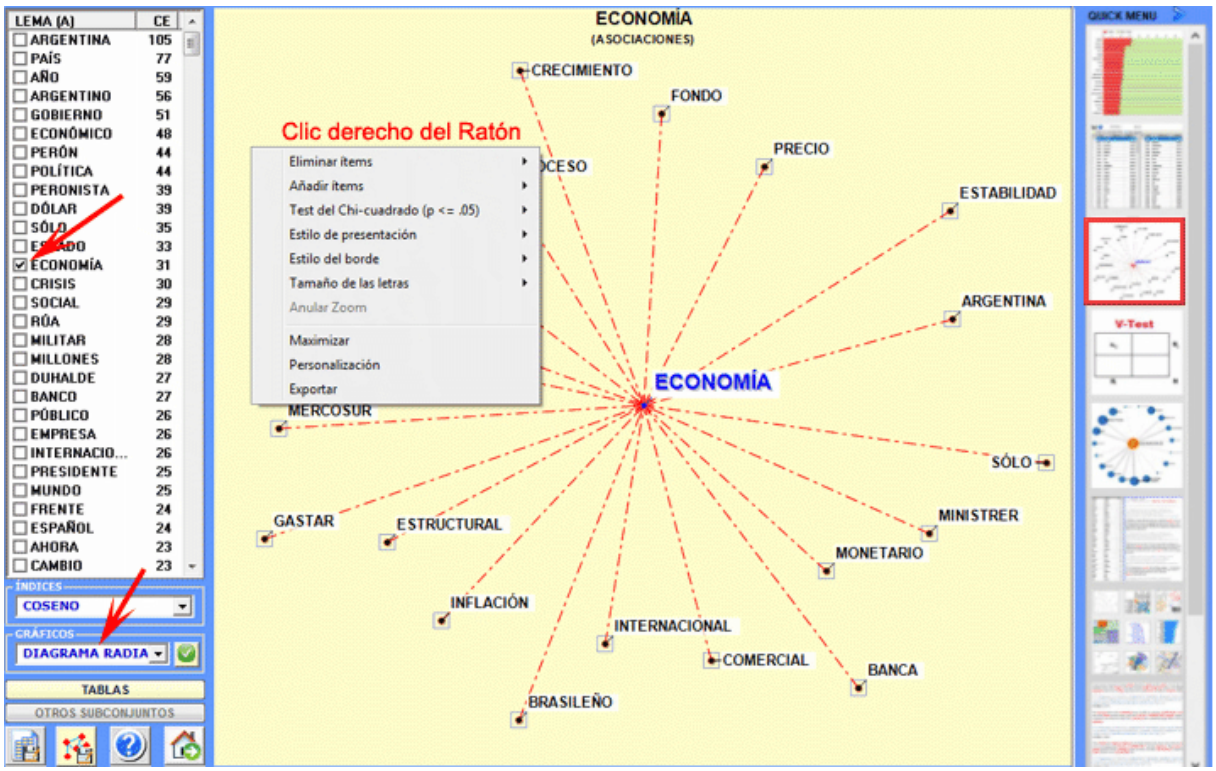


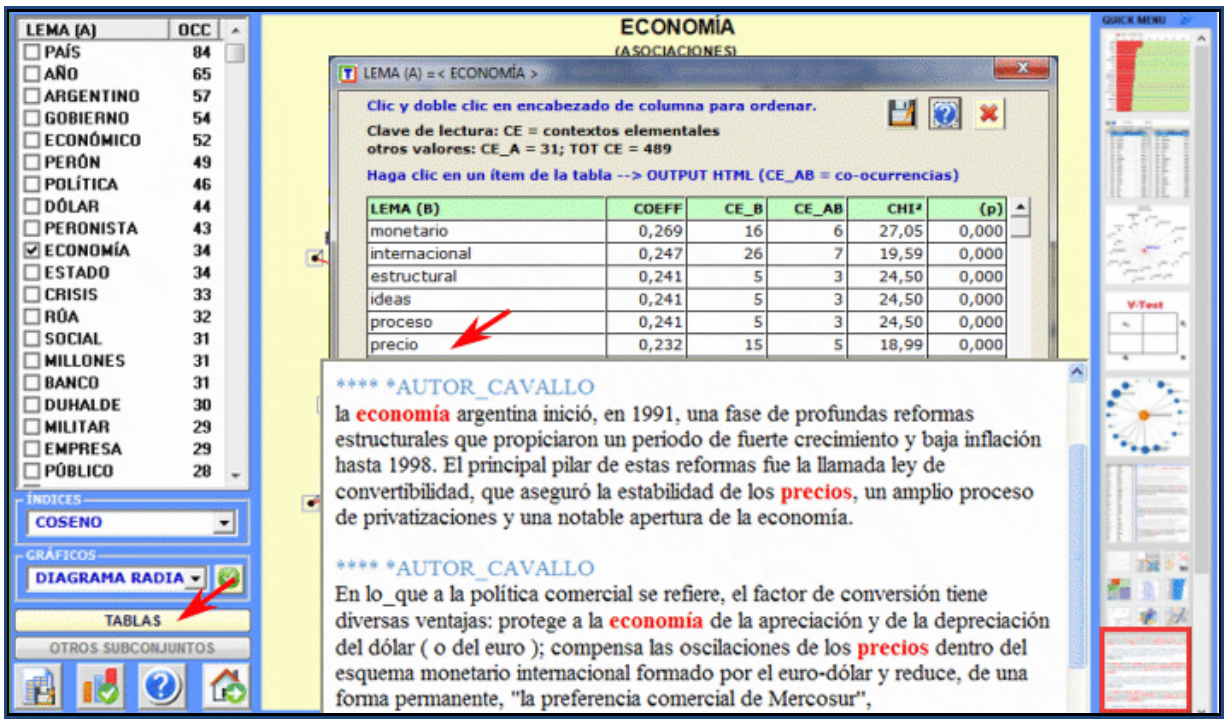
Según los tipos de relaciones a analizar, las funciones **T-LAB** indicadas en este diagrama (casillas coloradas) usan uno o más de los siguientes instrumentos estadísticos: **Índices de Asociación**, **Test del Chi Cuadrado**, **Cluster Analysis, Multidimensional Scaling**, **Principal Component Analysis, t-SNE** y **Cadenas Markovianas.**

Aquí están algunos ejemplos (Nota: para más información sobre la interpretación de los resultados, véanse las secciones correspondientes en la guía/manual):

- Asociaciones de Palabras

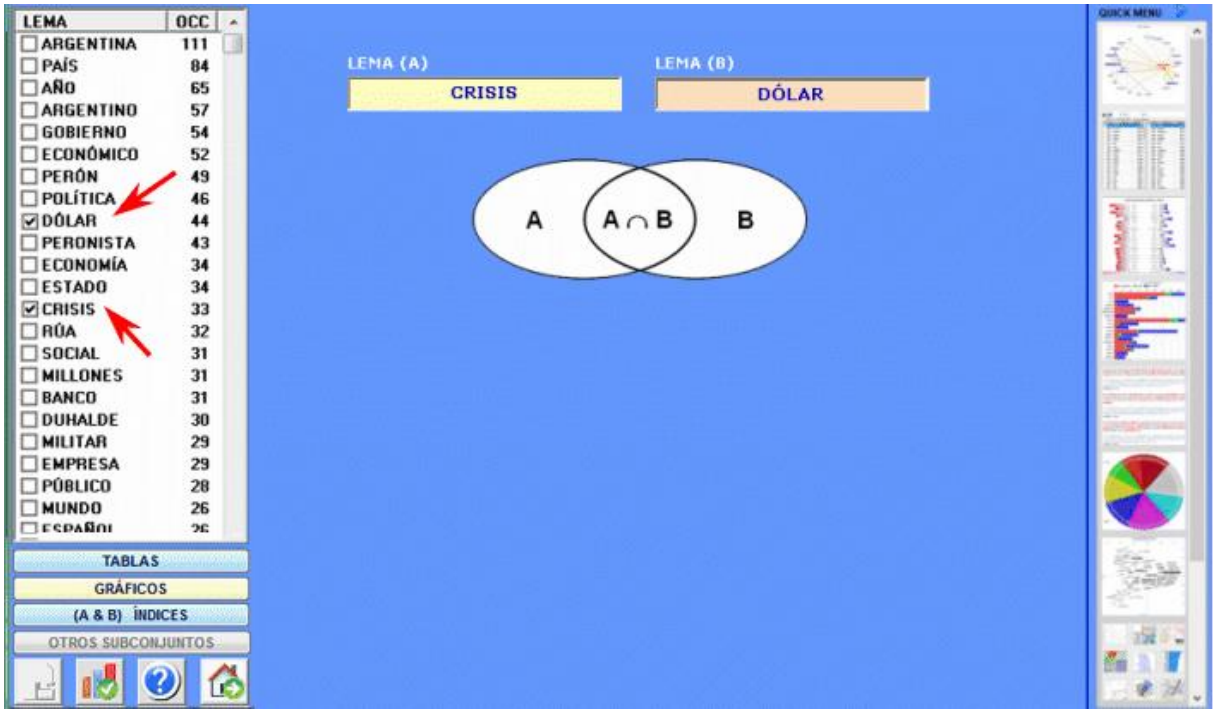
Esta herramienta de **T-LAB** nos permite comprobar como las relaciones de co-ocurrencia determinan el significado local de palabras seleccionadas.

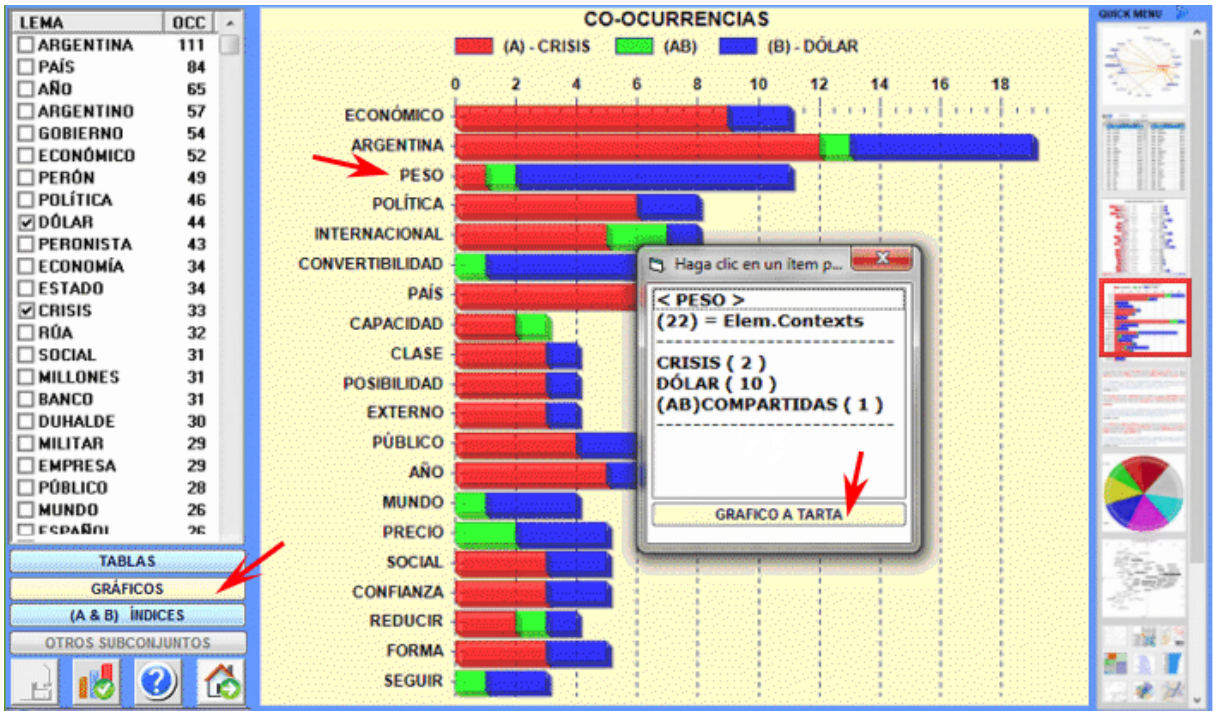




- Comparaciones entre Parejas

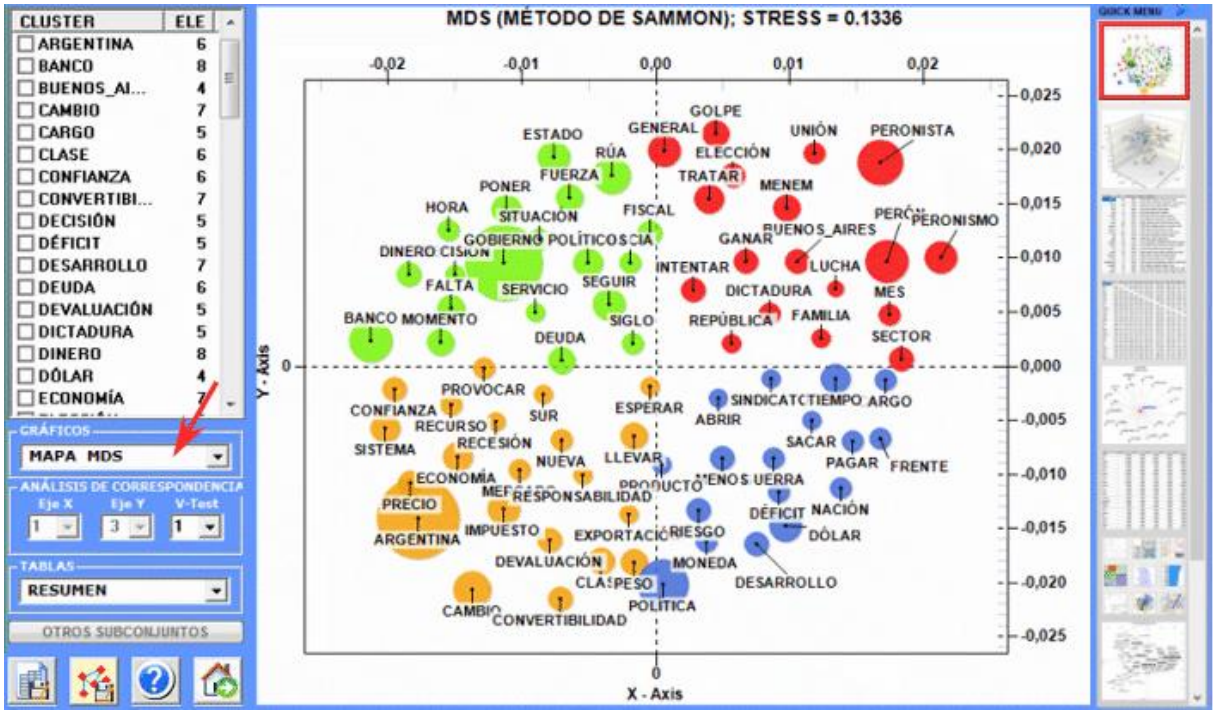
Esta herramienta de **T-LAB** nos permite comparar los conjuntos de contextos elementales (es decir contextos de co-ocurrencia) en los cuales los miembros de una pareja de palabras-clave están presentes.

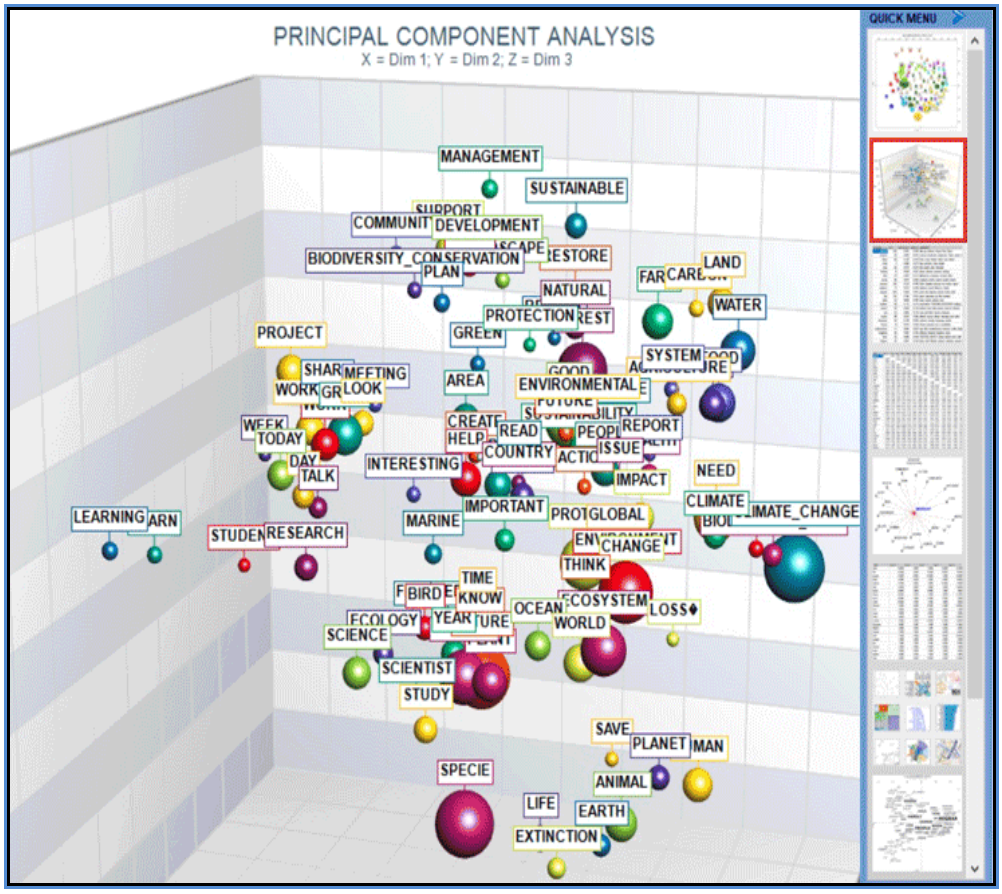




- Análisis de Co-Palabras

Esta herramienta de **T-LAB** nos permite trazar mapas de co-ocurrencias entre conjuntos de palabras clave.



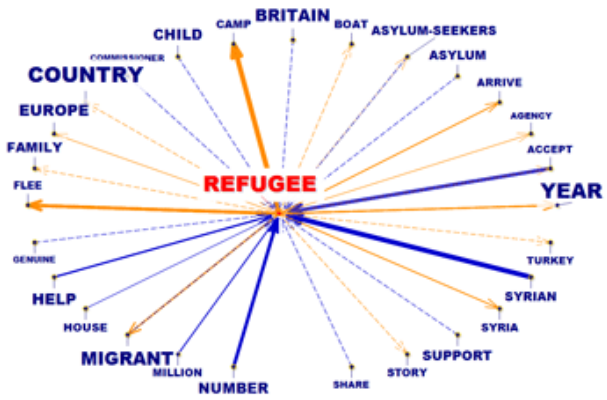


- Análisis de Secuencias y Análisis de Redes

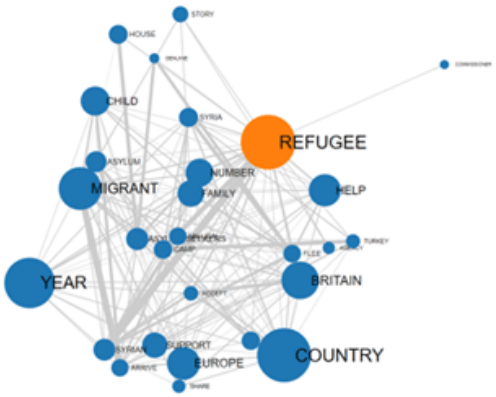
Esta herramienta de **T-LAB** tiene en cuenta las posiciones de las diferentes unidades lexicales dentro de las frases, permitiendonos así analizar y representar cualquier texto como si fuera una **red** de relaciones.

Una vez implementado este tipo de analisis, el usuario podrá explorar las relaciones que existen entre nodos de la red (esto es, las palabras clave) a diferentes niveles: a) dentro de las relaciones del tipo uno-a-uno; b) dentro de un "ego network"; c) dentro de las comunidades a las que pertenecen; d) dentro de la red formada por el texto analizado.

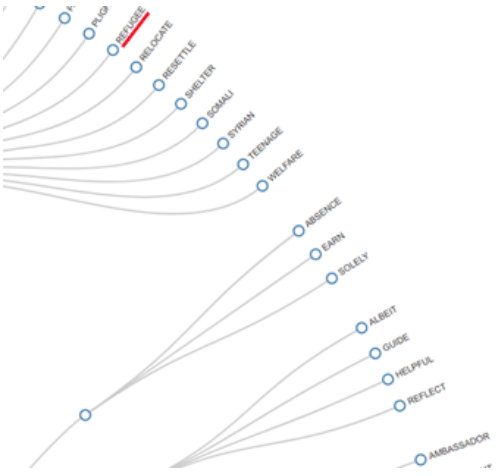
RELACIONES DEL TIPO UNO-AD-UNO



EGO-NETWORK

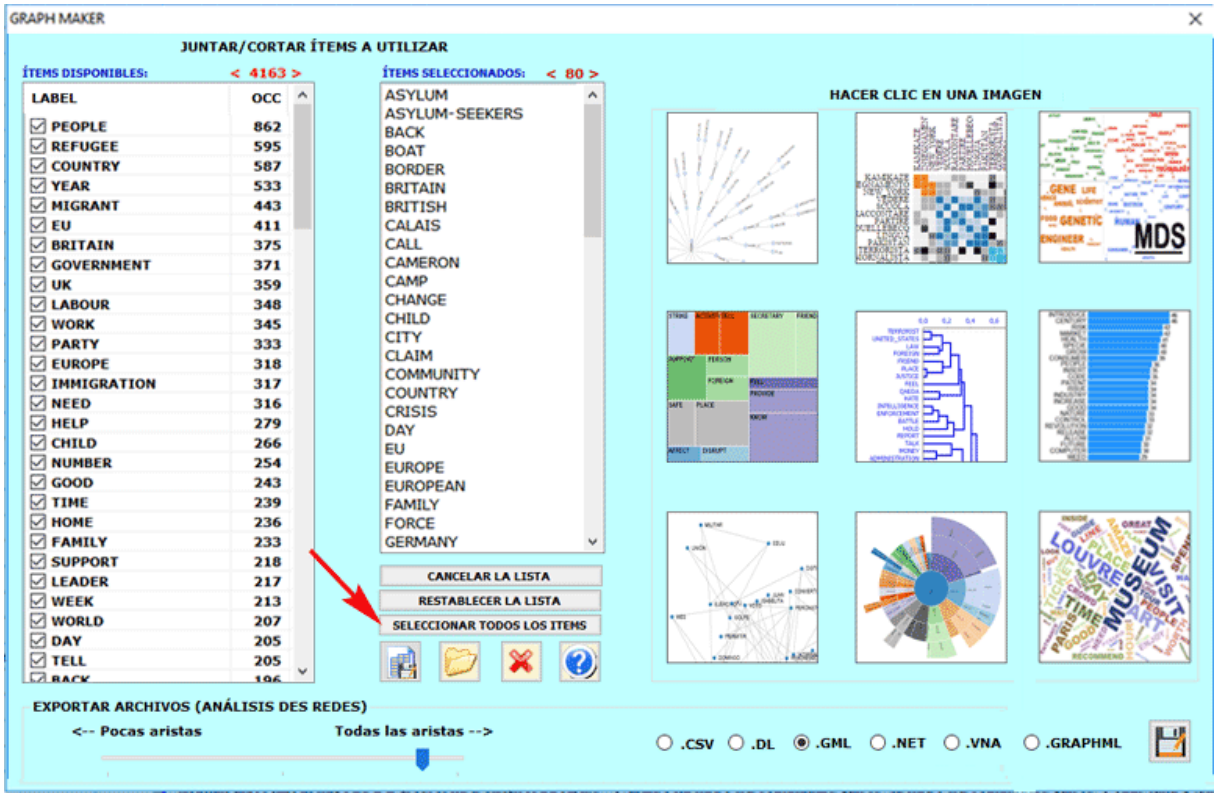


COMUNIDAD



RED ENTERA

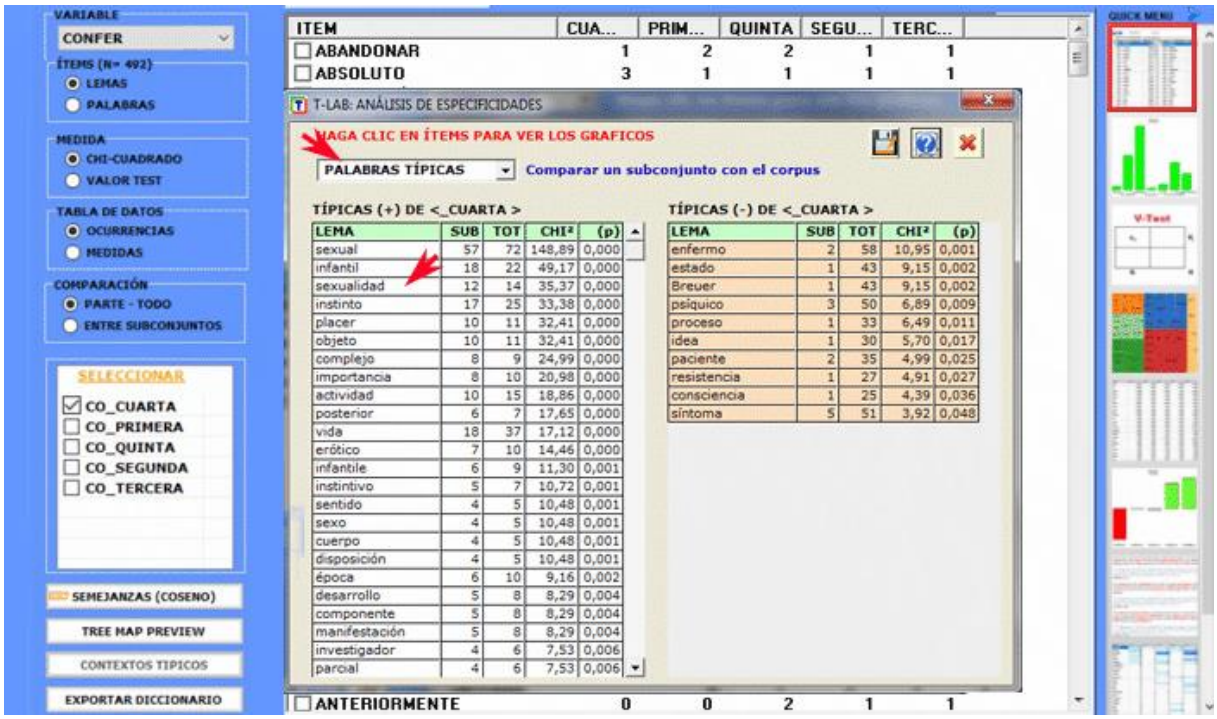
Además, cliqueando en la opción GRAPH MAKER el usuario podrá crear, en todo momento y a partir de los listados personalizados de palabras clave, diferentes tipos de graficos (véase abajo).

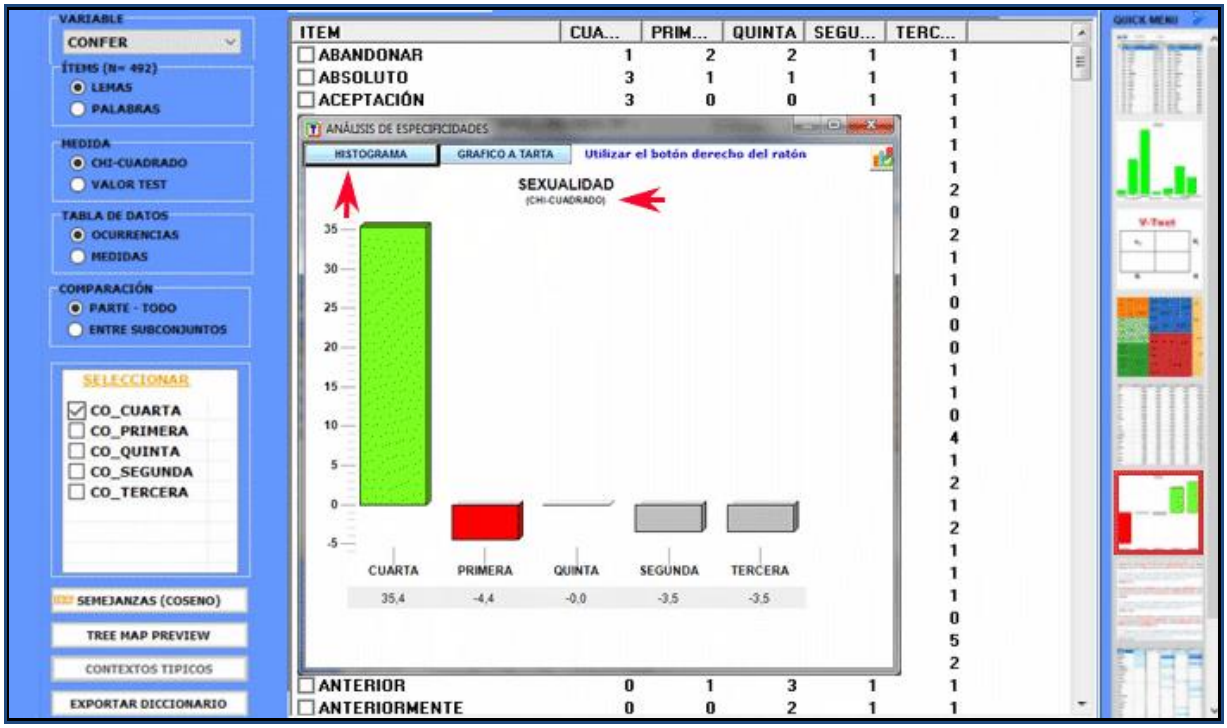


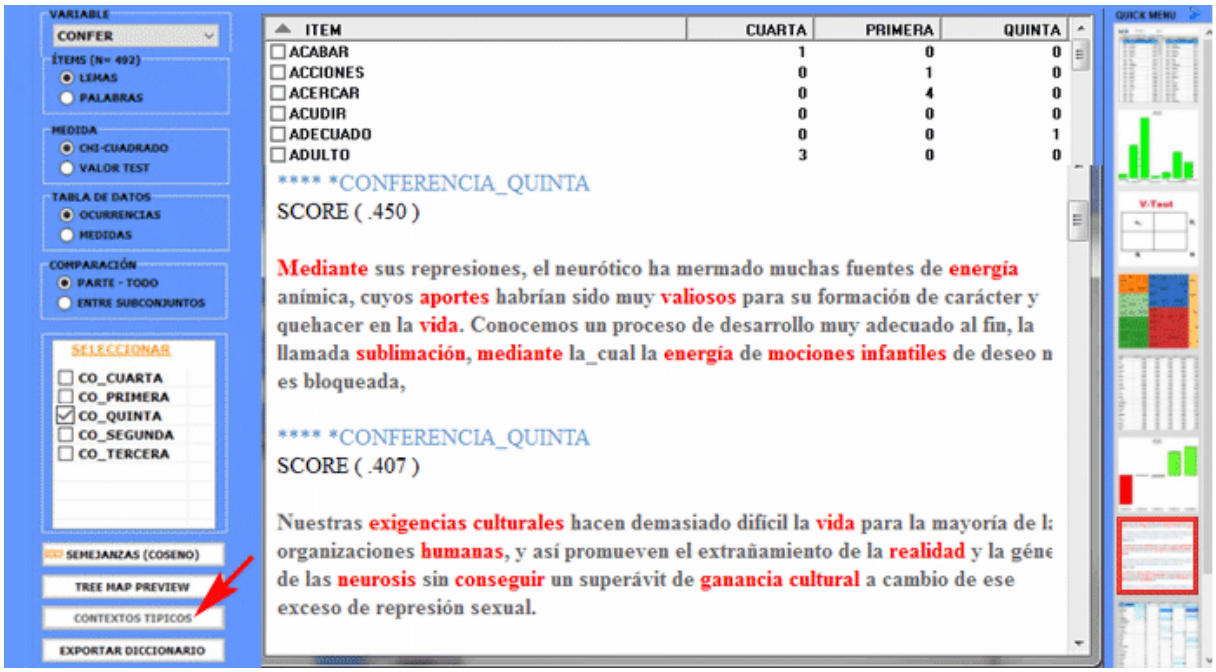
B : INSTRUMENTOS PARA ANÁLISIS COMPARATIVOS

Estos instrumentos permiten analizar varios tipos de relaciones entre las unidades de contexto.

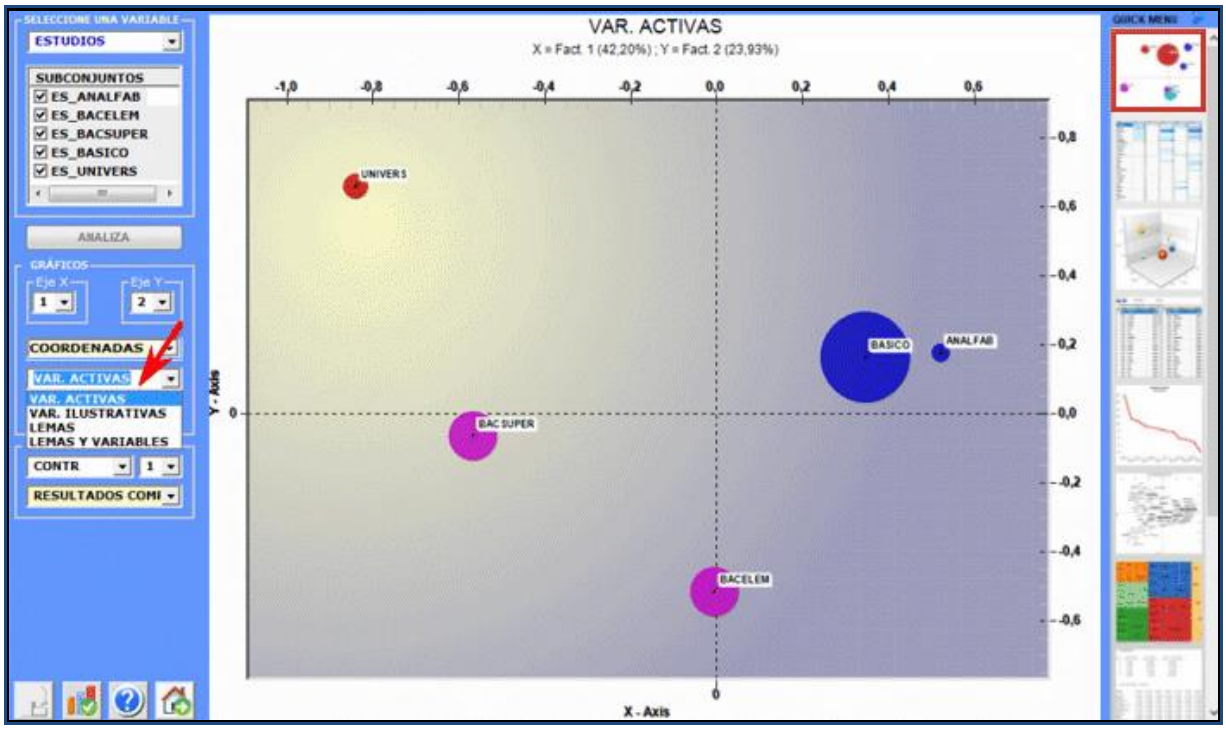
El **Análisis de las Especificidades** permite verificar cuáles palabras son "típicas" o "exclusivas" de cada subconjunto del corpus. Además, nos permite extraer los contextos típicos, es decir, los contextos elementales característicos, de cada uno de los subconjuntos analizados (p. ej.: las frases 'típicas' utilizadas por los líderes políticos).

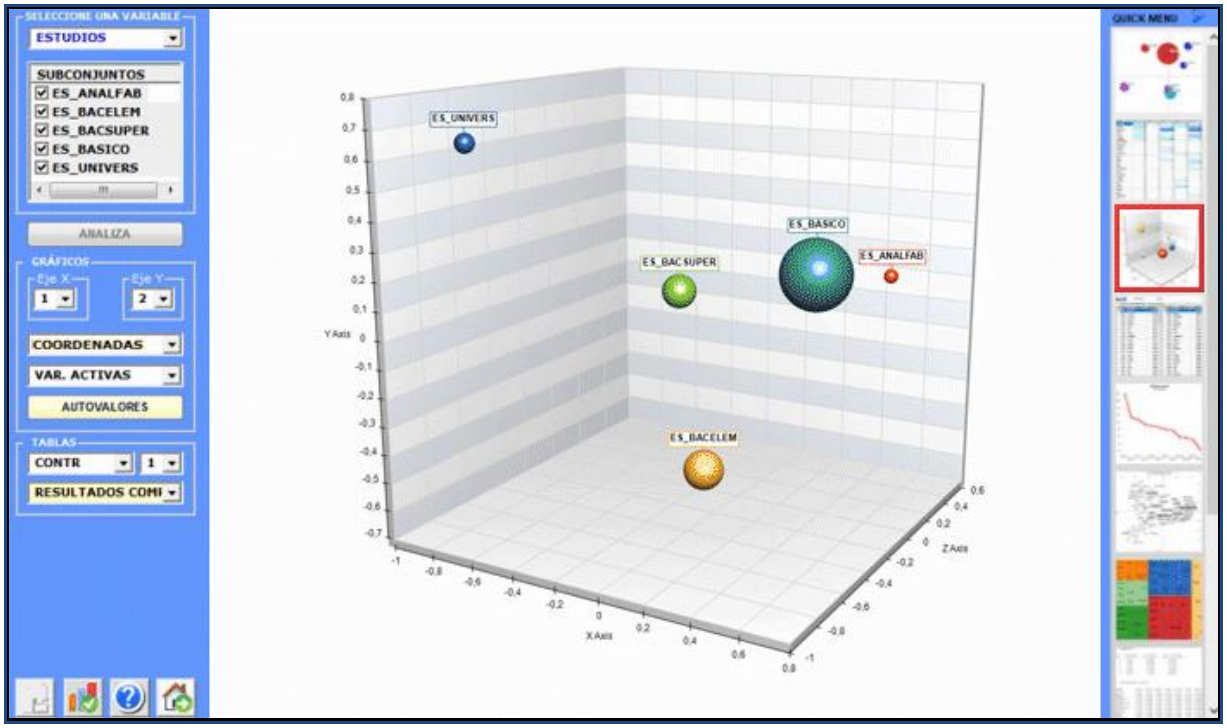

VARIABLE
CONFER
ÍTEMS (N= 492)
LEMAS
PALABRAS
MEDIDA
CHI-CUADRADO
VALOR TEST
TABLA DE DATOS
OCURRENCIAS
MEDIDAS
COMPARACIÓN
PARTE - TODO
ENTRE SUBCONJUNTOS
SELECCIONAR
CO_CUARTA
CO_PRIMERA
CO_QUINTA
CO_SEGUNDA
CO_TERCERA
SEMEJANZAS (COSENO)
TREE MAP PREVIEW
CONTEXTOS TIPICOS
EXPORTAR DICCIONARIO
ITEM
CUA...
PRIM...
QUINTA
SEGU...
TERC...
ABANDONAR 1 2 2 1 1
ABSOLUTO 3 1 1 1 1
T-LAB: ANÁLISIS DE ESPECIFICIDADES
HAGA CLIC EN ÍTEMS PARA VER LOS GRAFICOS
PALABRAS TÍPICAS
Comparar un subconjunto con el corpus
TÍPICAS (+) DE <_CUARTA >
LEMA SUB TOT CHI² (p)
sexual 57 72 148,89 0,000
infantil 18 22 49,17 0,000
sexualidad 12 14 35,37 0,000
instinto 17 25 33,38 0,000
placer 10 11 32,41 0,000
objeto 10 11 32,41 0,000
complejo 8 9 24,99 0,000
importancia 8 10 20,98 0,000
actividad 10 15 18,86 0,000
posterior 6 7 17,65 0,000
vida 18 37 17,12 0,000
erótico 7 10 14,46 0,000
infantile 6 9 11,30 0,001
instintivo 5 7 10,72 0,001
sentido 4 5 10,48 0,001
sexo 4 5 10,48 0,001
cuerpo 4 5 10,48 0,001
disposición 4 5 10,48 0,001
época 6 10 9,16 0,002
desarrollo 5 8 8,29 0,004
componente 5 8 8,29 0,004
manifestación 5 8 8,29 0,004
investigador 4 6 7,53 0,006
parcial 4 6 7,53 0,006
TÍPICAS (-) DE <_CUARTA >
LEMA SUB TOT CHI² (p)
enfermo 2 58 10,95 0,001
estado 1 43 9,15 0,002
Breuer 1 43 9,15 0,002
psíquico 3 50 6,89 0,009
proceso 1 33 6,49 0,011
idea 1 30 5,70 0,017
paciente 2 35 4,99 0,025
resistencia 1 27 4,91 0,027
consciencia 1 25 4,39 0,036
síntoma 5 51 3,92 0,048
ANTERIORMENTE 0 0 2 1 1


VARIABLE
CONFER
ÍTEMS (N= 492)
LEMAS
PALABRAS
MEDIDA
CHI-CUADRADO
VALOR TEST
TABLA DE DATOS
OCURRENCIAS
MEDIDAS
COMPARACIÓN
PARTE - TODO
ENTRE SUBCONJUNTOS
SELECCIONAR
CO_CUARTA
CO_PRIMERA
CO_QUINTA
CO_SEGUNDA
CO_TERCERA
SEMEJANZAS (COSENO)
TREE MAP PREVIEW
CONTEXTOS TIPICOS
EXPORTAR DICCIONARIO
ITEM
CUA...
PRIM...
QUINTA
SEGU...
TERC...
ABANDONAR 1 2 2 1 1
ABSOLUTO 3 1 1 1 1
ACEPTACIÓN 3 0 0 1 1
ANÁLISIS DE ESPECIFICIDADES
HISTOGRAMA
GRAFICO A TARTA
Utilizar el botón derecho del ratón
SEXUALIDAD
(CHI-CUADRADO)
35
30
25
20
15
10
5
0
-5
CUARTA
PRIMERA
QUINTA
SEGUNDA
TERCERA
35,4
-4,4
-0,0
-3,5
-3,5
ANTERIOR 0 1 3 1 1
ANTERIORMENTE 0 0 2 1 1

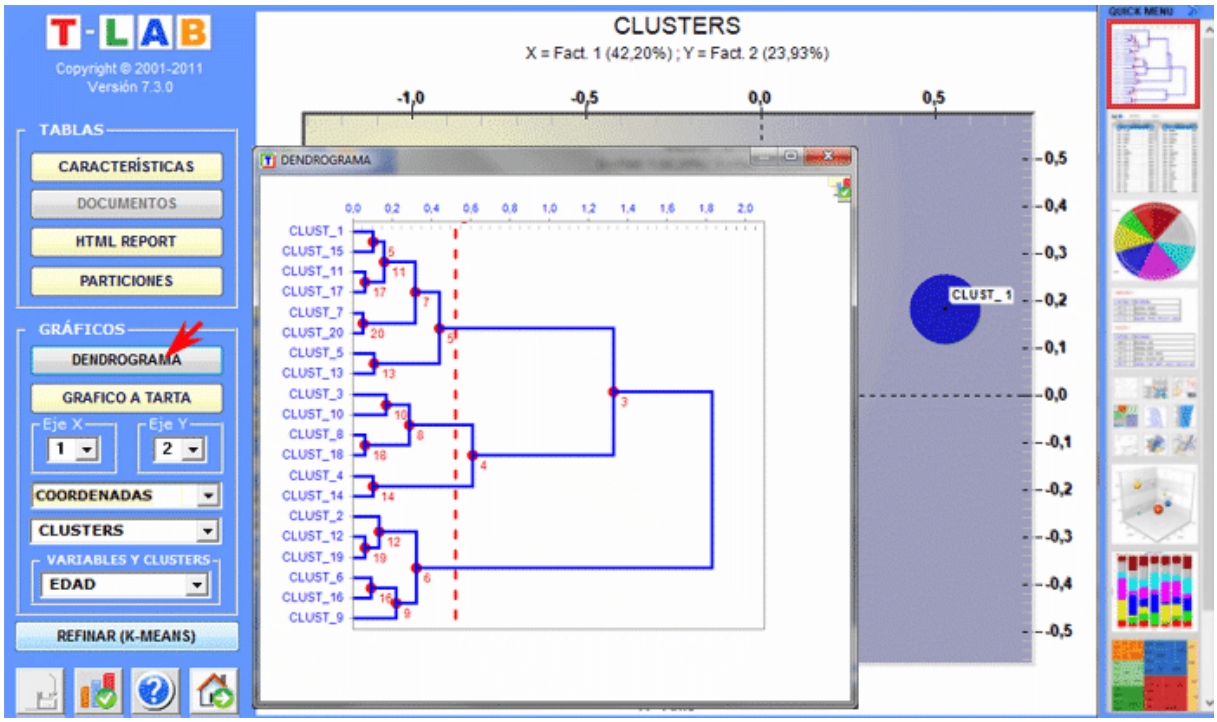


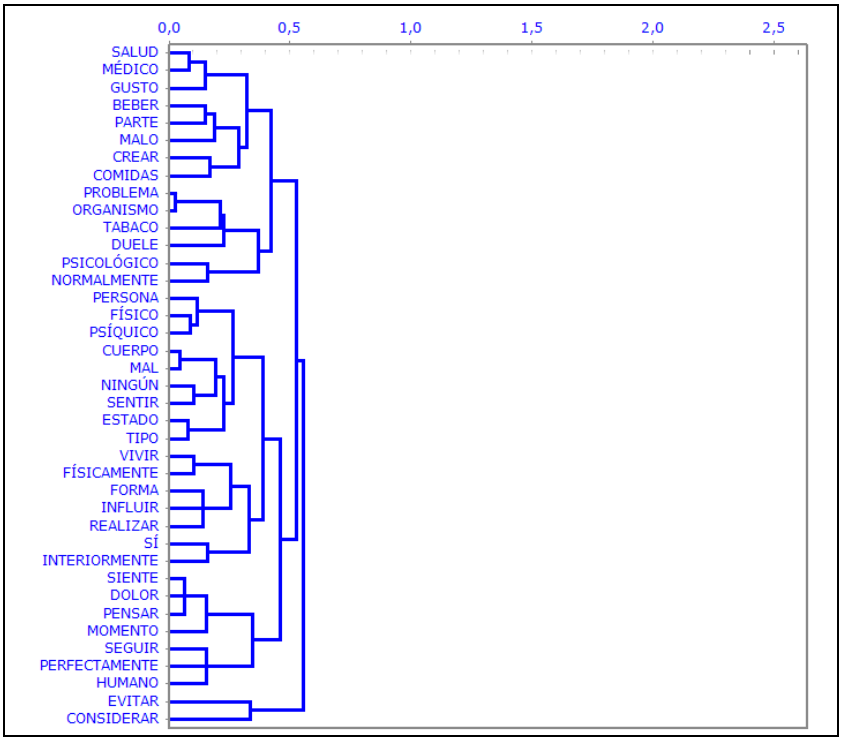
El **Análisis de Correspondencias** permite explorar varios tipos de relaciones (semejanzas y diferencias) entre grupos de unidades de contexto.





El **Cluster Analysis** permite encontrar grupos de unidades de texto que presentan dos características complementarias: máxima homogeneidad interna y máxima heterogeneidad entre cada clúster y todos los demás. Se puede implementar recurriendo a múltiples técnicas y requiere, previamente, un análisis de las correspondencias o un SVD.





C : INSTRUMENTOS PARA ANÁLISIS TEMÁTICOS

Estos instrumentos permiten individuar, examinar y trazar el mapa de los "temas" que emergen de los textos analizados.

Puesto que "**Tema**" es una palabra polisémica, cuando se usa software para análisis temático es útil hacer referencia a algunas definiciones operativas.

En los instrumentos **T-LAB**, "tema" es una etiqueta usada para indicar cuatro diferentes entidades:

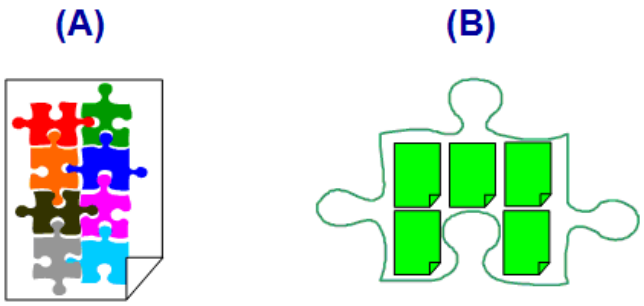
1- un **clúster temático de unidades de contexto** caracterizados por los mismos modelos de palabras clave (ver los instrumentos **Análisis Temático de Contextos Elementales** y **Clasificación Temática de Documentos**);

2- un **grupo temático de palabras-clave** clasificadas en términos de pertenencia a una misma categoría (véase la herramienta **Clasificación Basada en Diccionarios**);

3- un **componente de un modelo probabilista** que representa cada unidad de contexto (contextos elementales o documentos) generado de una mezcla de "temas" (ver los instrumentos **Modelización de los Temas Emergentes** y **Textos y Discursos como Sistemas Dinámicos**);

4- una **específica palabra clave** ("temática") usada para extraer un conjunto de contextos elementales. Esta palabra está asociada con un específico conjunto de palabras preseleccionadas por el usuario (ver el instrumento **Contextos Clave de Palabras Temáticas**).

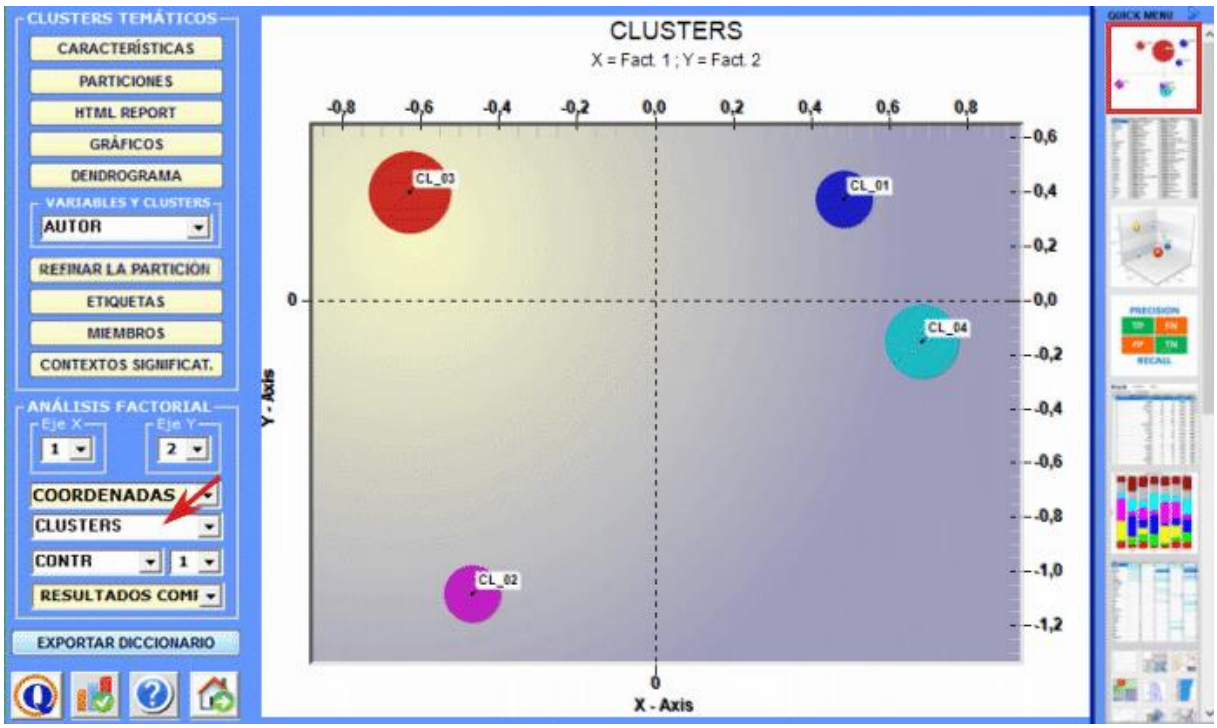
Por ejemplo, según el tipo de herramienta que estemos utilizando, un documento concreto puede ser analizado bien en términos de co-presencia de varios 'temas' en un único documento (véase 'A' abajo) o bien como parte de un conjunto de documentos que conciernen el mismo 'tema' (véase 'B' abajo). De hecho, en el caso 'A', cada tema puede corresponder a una palabra o frase mientras que, en el caso 'B', un tema puede representar una etiqueta asignada a un conjunto de documentos que presentan los mismos patrones de palabras-clave.

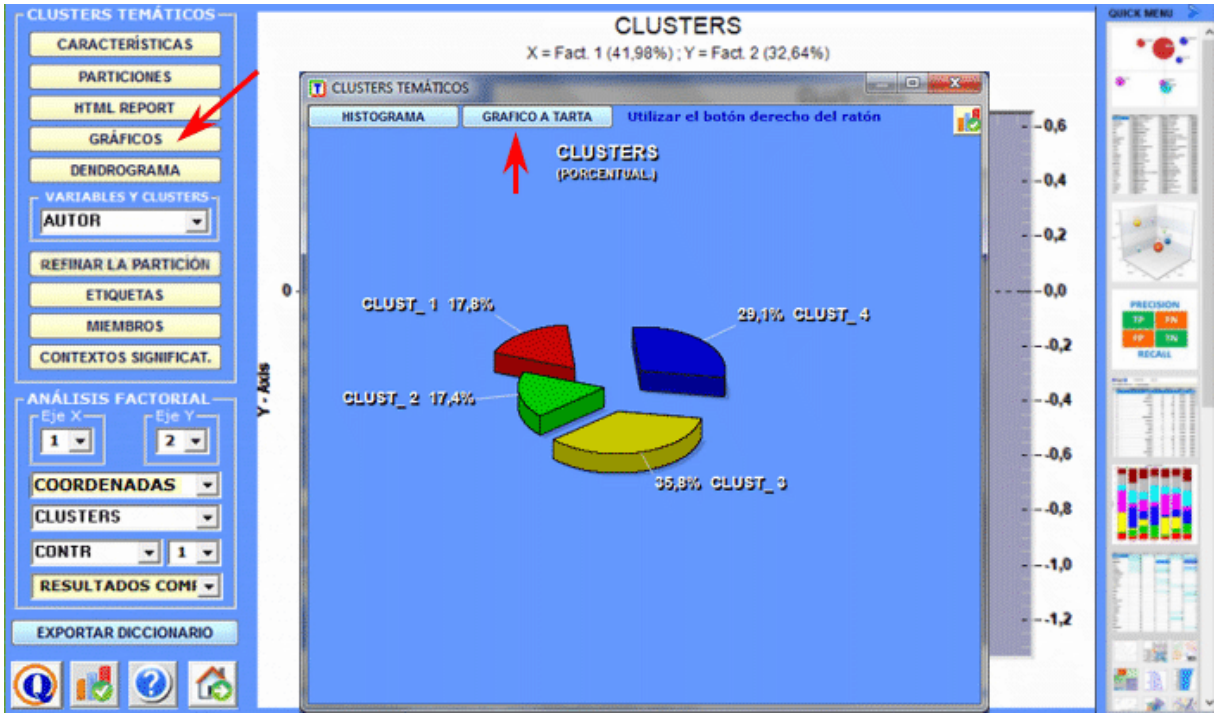


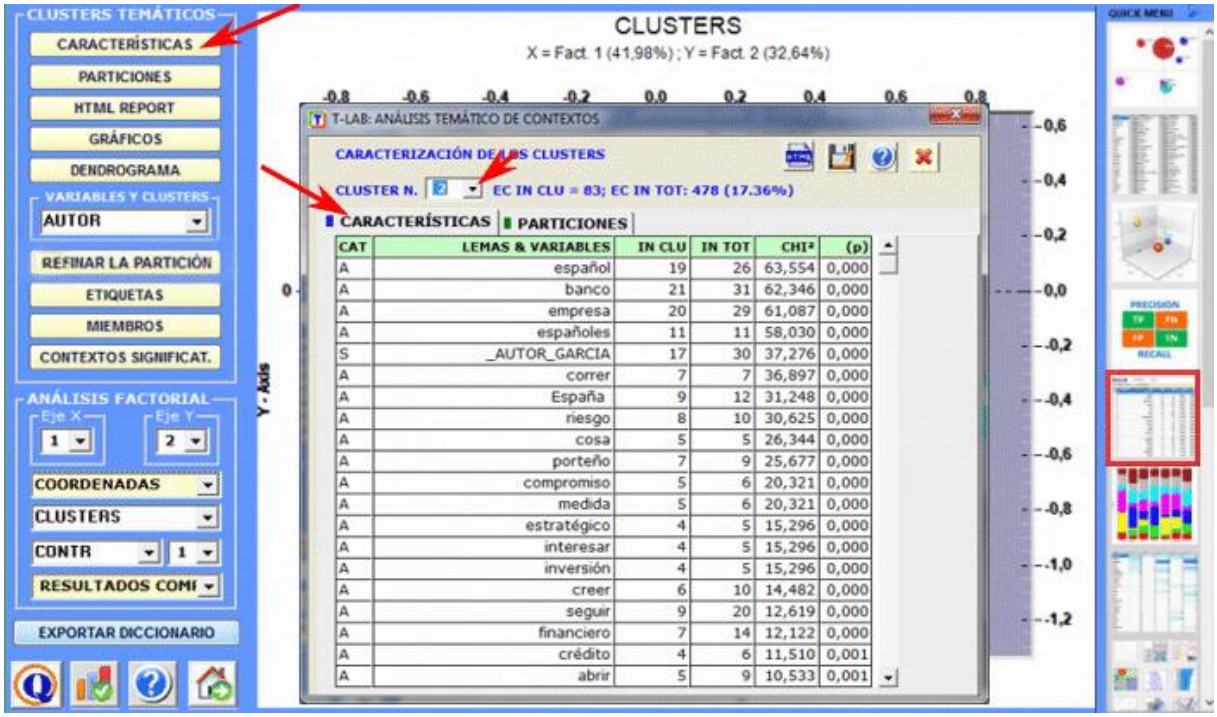
Más en detalle, **T-LAB** 'extrae' los temas utilizando las siguientes metodologías:

1 - tanto el **Análisis Temático de Contextos Elementales** como la **Clasificación Temática de Documentos** funcionan en la siguiente manera:

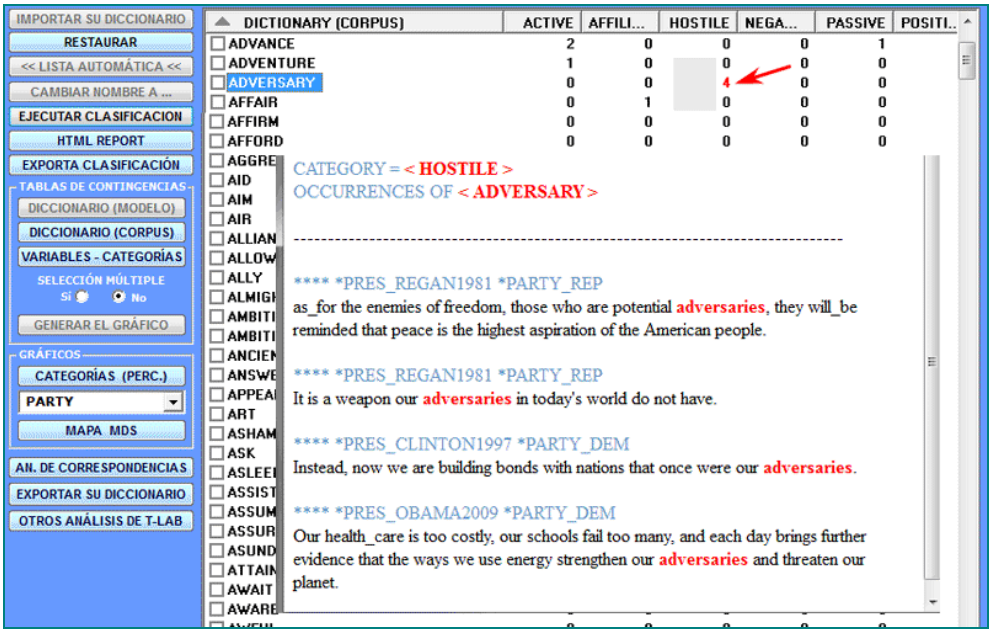
a- realizan un **análisis de co-ocurrencias** para obtener los clusters temáticos de unidades de contexto;
b- realizan un **análisis comparativo** para confrontar los perfiles de los distintos clusters;
c- generan varios tipos de gráficos y tablas (ver a continuación);
d- permiten archivar las **nuevas variables** obtenidas (clusters temáticos) y utilizarlas en análisis posteriores.

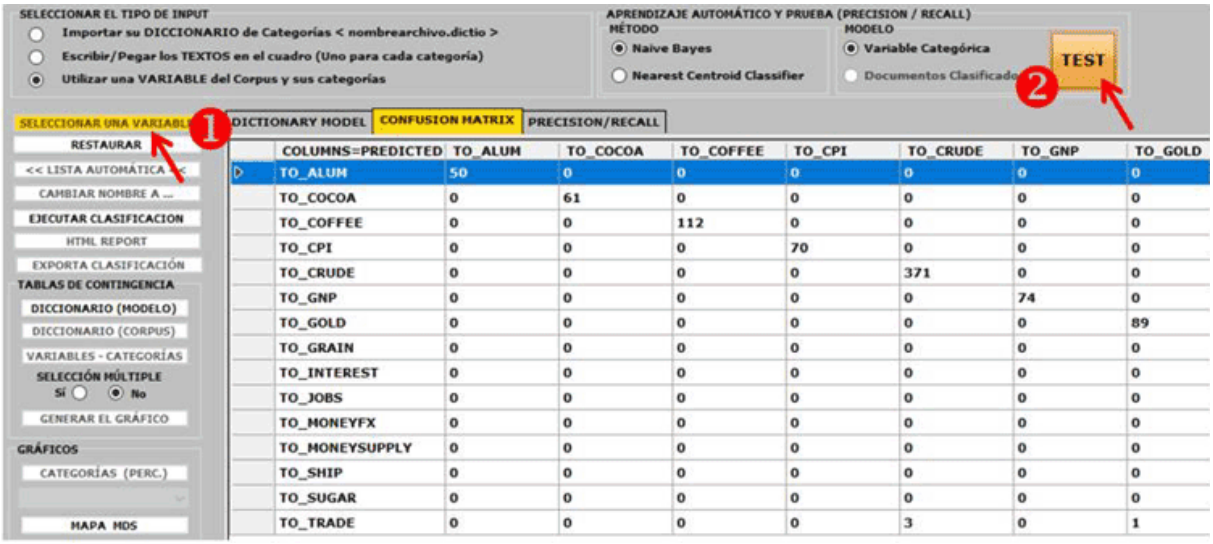



CLUSTERS TEMÁTICOS
CARACTERÍSTICAS
PARTICIONES
HTML REPORT
GRÁFICOS
DENDROGRAMA
VARIABLES Y CLUSTERS
AUTOR
REFINAR LA PARTICIÓN
ETIQUETAS
MIEMBROS
CONTEXTOS SIGNIFICAT.
ANÁLISIS FACTORIAL
Eje X
Eje Y
COORDENADAS
CLUSTERS
CONTR
RESULTADOS COMI
EXPORTAR DICCIONARIO
CLUSTERS
X = Fact. 1 (41,98%) ; Y = Fact. 2 (32,64%)
CLUSTERS TEMÁTICOS
HISTOGRAMA
GRAFICO A TARTA
Utilizar el botón derecho del ratón
CLUSTERS
(PORCENTUAL)
CLUST_1 17,8%
29,1% CLUST_4
CLUST_2 17,4%
35,8% CLUST_3
Y - Axis
QUICK MENU

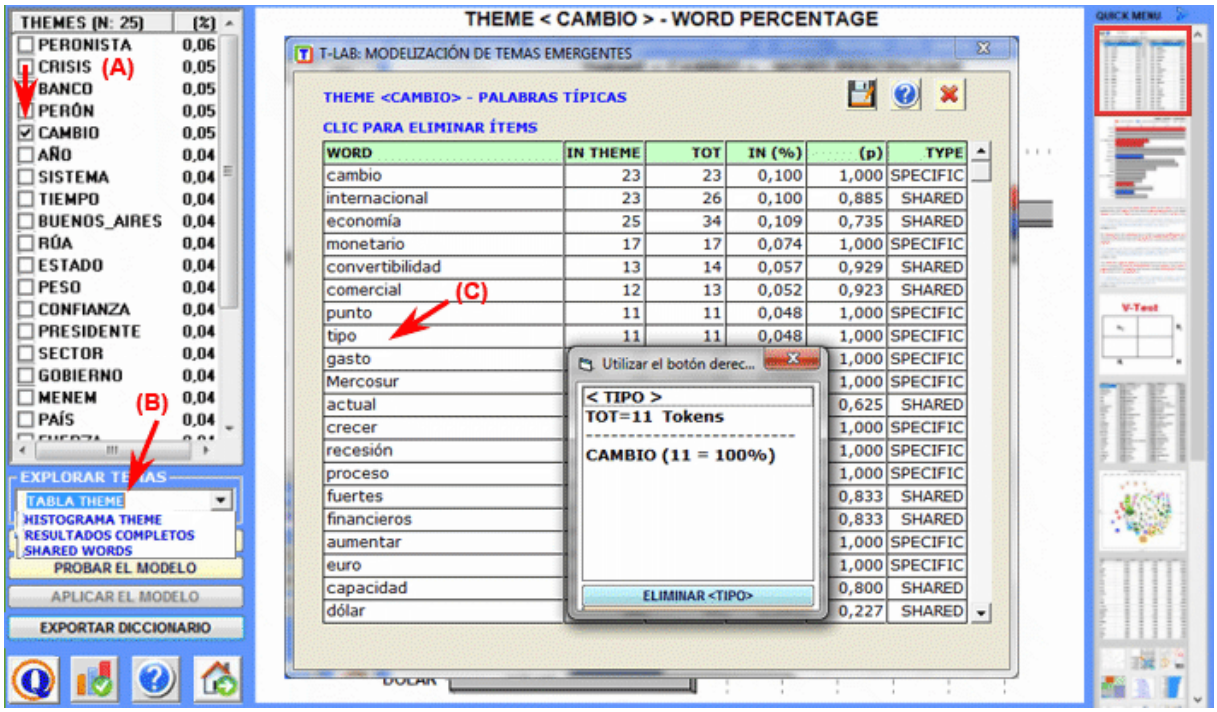

CLUSTERS TEMÁTICOS
CARACTERÍSTICAS
PARTICIONES
HTML REPORT
GRÁFICOS
DENDROGRAMA
VARIABLES Y CLUSTERS
AUTOR
REFINAR LA PARTICIÓN
ETIQUETAS
MIEMBROS
CONTEXTOS SIGNIFICAT.
ANÁLISIS FACTORIAL
EXPORTAR DICCIONARIO
CLUSTERS
X = Fact. 1 (41,98%) ; Y = Fact. 2 (32,64%)
T-LAB: ANÁLISIS TEMÁTICO DE CONTEXTOS
CARACTERIZACIÓN DE LOS CLUSTERS
CLUSTER N. 2
EC IN CLU = 83; EC IN TOT: 478 (17.36%)
CARACTERÍSTICAS
PARTICIONES
CAT | LEMAS & VARIABLES | IN CLU | IN TOT | CHI² | (p)
A | español | 19 | 26 | 63,554 | 0,000
A | banco | 21 | 31 | 62,346 | 0,000
A | empresa | 20 | 29 | 61,087 | 0,000
A | españoles | 11 | 11 | 58,030 | 0,000
S | _AUTOR_GARCIA | 17 | 30 | 37,276 | 0,000
A | correr | 7 | 7 | 36,897 | 0,000
A | España | 9 | 12 | 31,248 | 0,000
A | riesgo | 8 | 10 | 30,625 | 0,000
A | cosa | 5 | 5 | 26,344 | 0,000
A | porteño | 7 | 9 | 25,677 | 0,000
A | compromiso | 5 | 6 | 20,321 | 0,000
A | medida | 5 | 6 | 20,321 | 0,000
A | estratégico | 4 | 5 | 15,296 | 0,000
A | interesar | 4 | 5 | 15,296 | 0,000
A | inversión | 4 | 5 | 15,296 | 0,000
A | creer | 6 | 10 | 14,482 | 0,000
A | seguir | 9 | 20 | 12,619 | 0,000
A | financiero | 7 | 14 | 12,122 | 0,000
A | crédito | 4 | 6 | 11,510 | 0,001
A | abrir | 5 | 9 | 10,533 | 0,001
Y - Axis
QUICK MENU

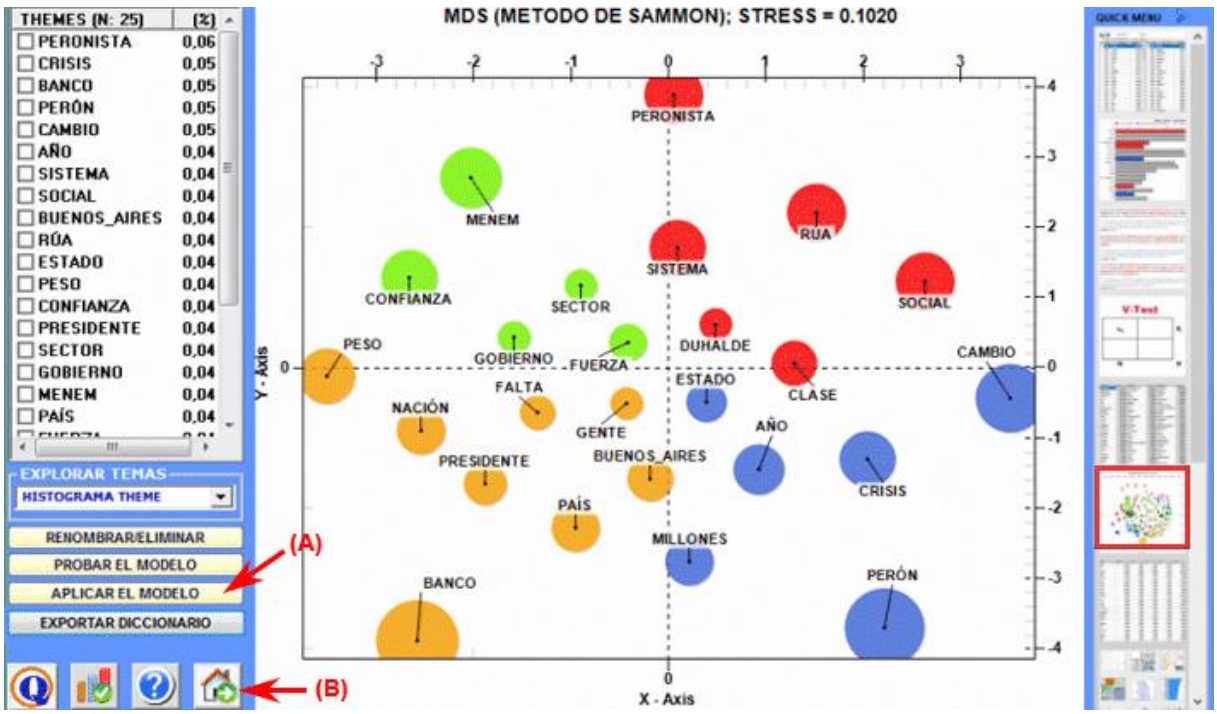
2 - a través de la herramienta **Clasificación Basada en Diccionarios**, podremos fácilmente construir/testar/aplicar modelos (p. ej.: Diccionarios de categorías) tanto para el análisis de contenido clásico como para el sentiment analysis. De hecho, esta herramienta nos permite implementar una clasificación automática de tipo top-down de las unidades lexicales (es decir, palabras y lemas) y también de las unidades de contexto (es decir, frases, párrafos y pequeños documentos).



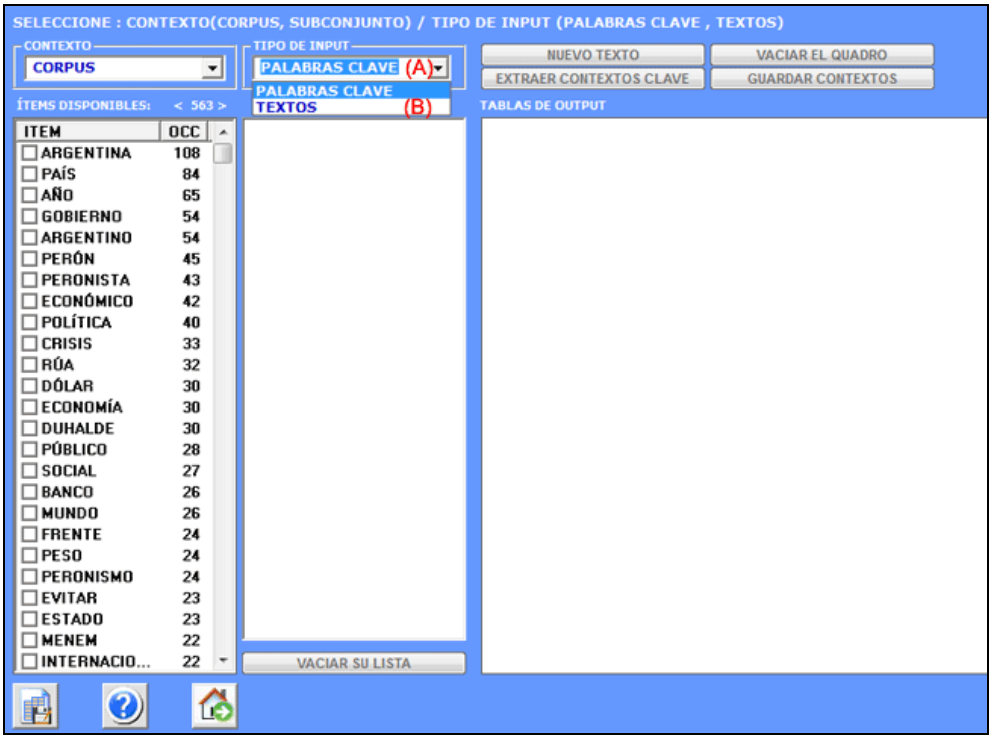


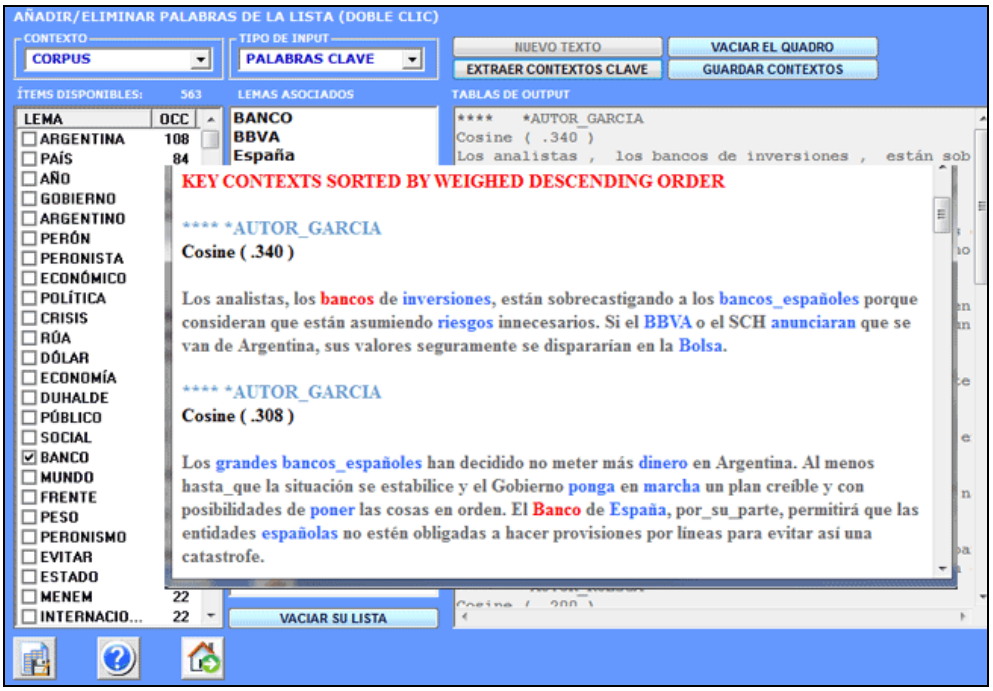
3 - mediante la herramienta **Modelización de los Temas Emergentes** (véase abajo), los componentes de la 'mixtura' temática pueden ser descritos a través de su vocabulario característico y pueden ser utilizados para construir tablas para el análisis cualitativo y/o para la clasificación automática de las unidades de contexto (es decir, contextos elementales o documentos).





4 - La herramienta **Contextos Clave de Palabras Temáticas** se puede utilizar para alcanzar dos objetivos: (a) extraer listados de unidades de contexto (es decir, contextos elementales) que permitan profundizar el valor temático de palabras-clave especificas; (b) extraer grupos de unidades de contexto que resulten ser similares a algún texto de 'ejemplo' escogido por el usuario.

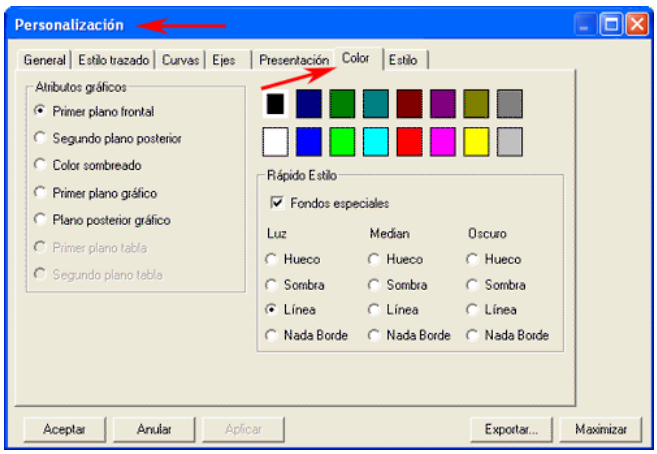


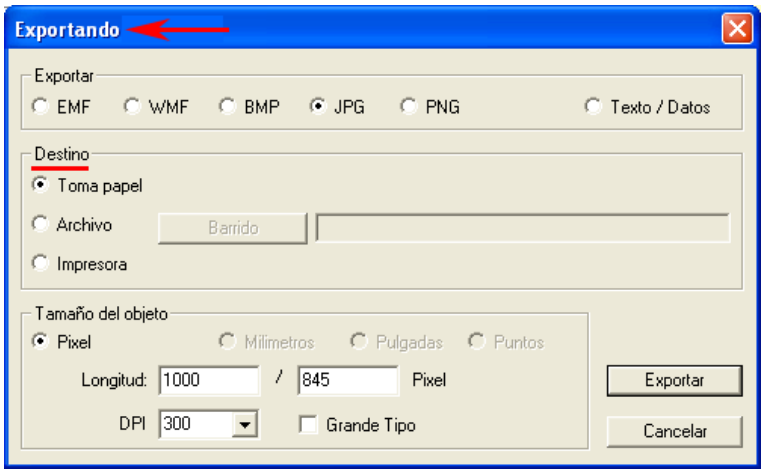


6 - LA INTERPRETACIÓN DE LOS OUTPUT consiste en la consulta de las tablas y de los gráficos producidos por **T-LAB**, en la eventual personalización de su formato y en el hacer inferencias sobre el significado de las relaciones en los mismos representados.

En el caso de las **tablas**, según los casos, **T-LAB** permite exportarlas en filas con las siguientes extensiones: **.DAT**, **.TXT**, **.CSV**, **.XLXS**, **.HTML**. Esto significa que, utilizando cualquier editor de textos y/o de cualquier aplicativo de la suite Microsoft Office, el usuario puede, fácilmente, importarlos y reelaborarlos.

Todos los **gráficos** y **tablas** pueden ser maximizados (hacer clic con el botón izquierdo y arrastre), personalizados y exportados en diferentes formatos (hacer clic con el botón derecho del ratón para ver los pop up menús).



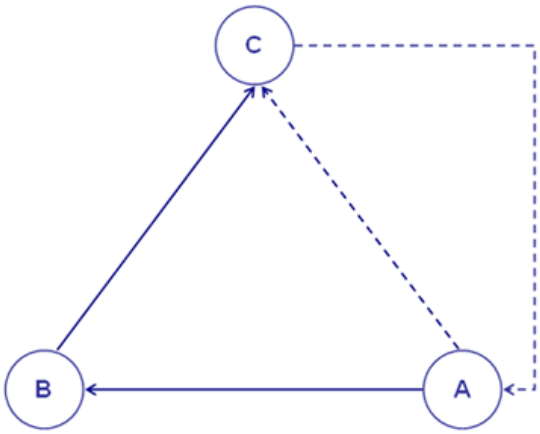


En un paper citado en **Bibliografía** y disponible en el sitio https://www.tlab.it (Lancia F.: 2007) se mencionan algunos criterios generales para la interpretación de los outputs **T-LAB**. En el mismo se propone la hipótesis que los output de las elaboraciones estadísticas (tablas y gráficos) son un tipo particular de textos, es decir son objetos multi-semióticos caracterizados por el hecho que las relaciones entre los signos y los símbolos están ordenadas por medidas que redireccionan a **códigos** específicos.

En otros términos, tanto en el caso de textos escritos en lenguaje natural como en los escritos en el lenguaje de la estadística, la posibilidad de hacer inferencias sobre las relaciones que organizan las **formas del contenido** está garantizada por el hecho de que las relaciones entre las **formas de la expresión** no son casuales (random); de hecho, en el primer caso (lenguaje natural) las unidades significantes se subsiguen ordenadas según un modo lineal (una tras otra en la cadena del discurso), mientras que en el segundo caso (tablas y gráficos) los principios de ordenación están constituidos por las medidas que determinan la organización de los **espacios semánticos** multidimensionales.

Si bien los espacios semánticos representados en los mapas **T-LAB** son muy variados y cada uno de esos requiere procedimientos de interpretación específicos, se puede suponer que - en general - la lógica del proceso inferencial es la siguiente:

A - sacar cualquier relación significativa entre las unidades "presentes" en el plano de la expresión (por ej. entre "datos" de tablas y/o entre "etiquetas" de gráficos);
B - explorar y comparar los componentes semánticos de las mismas unidades y los contextos a los que están mentalmente y culturalmente asociadas (plano del contenido);
C - construir algunas hipótesis o algunas "categorías" de análisis que, en el contexto definido por el corpus den cuenta de las relaciones entre formas de la expresión y formas del contenido.



Actualmente, las opciones de **T-LAB** presentan las siguientes **restricciones**:

- tamaño del corpus: máximo 90Mb (casi 55.000 páginas en formato ASCII);
- documentos primarios: máximo 30.000 (o 99.999 textos cortos Max. 2.000 caracteres cada uno; Eje. respuestas a preguntas abiertas, mensajes de Twitter etc.);
- variables categóricas: máximo 50, cada una con máximo 150 modalidades;
- modelización de los temas emergentes: máximo 5.000 unidades lexicales (*) por 5.000.000 ocurrencias;
- análisis temático de contextos elementales: máximo 300.000 filas (unidades de contexto) por 5.000 columnas (unidades lexicales);
- clasificación temática de documentos: máximo 99.999 filas (unidades de contexto, documentos) por 5.000 columnas (unidades lexicales);
- análisis de especificidades (unidades lexicales x categorías de una variable): máximo 10.000 filas por 150 columnas;
- análisis de correspondencias (unidades lexicales x categorías de una variable): máximo 10.000 filas por 150 columnas;
- análisis de correspondencias (unidades de contexto x unidades lexicales): máximo 10.000 filas por 5.000 columnas;
- análisis de correspondencias múltiples (contextos elementales x categorías de dos o más variables): máximo 150.000 filas por 250 columnas;
- descomposición de valores singulares (SVD)): máximo 300.000 filas por 5.000 columnas;
- cluster analysis que utiliza los resultados de un precedente análisis de correspondencias (o SVD): máximo 10.000 filas (unidades lexicales o contextos elementales);
- asociaciones de palabras, comparaciones entre parejas de palabras-clave y análisis de co-palabras: máximo 5.000 unidades lexicales;
- análisis de secuencias: máximo 5.000 unidades lexicales (o categorías) por 3.000.000 ocurrencias.

(*) En **T-LAB**, las 'unidades léxicales' son palabras, palabras múltiples, lemas y categorías semánticas. Así que, cuando se aplica la lematización automática, 5.000 unidades léxicales corresponden a cerca de 12.000 palabras.