## **T-LAB 10**

# **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.

T-LAB: TOOLS FO	R TEVT ANALVSK	- D X
	is Other tools License Help	
	AVAILABLE OPTIONS - MENU	
	START A NEW SESSION Select a T-LAB demo file	
CO-OCCUR	RENCE ANALYSIS  Import a single file (.txt, .doc, .docx, .pdf, .rtf)  Prepare/Import more files or tables (Corpus Builder)	
THEMATIC		
		Check / Change - Key terms list
		Seginner 🖲 Expert
		TNGLISH V Interface
		🤭 🔿 Silver 🛞 Blue 🔿 Dark
	Image: 10 User ID: MYTLAB20           C 2001-2023           OUR FILE WITHOUT VARIABLES	×
-		Browse
File:	Accepted file types: .TXT, .DOC, .DOCX, .PDF, .RTF (Other Formats: Copy/Paste your TEXT in the box b	
	N.B.: In the case of PDF files, make sure that they are not in an image-only format.	
(editable)	The pictures of airplanes flying into buildings, fires burning, huge structures collapsing have filled us with or sadness and a quiet, unyielding anger. These acts of mass murder were intended to frighten our nation into chaos and retreat. But they have fail strong. A great people has been moved to defend a great nation. Terrinist attacks can shake the foundations of our biggest buildings, but they cannot touch the foundation These acts shafter steel, but they cannot dent the steel of American resolve. America was targeted for attack because were the bhightest beacon for freedom and opportuntly in the ver Today, our nation saw evil, the very worst of human nature, and we responded with the best of America, our rescue workers, with the caring for strangers and neighbors who came to give blood and help in any immediately following the first attack. I implemented our government's emergency response plans. Our mil and it's prepared. Our emergency teams are working in New York City and Washington, D.C., to help with afforts. Our first priority is to get help to these who have been injured and to take every precaution to protect our and around the world from further attacks. The functions of our government continue without interruption. Federal agencies in Washington which his today are recogening for essential personnel toright and will be open for business tomorow. Our financial institutions remain strong, and the American economy will be open for business as well. The search is underway for those who are behind these evil acts. I've directed the full resources for our intelligence and law enforcement communities to find those respon to justoce. We will make no distinction between the terroints who committed these acts and those who he lappreciste so very much the members of Corgress with have joinder me is strongly conderming these a behalf of the American people, I thank the many wold leaders who have called to offer their condolence Corpus name without any extension MyConpus	ed. Our country is n of America. world. with the daring of way they could. tary is powerful, local rescue citizens at home ad to be evacuated sible and bring then stort hem. tacks. And on
Path :		Browse
	Select the language of your corpus	Cancel
ß	🛟 🕕 🕕 🤹 🧐 🛑 Choose a Language 👻	Save as
	ENG ITÀ FRA ESP POR DEU OTHER LANGUAGES	Import 4



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**.

T-LAB: TOOLS FOR TEXT ANALYSIS     File Lexical Tools Other tools License Help		- 0 X
START A NEW SESSION	AVAILABLE OPTIONS - MENU  Select a T-LAB demo file	
CO-OCCURRENCE ANALYSIS	Import a single file (.txt, .doc, .docx, .pdf, .rtf)	
THEMATIC ANALYSIS	<ul> <li>Prepare/Import more files or tables (Corpus Builder)</li> <li>Open an existing project (from folder)</li> </ul>	
COMPARATIVE ANALYSIS	Open an existing project from the < My Projects > list	
		Check / Change - Key terms list
		🧟 🔿 Beginner 🛞 Expert
		ENGLISH ~ Interface
		🤣 🔿 Silver 🛞 Blue 🔿 Dark
Copyright © 2001-2023	USER ID: MYTLAB20	Q @

	Open Table	Gather your Texts	Open Corpus Variable	e Manager 🤌 Text Screening 🎽	Gear al Oreck/Generate.	Analize	Save_	<b>U</b>
_	.CSV		T Open		and the second			
	SAV JSON	Add Id	IN CONTRACTOR					
	XML		_ 😋 🔵 💌 🎍 🕨 test		- + Search test	9	Variabl	
	XLS XLSX	Variable	Organize  New fold	der	· ·	0 0	DER	
	ACCDB, MDI	GENDER					AT	
-	00002	MALE	🔶 🚖 Favorites	Name	Date modified	Type		
	COUCH	MALE	M Desktop	🚳 example01.csv	12/01/2013 10:02	File con v		
	00004	FEMALE	Bownloads	example02.mdb	13/01/2013 12:18	Applicazie		
	00005	FEMALE	Recent Places	(in example03.xls)	13/01/2013 12:18	Foglio di L		
	00006	MALE		(2) example04.xlsx	13/01/2013 12:18	Foglio di I		
	00007	FEMALE	Libraries E	example05.accdb	13/01/2013 12:18	ACCDB F		
	00008	FEMALE	Documents					
	00009	FEMALE	A Music					
	00010	FEMALE	Pictures					
	00011	MALE	Videos					
	Ø 00012	FEMALE						
	00013	FEMALE	Homegroup					
	00014	MALE	romegroup					
-	1001L	EEMILE						
dhe	umber_00001 alth; happiness	GENDER_MALE "AGE_55TO	90 🐘 Computer 💌	•			ext	
		GENDER_MALE *AGE_30TO	File	name:	<ul> <li>Table Files ("ads, "adsx, "</li> </ul>	csv, *.r 🔻		No. of Concession, Name
-	ss in people an	und me, contented family, wou	blu		Open 💌 🕜	ancel		



	Open Corpus_	Variable Manager 👌 Clear a	Check/Generate	Save_ 🕢 H
DDING OPTIONS	File Name: mbers Responses bt		_	
	Variable Variable	• My Texts •	Variab	los
cord Sel IDnumber GENDER	A T Open	error manifer tige:	GENDER	
T Add Records	🕤 🚱 🖉 🔹 🗼 test		✓ 4y Search test	Q
	Organize - New fold	der	<u> </u> = •	
T-LAB"	Favorites	Name	Date modified	Туре
L.	5 Desktop	file01.TXT	05/03/2002 22:12	Text D
	Downloads	file02.TXT	05/03/2002 22:12	Text D
Add your Documents	Recent Places	file03.TXT	05/03/2002 22:12	Text D
	5	file04.DOC	05/03/2002 22:12	Docun =
Add Empty Record	🙀 🤤 Libraries 🛛 🗉	file05.DOC	05/03/2002 22:12	Docun
	Documents	file06.RTF	05/03/2002 22:12	RTF (R
import text from URL: (http://)	Music	file07.RTF	05/03/2002 22:12	RTF (R
3	E Pictures	file08.DOCX	05/03/2002 22:12	Docun
	Videos	file09.DOCX	05/03/2002 22:12	Docun
5		file10.DOCX	05/03/2002 22:12	Docun
	Homegroup	i file11.TXT	05/03/2002 22:12	Text D
	-	file12.TXT	05/03/2002 22:12	Text D 📼
Oose	Computer *	· • •		

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

File Lexical Tools Other tools License He	4p	
START A NEW SESSION CO-OCCURRENCE ANALYSIS THEMATIC ANALYSIS COMPARATIVE ANALYSIS	AVAILABLE OPTIONS - MENU Select a T-LAB demo file Import a single file (.txtdoc, .docxpdf, .rtf) Prepare/Import more files or tables (Corpus Builder) Open an existing project (from folder) Open an existing project from the < My Projects > list	Check / Change - Key terms list Check / Check / Change - Key terms list Check / Check / Check / Change - Key terms list Check / Check / Check / Change - Key terms list Check / Check /



#### 2 - Select any corpus to analyse

T-LAB: TOOLS FOR TEXT ANALYSIS						-		
ile Lexical Tools Other tools License Hel	p							
	A	AILABI	LEOPTION	S - MENU				
		e		4				
START A NEW SESSION		Selec	t a T-LAB	demo file				
	ala	ONAN	ITEM = SELI	CT THE FILE TO	BE INPORTED			
CO-OCCURRENCE ANALYSIS								
		1	ANGUAGE	<ul> <li>NAME</li> </ul>	DIMENSION	DESCRIPTION	1	
THEMATIC ANALYSIS		<b>.</b>	HINESE	中国会议代表	100 10	中華人民共和國主法研境部沒有的中國信討性總管化的政策并行出		
		<b>III</b> (	HINESE	论语二十篇全	81 Kb	"《论语》二十篇全文"语科库由广东外语外较大学王晓雯副教授整理完		
		<b>III</b> (	HINESE	新冠肺炎溃疡	11233 Kb	这个数据包括10.000条随机抽取的跟新冠肺炎相关的液体。数据自复!		
COMPARATIVE ANALYSIS		- 0	EUTSCH	Bundeskanzler.txt	661 Kb	Neun Eröffnungsreden deutscher Bundeskanzler: Kohl, Schröder und Merke		
		- 0	EUTSCH	Freud1909.txt	105 Kb	Fünf Vorträge über Psychoanalyse von Sigmund Freud an der Clark Universi		
		- 0	EUTSCH	Zivilcourage bit	74 Kb	Sammlung von achtzehn Zeitungsartikeln über einen realen Fall (September	list	
		90 E	NGLISH	Biotech.bt	100	24 shot atticles concerning the biotech and genetically modified organisms,		
		1990 E	NGLISH	Bush_Sept11.bt	148 Hb	Bush's speeches, delivered between September 11th and October 7th 2001	nt	
		88 E	NGLISH	Debate txt	101 Kb	U.S. Presidential debate between H. Clinton and D. Trump - October 2016	"	
		1983 E	NGLISH	Responses bit	146 Kb	1043 responses to the open-ended question What things are very important		
		200 E	SPAÑOL	argentina bit	127 Kb	15 artículos de opinión publicados por El Mundo que analizan la orisis econv		
		222 E	SPAÑOL	cadiz bit	103 Kb	constitución política de la monarquía española de 1812		
		20 E	SPAÑOL	respuestas bd	71 Kb	308 respuestas a la pregunta abierta ¿Qué significa para usted la expresión		
		11 F	RANÇAIS	DévelDurabitxt	204 Kb	sept articles concernant le développement durable	Dark	ĸ
		11 F	RANÇAIS	Palestine bit	137 Kb	dix articles concernant le conflit israélo palestinien		-
		88 F	RANÇAIS	Résponses_fr.bt	144 Kb	1008 résponses à la question ouverte 'Quelle est la chose la plus importante		
10 Version 10.2.2		88 I	TALIANO	governitot	227 Kb	dichiarazioni programmatiche di cinque governi italiani: Berlusconi 1 (maggio	-	
							14	

#### **3** - Click "ok" in the first Setup window

CORPUS	
NAME :	Bush_Sept11.txt
DIMENSION :	151 Kb
PATH :	C:\Users\\Documents\T-LAB PLUS\Demo_en\
TEXTS:	22 PRIMARY DOCUMENTS
VARIABLES :	1
IDNUMBERS :	Absent
LANGUAGE:	< ENGLISH >
	AUTOMATIC LEMMATIZATION 🛶 Yes 💿 No 📿
For more information c	lick on the (?) button



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

T-LAB: TOOLS FOR TEXT ANALYSS  File Lexical Tools Other tools License Help  START A NEW SESSION  CO-OCCURRENCE ANALYSIS  TUELLATIC ANALYSIS	ANALYSIS TOOLS Word Associations Co-Word Analysis Comparison between Word Pairs	x
COMPARATIVE ANALYSIS     COMPARATIVE ANALYSIS	Sequences and Network Analysis Concordances Co-occurrence Toolkit	Check / Change - Key terms list (N = 125)
Copyright © 2001-2023	USER ID: MYTLAB20	Interface       Image: Silver       Image: Silver

#### 5 - Check the results







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





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 s**election.

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.



	without any e	xtension					
MyCorpus							
Select the la	anguage of yo	ur corpus (ava	ailable langua	ges: 36)			
						0	
ৰ চ			۲	9	$\mathbf{e}$	Choose a Language	•
ENG		FRA	ESP	POR	DEU	Choose a Language	-
0	0	0	0	0	0	Arabic	
	Cano	sel				Bengali	
						Bulgarian	
						Catalan	
						Croatian	_
						Chinese	

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

CORPUS	
NAME: Bush_Sept	11.txt
DIMENSION: 151 Kb	
PATH: C:\Users\\D	ocuments\T-LAB PLUS\Demo_en\
TEXTS: 22 PRIMAR	YDOCUMENTS
VARIABLES : 1	
IDNUMBERS: Absent	
LANGUAGE: < ENGLISH >	
For more information click on the (? SHOW MORE OPTIONS	· · · · · · · · · · · · · · · · · · ·
AUTOMATIC LEMMATIZATION	STOP-WORD CHECK
>> ENGLISH Yes ()	Basic 🖲
No O	O No Advanced O
TEXT SEGMENTATION (ELEMENTARY CONTEXTS)	MULTI-WORD CHECK
Sentences O	No O
Chunks 🖲	Basic 🖲
Paragraphs ()	Advanced O
KEY-TERM SELECTION (IMPORTANCE	ORDER)
METHOD: O TF-IDF	AUTOMATIC LIST (MAX ITEMS)
CHI SQUARE	3000 ≑
O OCCURRENCES	WITH OCCURRENCE VALUE > = 4
OPTIONS FOR SOCIAL MEDIA DATA	
	t '#' from words (e.g. '#art' = '# art')   tags as they are (e.g. '#art' = '#art')
	ingla as mol are folly wait - wait )

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.



**3** - **THE USE OF LEXICAL TOOLS** allows us to verify the correct **recognition** of the lexical units and to customize their **classification**, that is to verify and to modify the automatic choices made by **T-LAB**.

The procedures of the various interventions are illustrated in the corresponding help sections (and in the manual).

In particular the user is requested to refer to the corresponding help section for a detailed description of the **Dictionary Building** process (see below). In fact any change concerning the dictionary entries affects both the occurrence and the co-occurrence computation.

(1) -		Y (STOP-WORD EXCLUDED)> RENAME - G			
	KEY WORDS SELECTION	TIONARY BUILDING	RY		LOOK UP ALPHABETICALLY
Automatic	WORD	ІТЕМ	OCC	INF ^	word
O Customized	adamantly	ADAMANTLY	1	LEM	
	adapt	ADAPT	1	LEM	RENAME AND GROUP
KEYWORDS All checked	adapted	ADAPT	7	LEM	Rename / Group
327	addressed	ADDRESS	1	OMO	ELEMENTS advisers ADVISER
All listed	Adelaide	ADELAIDE	1	LEM	advisers ADVISER advising ADVISE
334 💽	adequately	ADEQUATELY	1	LEM	Key-Word-in-Context
Checked only	administration	ADMINISTRATION	8	LEM	Remove selected Item
327	admits	ADMIT	1	LEM	Remove all Items
	adolescents	ADOLESCENT	1	LEM	
CHECK ALL	adopt	ADOPT	2	LEM	
UNCHECK ALL	adoption	ADOPTION	1	LEM	LABEL (RENAME)
8 8 🕀	adult	ADULT	1	LEM	ADVISE 🔗
	adults	ADULT	7	LEM	
HANGE THRESHOLD	advance	ADVANCE	4	DIS	IMPORT A DICTIONARY
ANAGEMENT	advancing	ADVANCE	1	LEM	DISCARDED LEMMAS
MPORT YOUR LIST	advantage	ADVANTAGE	2	LEM	
FILING	advantages	ADVANTAGE	2	LEM	
PILING	advising	ADVISE	1	LEM	
	advisers	ADVISE	1	LEM	
	advocates	ADVOC Selected Item = ADVISERS	7	OMO	
	affect	AFFEC Remove from list	12	OMO	
us	affiliations	AFFILI Rename as	4	LEM	
rs in the second se	affluence	AFFLUI Group with	2	LEM	
	afford	AFFOR Key-Word-in-Context	2	LEM	
IENTARY CONTEXTS	age	AGE	18	DIS	
· (4)	ageing	AGE	1	LEM	
	agencies	AGENCY	1	LEM	
MERTIZED CORPUS	agency	AGENCY	1	LEM	
Save the table as xis file	agenda	AGENDA	3	LEM	
	age-old	AGE-OLD	2	NCL	
Save the table as .csv file	agree	AGREE	7	LEM	
Save as Dictio.diz file	agribusiness	AGRIBUSINESS	2	LEM	
	agricultural	AGRICULTURAL	18	LEM	
	agriculture	AGRICULTURE	20	LEM	
	agrochemical	AGROCHEMICAL	4	NCL	

N.B.: When the user, without losing any lexical information, intends to apply coding schemes which group words or lemmas in a few categories (i.e. from 2 to 50) it is advisable to work with the **Dictionary-Based Classification** tool included in the **Thematic Analysis** sub-menu (see below).

SELECT THE INPUT TYPE			1		
	ONARY of Categories < filename.dictio >				
	EXTS in the box below (One for each Categ	ory)			
Use a VARIABLE of	your Corpus and its coded categories				
IMPORT YOUR DICTIONARY	DICTIONARY (MODEL)	ACTIVE	AFFILI	HOSTILE	NEGA
RESET	ABILITY	0	0	0	0
< AUTOMATIC LIST <<	ABLE	0	0	0	0
CANTOMATIC LIST SC	ABSENCE	0	0	0	0
RENAME CATEGORIES	ABUSE	1	0	1	1
EXECUTE CLASSIFICATION		0	0	0	0
HTML REPORT		0	0	0	1
		!	0	0	0
EXPORT CLASSIFICATION		!	0	0	0
CONTINGENCY TABLES			0	0	0
DICTIONARY (MODEL)			0	0	0
DICTIONARY (CORPUS)			0	0	0
			0	0	0
VARIABLES - CATEGORIES			ñ	n	0
MULTIPLE SELECTION	AFFAIB	1	1	ň	0 0
Yes 💭 💿 No	AFFORD	. 0	. 0	Ő	ů 0
PLOT YOUR CHART		1	0	0	1
[]		1	1	0	0
GRAPHS		1	0	0	0
CATEGORIES (PERC.)	ALLY	0	2	0	0
PARTY +	AMBITION	0	0	0	0
MDS MAP	ANSWER	1	0	0	0
MUS MAP	APPOINT	1	0	0	0
CORRESPONDENCE ANALYSIS	APPRECIATE	0	1	0	0
	ARGUMENT	1	0	1	1
EXPORT YOUR DICTIONARY	ARISE	1	0	0	0



**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.

APPLY AUTOMATIC	ITEM = TAGGED KEY	WORDS (LEMMAS OR CATEGORIES) - SELECT - RENAME - GROUP	
YOUR PROFILE	KEY WORDS SELEC	TION 🐻 DICTIONARY BUILDING 📓 CORPUS VOCABULARY	-
		ITEM	
Q O Customized	V	GENETIC	171 TEMS
	V	GENE Right Click	168 RENAME AND GROUP
All checked	▶ 🗹	ENGINEER Selected Item =ENGINEER	147 E Rename Group
327 💿	13	GENETICALLY Remove from list	122 ELEMENTS engineered (109)
	<b>V</b>	CROP Rename as	122 engineering (23)
B34 🕥		HUMAN Group with	109 engineer (13) engineers (2)
Checked only		LIFE Key-Word-in-Context	104
227	V	PLANT	101
		FOOD	88
CHECK ALL	V	SCIENCE	71
UNCHECK ALL		COMPANY	70 LABEL (RENAME)
8 8 😌	V	ANIMAL	68 ENGINEER 📝
CHANGE THRESHOLD		SCIENTIST	68
		TECHNOLOGY	66 IMPORT A DICTIONARY
		TRANSGENIC	
IMPORT YOUR LIST		PRODUCT	58
FILING	V	HERBICIDE	57
RESTORE	V	CHILD	55
	V	CREATE	53
		ORGANISM	53
		BIOTECH	51
TEXTS		ENVIRONMENT	49
24	v	RIFKIN	47
ELEMENTARY CONTEXTS		INTRODUCE	46
520		CENTURY	46
	<b>V</b>	YEAR	46
LEMMATIZED CORPUS		MARKET	42
4 🕜 🏠		RISK	42
		HEALTH	41
		GROW	40
		SPECIE	40
		CONSUMER	39
	<b>V</b>	PEOPLE	36
		INSERT	35
		CODE	35
		INDUSTRY	34
		ISSUE	34
	V	INCREASE	34
		GOOD PATENT	34
			34 -

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.



LEMMA (A)	^ 330				TERROR	RIST			www.energe		GUICK MENU	
PEOPLE	177 📃		(ASSOCIATIONS)									
AMERICA	166		T LEMMA (A) = < TERRORIST >							4		
AMERICAN	150											
□ NATION	134			Click and double click column headings to sort.								
WORLD	125			Key: EC = elementary contexts								
✓ TERRORIST	114			other values: EC_A = 86; TOT EC = 734								
U WORK	108			Click on a item of the table> HTML OUTPUT (EC_AB = co-occurrences)								
COUNTRY	90			LEMMA (B)	COEFF	EC B	EC_AB	CHI3	(p) +		7.0-	
GREAT	79	E	NF	harbour	0,329	13	11	67,99	0,000		-8502a -	
L KNOW	77	4		justice	0,265	20	11	37,24	0,000	1	Sec. at	
WAR	77			activity	0,260	11	8	40,19	0,000		"per	
HELP	68			organization	0,259	25	12	32,94	0.000		W-Test	
FREEDOM	64	ſ	**** *BUSH SEP24									
	54											
	52		As I told the American people, we will'direct every resource at our command to win the								• •	
	51		war against terrorists, every means of diplomacy, every tool of intelligence, every									
	48		instrument of law enforcement, every financial influence. We will'starve the terrorists of								10 <b>-</b>	
THANK	48		funding, turn them against each other, root them out of their safe hiding places and bring								41 -	
	46		them	to justice.						-	****	
	46 -										A CONTRACTOR	
INDICES				*BUSH_SEP24								
COSINE	-			ave established a foreign te							A CONTRACTOR	
- CRAPHS				ury to identify and investig							and the second second	
RADIAL DIAGRA				ist networks. It will'bring to							122 2 2	
TRADIAL DIAGRA			enfore	cement and financial regulat	tory agencies to	accompli	sh two go:	als:				
TABLES	-	-		-			-				2 10 22	
SELECT A SUE	PET		****	*BUSH_SEP24							an and the second	
SELECT A SUB	JSE I		we re	working closely with the U	nited Nations, t	he EU and	l through t	he G-7/G	-8 structure		A CONTRACTOR OF THE OWNER	
			to lim	it the ability of terrorist or	ganizations to ta	ake advant	age of the	internatio	onal		approximation of	
is interview of the united States has signed, but not yet ratified, two international								and the second				
										~	and the second	

- 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 keyterms) 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.



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).



63 >         SELECTED ITURES           ASYLUM         ASYLUM           62         ASYLUM-SEE           95         BOAT           87         BORDER           33         BRITIAIN           43         BRITISH           11         CALAIS           75         CALL           71         CAMP           48         CHANGE           45         CHILD           33         CLAIM           17         COMMUNITY	EKERS					MDS
CC         ASYLUM-SEE           62         BACK           95         BOAT           87         BORDER           33         BRITAIN           43         BRITISH           11         CALAIS           75         CALL           71         CAMERON           59         CAMP           48         CHANGE           45         CHILD           33         CITY           18         CLAIM           17         COMMUNITY						MDS
62         BACK           95         BOAT           87         BORDER           33         BRITIAIN           43         BRITISH           11         CALAIS           75         CALL           71         CAMPRON           59         CAMP           48         CHANGE           45         CHILD           33         CITY           18         CLAIM           17         COMMUNITY						MDS
95         BOAT           87         BORDER           33         BRITIAIN           43         BRITISH           11         CALAIS           75         CALL           71         CAMERON           59         CAMP           48         CHAINGE           45         CHILD           33         CITY           18         CLAIM           17         COMMUNITY						MDS
87         BORDER           33         BRITAIN           43         BRITSH           11         CALAIS           75         CALL           71         CAMERON           59         CAMP           48         CHANGE           45         CHILD           33         CITY           18         CLAIM           17         COMMUNITY						MDS
33         BRITAIN           43         BRITISH           43         BRITISH           11         CALAIS           75         CALL           71         CAMERON           59         CAMP           48         CHANGE           45         CHILD           33         CITY           18         CLAIM           17         COMMUNITY						MDS
43         BRITISH           11         CALAIS           75         CALL           71         CAMERON           59         CAMP           48         CHANGE           45         CHILD           33         CITY           16         CLAIM           17         COMMUNITY						MDS
11         CALAIS           75         CALL           71         CAMERON           59         CAMP           48         CHANGE           45         CHILD           33         CITY           18         CLAIM           12         COMMUNITY						MDS
75 CALL 71 CAMERON 59 CAMP 48 CHANGE 45 CHILD 33 CITY 18 CLAIM 17 COMMUNITY						MDS
59 CAMP 48 CHANGE 45 CHILD 33 CITY 18 CLAIM 17 COMMUNITY		Pier Root No. 1 and	E		1000000 1000000 100000	
48 CHANGE 45 CHILD 33 CITY 18 CLAIM 17 COMMUNITY		1966 <b>- 1966 - 1</b> 63	C'NO' THEN			.2
45 CHILD 33 CITY 18 CLAIM 17 COMMUNITY		1766 <b>2567 36</b> 63	ener meno	10 17 10 10 1000 0000 DD	- ******	. 2
CITY CLAIM CLAIM COMMUNITY		1764 A1001 304 H.S	E.M. BIENS	10 12 14 15 10000 000 000		2
18 CLAIM COMMUNITY				and the second second	March C	
17 COMMUNITY				1.00 mm	44()(	and a
		SERVE MALOR		1998 E.h	CONCIMIN	1
		eatica est			PALES	8
16 COUNTRY		2011 A.K.S	ria i		NOVEM NOVEM NO	8
79				·····:::::::::::::::::::::::::::::::::	MCOR CRUMES MUNICIPAL	1
					1011A0 A.C.M F.C.M	1
		and the second second			COMP//TR HEID	5
		10/4		1000	INSIDE TO ON	EAT CO
r vnut	~		- 84	1	6000	5
				Allo -	and a back	
	T YOUR LIST			10 million (10 mil	Store C	15
13 RESTORE YO	UR FIRST LIST	and the second sec	ALT: Coor		545	20
		100 1000		Con La Sala	SMES	1 Sector
05 SELECT	ALL HUNS		S/II Z		2000	1001
05 101 102	¥ 🕗		. and	·	RECOMMEND	1 281
as 🖌 🔛 🛀						
	79 CRISIS 66 EU 54 EUROPE 43 EUROPEAN 39 FAMILY 36 FORCE 33 GERMANY 18 17 CLEAN OL 50 S 50 S	79 CRISIS 66 DAY 54 EU 54 EUROPE 43 EUROPEAN 39 FAMILY 36 FORCE 33 GERMANY ~ 18 17 CLEAN OUT YOUR LIST 13 RESTORE YOUR FIRST LIST 05 SLLECT ALL ITEMS 05 05 CLEAN OUT YOUR FIRST LIST 05 CLEAN OUT YOUR FIRST LIST 10 CLEAN OUT YOUR FIRST FIR	79 CRISIS 66 EU 54 EUROPE 43 EUROPEAN 79 FAMILY 18 CLEAN OUT YOUR LIST 17 CLEAN OUT YOUR LIST 17 SELECT ALL ITEMS 05 SELECT ALL ITEMS 05 ALL ITEMS 17 SELECT ALL ITEMS 18 CLEAN OUT YOUR LIST 19 CLEAN OUT YOUR LIST 10 CLEAN OUT YOUR LI	79     CRISIS       66     DAY       66     EU       54     EUROPE       39     FAMILY       36     FORCE       33     GERMANY       18     CLEAN OUT YOUR LIST       17     RESTORE YOUR FIRST LIST       07     SELECT ALL ITEMS       05     Image: Construction of the second s	79     CRISIS       66     EU       54     EUROPE       39     FAMILY       36     FORCE       33     GERMANY       18     CLEAN OUT YOUR LIST       17     RESTORE YOUR FIRST LIST       07     SELECT ALL TIPUS       05     Image: Construction of the second	79     CRISIS       66     EU       54     EUROPE       39     FAMILY       76     FORCE       33     GERMANY       18     CLLAN OUT YOUR LIST       18     EUCOUR FIRST LIST       19     SELICT ALL ITIDIS       05     SELICT ALL ITIDIS       05     SELICT ALL ITIDIS       06     SELICT ALL ITIDIS       07     SELICT ALL ITIDIS       08     Second

#### **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).



ARIABLE	ITEM			CON		FAI	гн	FAMILY	OPP	P/	BT	POL		
CHAPTER V		M	_	COL	1	1 /41	3	0	011		4	1.01	1	-6
ITENS (N= 1052)								-	-					
( LEMMAS	ABIDE				1		1	0	0		U		0	
O WORDS	ABILITY	,			0		4	2	10		1		2	
<u> </u>	ABLE	T-LAB: SPECIFICITY A	NALVS	2									X	-
MEASURE		T DID DI CONTENTITI		- and the second second	and the second second	and the second	(All of the local data	No. of Concession, Name		and the	in a state of the later	and the second	Statute Server	
CHI SQUARE		CLICK ON ITEMS	TO PLO	OT CH	ARTS								X	
O TEST VALUE	ABR										-			
	ABSI	TYPICAL WORDS		-	Compa	re a su	ib-se	t with the	whole cor	pus				
DATA TABLE														
OCCURRENCES	ABS	TYPICAL (+) OF <	_	_	A CONTRACTOR OF				(-) OF <_(					
MEASURES	ABS	LEMMA	SUB	TOT	CHI3	(p)	1	LEMMA		SUB		CHI2	(p) -	
OMPARISON	ABU	tax	56		112,63			Black	Sector Sector	5	171	27,50		
PART - WHOLE		economy	56		103,88			Senate		3	138	24,37		
		worker	65		102,14			politics		1	98	19,83		
BETWEEN SUB-SETS		company	35	49	91,58			right		12	181	16,87		
		health-care	29 37	38	84,70			mother	010000000	1	73	14,17		
SELECT A		market	22		64,10			woman		5		13,95		
		Research	19	29	63,62			religious		14	71	13,72		
CH_CONSTIT		engineer competitive	17	20	58,93			Michelle		3	87	13,12		
CH_FAITH		wage	27	44		0,000		White		6	112	12,79		
CH_FAM		Teachers	21	30	53,95			War		8	128	12,79		
CH_OPPORTU		education	29	50	52,08			military		1	61		0,001	
CH_POLITICS		saving	13	14	51,56			Political		5	95	11,00	and the owner of the local division of the l	
CH RACE		insurance	22	33	51,06			life		4	86	10,90		
		cost	35	67	50,97			Power		7	110	10,70		
		Private	21	31	50,13			democrat		3	75	10,42		
CH_VALUES V		system	44	96	48,01			Dav		12	145	10,00		
		Investments	14	17	46,18			Senator	Part in the second	6	97		0,002	
SIMILARITIES (COSINE)	ACKI	federal	27	48	45,67			Irag		1	49		0,003	
	ACKI	economic	40	86	45,15			minority		1	49		0,003	
TREE MAP PREVIEW		create	18	26	44,62			faith	Distances -	3	67		0,003	8
		retirement	11	12	42,79	0,000		think	100000000	9	115	8,64	0,003	
TYPICAL CONTEXTS		employer	19	31	37,87	0,000		vote	instituters i	6	91	8,52	0,004	
EXPORT DICTIONARY	ALT	marketplace	12	15	37,80	0,000	-	Republica	in	9	112	8,09	0,004	
CAPOAT DICTIONANT	A CONTRACTOR OF THE OWNER		-				-							





CHAPTER ~	A ITEM	CONSTITU	FAITH	FAMILY	OPPORTU	PARTIES				
TENS (N= 1052)	CHRIST	0	8	0	0	1				
WORDS	CHRISTIAN	0	36	0	0	3				
O WORDS		0	6	0	0	1				
RASURE		0	1	0	2	2				
CHI SQUARE	**** *CHAPTER_Six *THEME F	AITH								
TEST VALUE	SCORE ( .233 )									
ATATABLE										
OCCURRENCES	of course organized religion does	n thave a monopoly on	virtue and	ano not no.	od he religion	to make moral				
MEASURES	-				of the second					
~	appeal to a common good. But we		g such claim	s or appea	15 - of aband	on any reference				
ORPARISON	religious traditions — in_order to	avoid giving offense.								
PART - WHOLE										
O BETWEEN SUB-SETS	**** *CHAPTER_Six *THEME_F	AITH								
	SCORE (.231)									
SELECT A										
CH_CONSTIT	If I am opposed_to abortion for re	eligious reasons and see	k to pass a l	aw banning	the practice,	I cannot simply				
CH FAITH	teachings of my church or invoke	God's will and expect th	at argument	to carry th	he day. If I wa	nt others to liste				
CH FAMILY	I have to explain why abortion vio									
CH_OPPORTU	at all.									
CH_POLITICS	art_and									
CH_RACE	**** SCUADTED SA STUENE E	ATTL								
CH_REPDEN	**** *CHAPTER_Six *THEME_F	Attri								
CH_VALUES	SCORE (.211)									
Cleu waara		3 10 10 10	11 115							
SIMILARITIES (COSINE)	The willingness of many who oppo									
and when the second	principle for the sake of practical	considerations; the will	ingness of ev	en the mo:	st ardent proc	hoice advocates				
TREE MAP PREVIEW	some restrictions on late-term ab	ortion marks a recogniti	on that a fet	us is more	than a body p	art and that soci				
TYPICAL CONTEXTS	interest in its development.	• • • • • • • • • • • • • • • • • • •			Average and a second second					
EXPORT DICTIONARY	COMFORT	1	1	2	1	2				

**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.









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	DICTIONARY (CORPUS)	ACTIVE	AFFILI	HOSTILE	NEGA	PASSIVE	POSITI ^	
RESET	ADVANCE	2	0	0	0	1		
AUTOMATIC LIST <		1 0	0	0	🖌 0 0	0		
RENAME CATEGORIES	AFFAIB	0	1	0	0	0		
EXECUTE CLASSIFICATION	AFFIRM AFFORD	0	0	0	0	0		
HTML REPORT	AGGRE CATEGORY = < HOSTILE	U	U	U	U	U	- m 11	
EXPORT CLASSIFICATION			~					
CONTINGENCY TABLES		VERSARI	2					
DICTIONARY (MODEL)								
DICTIONARY (CORPUS)								
VARIABLES - CATEGORIES	ALLY **** *PRES_REGAN1981 *PARTY_REP							
MULTIPLE SELECTION	ALMIGH as for the enemies of freedom these who are notential advergences they will be							
Yes 🔍 🌻 No		(MBITI						
PLOT YOUR CHART								
GRAPHS		**** *PRES_REGAN1981 *PARTY_REP						
CATEGORIES (PERC.)	APPEAI It is a weapon our adversari	<mark>es</mark> in today's	world do r	not have.				
PARTY	ASHAM **** *PRES_CLINTON19	07 *DA DTV	DEM					
MDS MAP					and a decourt			
		bonds with f	lauons that	once were	our advers	aries.		
CORRESPONDENCE ANALYSIS	ASSIST	*DADTY 1	NTM .					
EXPORT YOUR DICTIONARY						C. A.		
FURTHER T-LAB ANALYSES	- ACUND							
	ATTAIN evidence that the ways we use	e energy stre	nginen our	auversari	es and mrea	uen our		
	AWAIT planet.							
							1.4	

SELECT THE INPUT TYPE Import your DICTIONA Type/Paste your TEXTS Use a VARIABLE of you SELECT A VARIABLE	METHO	D ive Bayes arest Centroid Clas	MOI	ND TEST (PRECISION / RECALL) HODEL Categorical Variable Classified Document 2 TEST					
RESET			EDICTED TO ALUM		TO COFFEE	TO_CPI	TO CRUDE	TO GNP	TO GOLD
<< AUTOMATIC LIST <<	Þ	TO_ALUM	50	0	0	0	0	0	0
RENAME CATEGORIES		TO_COCOA	0	61	0	0	0	0	0
EXECUTE CLASSIFICATION		TO_COFFEE	0	0	112	0	0	0	0
HTML REPORT		TO_CPI	0	0	0	70	0	0	0
EXPORT CLASSIFICATION		TO_CRUDE	0	0	0	0	371	0	0
ONTINGENCY TABLES		TO_GNP	0	0	0	0	0	74	0
DICTIONARY (HODEL)		TO_GOLD	0	0	0	0	0	0	89
DICTIONARY (CORPUS)		TO_GRAIN	0	0	0	0	0	0	0
VARIABLES - CATEGORIES MULTIPLE SELECTION		TO INTEREST	0	0	0	0	0	0	0
Ye O O No		TO JOBS	0	0	0	0	0	0	0
PLOT YOUR CHART		TO MONEYFX	-	0	0	0	0	0	0
RAPHS		TO_MONEYSU	-	0	0	0	0	0	0
CATEGORIES (PERC.)		TO_SHIP	0	0	0	0	0	0	0
		TO_SUGAR	0	0	0	0	0	0	0
HDS MAP		TO_TRADE	0	0	0	0	3	0	1



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).

THEMES (N: 25)	(%) ^		THEME < F	AMILY > - WOP	RD PERC	ENTAGE			QUICK MENU	2
FAMILY	0,05		- SPECIFIC WORD	S SHAL	RED WORD	S (HIGHT P	ROBABILI	TY)	summer or man	- î
PLAN (A)	0,05	6								
WAR	0,05		T-LAB: MODELING OF EMERGING THEMES							
BLACK	0,05		EAM THEME <family> - TYPICAL WORDS</family>							
BEPUBLICAN	0,05	FAM								
SENATE	0,05	MICHE	CLICK ON ITEMS TO ELIM	CLICK ON ITEMS TO ELIMINATE						
RIGHT	0,04		WORD	IN THEME	COLOR TOT	IN (%)	са се се <b>(р)</b>	түре 🔺		
	0,04	CAT	family	124	125	0,100	0,992	SHARED	and a state of the	-
DEBATE	0,04	FATH	Michelle	87	87	0,070		SPECIFIC	Trates and	
	0,04	нс	life	75	86	0,060	0,872		The state of the s	
FAITH	0,04	MA	Father	44	45	0,035	0,978		and the second second	
	0.04	G	home	47	61	0,038	0,770		V-Test	
MEN_WOMEN	0,04		Malia (C)	28	28	0,023	and the second se	SPECIFIC		
AMERICAN	0.04		girl	📑 Use th	e right click o	f th		SPECIFIC		
FORCE	0.04	DAUGHT	0,421 SHARE							'
SCHOOL	0,04	SAS	daughter	TOT=145 Tokons				SPECIFIC	THE REAL	
U WORK	0,04(B)	PARE	Sasha	asha 1,000 SP					The second	
CONCONTRACT	100	LEA	parent		r (61 = 42			SHARED	1 1 1	1
	-		leave		E(15 = 10)		0,871		7 85 80	
EXPLORE THEMES		F	firm	SCHOO	0L (69 = 4	8%)		SPECIFIC	Sila	
TABLE THEME BAR CHART THEME	· ·	LC	look				0,474		10 . Star	1
COMPLETE RESULTS		SCHEDU	schedule				0,769		1 - Martin	-
SHARED WORDS	C1	DINN	dinner	_				SPECIFIC		
			laugh		REMOVE <d <="" th=""><th>(Y&gt;</th><th></th><th>SPECIFIC</th><th></th><th>1</th></d>	(Y>		SPECIFIC		1
APPLY THE MO	DEL	LAU	saw	2		0.010	0,647			
EXPORT DICTION	IARY	s	morning	16	18	0,013	0,889			111
		MORN	bag	14	14	0,011	1,000	SPECIFIC -		
0 🛃 0		E								





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) / INPU	T TYPE (KEY-	TERMS , TEXTS)	
CONTEXT-		Г ІМРИТ ТҮРЕ	4	NEW TEXT	CLEAN OUT YOUR BOX
CORPUS	-	KEY-TERMS	(A) -	EXTRACT KEY CONTEXTS	SAVE CONTEXTS
AVAILABLE ITEMS:	< 475 >	KEY-TERMS		OUTPUT TABLES	SAVE CONTEXTS
The second s		TEXTS	(B)	COTPOT TABLES	
ITEM	<u> </u>				
	177 🛄				
	166				
	150				
	134				
	125				
TERRORIST	114				
WORK	108				
	90				
GREAT	79				
KNOW	77				
WAR	77				
HELP	68				
FREEDOM	64				
TERRORISM	54				
UNDERSTAND	52				
TIME 1	51				
STRONG	48				
THANK	48				
□ NEED	46				
GOVERNMENT	46	1			
🗆 GOOD	45				
- CIIDE	46	CLEAN OUT Y	TOURLIST		
	6				



**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)







Exporting	X
Export CEMF CWMF CBMP © JPG CPNG	◯ Text / Data
Export Destination ClipBoard File Browse C Printer	
Export Size       Pixels     Millimeters     Inches     Points       Width:     1000     /     845     Pixels       DPI:     300     Image Font	Export Cancel

Some general criteria for the interpretation of the **T-LAB** outputs are illustrated in a paper quoted in the **Bibliography** and are available from <u>https://www.tlab.it</u> 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);

 $\mathbf{B}$  – 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).