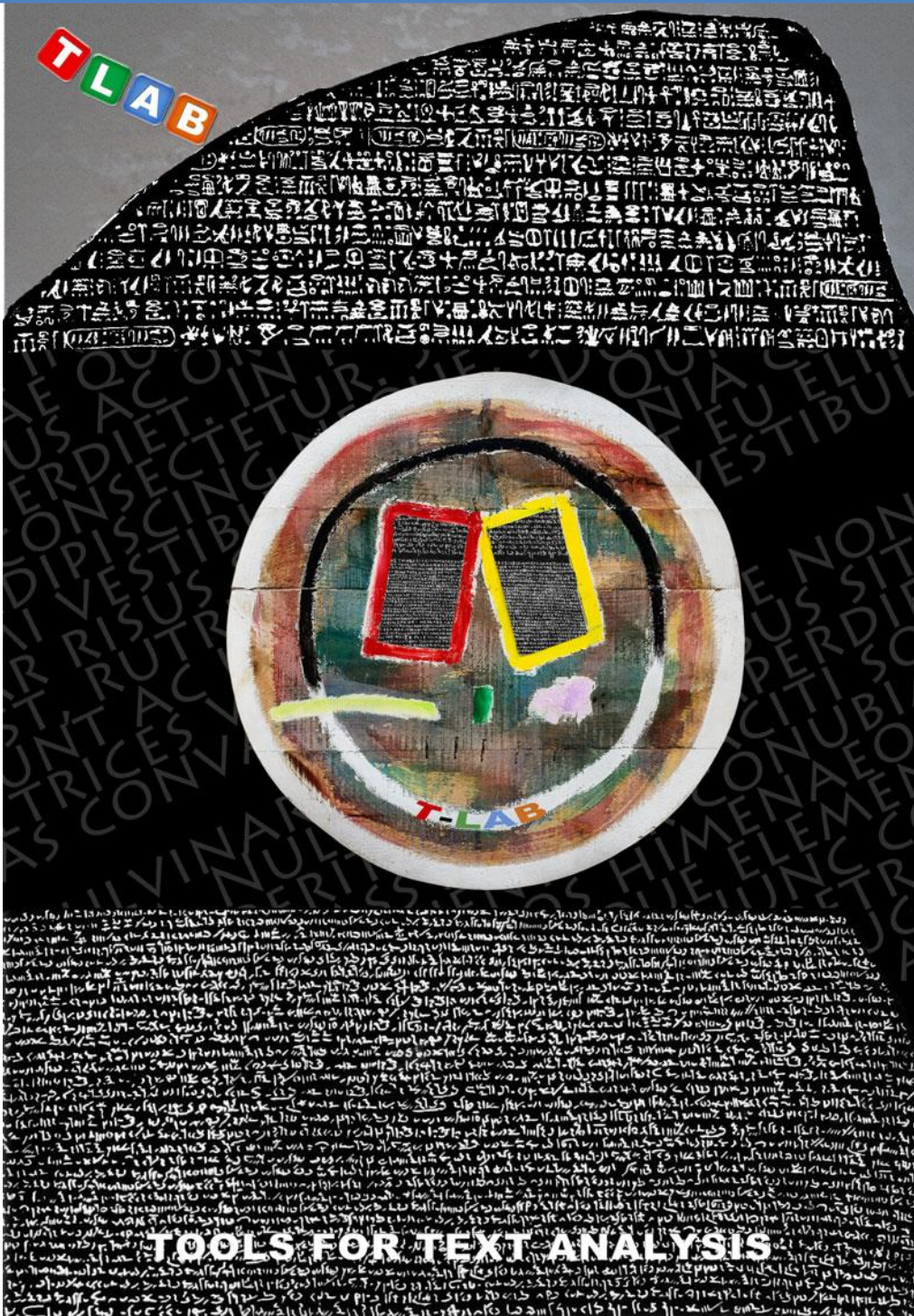


Quick Introduction



Tools for Text Analysis

Copyright © 2001-2024
T-LAB by Franco Lancia
All rights reserved.

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

T-LAB is a registered trademark

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

What T-LAB does and what it enables us to do

(Excerpt from the User's Manual)

T-LAB software is an all-in-one set of **linguistic, statistical and graphical tools for text analysis** which can be used in research fields like Content Analysis, Sentiment Analysis, Semantic Analysis, Thematic Analysis, Text Mining, Perceptual Mapping, Discourse Analysis, Network Text Analysis, Document Clustering, Text Summarization.



In fact **T-LAB** tools allow the user to easily manage tasks like the following:

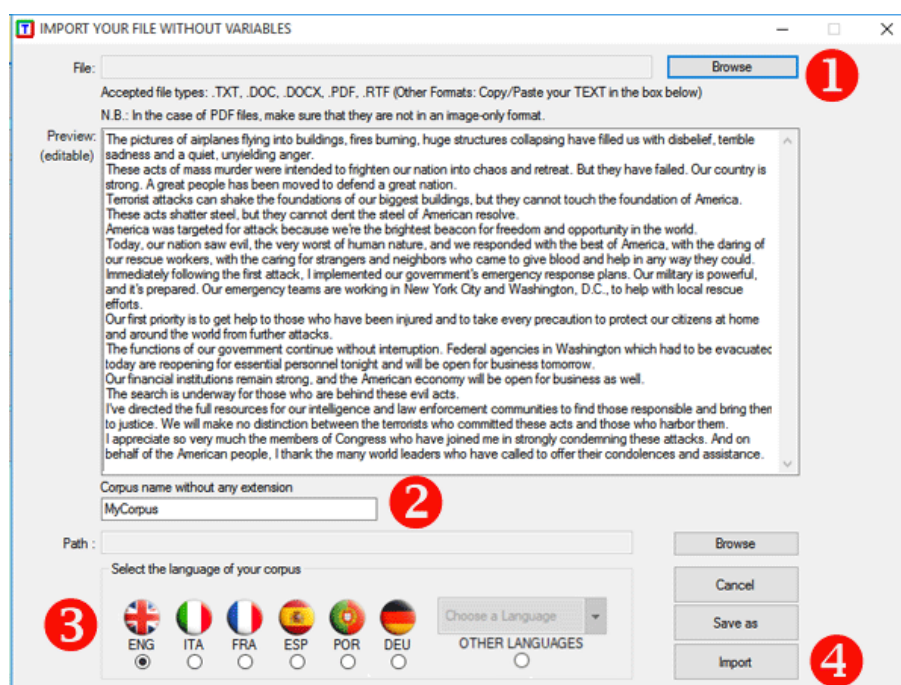
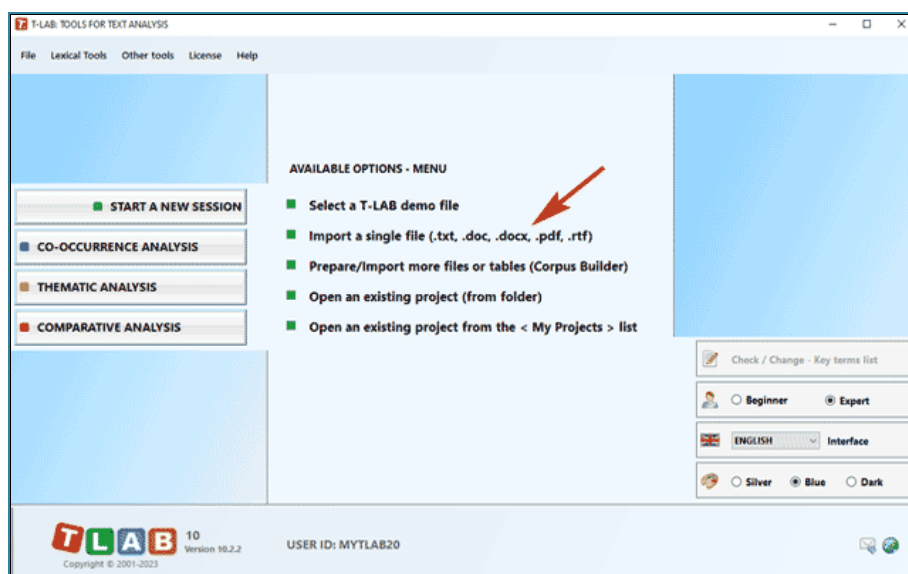
- measure, explore and map the **co-occurrence relationships** between key-terms;
- perform either unsupervised or supervised clustering of textual units and documents, i.e. perform a **bottom-up clustering** which highlights **emerging themes** or a perform **top-down classification** which uses a set of **predefined categories**;
- check the **lexical units** (i.e. words or lemmas), **context units** (i.e. sentences or paragraphs) and **themes** which are typical of specific text subsets (e.g. newspaper articles from specific time periods, interviews with people belonging to the same category);
- apply categories for **sentiment analysis**;
- perform various types of **correspondence analysis** and **cluster analysis**;
- create **semantic maps** that represent **dynamic** aspects of the discourse (i.e. sequential relationships between words or themes);
- represent and explore any text as a **network**;
- obtain measures and graphical representations concerning **texts and discourses** treated as **dynamic systems**;
- customize and apply various types of **dictionaries** for both lexical and content analysis;
- perform **concordance** searches;
- analyse all the **corpus** or its **subsets** (e.g. groups of documents) by using various key-term lists;
- create, explore and export numerous **contingency tables** and **co-occurrences matrices**.

The **T-LAB** user interface is very **user-friendly** and various types of texts can be analysed:

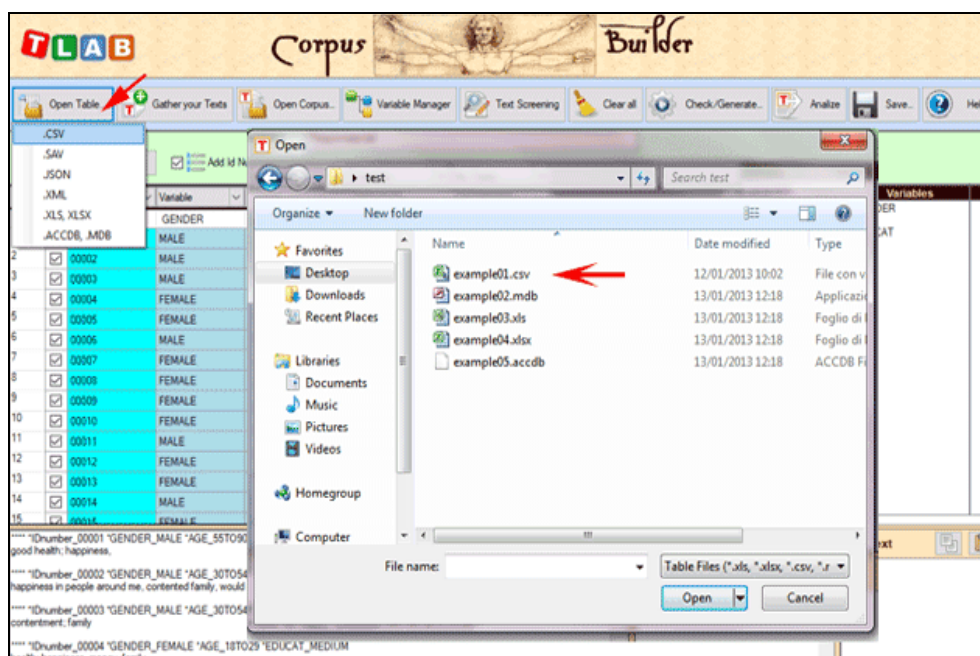
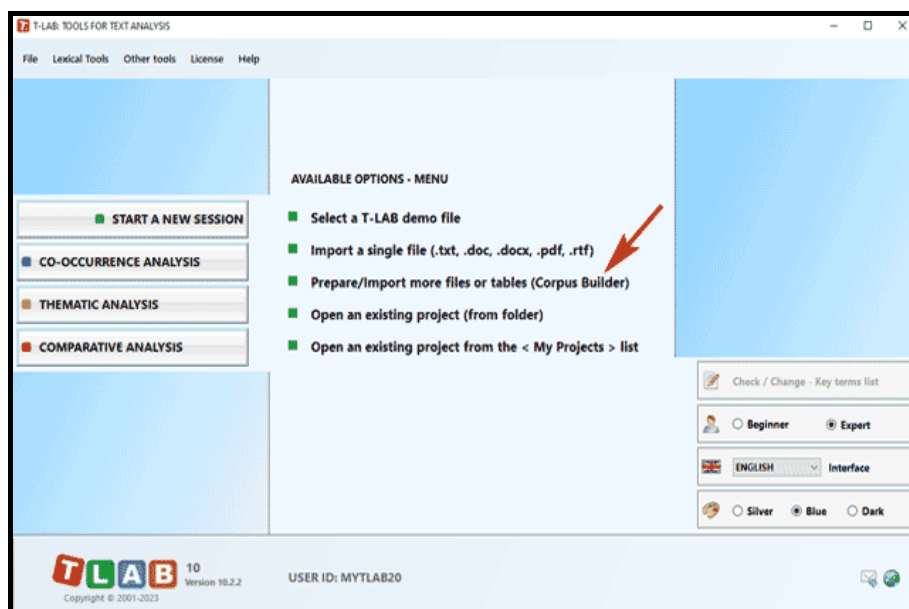
- a single text (e.g. an interview, a book, etc.);
- a set of texts (e.g. a set of interviews, web pages, newspaper articles, responses to open-ended questions, Twitter messages, etc.).

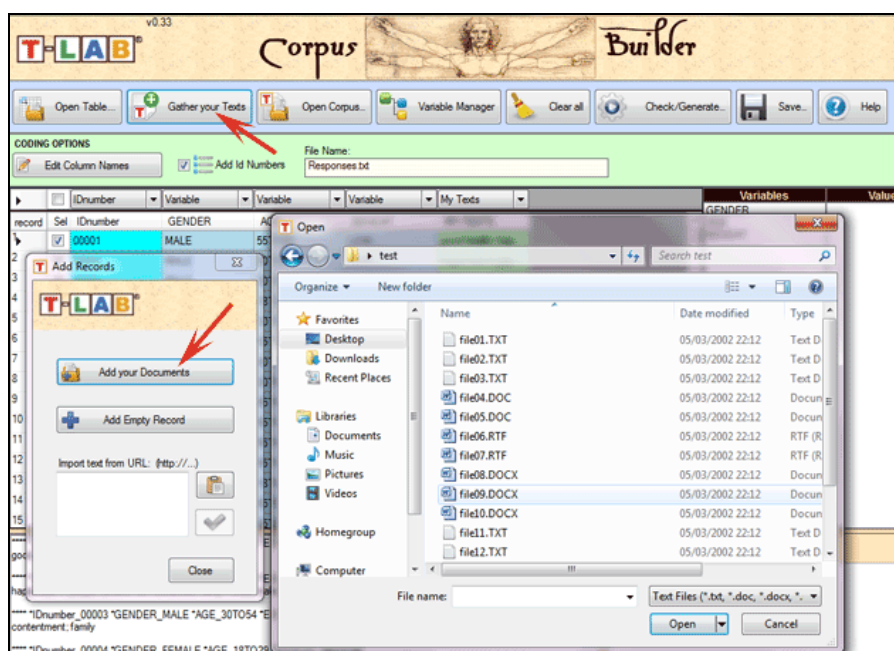
All texts can be encoded with categorical **variables** and/or with **Unique Identifiers** that correspond to context units or cases (e.g. responses to open-ended questions).

In the case of a **single document** (or a corpus considered as a single text) **T-LAB** needs no further work: just select the 'Import a single file...' option (see below) and proceed as follows.



When, on the other hand, the corpus is made up of various texts and/or categorical variables are used the **Corpus Builder** tool (see below) must be used. In fact, such a tool automatically transforms any textual material and various types of files (i.e. up to ten different formats) into a corpus file ready to be imported by **T-LAB**.

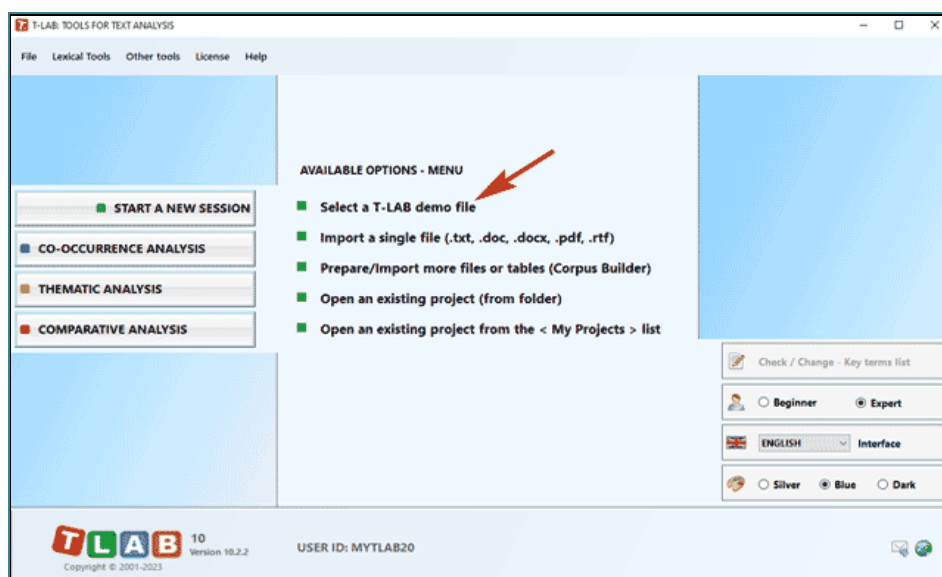




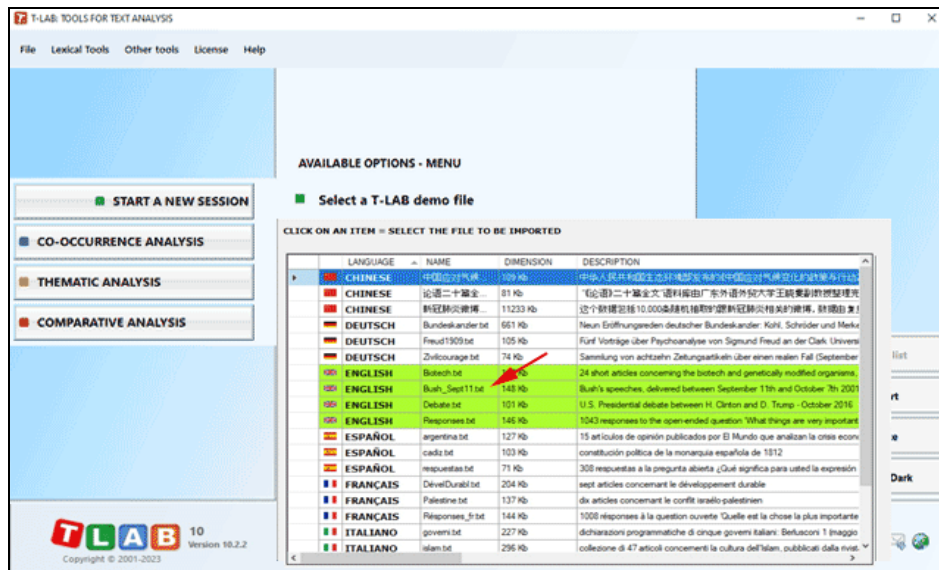
N.B.: At the moment, in order to ensure the integrated use of various tools, each corpus file shouldn't exceed 90 Mb (i.e. about 55,000 pages in .txt format). For more information, see the 'Requirements and Performances' section of the Help/Manual.

Six steps are that is required to perform a quick check of the software functionalities:

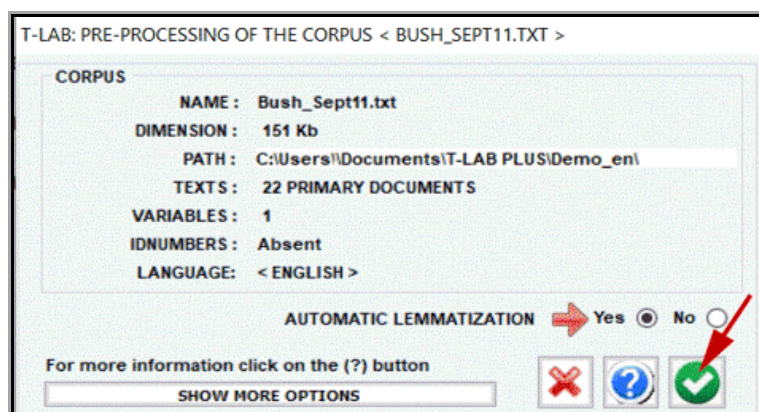
1 – Click on the 'Select a T-LAB demo File' option



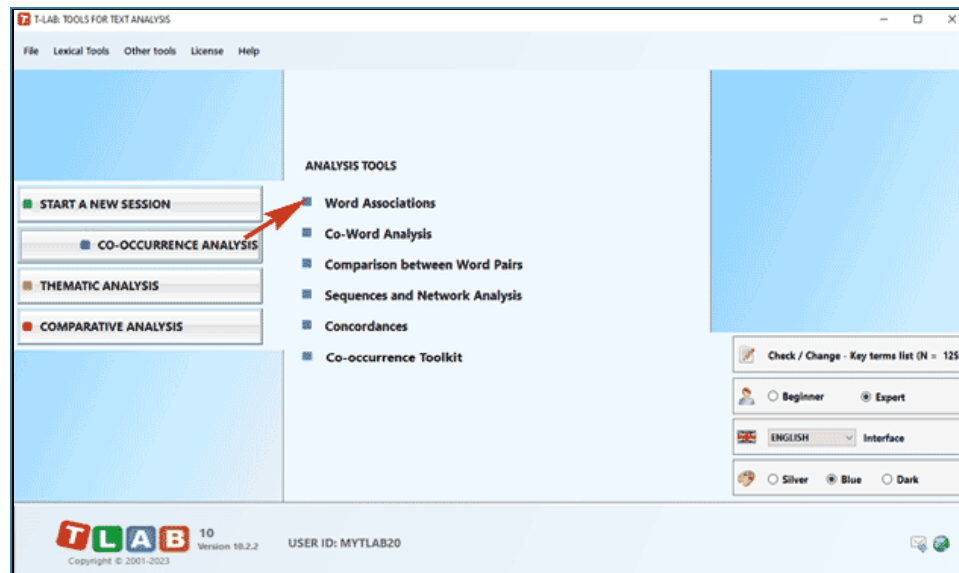
2 - Select any corpus to analyse



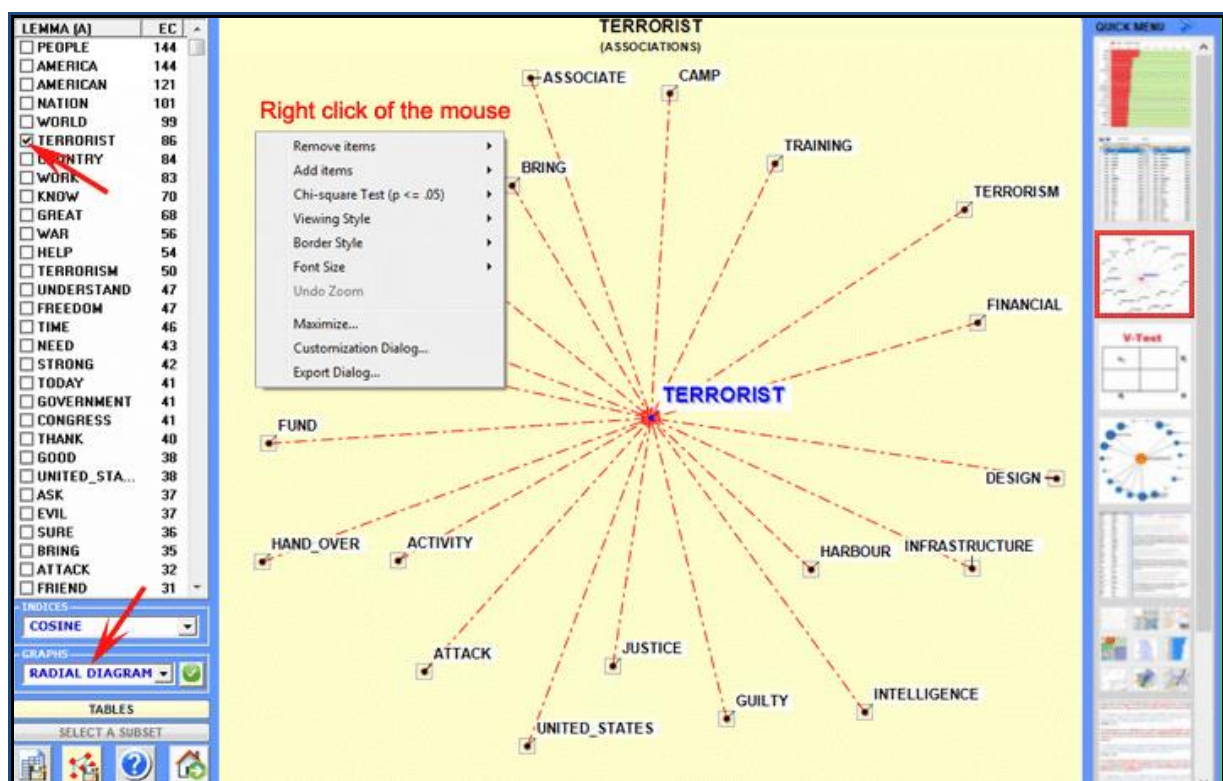
3 - Click "ok" in the first Setup window

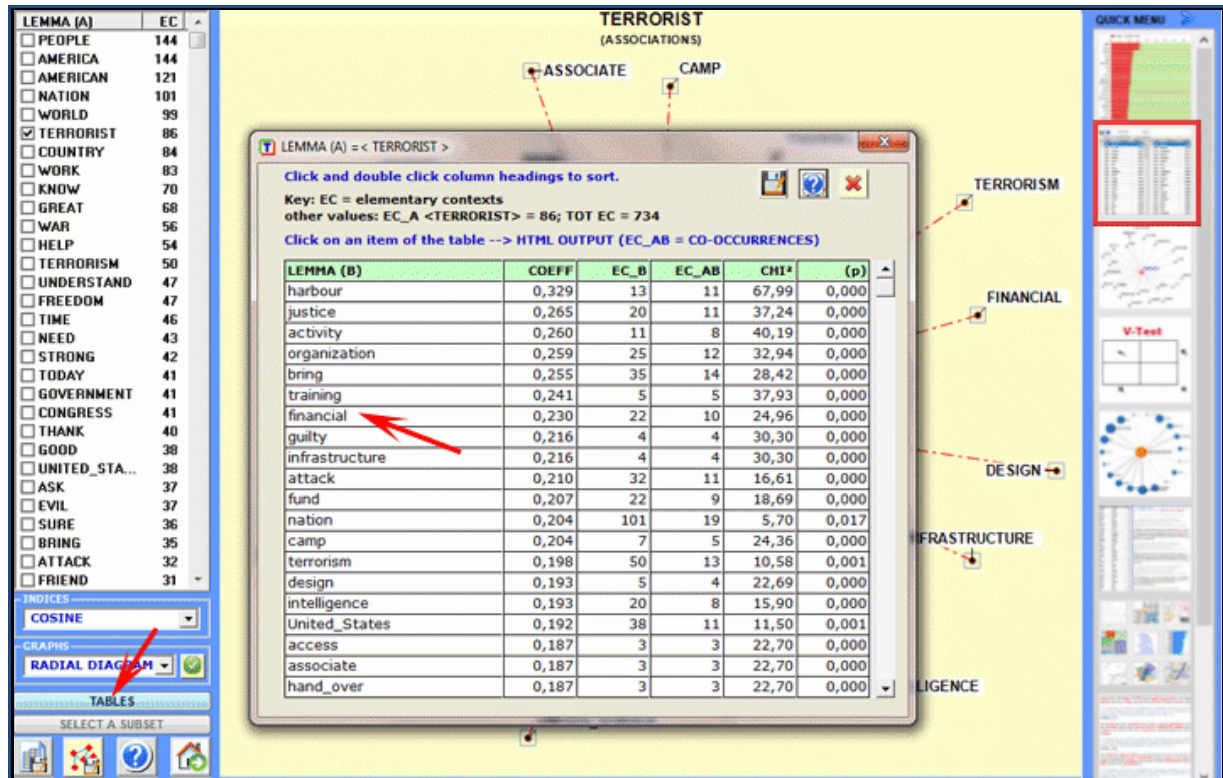


4 - Select a tool from one of the "Analysis" sub-menus



5 - Check the results





6 - Use the contextual help function to interpret the various graphs and tables

T-LAB Help

Nascondi Indietro Stampa Opzioni

Word Associations

This **T-LAB** tool allows us to check how **co-occurrence** relationships determine the local meaning of selected words.

On the left there is the table with the **key-term** list and their occurrence values within the whole corpus or a **subset** of it.

On user request (a simple click), for each key-term **T-LAB** shows the **lexical units** that share co-occurrence contexts with that key-word.

The selection is carried out by the computation of an **Association Index** (Cosine, Dice and Jaccard).

For each query, **T-LAB** produces graphs

The following information is provided to help the user to better understand what **T-LAB** does and how to make full use of it.

Apart from the user interface, the **T-LAB** system is organized into two main components:

- the **database**, the "place" where the input **corpus** (the text or the set of texts to be analysed) is represented as a set of **tables** in which the **analysis units**, their characteristics and their mutual relationships are recorded.
- the **algorithms**, which are subsets of **instructions** that allow us to use the interface, to consult and modify the database, to produce further tables with the available data, to perform **statistical computations** and to produce **outputs** that represent the relationships between the analysed data.

To understand how **T-LAB** works and how it can be used, it is essential to have a clear idea as to which **analysis units** are filed in its database and what statistical **algorithms** are used in the various analyses. In fact, the analysed data tables always consist of rows and columns the headings of which correspond to the analysis units filed in the database, while the algorithms regulate the processes that make it possible to detect significant relationships between the data and to extract useful information.

The **analysis units** used in **T-LAB** are of two types: **lexical units** and **context units**.

A - the **lexical units** are words and multi-words, filed and classified on the basis of a criterion. More precisely, in the **T-LAB** database each lexical unit consists of a classified record with two fields: **word** and **lemma**. In the first field ("**word**"), the words are listed as they appear in the corpus, while in the second ("**lemma**") the **labels** attributed to groups of lexical units are listed and classified according to linguistic criteria (e.g. **lemmatization**) or by dictionaries and semantic grids defined by the user.

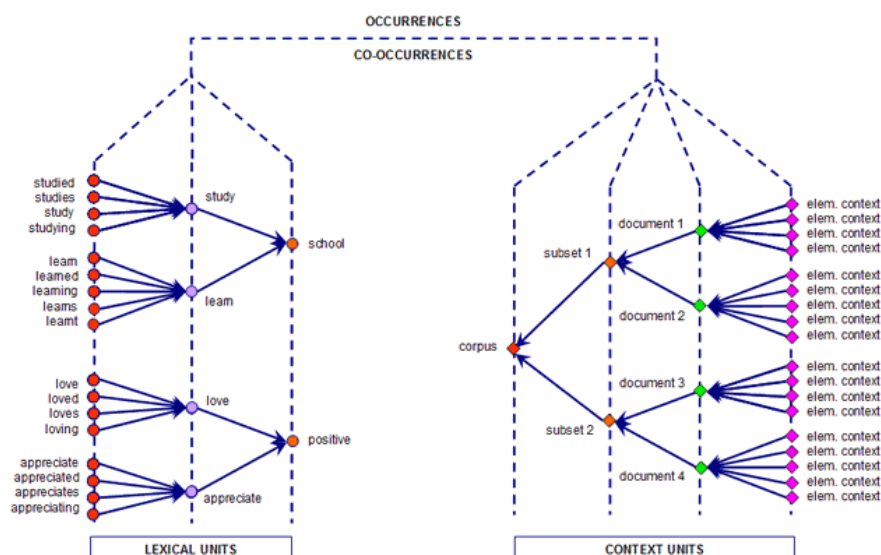
B - the **context units** are portions of text that the corpus can be divided into. More precisely, according to **T-LAB** logic, there can be three types of context units:

B.1 primary documents, which correspond to the "natural" subdivision of the corpus (e.g. interviews, articles, answers to open-ended questions, etc.), that is the **initial context** defined by the user;

B.2 elementary contexts, which correspond to syntagmatic units (i.e. fragments, sentences, paragraphs) in which each primary document can be subdivided;

B.3 corpus subsets, which correspond to groups of primary documents which lead to the same category (e.g. interviews with "men" or "women", articles in a specific year or a particular magazine and so on), including **thematic clusters** of documents or elementary contexts obtained by using the corresponding **T-LAB** tools.

The picture below illustrates the possible relationships between lexical and context units which **T-LAB**, through statistical and graphical tools (see section 5 below), allows us to analyse.

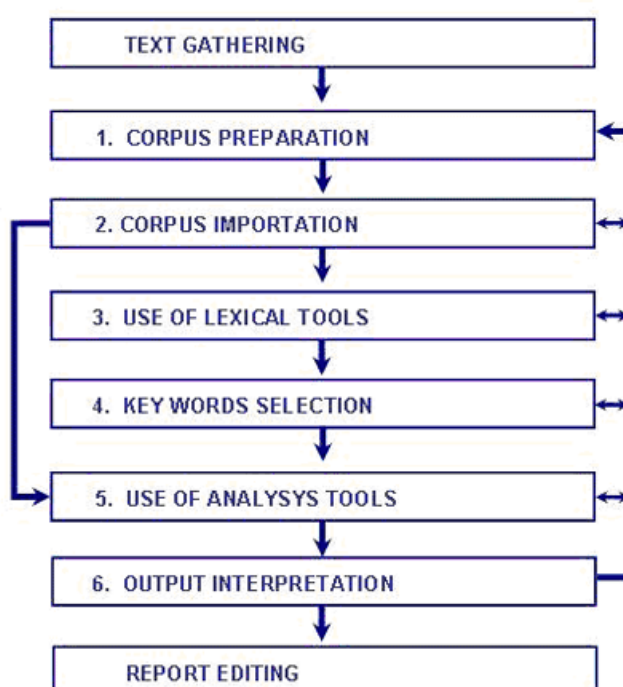


Starting from this database organization, **T-LAB** makes it possible - in automatic mode - to explore and to analyse the relationships between the analysis units of the whole **corpus** or its **subsets**.

In **T-LAB**, the selection of any analysis tool (click of the mouse) always activates a semi-automatic process that, with a few simple operations, generates an input table, applies some statistical algorithms and produces some outputs.

Let's consider how a typical work **project** which uses **T-LAB** can be managed. Hypothetically, each project consists of a set of analytical activities (operations) which have the same **corpus** as their subject and are organized according to the user's **strategy** and **plan**. It then begins **gathering the texts** to be analysed, and concludes with a **report**.

The succession of the various phases is illustrated in the following diagram:



N.B.:

- The six numbered phases, from the corpus preparation to the interpretation of the outputs, are supported by **T-LAB** tools and are always reversible;
- By using **T-LAB automatic settings** it is possible to avoid two phases (3 and 4); however, in order to achieve high quality results, their use is, nevertheless, advisable.

Now let's try to comment on the various steps.

1 - CORPUS PREPARATION: transformation of the texts to be analysed in a file (**corpus**) that can be processed by the software.

In the case of a single text (or a corpus considered as a single text) **T-LAB** needs no further work. When, on the other hand, the corpus is made up of various texts and/or categorical variables are used the Corpus Builder tool must be used, which automatically transforms any textual material and various types of files (i.e. up to eleven different formats) into a corpus file ready to be imported by **T-LAB**.

N.B.:

- At the end of the corpus preparation phase it is recommended that a new folder be created which contains only the corpus to be imported;
- When analysing any corpus, it is recommended that the working files (i.e. the working folder of the corpus) reside on a hard disk of the computer where **T-LAB** is installed. Otherwise, the various procedures could slow down and the software may report errors.

2 - CORPUS IMPORTATION: a series of **automatic processes** that transform the corpus into a set of tables integrated in the **T-LAB database**.

During the pre-processing phase, **T-LAB** carries out the following treatments: Corpus **Normalization**; **Multi-Word** and **Stop-Word** detection; **Elementary Context** segmentation; Automatic **Lemmatization** or **Stemming**; **Vocabulary** building; **Key-Terms** selection.

Here is the complete list of the languages for which specific pre-processing options are available.

LEMMATIZATION: Catalan, Croatian, English, French, German, Italian, Latin, Polish, Portuguese, Romanian, Russian, Serbian, Slovak, Spanish, Swedish, Ukrainian.

STEMMING: Arabic, Bengali, Bulgarian, Czech, Danish, Dutch, Finnish, Greek, Hindi, Hungarian, Indonesian, Marathi, Norwegian, Persian, Turkish.

WORD SEGMENTATION: Chinese and Japanese.

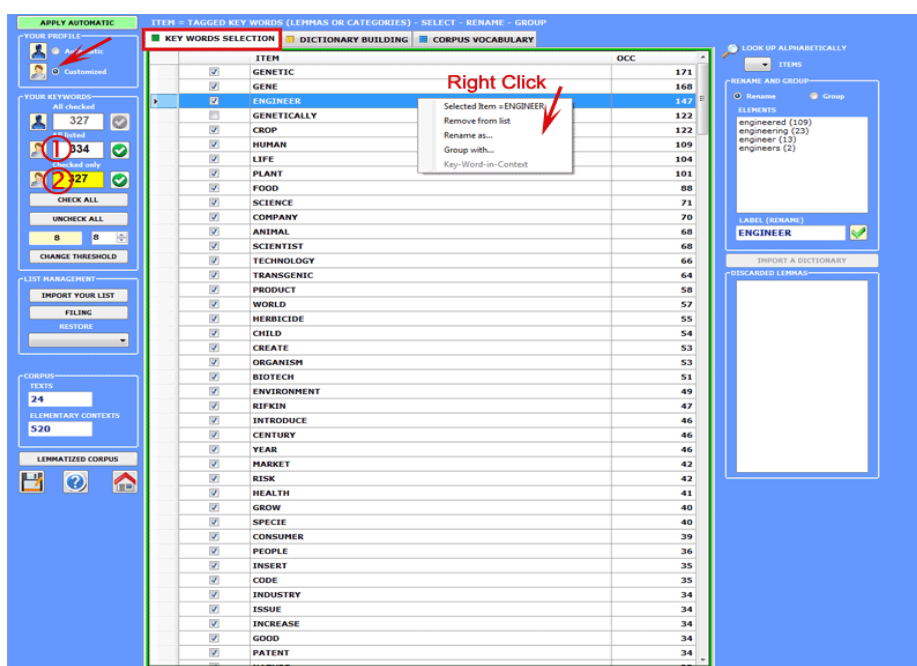
In any case, without automatic lemmatization and / or by using customized dictionaries the user can analyse texts in **all languages**, provided that words are separated by spaces and / or punctuation.

The setup form in which the user can select the pre-processing options which fit his needs is the following:

N.B.: As the pre-processing options determine both the kind and the number of analysis units (i.e. context units and lexical units), different choices determine different analysis results. For this reason, all **T-LAB** outputs (i.e. charts and tables) shown in the user's manual and in the on-line help are just indicative.

4 - THE KEY-WORD SELECTION consists of the arrangement of one or more lists of lexical units (words, lemmas or categories) to be used for producing the data tables to be analysed.

The **automatic settings** option provides the lists of the **key-words** selected by **T-LAB**; nevertheless, since the choice of the analysis units is extremely relevant in relation to subsequent elaborations, the use of **customized settings** (see below) is highly recommended. In this way the user can choose to modify the list suggested by **T-LAB** and/or to arrange lists that better correspond to the objectives of his research.



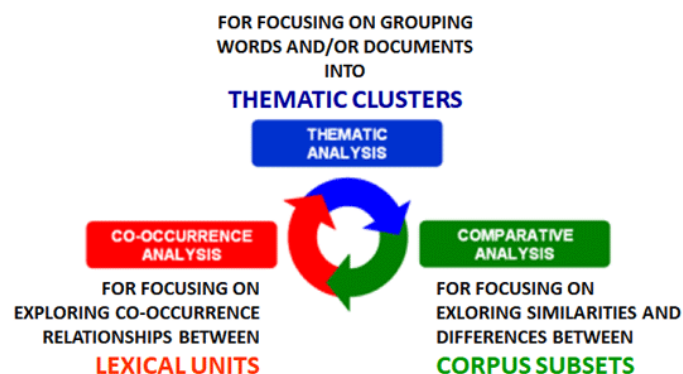
In any case, while creating these lists, the user can refer to the following criteria:

- check the quantitative (total of the occurrences) and qualitative **importance** of the various items;
- check the **limitations** of the analytical tools that you intend to use (see at the end of this chapter);
- check whether the set of items is compatible with your own research **strategies** (see item : 5 to follow).

5 - THE USE OF ANALYSIS TOOLS allows the user to obtain outputs (tables and graphs) that represent **significant relationships** between the analysis units and enables the user to make **inferences**.

At the moment, **T-LAB** includes twenty different analysis tools each of them having its own specific logic; that is, each one generates specific tables, uses specific algorithms and produces specific outputs.

Consequently, depending on the structure of texts to be analysed and on the goals to be achieved, the user has to decide which tools are more appropriate for their analysis strategy every time.

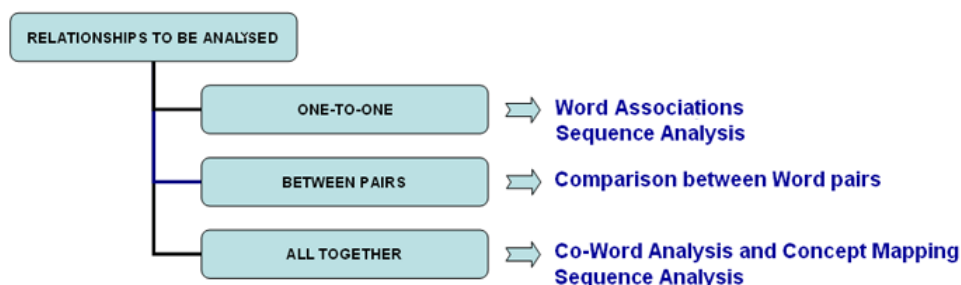


N.B.: Besides the distinction between tools for **co-occurrence**, **comparative** and **thematic** analysis, it can be useful to consider that some of the latter allow us to obtain new corpus subsets which can be included in further analysis steps.

Even though the various **T-LAB** tools can be used in any order, there are nevertheless three ideal starting points in the system which correspond to the three ANALYSIS sub-menus:

A : TOOLS FOR CO-OCCURRENCE ANALYSIS

These tools enable us to analyse different kinds of relationships between lexical units (i.e. words or lemmas)

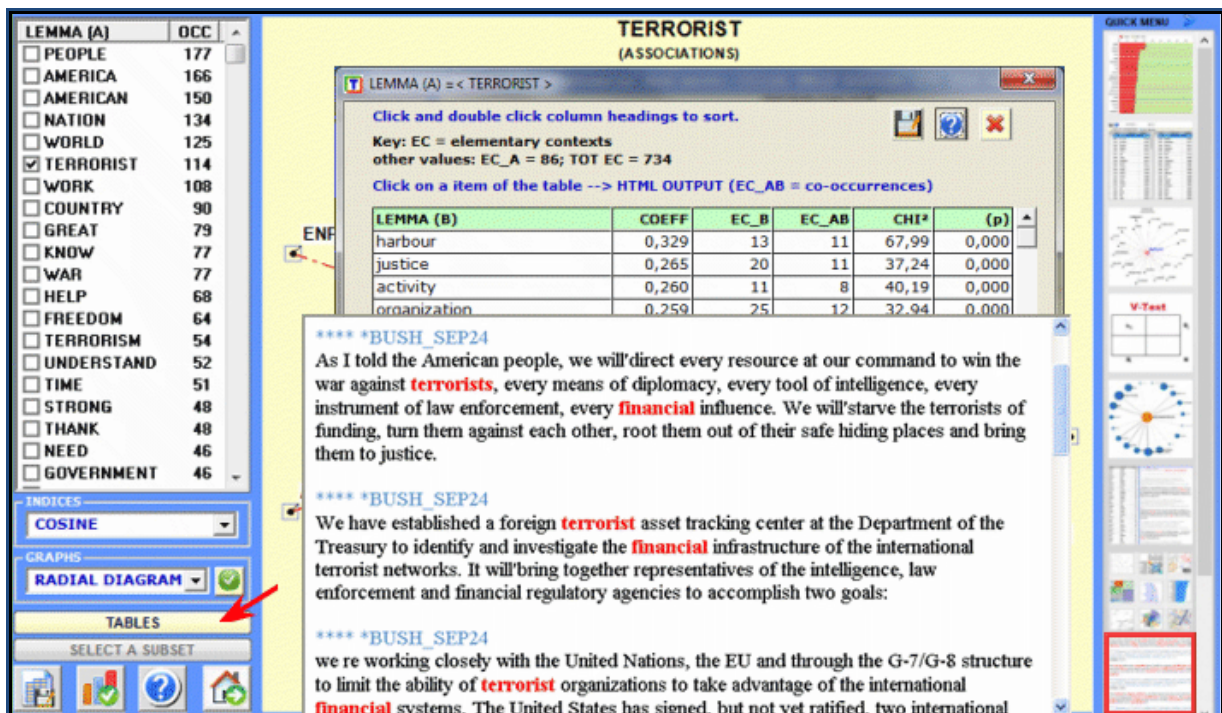
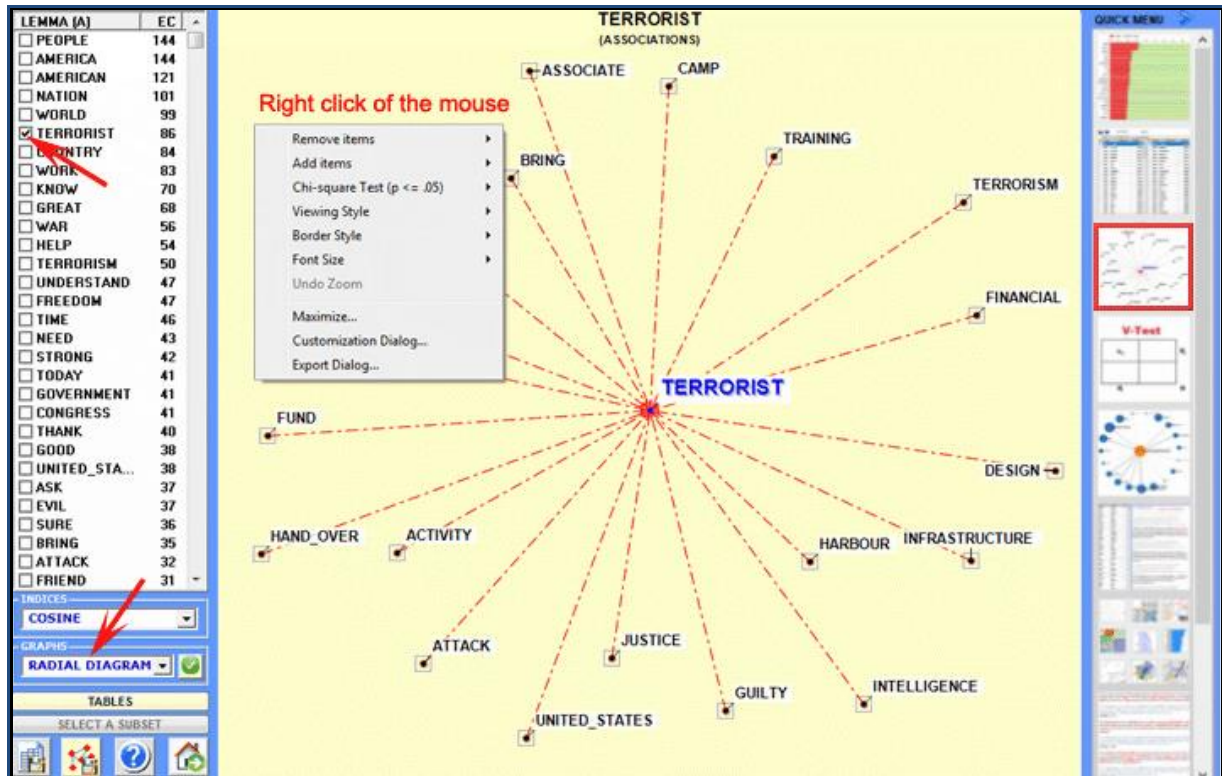


According to the types of relationships to be analysed, the **T-LAB** options indicated in this diagram use one or more of the following statistical tools: **Association Indexes**, **Chi Square Tests**, **Cluster Analysis**, **Multidimensional Scaling**, **Principal Component Analysis**, **t-SNE** and **Markov chains**.

Here are some examples (N.B.: for more information on how to interpret the outputs please refer to the corresponding sections of the help/manual).

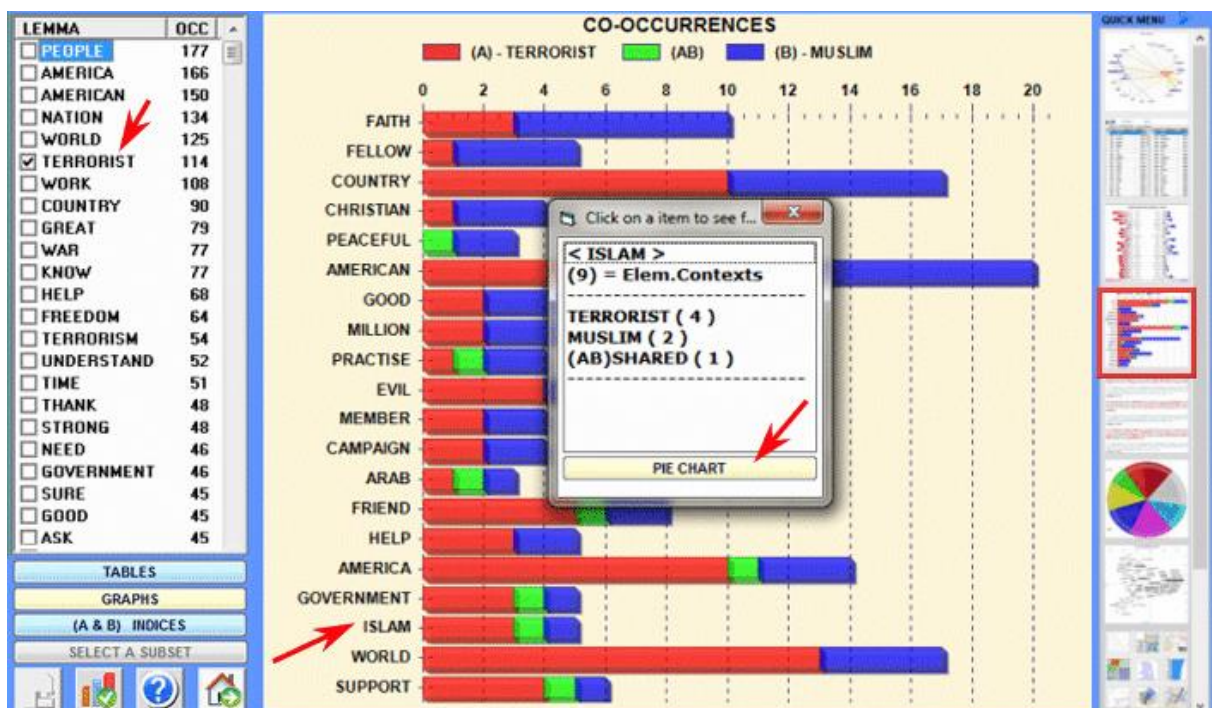
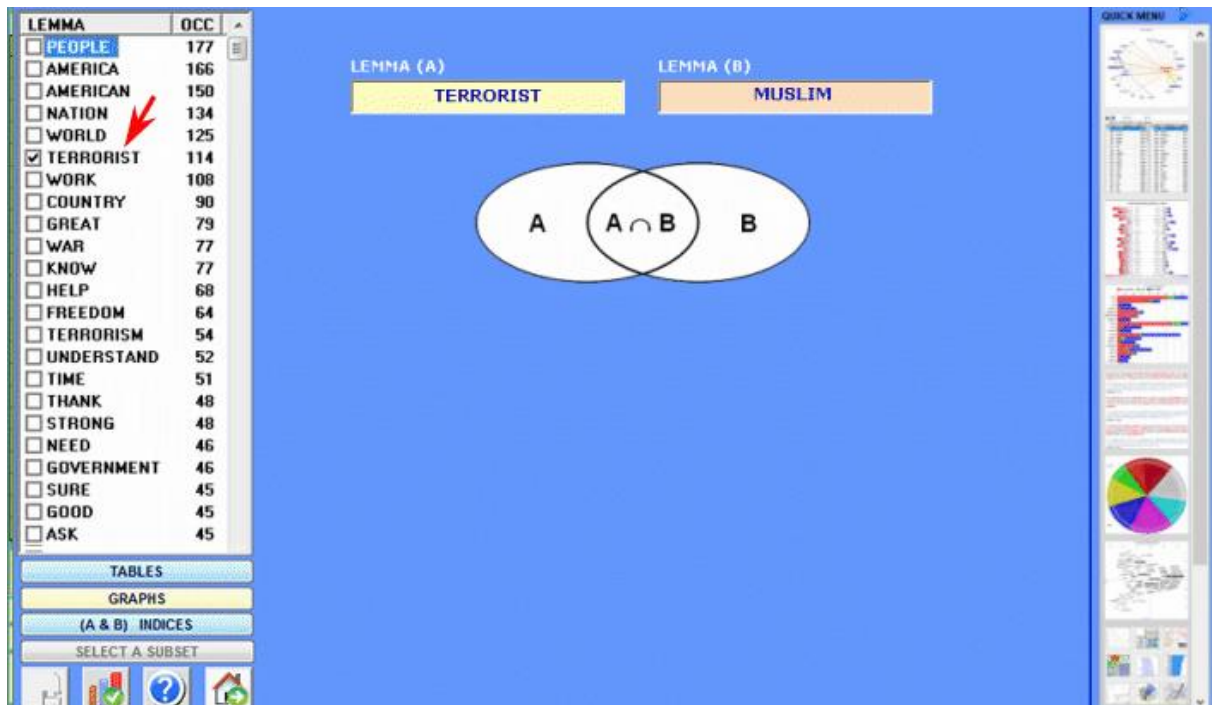
- Word Associations

This **T-LAB** tool allows us to check how **co-occurrence** relationships determine the **local meaning** of selected words.



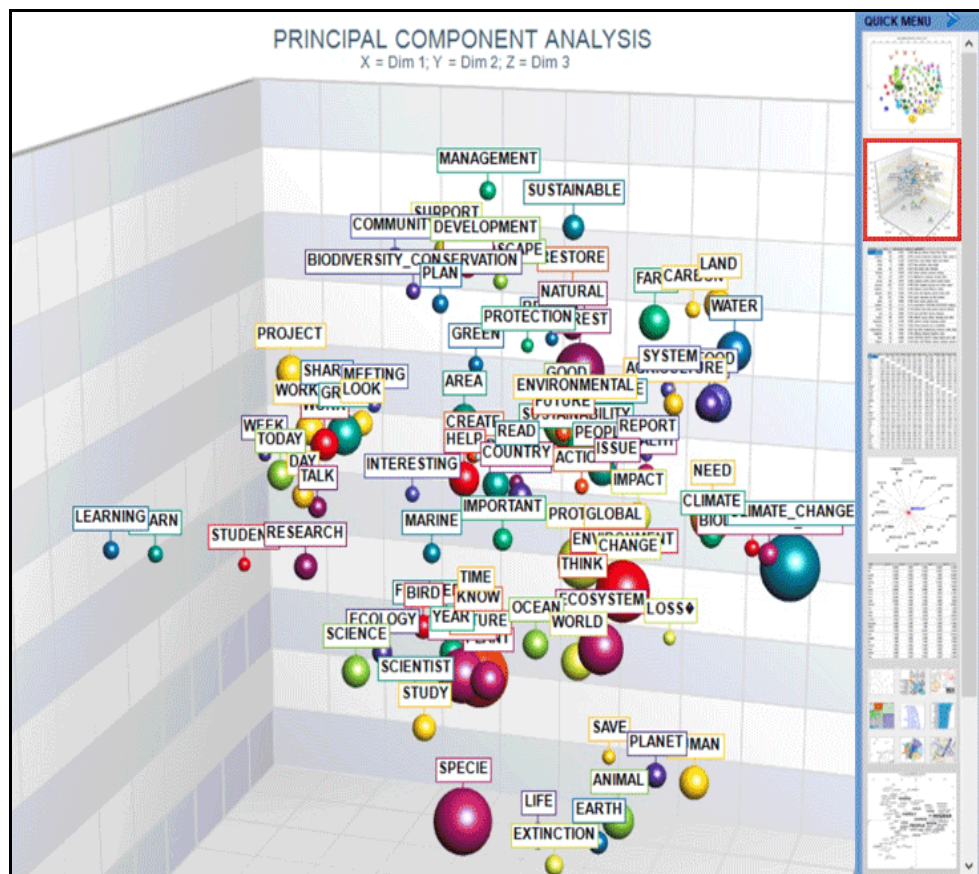
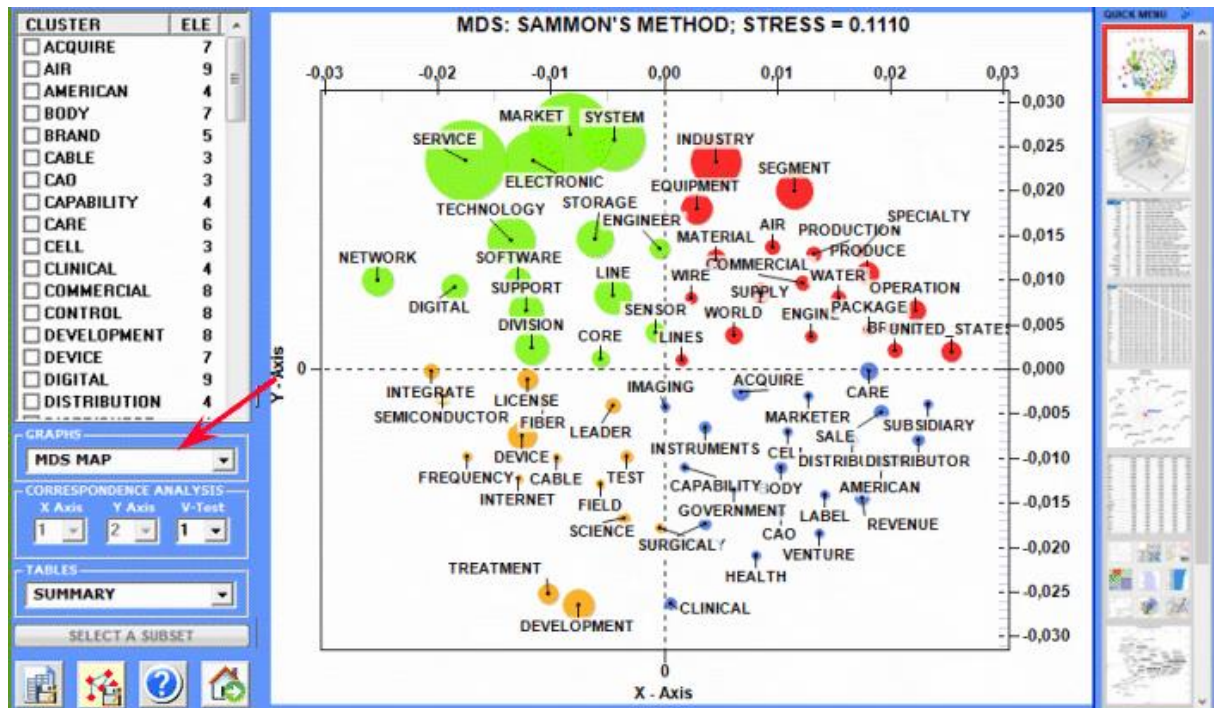
- Comparison between Word Pairs

This **T-LAB** tool allows us to compare sets of elementary contexts (i.e. co-occurrence contexts) in which the elements of a pair of key-words are present.



- Co-Word Analysis

This **T-LAB** tool allows us to find and map co-occurrence relationships within (and between) sets of key-words.

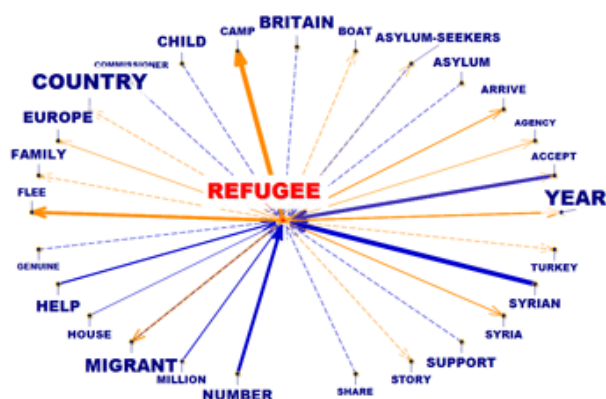


- Sequence and Network Analysis

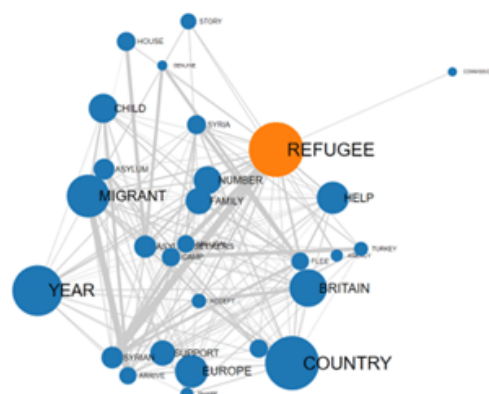
This **T-LAB** tool, which takes into account the positions of the various lexical units relative to each other, allows us to represent and explore any text as a network.

That means that the user is allowed to check the relationships between the ‘nodes’ (i.e. the key-terms) of the network at different levels: a) in one-to-one connections; b) in the ‘ego’ networks; c) within the ‘community’ to which they belong; d) within the entire text network.

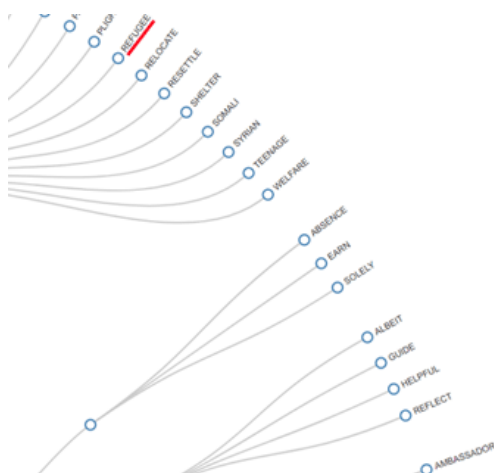
ONE-TO-ONE



EGO-NETWORK



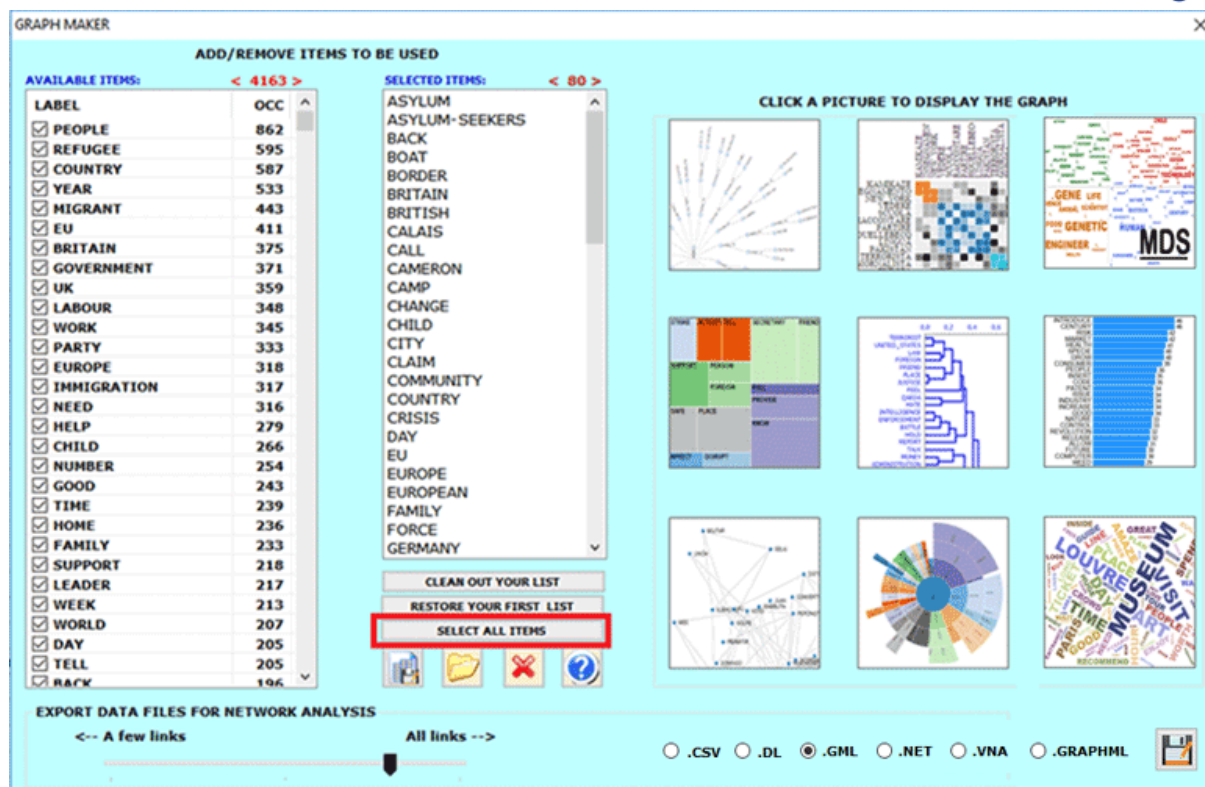
COMMUNITY



ENTIRE NETWORK

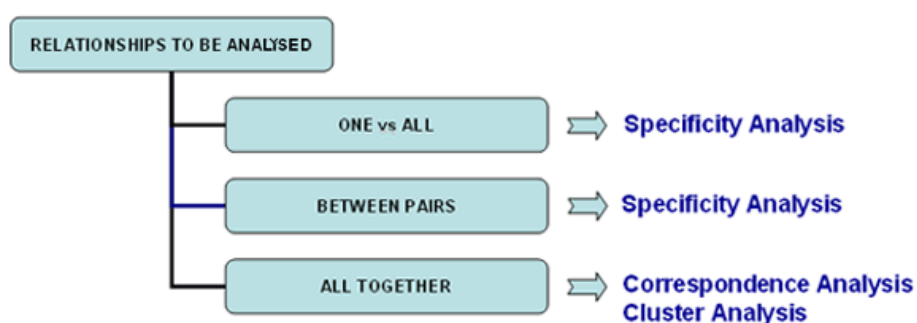


Moreover, by clicking the **GRAPH MAKER** option, the user is allowed to obtain various types of graphs by using customized lists of key words (see below).

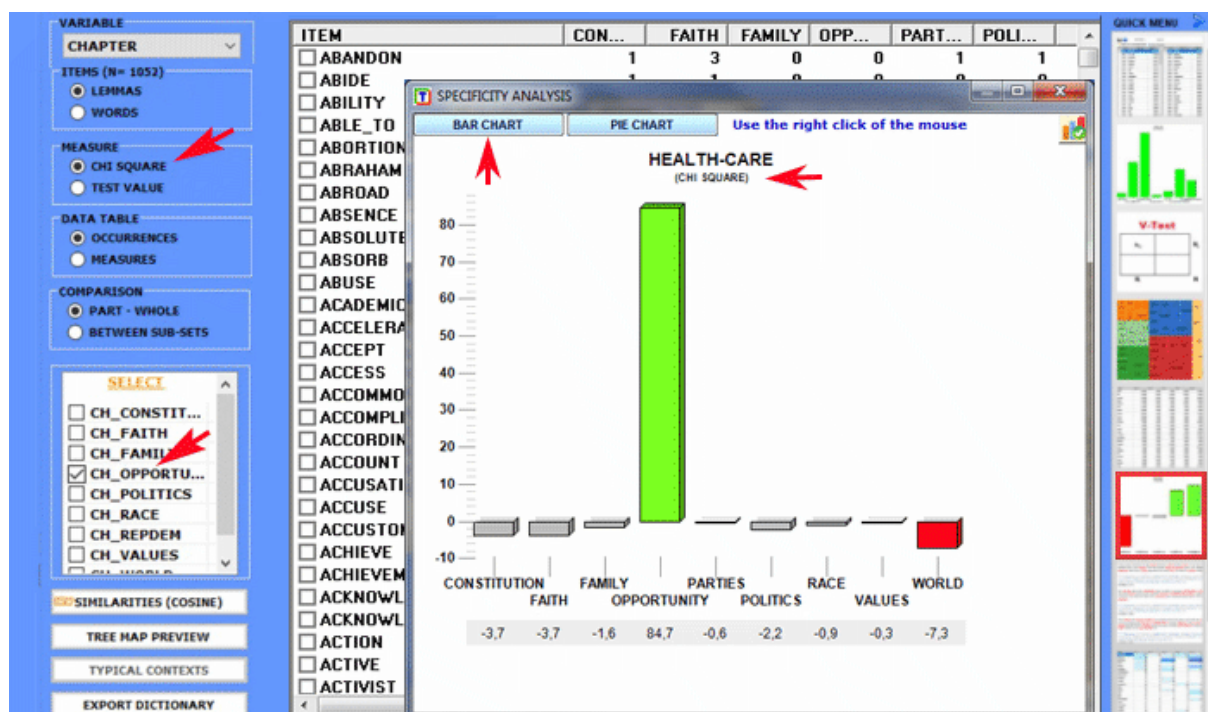
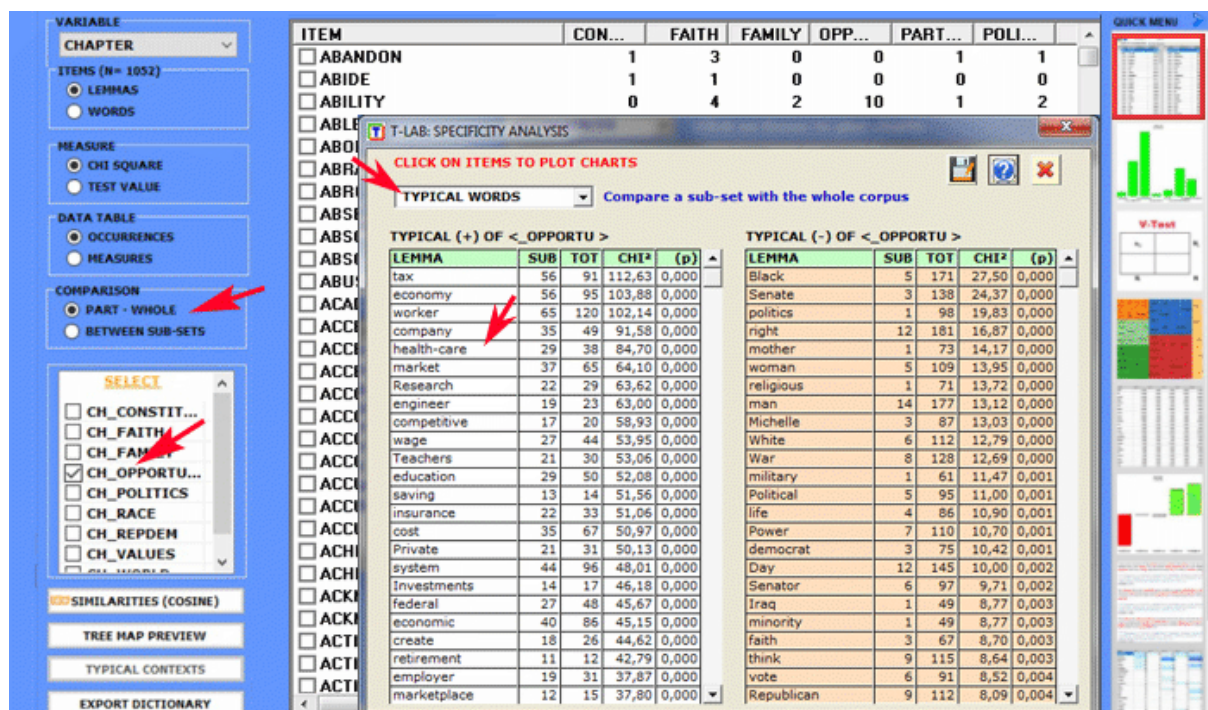


B : TOOLS FOR COMPARATIVE ANALYSIS

These tools enable us to analyse different kinds of relationships between context units (e.g. documents or corpus subsets)

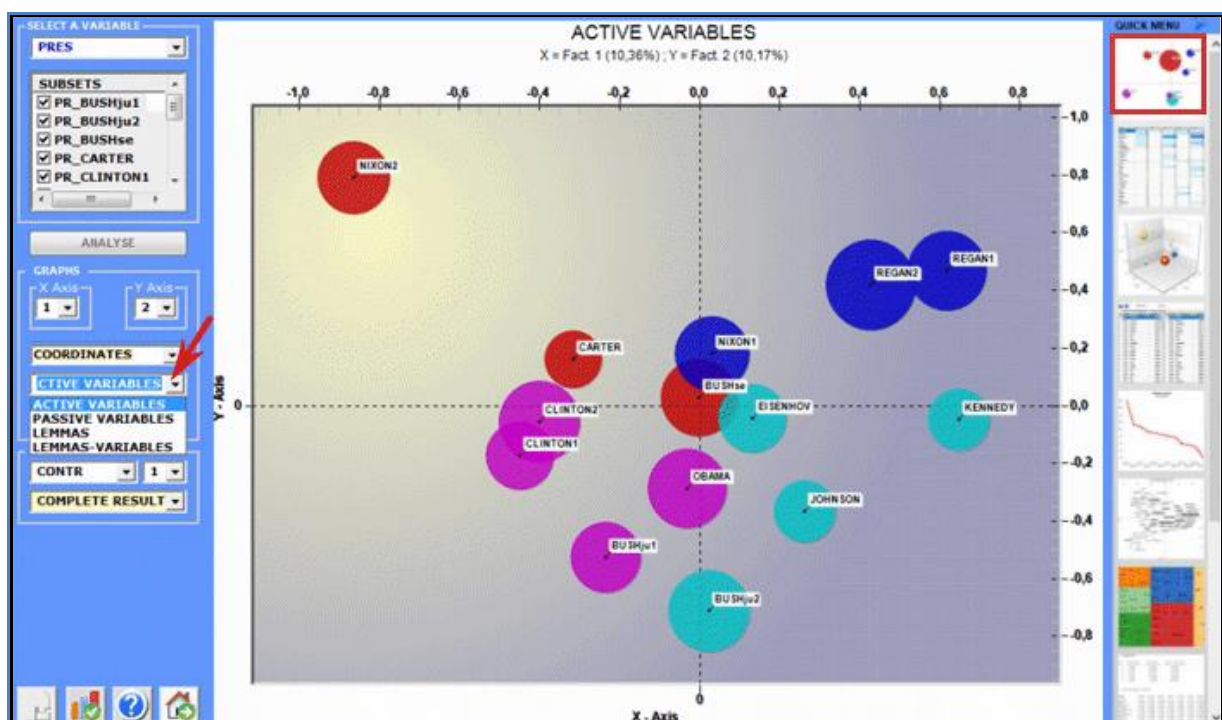


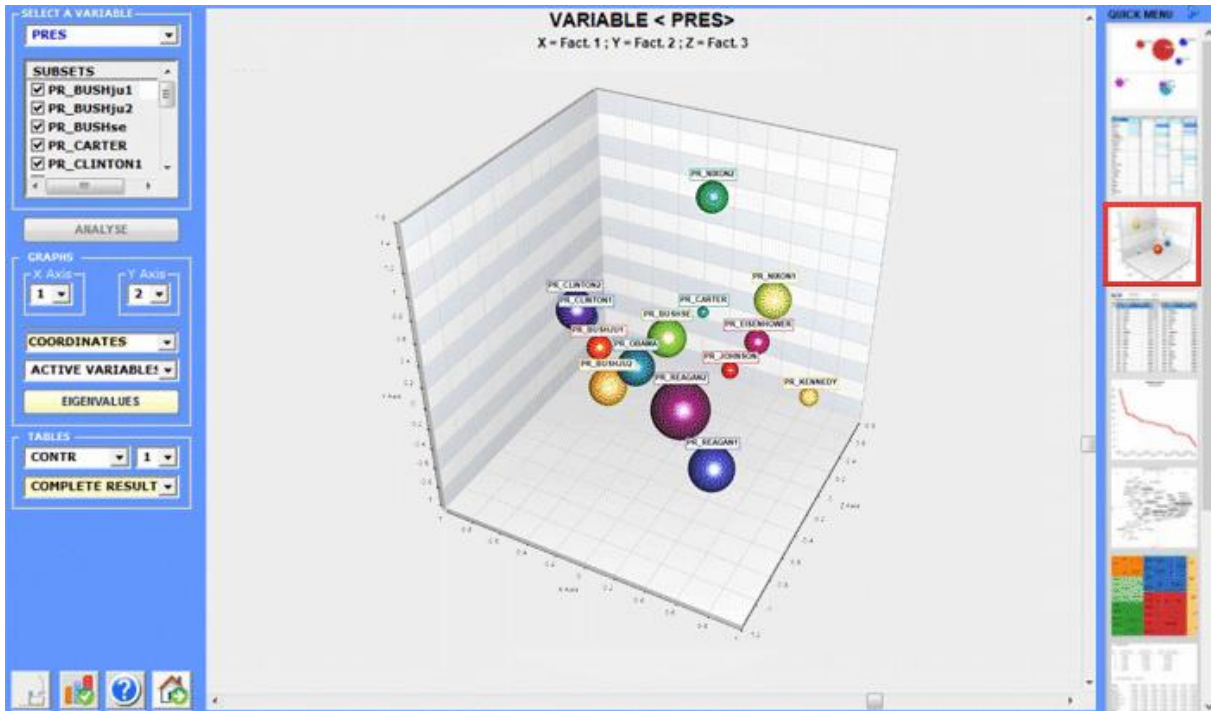
Specificity Analysis enables us to check which words are **typical** or **exclusive** of a specific corpus subset, either comparing it with the rest of the corpus or with another subset. Moreover it allows us to extract the **typical contexts** (i.e. the characteristic elementary contexts) of each analysed subset (e.g. the 'typical' sentences used by any specific political leader).



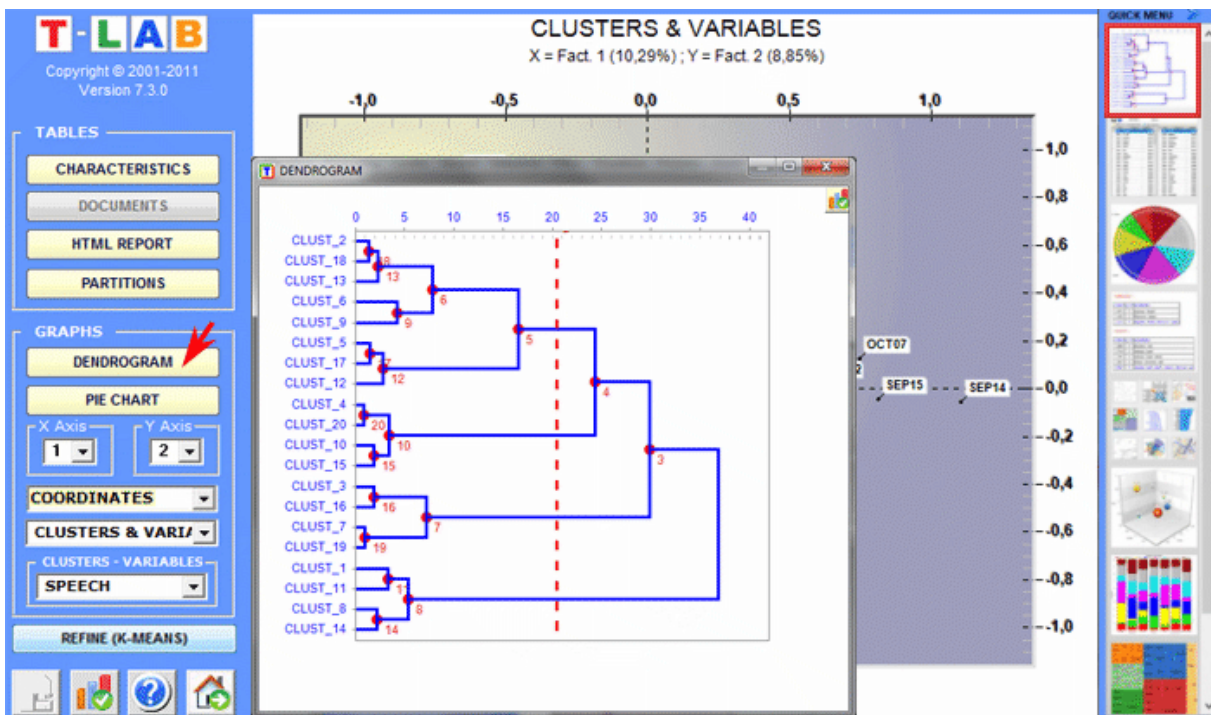


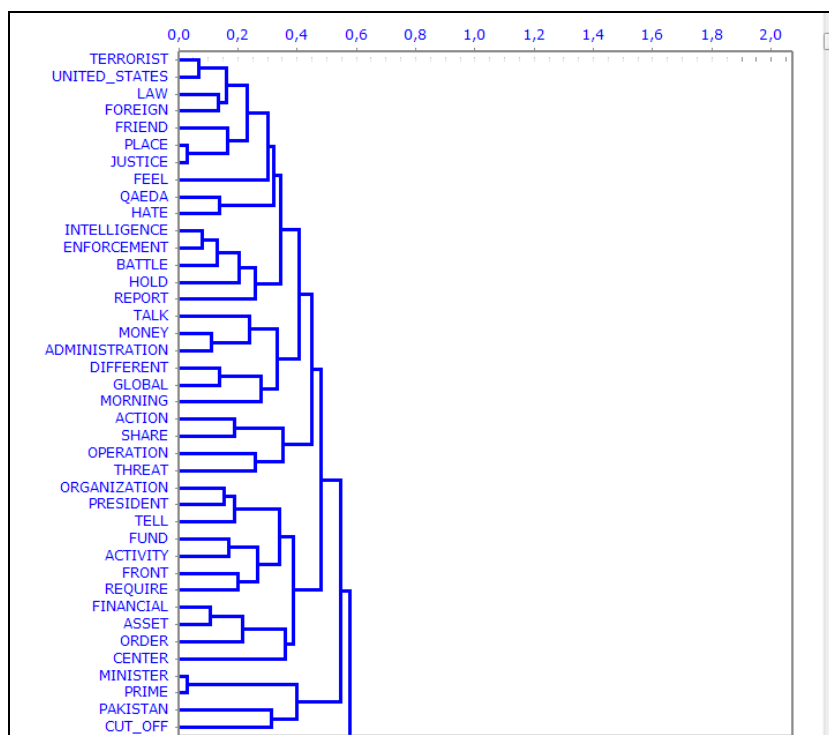
Correspondence Analysis allows us to explore similarities and differences between (and within) groups of context units (e.g. documents belonging to the same category).





Cluster Analysis, which can be carried out using various techniques, allows us to detect and explore groups of analysis units which have two complementary features: high internal (within cluster) homogeneity and high external (between cluster) heterogeneity.





C : TOOLS FOR THEMATIC ANALYSIS

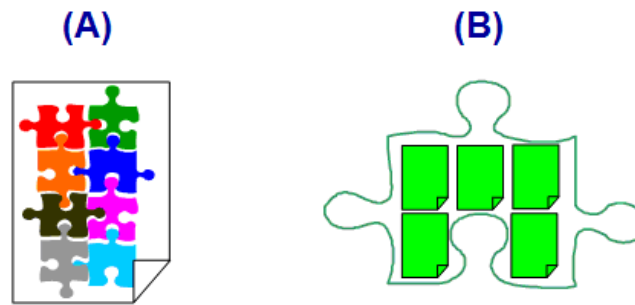
These tools enable us to discover, examine and map “themes” emerging from texts.

As **theme** is a polysemous word, when using software tools for thematic analysis we have to refer to operational definitions. More precisely, in these **T-LAB** tools, “theme” is a label used to indicate four different entities:

- 1- a **thematic cluster of contexts units** characterized by the same patterns of key-words (see the Thematic Analysis of Elementary Contexts, Thematic Document Classification and Dictionary-Based Classification tools);
- 2- a **thematic group of key terms** classified as belonging to the same category (see the Dictionary-Based Classification tool);
- 3 – a **mixture component** of a probabilistic model which represents each context unit (i.e. elementary context or document) as generated from a fixed number of topics or “themes” (see the Modeling of Emerging Themes and the Texts and Discourses treated ad Dynamic Systems tools).
- 4- a **specific key term** used for extracting a set of elementary contexts in which it is associated with a specific group of words pre-selected by the user (see the Key Contexts of Thematic Words tool).

For example, depending on the tool we are using, a single document can be analysed as composed of various ‘themes’ (see ‘A’ below) or as belonging to a set of documents concerning the same ‘theme’ (see ‘B’ below). In fact, in the case of ‘A’ each theme can correspond to a word or to a

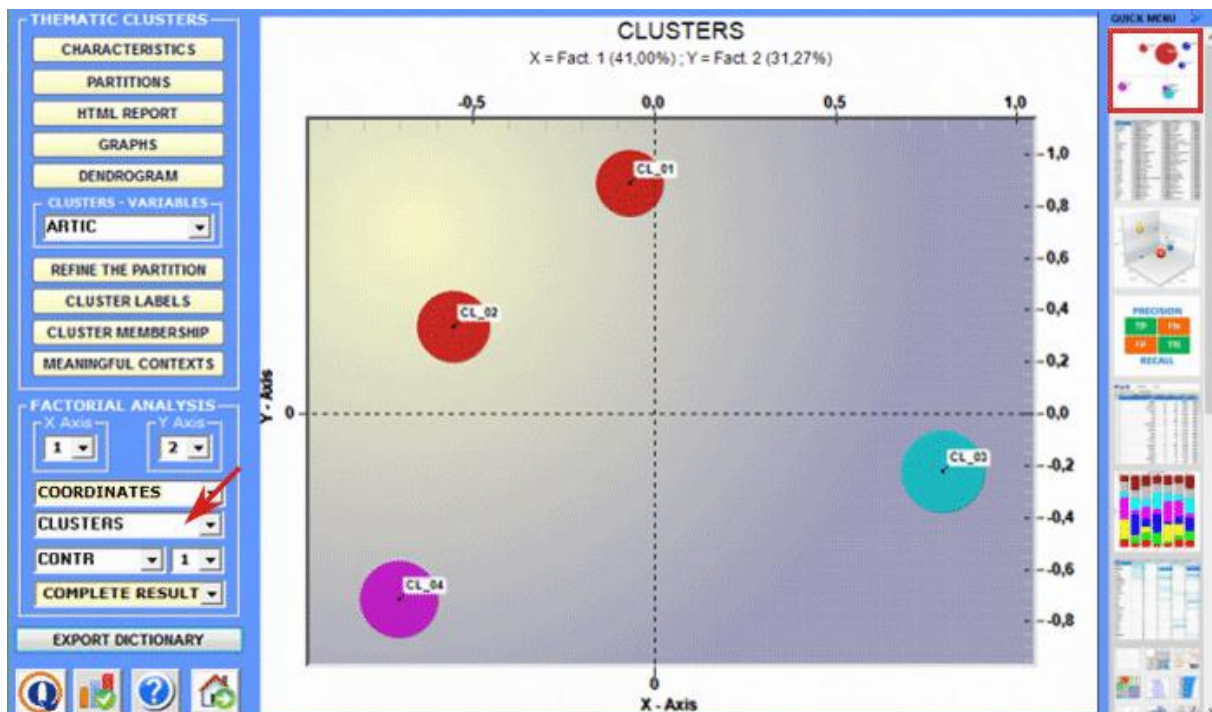
sentence, whereas in the case of 'B' a theme can be a label assigned to a cluster of documents characterized by the same patterns of key-words.

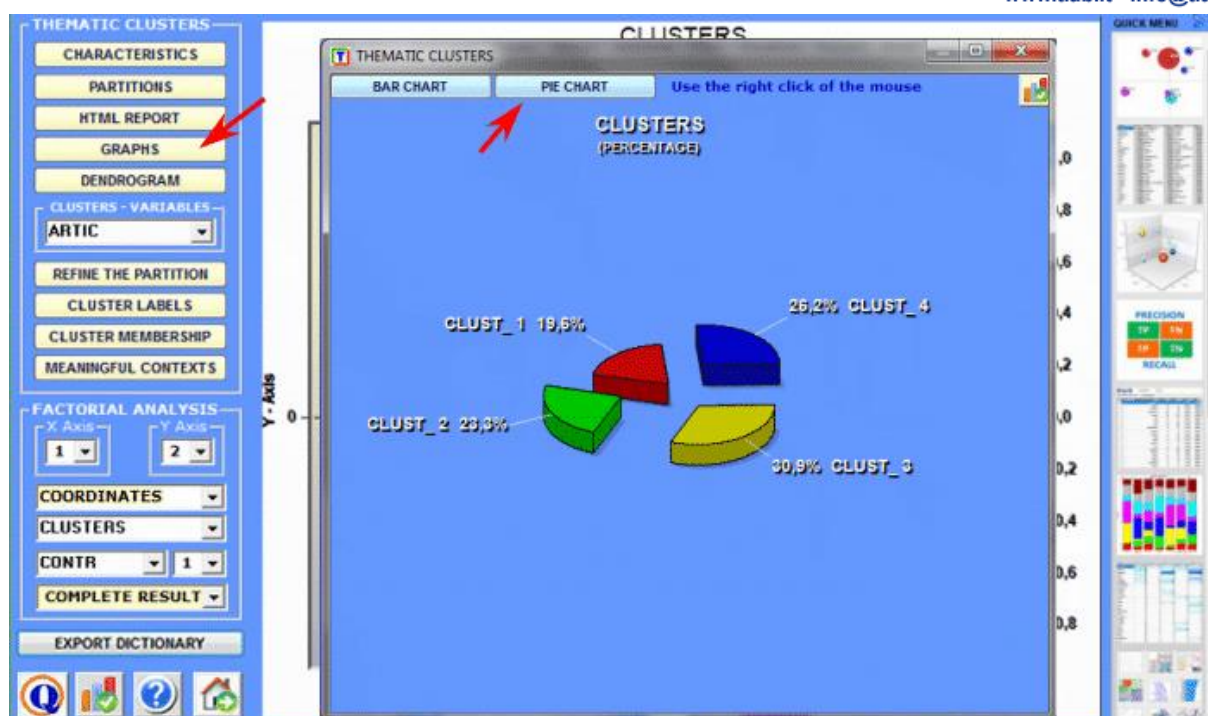


In detail, the ways how **T-LAB** 'extracts' themes are the following:

1 - both the **Thematic Analysis of Elementary Contexts** and the **Thematic Document Classification** tools, when performing an unsupervised clustering, work in the following way:

- a - perform **co-occurrence analysis** to identify thematic clusters of context units;
- b - perform **comparative analysis** of the profiles of the various clusters;
- c - generate various types of graphs and tables (see below);
- d - allow you to save the **new variables** (thematic clusters) for further analysis.





THEMATIC CLUSTERS

CHARACTERISTICS

PARTITIONS

HTML REPORT

GRAPHS

DENDROGRAM

CLUSTERS - VARIABLES

ARTIC

REFINE THE PARTITION

CLUSTER LABELS

CLUSTER MEMBERSHIP

MEANINGFUL CONTEXTS

FACTORIAL ANALYSIS

X Axis: 1 Y Axis: 2

COORDINATES

CLUSTERS

CONTR: 1

COMPLETE RESULT

EXPORT DICTIONARY

CLUSTERS

X = Fact. 1 (41,00%); Y = Fact. 2 (31,27%)

CLUSTER CHARACTERISTICS

CLUSTER N. 2 EC IN CLU = 159; EC IN TOT: 515 (30.87%)

CAT	LEMMAS & VARIABLES	IN CLU	IN TOT	CHI²	(p)
A	crop	102	122	117,886	0,000
A	herbicide	55	55	95,876	0,000
A	transgenic	61	64	95,783	0,000
A	plant	80	105	71,807	0,000
A	weed	31	31	53,876	0,000
A	resistant	26	26	45,158	0,000
A	virus	24	24	41,674	0,000
A	pest	21	21	36,451	0,000
A	Corn	19	19	32,971	0,000
A	relative	19	19	32,971	0,000
A	ecologist	17	17	29,493	0,000
A	Monsanto	22	25	28,533	0,000
A	create	38	53	28,297	0,000
A	seed	16	16	27,755	0,000
A	farmland	15	15	26,017	0,000
A	pollution	20	23	25,189	0,000
A	code	26	35	21,496	0,000
A	insert	26	35	21,496	0,000
A	chemical	22	28	21,314	0,000
A	company	44	70	20,956	0,000

2 - through the **Dictionary-Based Classification** tool we can easily build/test/apply models (e.g. dictionaries of categories or pre-existing manual categorizations) both for the classical qualitative content analysis and for the sentiment analysis. In fact such a tool allows us to perform an automated top-down classification of lexical units (i.e. words and lemmas) or context units (i.e. sentences, paragraphs and short documents) present in a text collection.

IMPORT YOUR DICTIONARY

RESET

<< AUTOMATIC LIST <<

RENAME CATEGORIES

EXECUTE CLASSIFICATION

HTML REPORT

EXPORT CLASSIFICATION

CONTINGENCY TABLES

DICTIONARY (MODEL)

DICTIONARY (CORPUS)

VARIABLES - CATEGORIES

MULTIPLE SELECTION

Yes ☐ No ☐

PLOT YOUR CHART

GRAPHS

CATEGORIES (PERC.)

PARTY

MDS MAP

CORRESPONDENCE ANALYSIS

EXPORT YOUR DICTIONARY

FURTHER T-LAB ANALYSES

DICTIONARY (CORPUS)	ACTIVE	AFFILI...	HOSTILE	NEGA...	PASSIVE	POSITI...
<input type="checkbox"/> ADVANCE	2	0	0	0	1	
<input type="checkbox"/> ADVENTURE	1	0	0	0	0	
<input checked="" type="checkbox"/> ADVERSARY	0	0	4	0	0	
<input type="checkbox"/> AFFAIR	0	1	0	0	0	
<input type="checkbox"/> AFFIRM	0	0	0	0	0	
<input type="checkbox"/> AFFORD	0	0	0	0	0	
<input type="checkbox"/> AGGRE						
<input type="checkbox"/> AID						
<input type="checkbox"/> AIM						
<input type="checkbox"/> AIR						
<input type="checkbox"/> ALLIAN						
<input type="checkbox"/> ALLOW						
<input type="checkbox"/> ALLY						
<input type="checkbox"/> ALMIGH						
<input type="checkbox"/> AMBITI						
<input type="checkbox"/> AMBITI						
<input type="checkbox"/> ANCIEN						
<input type="checkbox"/> ANSWER						
<input type="checkbox"/> APPEAL						
<input type="checkbox"/> ART						
<input type="checkbox"/> ASHAM						
<input type="checkbox"/> ASK						
<input type="checkbox"/> ASLEE						
<input type="checkbox"/> ASSIST						
<input type="checkbox"/> ASSUM						
<input type="checkbox"/> ASSUR						
<input type="checkbox"/> ASUND						
<input type="checkbox"/> ATTAIN						
<input type="checkbox"/> AWAIT						
<input type="checkbox"/> AWARE						

CATEGORY = < HOSTILE >
OCCURRENCES OF < ADVERSARY >

**** *PRES_REGAN1981 *PARTY_REP
as_for the enemies of freedom, those who are potential **adversaries**, they will be reminded that peace is the highest aspiration of the American people.

**** *PRES_REGAN1981 *PARTY_REP
It is a weapon our **adversaries** in today's world do not have.

**** *PRES_CLINTON1997 *PARTY_DEM
Instead, now we are building bonds with nations that once were our **adversaries**.

**** *PRES_OBAMA2009 *PARTY_DEM
Our health_care is too costly, our schools fail too many, and each day brings further evidence that the ways we use energy strengthen our **adversaries** and threaten our planet.

SELECT THE INPUT TYPE

☐ Import your DICTIONARY of Categories < filename.dictio >

☐ Type/Paste your TEXTS in the box below (One for each Category)

☒ Use a VARIABLE of your Corpus and its coded categories

LEARNING FROM LABELED DATA AND TEST (PRECISION / RECALL)

METHOD

☒ Naive Bayes

☐ Nearest Centroid Classifier

MODEL

☒ Categorical Variable

☐ Classified Document

SELECT A VARIABLE

RESET

<< AUTOMATIC LIST <<

RENAME CATEGORIES

EXECUTE CLASSIFICATION

HTML REPORT

EXPORT CLASSIFICATION

CONTINGENCY TABLES

DICTIONARY (MODEL)

DICTIONARY (CORPUS)

VARIABLES - CATEGORIES

MULTIPLE SELECTION

Yes ☐ No ☒

PLOT YOUR CHART

GRAPHS

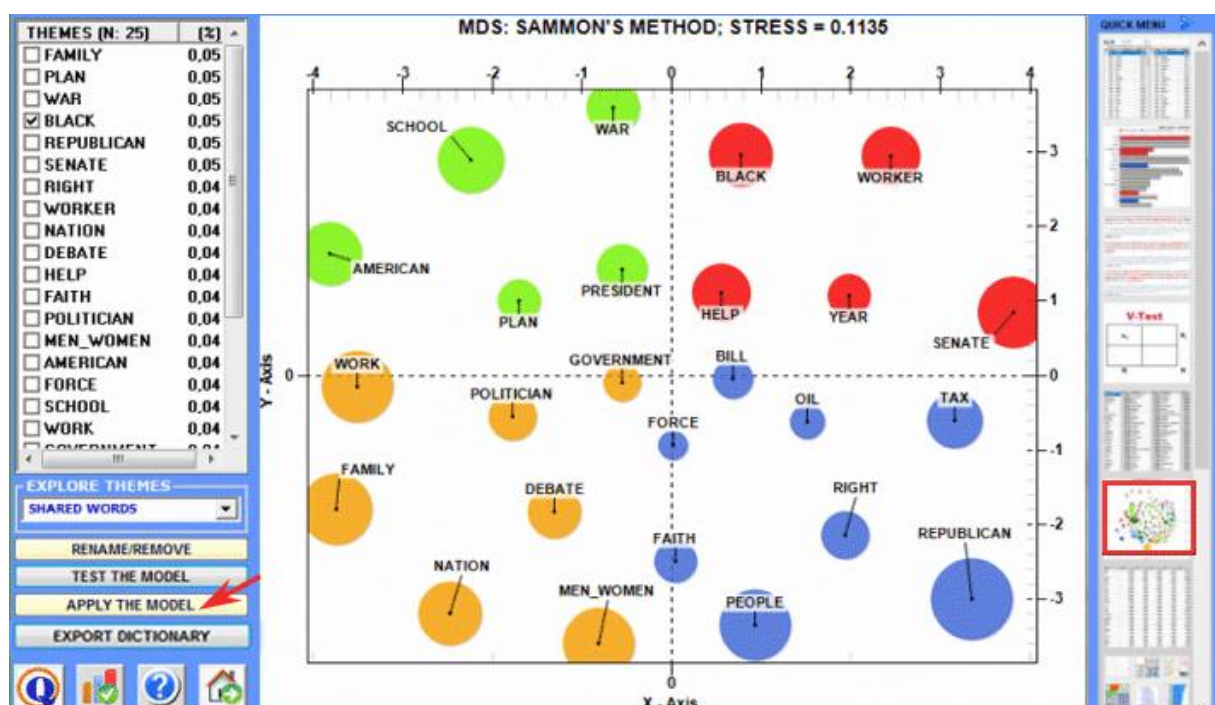
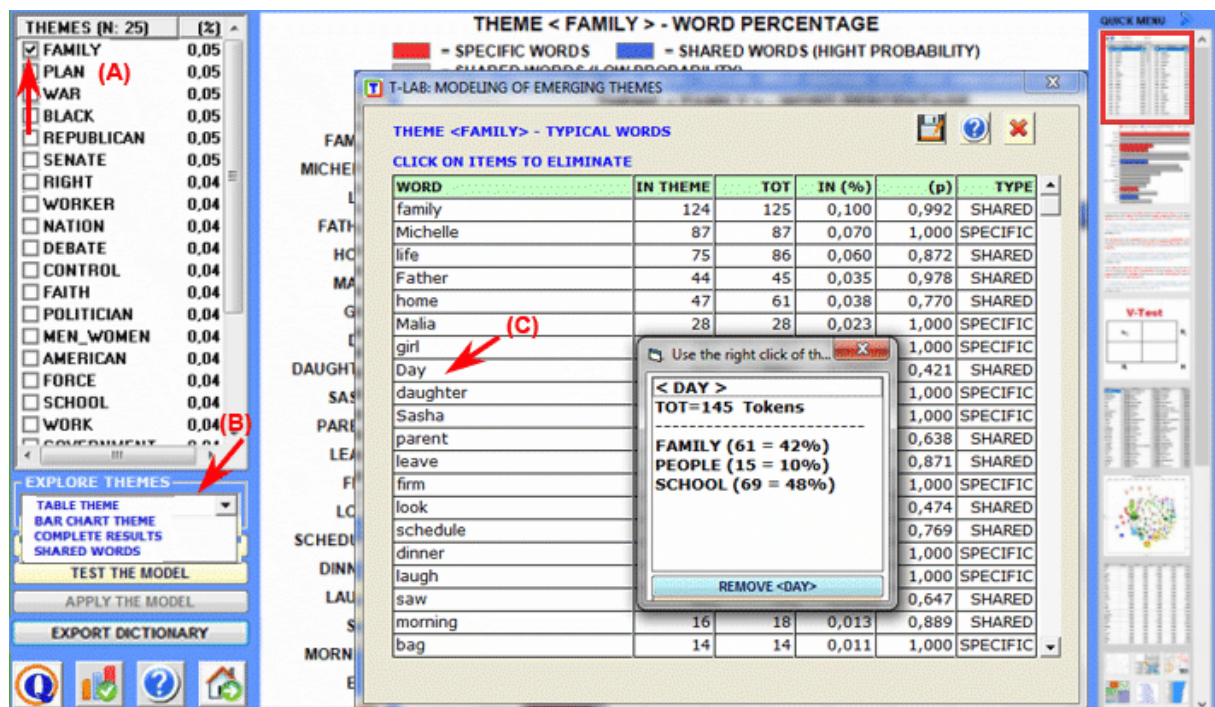
CATEGORIES (PERC.)

MDS MAP

COLUMNS=PREDICTED	TO_ALUM	TO_COCOA	TO_COFFEE	TO_CPI	TO_CRUDE	TO_GNP	TO_GOLD
TO_ALUM	50	0	0	0	0	0	0
TO_COCOA	0	61	0	0	0	0	0
TO_COFFEE	0	0	112	0	0	0	0
TO_CPI	0	0	0	70	0	0	0
TO_CRUDE	0	0	0	0	371	0	0
TO_GNP	0	0	0	0	0	74	0
TO_GOLD	0	0	0	0	0	0	89
TO_GRAIN	0	0	0	0	0	0	0
TO_INTEREST	0	0	0	0	0	0	0
TO_JOBS	0	0	0	0	0	0	0
TO_MONEYFX	0	0	0	0	0	0	0
TO_MONEYSUPPLY	0	0	0	0	0	0	0
TO_SHIP	0	0	0	0	0	0	0
TO_SUGAR	0	0	0	0	0	0	0
TO_TRADE	0	0	0	0	3	0	1

TEST

3 - through the **Modelling of Emerging Themes** tool (see below) the **mixture components** described through their characteristic vocabulary can be used for building a coding scheme for qualitative analysis and/or for the automatic classification of the context units (i.e. documents or elementary contexts).



4 - the **Key Contexts of Thematic Words** tool (see below) can be used for two different purposes:
(a) to extract lists of meaningful context units (i.e. elementary contexts) which allow us to deepen the thematic value of specific **key words**; (b) to extract context units which are the most similar to sample **texts** chosen by the user.

SELECT : CONTEXT(CORPUS, SUBSET) / INPUT TYPE (KEY-TERMS, TEXTS)

CONTEXT: **CORPUS** INPUT TYPE: **KEY-TERMS (A)** **TEXTS (B)**

AVAILABLE ITEMS: < 475 >

ITEM	OCC
<input type="checkbox"/> PEOPLE	177
<input type="checkbox"/> AMERICA	166
<input type="checkbox"/> AMERICAN	150
<input type="checkbox"/> NATION	134
<input type="checkbox"/> WORLD	125
<input type="checkbox"/> TERRORIST	114
<input type="checkbox"/> WORK	108
<input type="checkbox"/> COUNTRY	90
<input type="checkbox"/> GREAT	79
<input type="checkbox"/> KNOW	77
<input type="checkbox"/> WAR	77
<input type="checkbox"/> HELP	68
<input type="checkbox"/> FREEDOM	64
<input type="checkbox"/> TERRORISM	54
<input type="checkbox"/> UNDERSTAND	52
<input type="checkbox"/> TIME	51
<input type="checkbox"/> STRONG	48
<input type="checkbox"/> THANK	48
<input type="checkbox"/> NEED	46
<input type="checkbox"/> GOVERNMENT	46
<input type="checkbox"/> GOOD	45
<input type="checkbox"/> CHIDE	45

CLEAN OUT YOUR LIST

OUTPUT TABLES

NEW TEXT CLEAN OUT YOUR BOX
EXTRACT KEY CONTEXTS SAVE CONTEXTS

ADD/CUT WORDS FROM YOUR LIST (DOUBLE CLICK)

CONTEXT: **CORPUS** INPUT TYPE: **KEY-TERMS**

AVAILABLE ITEMS: < 475 >

ITEM	OCC
<input type="checkbox"/> MINORITY	5
<input type="checkbox"/> MISSION	21
<input type="checkbox"/> MISTAKE	14
<input type="checkbox"/> MOM	
<input type="checkbox"/> MOMENT	
<input type="checkbox"/> MONEY	
<input type="checkbox"/> MORNING	
<input type="checkbox"/> MOVE	
<input type="checkbox"/> MURDER	
<input checked="" type="checkbox"/> MUSLIM	
<input type="checkbox"/> NAME	
<input type="checkbox"/> NATION	
<input type="checkbox"/> NATIONAL	
<input type="checkbox"/> NECESSARY	
<input type="checkbox"/> NEED	
<input type="checkbox"/> NEIGHBOR	
<input type="checkbox"/> NETWORK	
<input type="checkbox"/> NEW_YORK	
<input type="checkbox"/> NIGHT	
<input type="checkbox"/> NUMBER	
<input type="checkbox"/> OCT	
<input type="checkbox"/> OFFICE	

LIST OF ASSOCIATED LEMMAS

American
Arab
campaign
Christian
contribution

KEY CONTEXTS SORTED BY WEIGHED DESCENDING ORDER

**** *SPEECH_SEP19
Cosine (.546)
as_a_matter_of_fact, Islam preaches peace. The Muslim faith is a peaceful faith. And there are millions of good Americans who practice the Muslim faith, who love their country as much as I love the country, who salute the flag as strongly as I salute the flag.

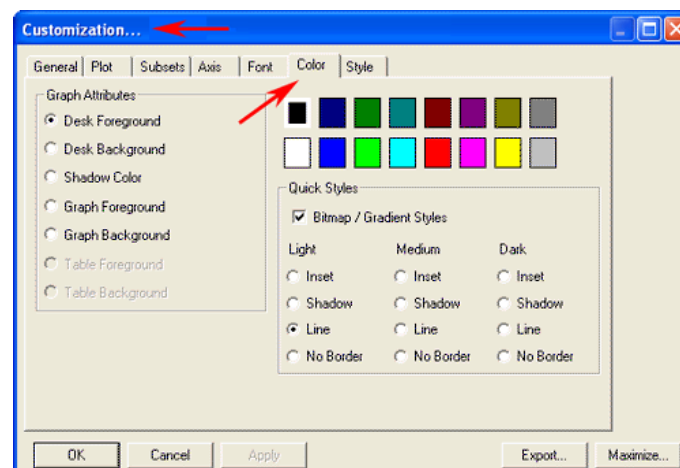
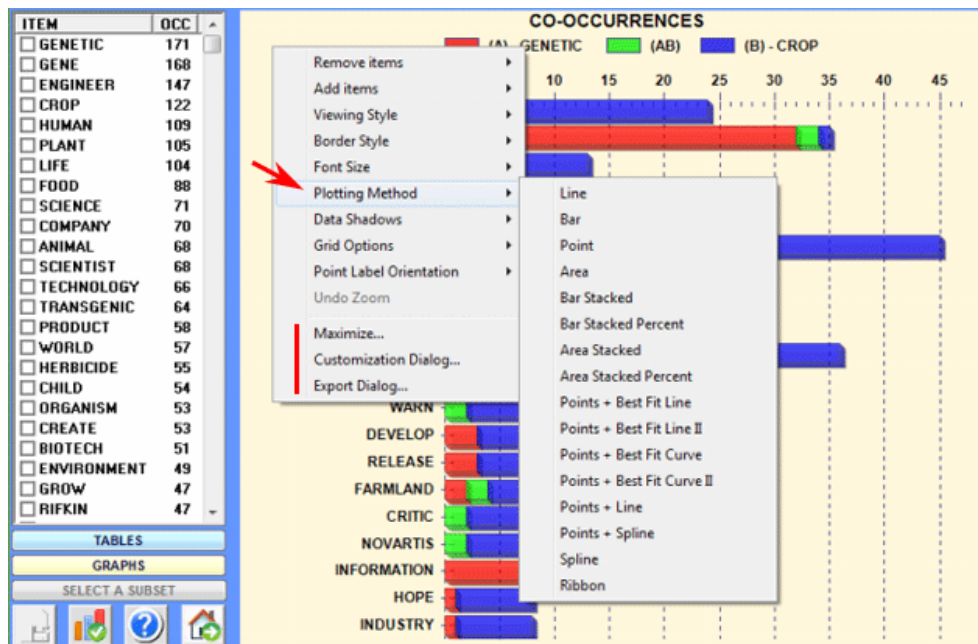
**** *SPEECH_SEP20
Cosine (.511)
I also want to speak tonight directly to Muslims throughout the world. We respect your faith. it's practiced freely by many millions of Americans and by millions more in countries that America counts as friends. Its teachings are good and peaceful, and those who commit evil in the name of Allah blaspheme the name of Allah.

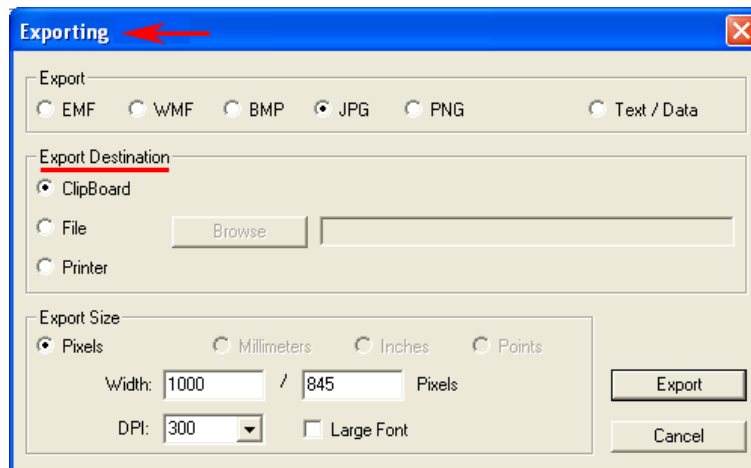
NEW TEXT CLEAN OUT YOUR BOX
EXTRACT KEY CONTEXTS SAVE CONTEXTS

6 - **INTERPRETATION OF THE OUTPUTS** consists in the consultation of the tables and the graphs produced by **T-LAB**, in the eventual customization of their format and in making inferences on the meaning of the relationships represented by the same.

In the case of **tables**, according to each case, **T-LAB** allows the user to export them in files with the following extensions: **.DAT**, **.TXT**, **.CSV**, **.XLXS**, **.HTML**. This means that, by using any text editor program and /or any Microsoft Office application, the user can easily import and re-elaborate them.

All **graphs and charts** can be zoomed (left-click and drag), maximized, customized and exported in different formats (right click to show popup menu)





Some general criteria for the interpretation of the **T-LAB** outputs are illustrated in a paper quoted in the **Bibliography** and are available from <https://www.tlab.it> website (Lancia F.: 2007). This document presents the hypothesis that the statistical elaboration outputs (tables and graphs) are particular types of texts, that is they are multi-semiotic objects characterized by the fact that the relationships between the signs and the symbols are ordered by measures that refer to specific **codes**.

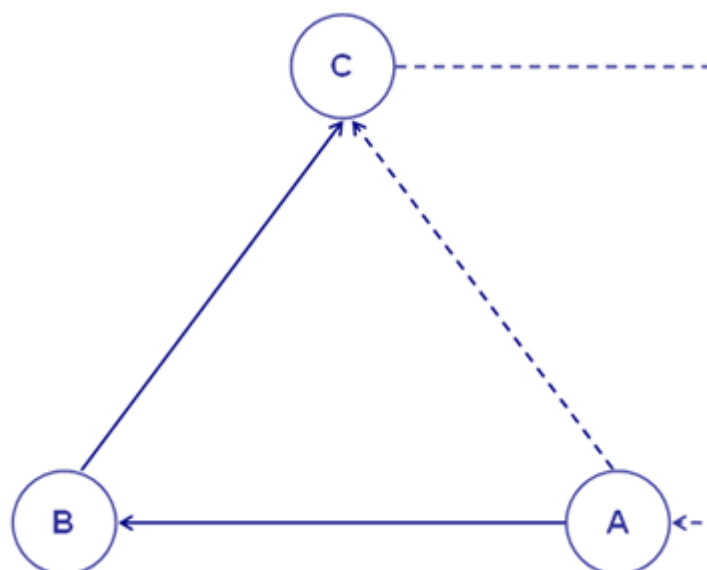
In other words, both in the case of texts written in "natural language" and those written in the "statistical language", the possibility of making inferences on the relationships that organize the **content forms** is guaranteed by the fact that the relationships between the **expression forms** are not random; in fact, in the first case (natural language) the significant units follow on and are ordered in a linear manner (one after the other in the chain of the discourse), while in the second case (tables and graphs) the organization of the multidimensional **semantic spaces** comes from statistical measures.

Even if the semantic spaces represented in the **T-LAB** maps are extremely varied, and each of them require specific interpretative procedures, we can theorize that - in general - the logic of the inferential process is the following:

A – to detect some significant relationships between the units "present" on the expression plan (e.g. between table and/or graph labels);

B – to explore and compare the semantic traits of the same units and the contexts to which they are mentally and culturally associated (content plan);

C – to generate some hypothesis or some analysis categories that, in the context defined by the corpus, give reason for the relationships between expression and content forms.



At present, **T-LAB** options have the following **restrictions**:

- corpus dimension: max 90Mb, equal to about 55,000 pages in .txt format;
- primary documents: max 30,000 (max 99,999 for short texts which do not exceed 2,000 characters each, e.g. responses to open-ended questions, Twitter messages, etc.);
- categorical variables: max 50, each allowing max 150 subsets (categories) which can be compared;
- modelling of emerging themes: max 5,000 lexical units (*) by 5,000,000 occurrences;
- thematic analysis of elementary contexts: max 300,000 rows (context units) by 5,000 columns (lexical units);
- thematic document classification: max 99,999 rows (i.e. documents) by 5,000 columns (lexical units);
- specificity analysis (lexical units x categories): max 10,000 rows by 150 columns;
- correspondence analysis (lexical units x categories): max 10,000 rows by 150 columns;
- correspondence analysis (context units x lexical units): max 10,000 rows by 5,000 columns;
- multiple correspondence analysis (elementary contexts x categories): max 150,000 rows by 250 columns;
- singular value decomposition: max 300,000 rows by 5,000 columns;
- cluster analysis that uses the results of a previous correspondence analysis (or SVD): max 10,000 rows (lexical units or elementary contexts);
- word associations, comparison between word pairs: max 5,000 lexical units;
- co-word analysis and concept mapping: max 5,000 lexical units;
- sequence analysis: max 5,000 lexical units (or categories) by 3,000,000 occurrences.

(*) In **T-LAB**, 'lexical units' are words, multi-words, lemmas and semantic categories. So, when the automatic lemmatization is applied, 5,000 lexical units correspond to about 12,000 words (i.e. raw forms).