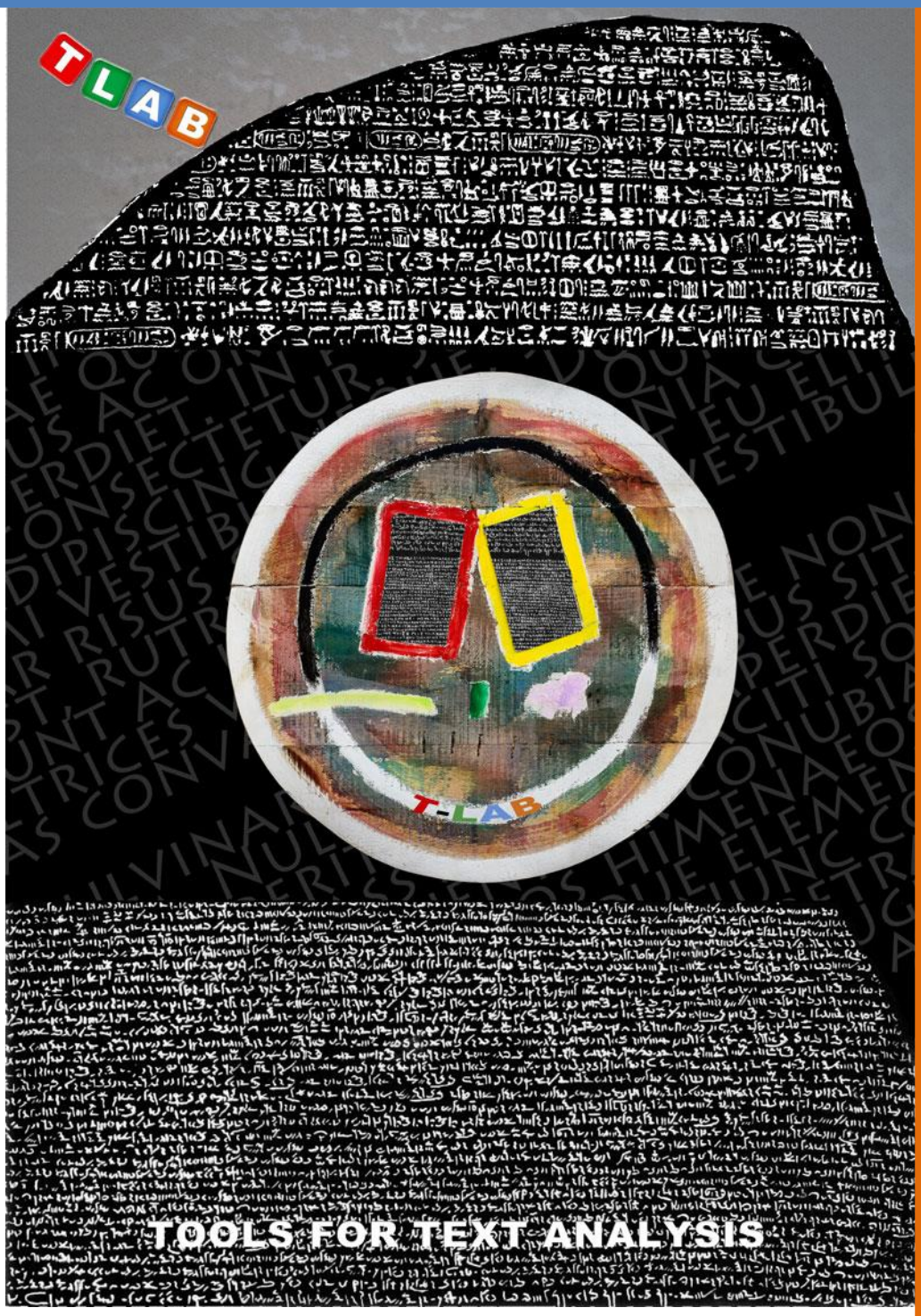
T-LAB 10

Guide de Démarrage Rapide



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

Outils pour l'Analyse de Textes

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

Ce que T-LAB fait et ce qu'il vous permet de faire

(Extrait du Manuel de l'Utilisateur)

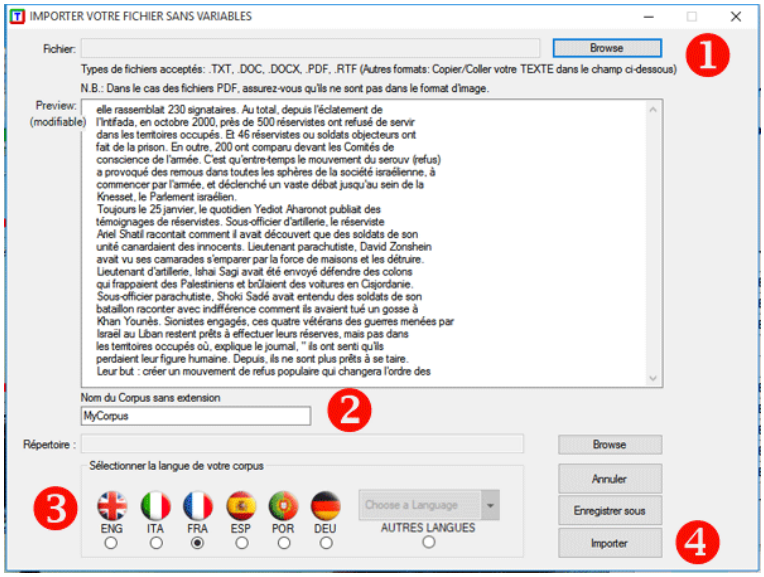
T-LAB est un logiciel composé par un ensemble d'**outils linguistiques, statistiques et graphiques pour l'analyse des textes** qui peuvent être utilisés dans les pratiques de recherche suivantes: Analyse du Contenu, Sentiment Analysis, Analyse Sémantique, Analyse Thématique, Text Mining, Perceptual Mapping, Analyse du Discours, Network Text Analysis.

En fait, au moyen des outils **T-LAB** les chercheurs peuvent facilement gérer les activités d'analyse suivantes:

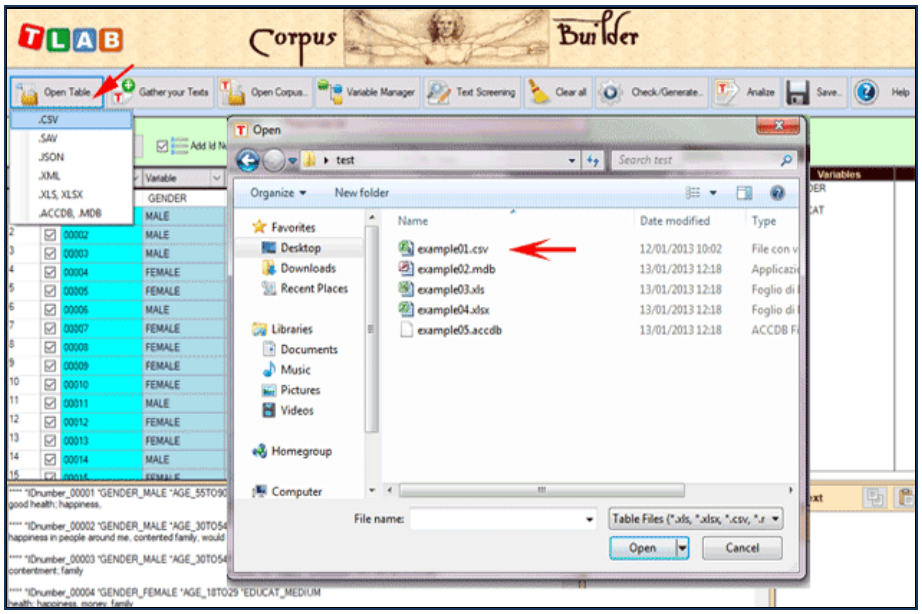
- explorer, mesurer et topographier les **relations de co-occurrence** entre mots-clés;
- réaliser une **classification automatique** d'unités de contexte et de documents, soit à travers une **approche bottom-up** (c'est-à-dire à travers l'analyse des thèmes émergents) soit à travers une **approche top-down** (c'est-à-dire à travers l'utilisation de catégories prédéfinies);
- vérifier quelles **unités lexicales** (c'est-à-dire mots ou lemmes), quelles **unités de contexte** (c'est-à-dire phrases ou paragraphes) et quels **thèmes** sont «typiques» de sous-ensembles de textes spécifiques (par exemple, les discours de certains leaders politiques, les interviews avec certaines catégories de personnes, etc.);
- appliquer des catégories pour la **sentiment analysis**;
- effectuer différents types d'**analyse des correspondances** et de **clusters analysis**;
- créer des **cartes sémantiques** qui représentent des **aspects dynamiques du discours** (c'est-à-dire des relations séquentielles entre les mots ou les thèmes);
- représenter et explorer un texte quelconque comme un **réseau** de relations;
- obtenir des mesures et des représentations graphiques concernant les **textes et discours** traités **comme des systèmes dynamiques**;
- personnaliser et appliquer **différents types de dictionnaires**, aussi bien pour l'analyse lexicale que pour l'analyse du contenu;
- vérifier les contextes d'occurrence (par ex., **concordances**) de mots et de lemmes;
- analyser tout le **corpus** ou seulement certains de ses **sous-ensembles** (par ex. des groupes de documents) en utilisant différentes listes de mots-clés
- créer, explorer et exporter différents types de **tableaux de contingence** et de **matrices de co-occurrences**.

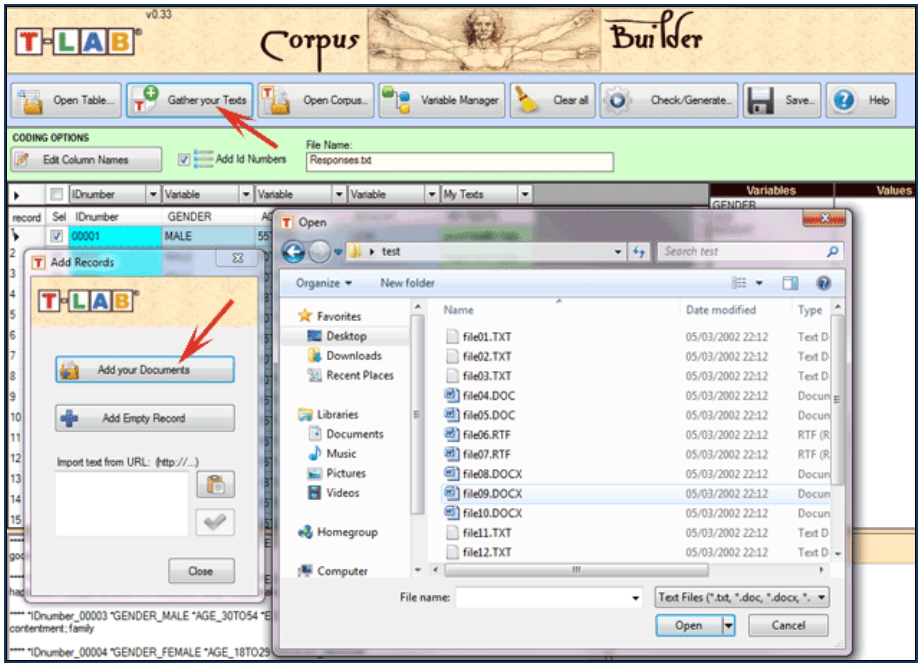
L'interface utilisateur est **très conviviale** et les textes à analyser peuvent être des plus variés:
- un seul texte (ex. une interview, un livre, etc.);
- un ensemble de textes (ex. diverses interviews, pages web, articles de journal, réponses à des questions ouvertes, messages Twitter, etc.).

Tous les textes peuvent être codifiés avec des **variables** catégorielles et peuvent inclure un identificateur (**Unique Identifier**) qui correspond à des unités de contexte ou à des cas (ex. réponses à des questions ouvertes).
Dans le cas d'un seul document (ou un corpus considéré comme un texte unique) **T-LAB** nécessite pas de travail supplémentaire: il vous suffit de sélectionner l'option 'Importer un fichier unique (voir ci-dessous).



Différemment, dans les autres cas il faut utiliser le module **Corpus Builder** (voir ci-dessous) qui transforme automatiquement des documents textuels et différents types de fichiers (c'est-à-dire jusqu'à dix différents formats) dans un corpus prêt à être importé par **T-LAB**.

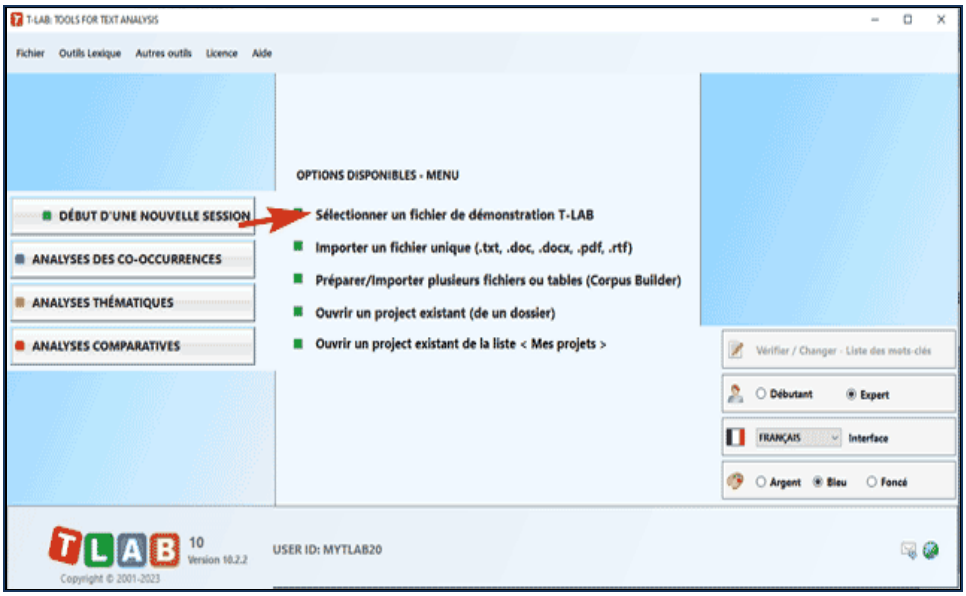




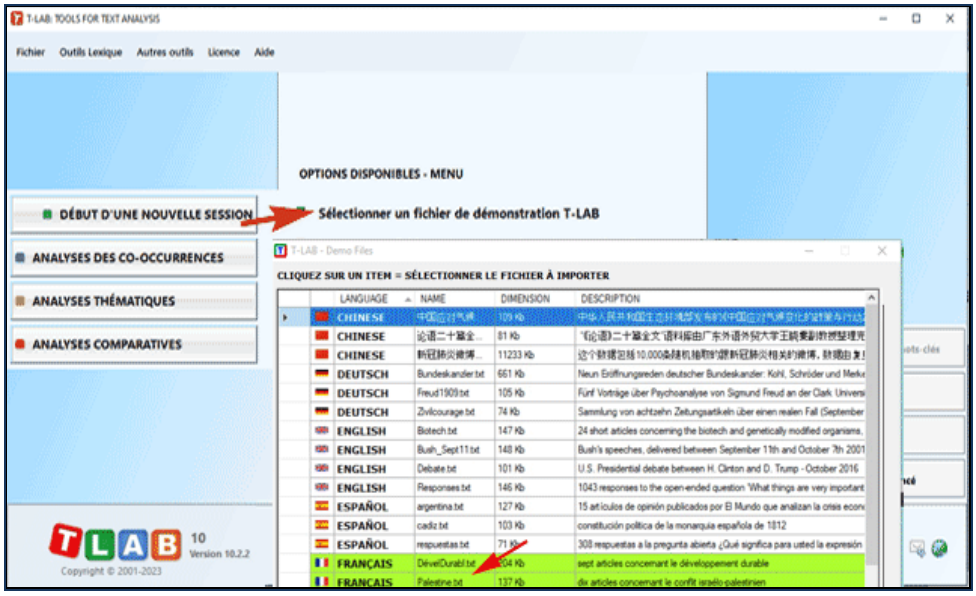
N.B. : En ce moment, afin d'assurer l'utilisation intégrée des différents outils, chaque fichier/corpus à analyser ne devrait pas dépasser 90 Mo (c'est-à-dire environ 55.000 pages au format .txt). Pour plus d'informations, voir la section 'Conditions et performances' du Manuel / Help.

Six étapes suffisent pour explorer rapidement les fonctions du logiciel:

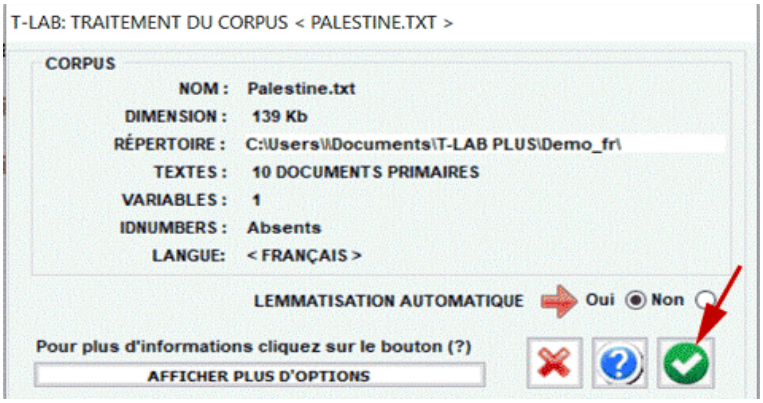
1 - Cliquer l'option 'Sélectionner un fichier de démonstration..'



2 - Sélectionner un corpus à analyser

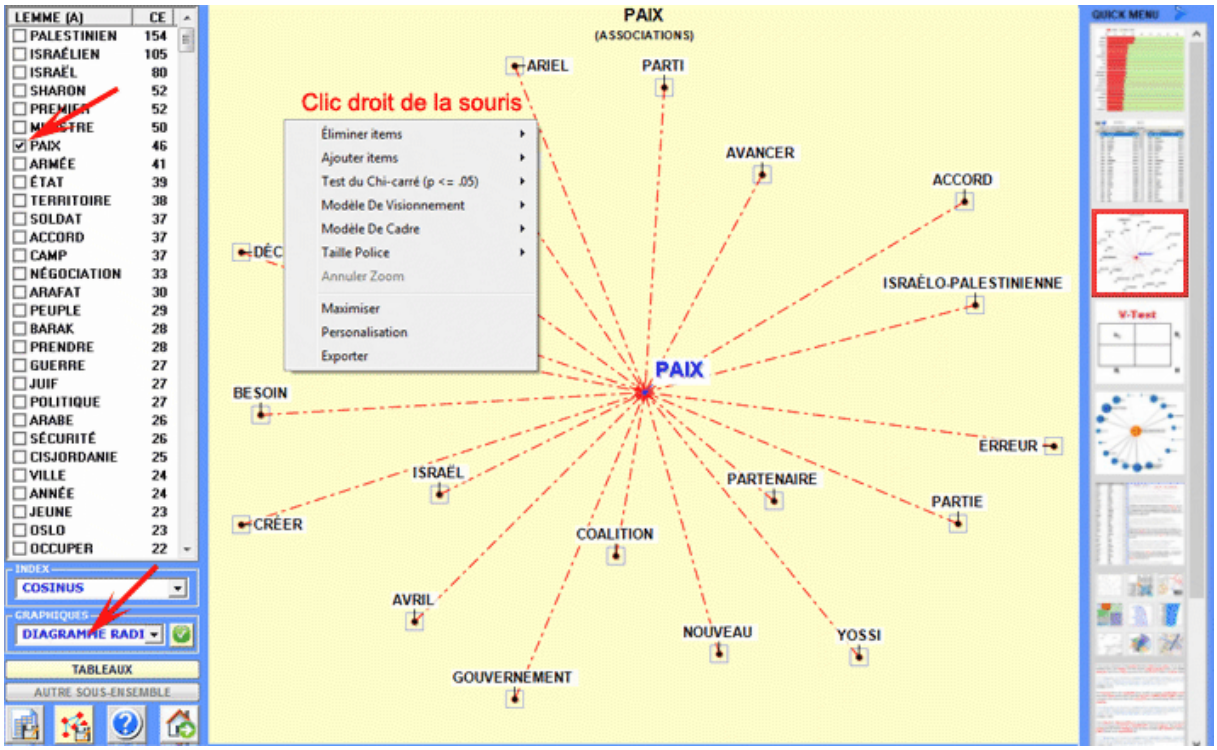


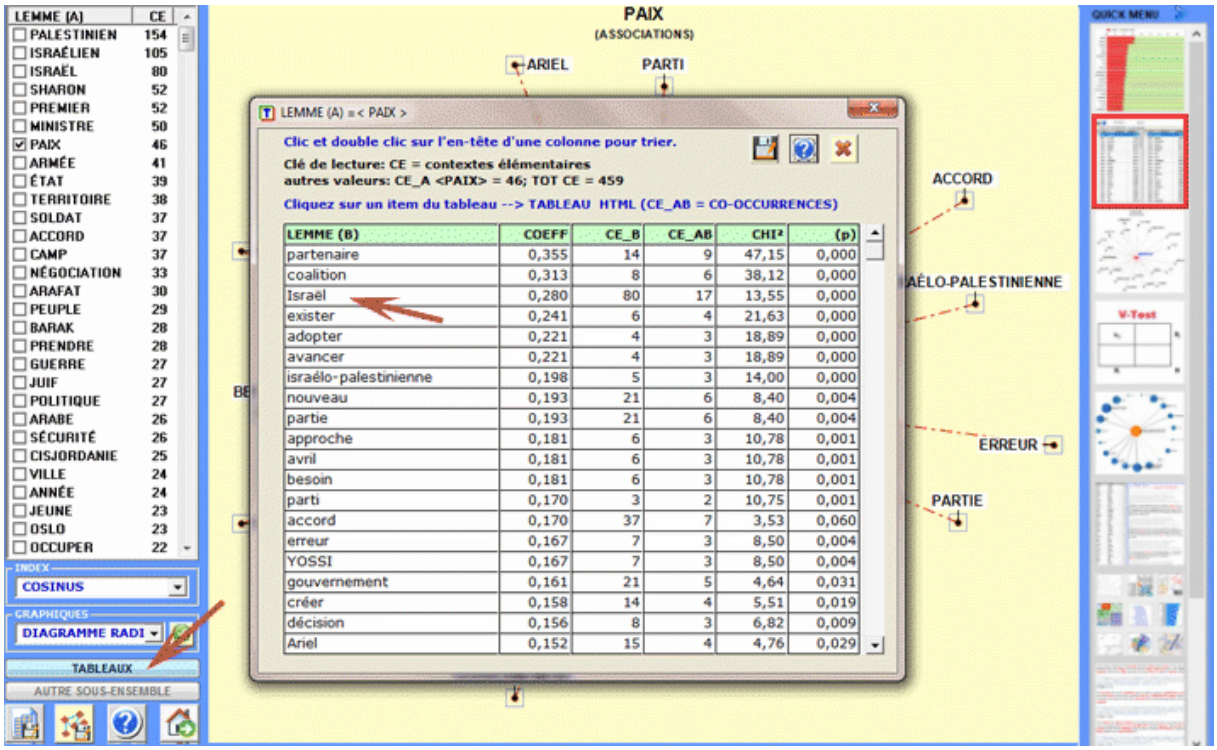
3 - Cliquer sur "ok" dans la première fenêtre de configuration



4 - Choisir un outil à l'intérieur d'un des sous-menus "Analyse"

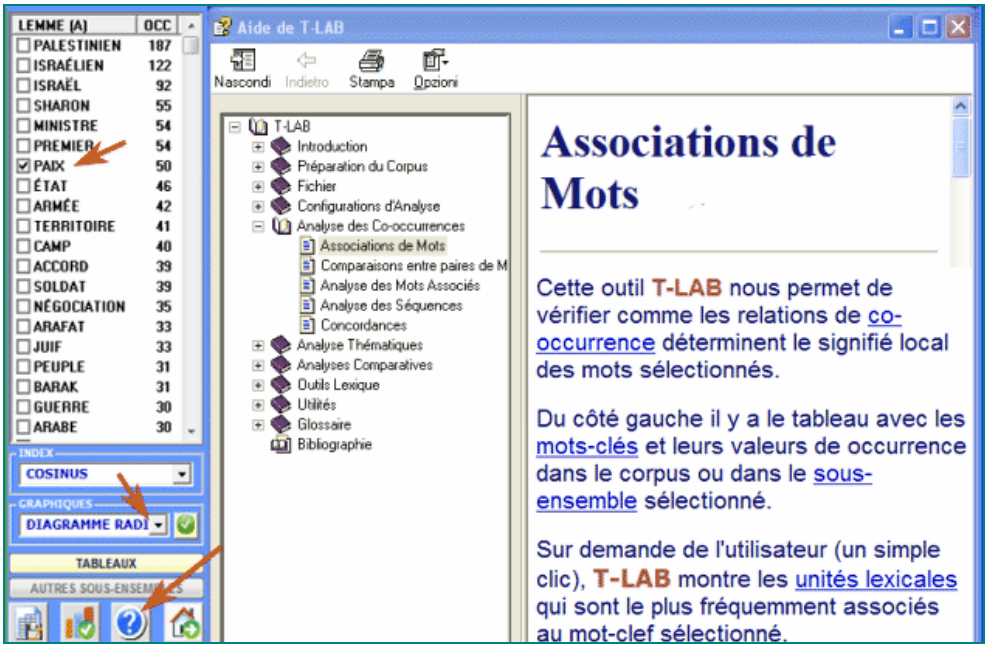
5 - Examiner les résultats





LEMME (B)	COEFF	CE_B	CE_AB	CHI²	(p)
partenaire	0,355	14	9	47,15	0,000
coalition	0,313	8	6	38,12	0,000
Israël	0,280	80	17	13,55	0,000
exister	0,241	6	4	21,63	0,000
adopter	0,221	4	3	18,89	0,000
avancer	0,221	4	3	18,89	0,000
israélo-palestinienne	0,198	5	3	14,00	0,000
nouveau	0,193	21	6	8,40	0,004
partie	0,193	21	6	8,40	0,004
approche	0,181	6	3	10,78	0,001
avril	0,181	6	3	10,78	0,001
besoin	0,181	6	3	10,78	0,001
parti	0,170	3	2	10,75	0,001
accord	0,170	37	7	3,53	0,060
erreur	0,167	7	3	8,50	0,004
YOSSI	0,167	7	3	8,50	0,004
gouvernement	0,161	21	5	4,64	0,031
créer	0,158	14	4	5,51	0,019
décision	0,156	8	3	6,82	0,009
Ariel	0,152	15	4	4,76	0,029

6 - Utiliser l'aide contextuelle pour interpréter les graphiques et les tableaux.



Cette section introductive fournit les informations essentielles afin de mieux comprendre ce que **T-LAB** fait et comment il peut être utilisé.

Du point de vue externe, l'utilisation du logiciel est organisée par l'**interface**, c'est-à-dire par le **menu principal**, par les **sous-menus** et les **fonctions** qui les composent.

D'un point de vue logique, en plus de l'interface usager, le système **T-LAB** est organisé par deux composantes principales:

- le **database**, c'est-à-dire le lieu informatique dans lequel le **corpus** en input (soit le texte ou l'ensemble des textes à analyser) est représenté comme un ensemble de **tableaux** dans lesquels sont enregistrées les **unités d'analyse**, leurs caractéristiques et leur relations réciproques.
- les **algorithmes**, c'est-à-dire des sous-ensembles d'**instructions** qui permettent d'utiliser l'interface usager, de consulter et modifier le database, de construire d'ultérieurs tableaux avec les données contenues dans ce dernier, d'effectuer des **calculs statistiques** et de produire des **outputs** qui représentent les relations entre les données analysées.

Pour comprendre comment **T-LAB** fonctionne et comment il peut être utilisé, il est fondamental de savoir clairement quelles unités d'analyse sont archivées dans son database et quels algorithmes statistiques sont utilisés dans les diverses analyses. En effet, les tableaux de données analysées sont toujours constitués de lignes et de colonnes dont les titres correspondent aux unités d'analyse archivées dans le database, alors que les algorithmes règlent les processus qui permettent de repérer des relations significatives entre les données et d'extraire des informations utiles.

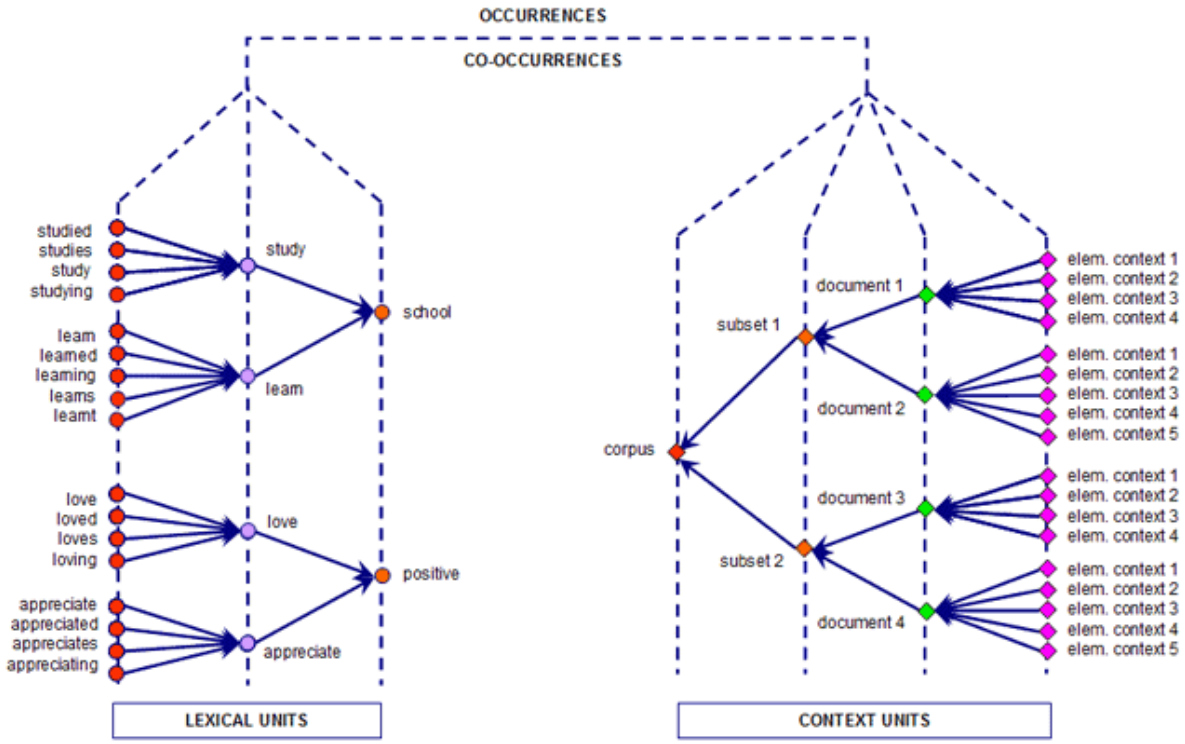
Les **unités d'analyse** de T-**LAB** sont de deux types: **unités lexicales** et **unités de contexte**.

A - les **UNITÉS LEXICALES** sont des mots, simples ou multiples, archivés et classifiés sur la base d'un critère. Plus précisément, dans le database **T-LAB** chaque unité lexicale constitue un record classifié avec deux champs: **mot et lemme**. Dans le premier champ, appelé **mot**, sont listés les mots ainsi qu'ils apparaissent dans le corpus, alors que dans le second, appelé **lemme**, sont listés les labels attribués à des groupes d'unités lexicales classifiées selon des critères linguistiques (ex. **lemmatisation**) ou au moyen de dictionnaires et de grilles sémantiques définies par l'utilisateur.

B - les **UNITÉS DE CONTEXTE** sont des portions de texte dans lesquelles le corpus peut être subdivisé. Plus exactement, dans la logique **T-LAB**, les unités de contexte peuvent être de trois types:

B.1 documents primaires, correspondants à la subdivision "naturelle" du corpus (ex. interviews, articles, réponses à des questions ouvertes, etc.), ou bien aux contextes initiaux définis par l'utilisateur;
B.2 contextes élémentaires, correspondants à des unités syntagmatiques (ex. fragments de texte, phrases, paragraphes) dans lesquelles chaque document primaire peut être subdivisé;
B.3 sous-ensembles du corpus, correspondants à des groupes de documents primaires reconductibles à la même catégorie (ex. interviews d' "hommes" ou de "femmes", articles d'une année particulière ou d'un titre particulier, et ainsi de suite), ou à clusters thématiques obtenus avec des instruments spécifiques de **T-LAB**.

Le diagramme suivant illustre les relations possibles entre les unités lexicales et les unités de contexte que **T-LAB** nous permet d'analyser.

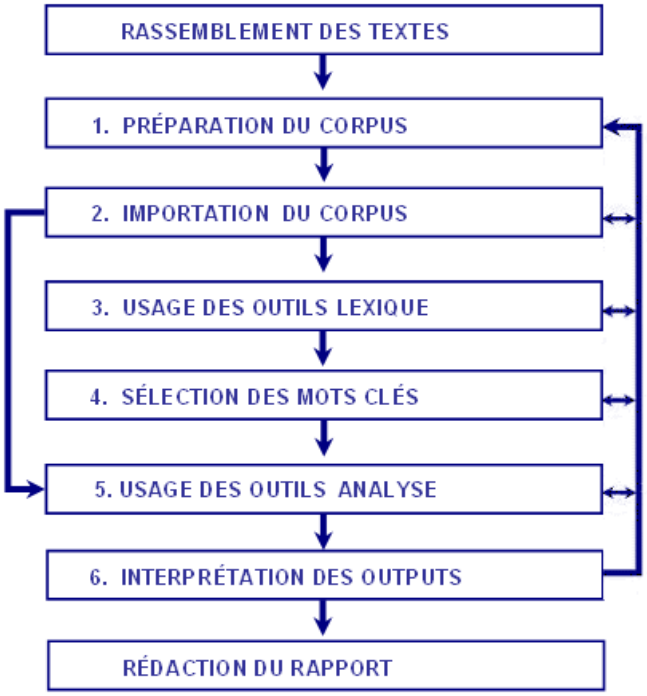


À partir de cette organisation du database, **T-LAB** permet - de façon automatique - d'explorer et d'analyser les relations entre les unités d'analyse de **tout le corpus** ou de ses **sous-ensembles**.

Dans **T-LAB**, la sélection d'un quelconque instrument d'analyse (clic de la souris) active toujours un processus semi-automatique qui, grâce à quelques simples opérations, génère un tableau input, applique un algorithme de type statistique et produit quelques outputs.

Un **projet** de travail "typique" dans lequel est utilisé **T-LAB** est constitué de l'ensemble des activités analytiques (opérations) qui ont pour objet le même **corpus** et est organisé par une **stratégie** et par un **plan** de l'utilisateur. Ainsi, il commence par le **rassemblement des textes** à analyser et s'achève par la **rédaction d'un rapport**.

La succession des diverses phases est illustrée dans le diagramme suivant:



N.B.:
- Les six phases énumérées, de la préparation du corpus à l'interprétation des outputs, sont supportées par des instruments **T-LAB** et sont toujours réversibles;
- Grâce aux **configurations automatiques T-LAB** il est possible d'éviter deux phases (3 et 4); toutefois, aux fins de la **qualité** des résultats, leur réalisation est fortement recommandée.

1 - La **PRÉPARATION DU CORPUS** consiste en la transformation des textes à analyser dans un fichier (**corpus**) qui peut être élaboré par le logiciel.

Dans le cas de textes uniques (ou corpus considéré comme texte unique) on n'a pas besoin d' autre travail. Autrement, si le corpus se compose de plusieurs documents primaires codifiés (**variables et modalités**), dans la phase de préparation on doit utiliser l'outil **Corpus Builder**, qui transforme automatiquement tout matériel textuel et divers types de fichiers (c.-à-d. jusqu'à dix formats différents) dans un fichier corpus prêt à être importé par **T-LAB.**

N.B.:

- au terme de la phase de préparation du corpus on recommande de créer un nouveau dossier de travail avec à l'intérieur le fichier corpus à importer ;

- durant les analyses il est recommandé de garder le corpus et le dossier de travail relatif sur un disque dur de l'ordinateur où **T-LAB** est installé. Dans le cas contraire, l'exécution des diverses procédures pourrait être ralentie et le logiciel pourrait signaler des erreurs.

2 - **L'IMPORTATION DU CORPUS** consiste en une série de **processus automatiques** qui transforment le corpus en un ensemble de tableaux intégrés dans le **database T-LAB**.

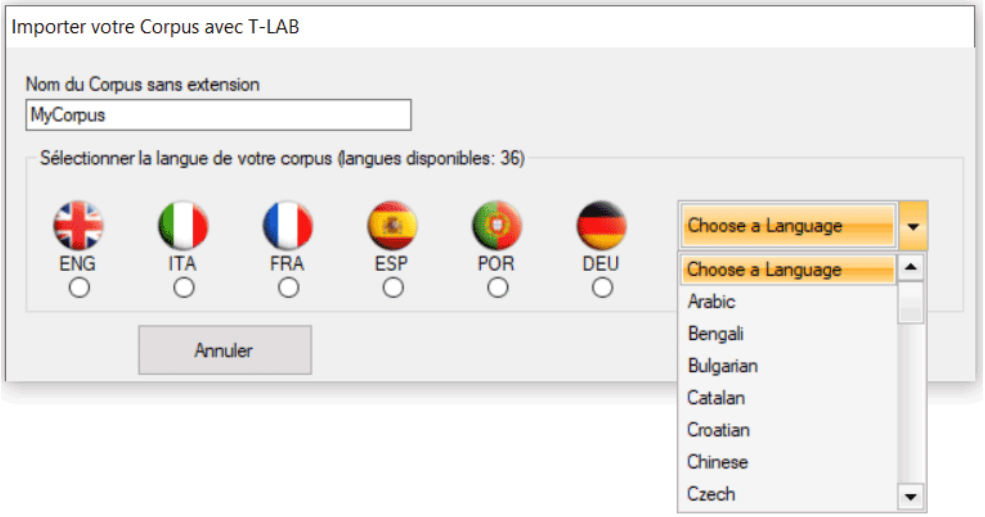
Pendant la phase d'importation du corpus, **T-LAB** effectue les traitements suivants: **normalisation** du corpus; détection des **multi-words** et des **stop-words**; segmentation des **contextes élémentaires**; **lemmatisation** automatique ou **stemming**; construction du **vocabulaire**; sélection des **mots-clés**.

De suite la liste complète des trente langues pour lesquelles la lemmatisation automatique ou bien le processus de stemming sont supportés par **T-LAB**.

LEMMATISATION: allemand, anglais, catalan, croate, espagnol, français, italien, latin, polonais, portugais, roumain, russe, serbe, slovaque, suédois, ukrainien.

STEMMING: arabe, bengali, bulgare, danois, hollandais, finlandais, grec, hindi, hongrois, indonésien, marathi, norvégien, persan, tchèque, turc.

En tout les cas, sans lemmatisation automatique et / ou en utilisant des dictionnaires personnalisés, l'utilisateur peut analyser textes dans **toutes les langues**, à condition que les mots soient séparés par des espaces et/ou des signes de ponctuation.

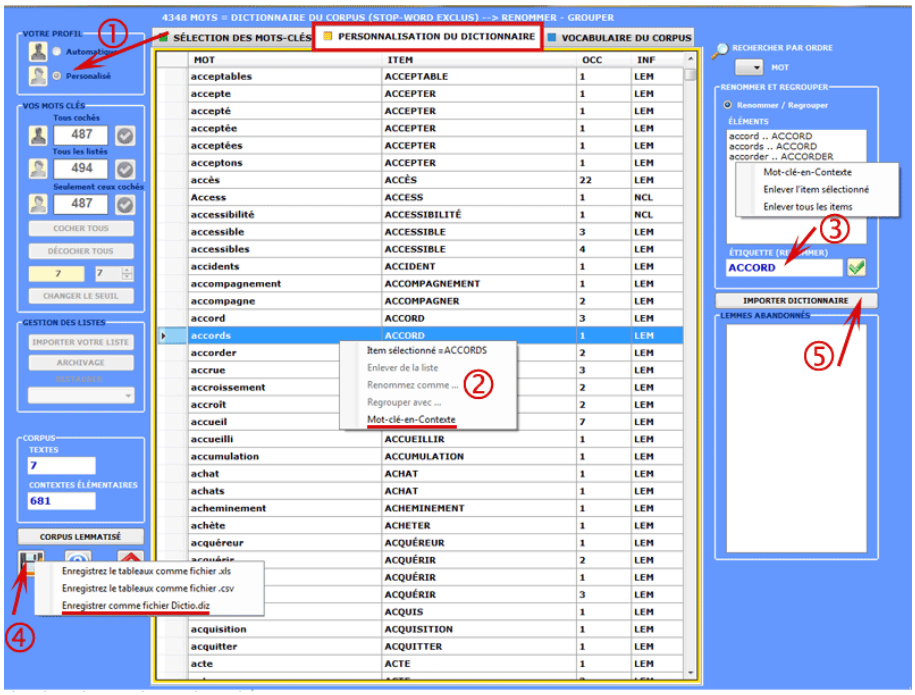


À partir de la sélection de la langue, l'intervention de l'utilisateur (options avancées) est requise afin de définir les choix indiqués dans la fenêtre suivante.

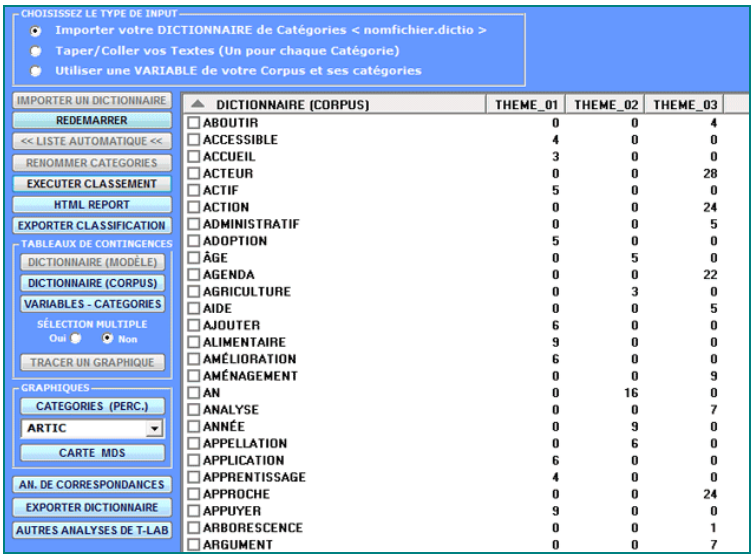
N.B. : Puisque les options de prétraitement déterminent le type et la quantité d'unités d'analyse (c.-à-d. des unités de contexte et des unités lexicales), les différents choix de l'utilisateur déterminent différents résultats de l'analyse. Pour cette raison, tous les outputs de **T-LAB** (c.-à-d. graphiques et tableaux) montrés dans le manuel et dans l'aide en ligne sont simplement indicatifs.

3 - L'UTILISATION DES OUTILS LEXIQUE est finalisée à la vérification de la correcte **reconnaissance** des unités lexicales et à personnaliser leur **classification**, c'est-à-dire à vérifier et à modifier les choix automatiques faits par **T-LAB**.

Les modalités des diverses interventions sont illustrées dans les rubriques de l'aide (et du manuel) correspondantes. En particulier on renvoie à la rubrique de l'aide (et du manuel) correspondante pour une description détaillée du processus **Personnalisation du Dictionnaire**. En effet, n'importe quel changement relatif aux voix du dictionnaire (par ex., le regroupement de deux ou plusieurs items) influe aussi bien sur le calcul des occurrences que sur celui des co-occurrences.

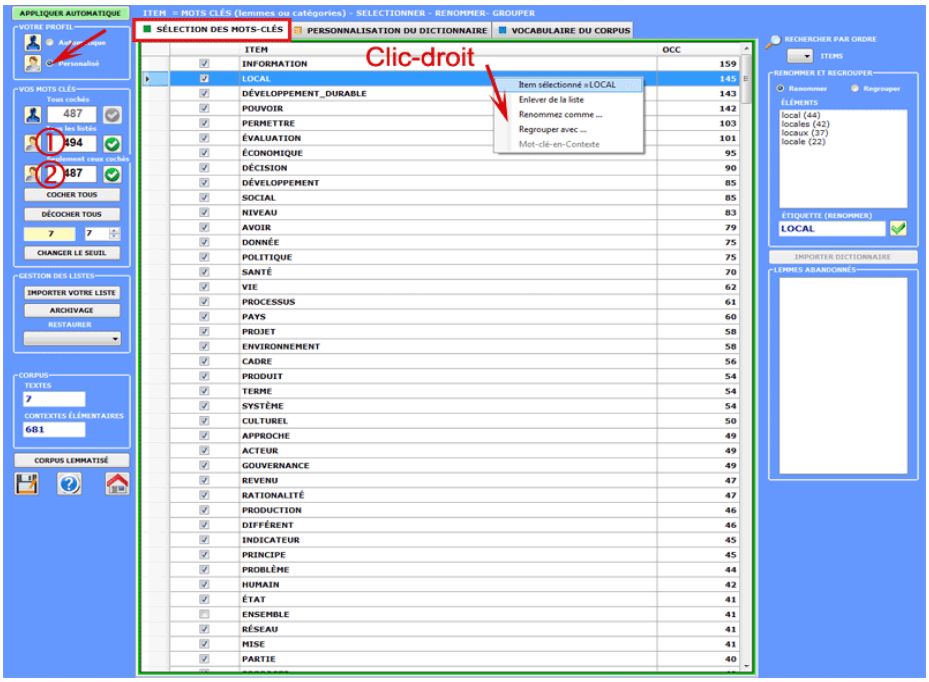


NB: Lorsque l'utilisateur, sans perdre aucune information lexicale, a l'intention d'appliquer des schémas de codage qui regroupent plusieurs mots ou lemmes dans peu de catégories (de 2 à 50), il est conseillé d'utiliser l'outil **Classification Basée sur des Dictionnaires** inclus dans le sous-menu **Analyse Thématique** (voir ci-dessous).



4 - LA SÉLECTION DES MOTS-CLÉS consiste en la prédisposition d'un ou de plusieurs listes d'unités lexicales (mots, lemmes ou catégories) à utiliser pour construire les tableaux données à analyser.

L'option **configurations automatiques** rend disponible des listes de **mots-clés** sélectionnés par **T-LAB**; toutefois, puisque le choix des unités d'analyse est extrêmement important aux fins des élaborations successives, on conseille vivement l'utilisation des **configurations personnalisées**.
De cette façon l'utilisateur pourra choisir de modifier la liste suggérée par **T-LAB** et/ou de construire des listes qui correspondent mieux à ses objectifs de recherche.



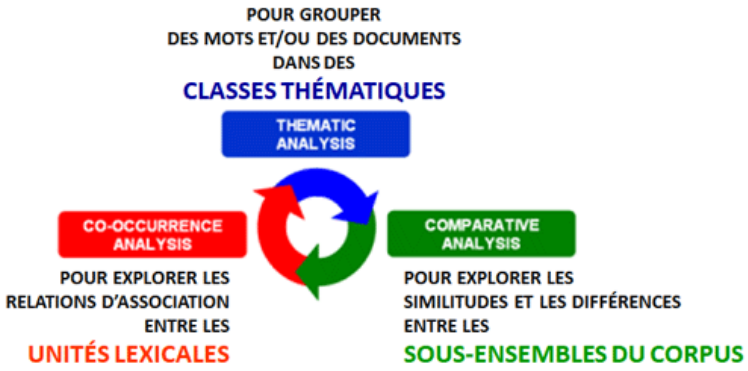
De toute façon, dans la construction de ces listes, valent les critères suivants:

- vérifier l'**importance** quantitative (total des occurrences) et qualitative (non banalité du sens) des divers items;
- vérifier les **limitations** (voir note à la fin de cette section) des instruments analytiques que l'on entend utiliser;
- vérifier si l'ensemble des items est compatible avec la propre **stratégie** de recherche (voir point suivant: 5).

5 - L'UTILISATION DES OUTILS D'ANALYSE est finalisée à la production d'outputs (tableaux et graphiques) qui représentent des **relations significatives** entre les unités d'analyse et qui permettent de faire des **inférences**.

Au moment actuel **T-LAB** inclut vingt différents outils d'analyse et chacun d'eux a sa propre logique; c'est-à-dire, chacun d'eux emploie des algorithmes spécifiques et produit des outputs spécifiques.

Par conséquent, selon le type de textes qu'il a l'intention d'analyser et des objectifs qu'il veut poursuivre, l'utilisateur doit décider de fois en fois quels sont les outils les plus appropriés pour sa stratégie d'analyse.

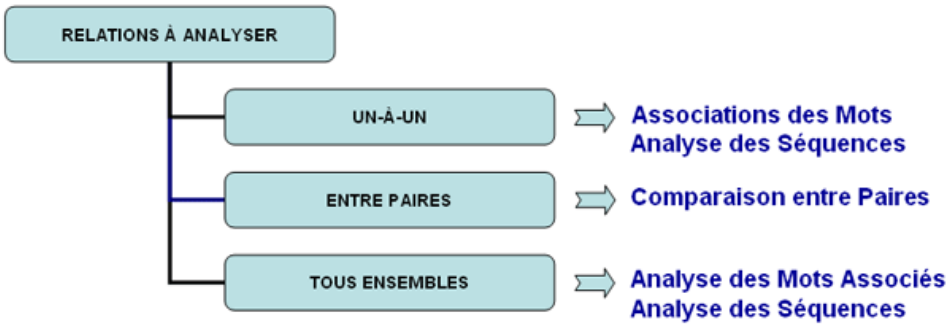


À cette fin, outre la distinction entre outils pour l'**analyse des cooccurrences**, pour l'**analyse comparative** et pour l'**analyse thématique**, il est utile de considérer que certains de ces derniers instruments permettent d'obtenir d'ultérieurs **sous-ensembles** fondés sur la similarité des contenus qui peuvent être inclus dans d'autres étapes de l'analyse.

Toutefois, compte tenu du fait que l'utilisation des outils **T-LAB** peut être circulaire et réversible, nous pouvons identifier trois points de démarrage (start points) qui correspondent aux trois sous-menus ANALYSE.

A : OUTILS POUR LES ANALYSES DE CO-OCCURRENCES

Ces outils nous permettent d'analyser différentes typologies de relations entre les mots.

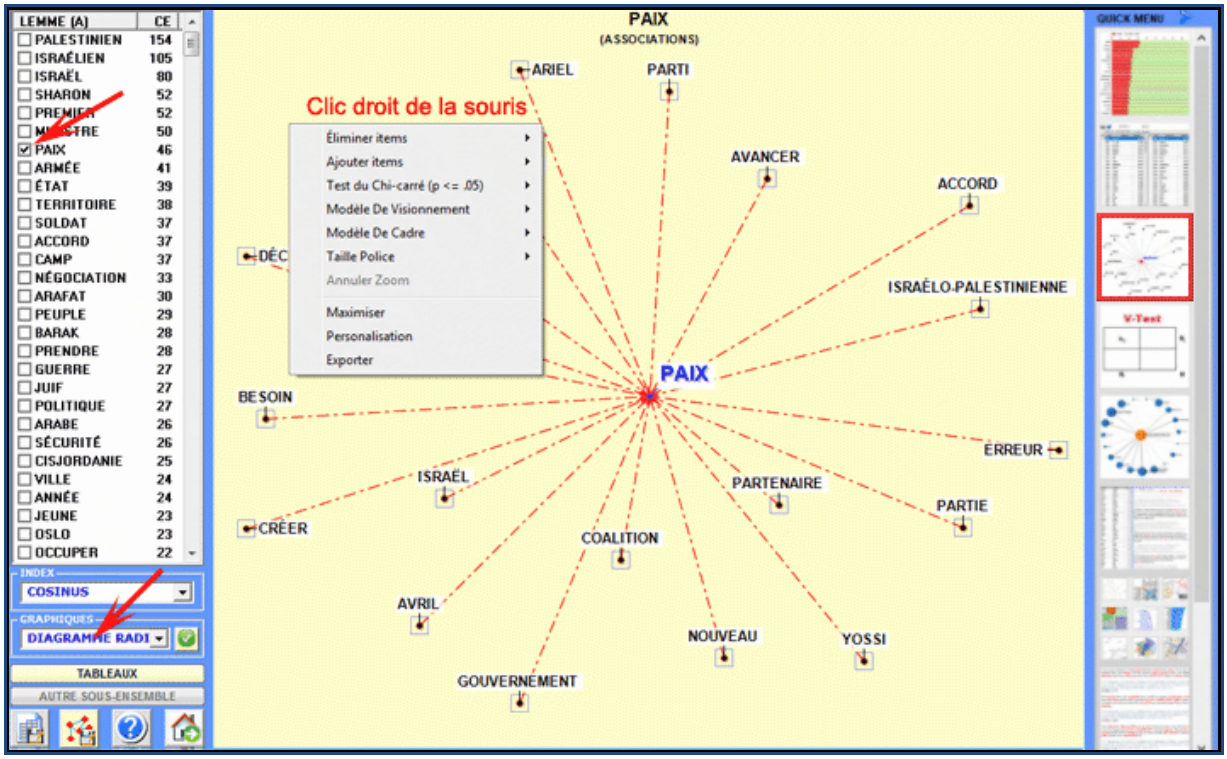


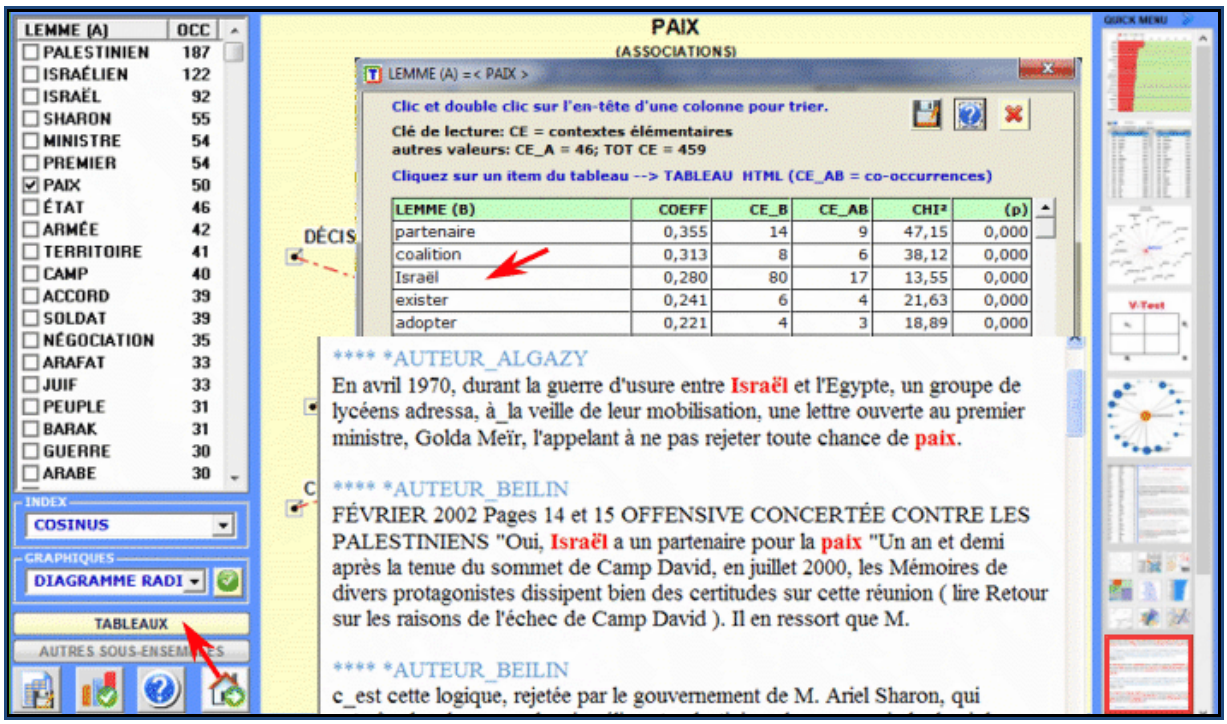
Selon les types de relations à analyser, les fonctions **T-LAB** indiquées dans ce diagramme (box colorés) utilisent un ou plusieurs des instruments statistiques suivants: **Indices d'Association**, **Test du Chi-Deux**, **Cluster Analysis**, **Multidimensional Scaling, Principal Component Analysis, t-SNE** et **Chaînes Markoviennes.**

Voici quelques exemples (N.B. : pour plus d'informations sur l'interprétation des outputs, veuillez vous référer aux sections correspondantes du guide / manuel):

- Associations des Mots

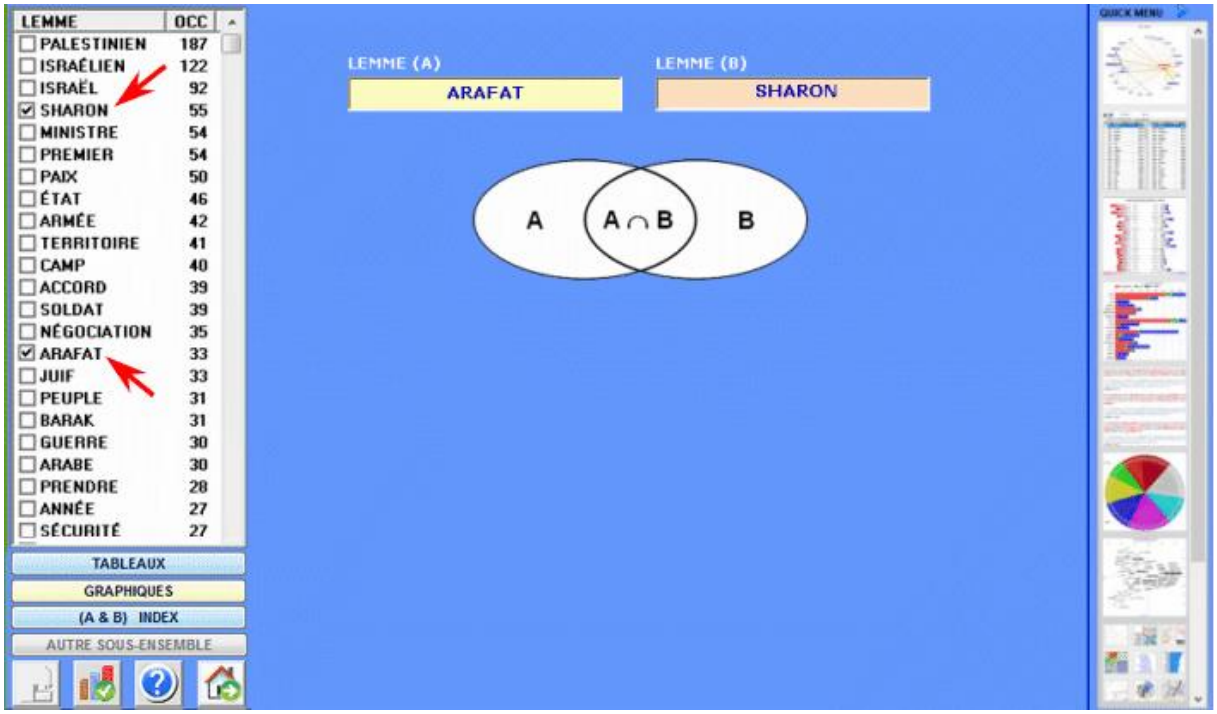
Cet outil **T-LAB** nous permet de vérifier comment les relations de **co-occurrence** déterminent le signifié local des mots sélectionnés.

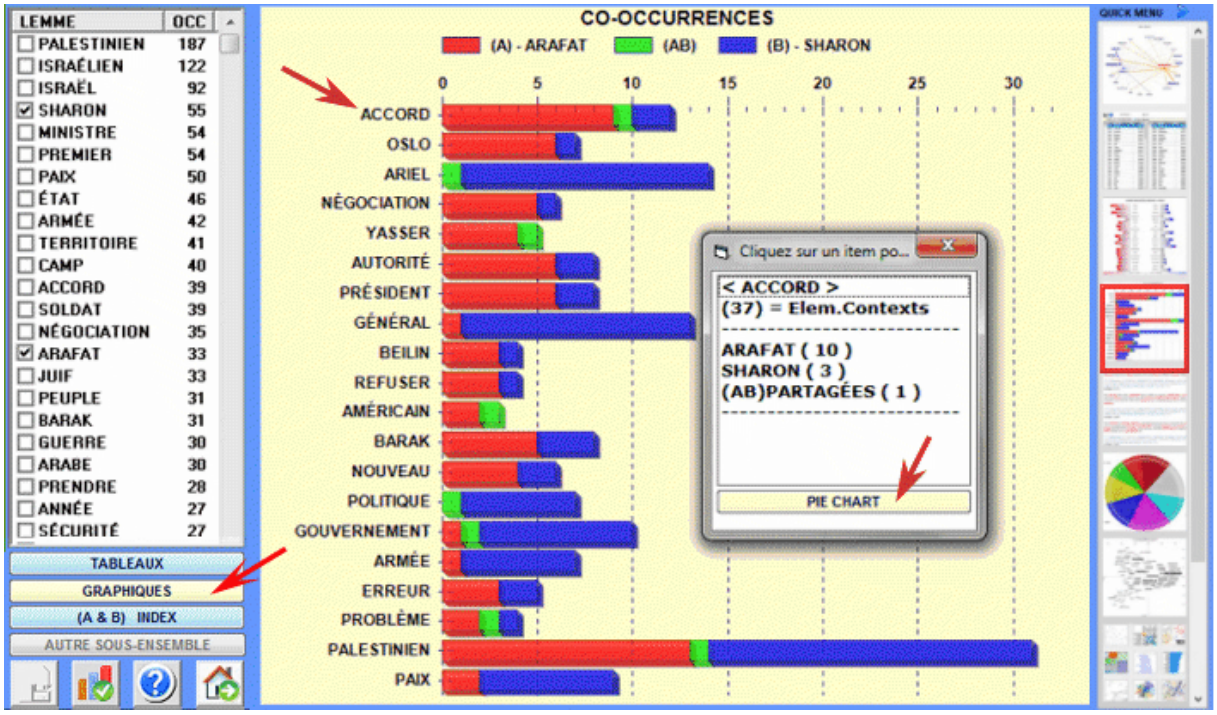




- Comparaison entre Paires

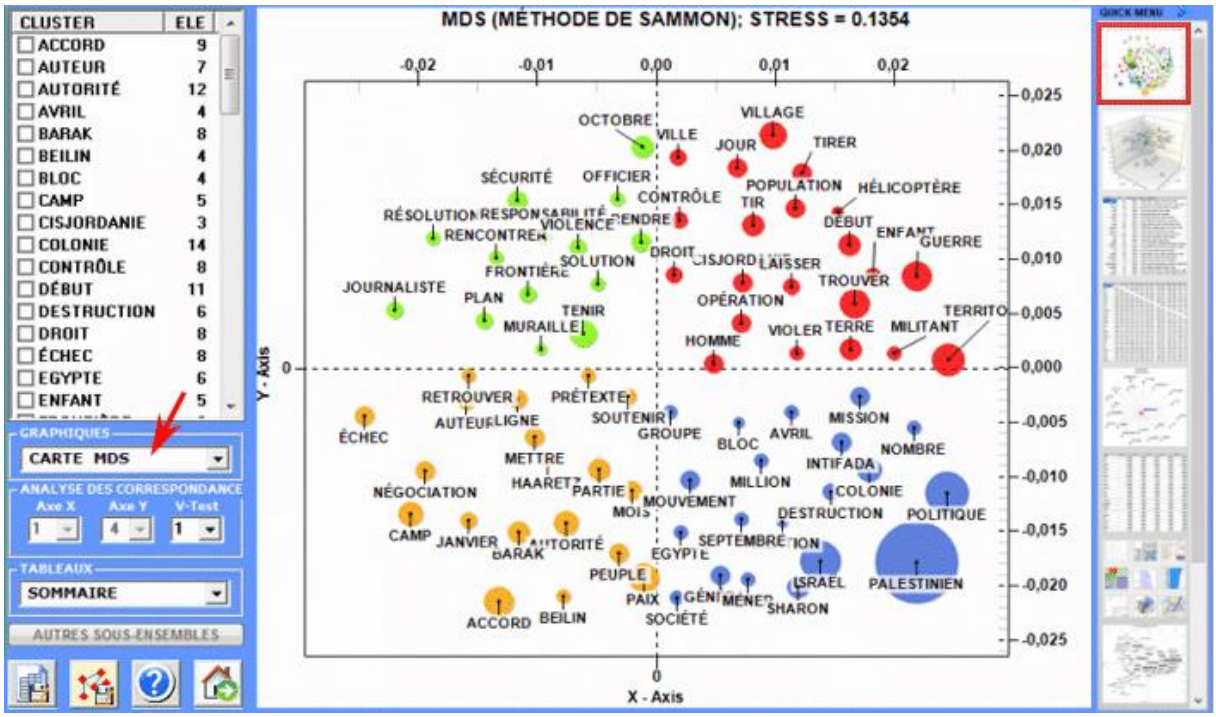
Cet outil **T-LAB** nous permet de comparer des ensembles de **contextes élémentaires** (c.-à-d. contextes de co-occurrence) dans lesquels sont présents les éléments d'une paire de **mots-clés**.

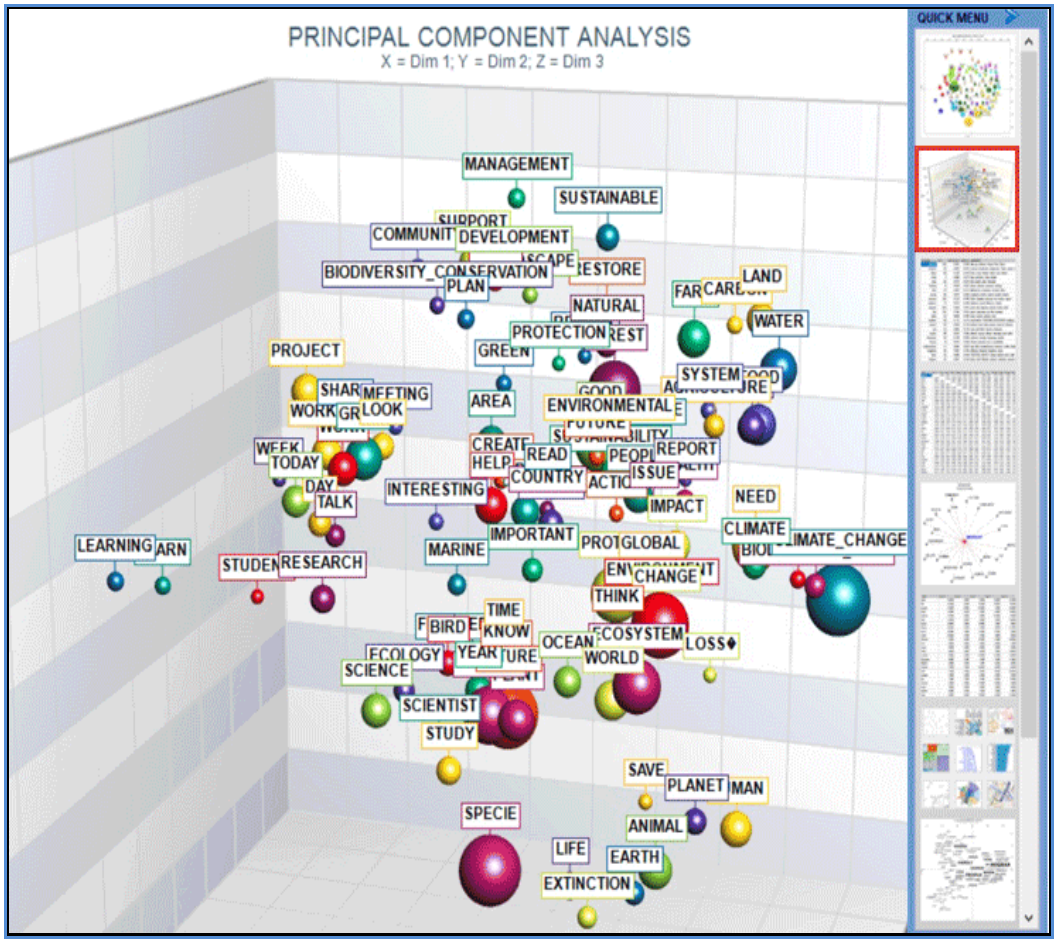




- Analyse des Mots Associés

Cet outil **T-LAB** nous permet de cartographier les relations de co-occurrence entre les **ensembles de mots-clés**.



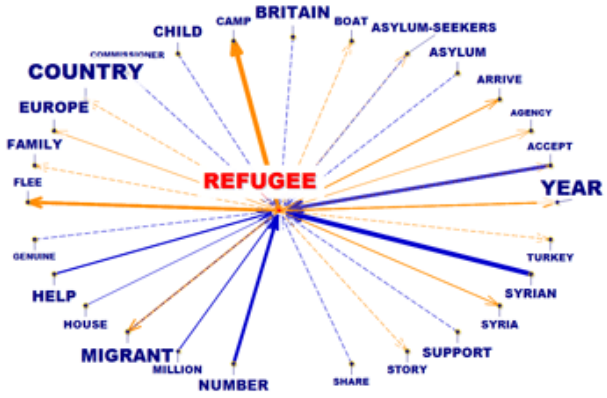


- Analyse des Séquences et Analyse des Réseaux

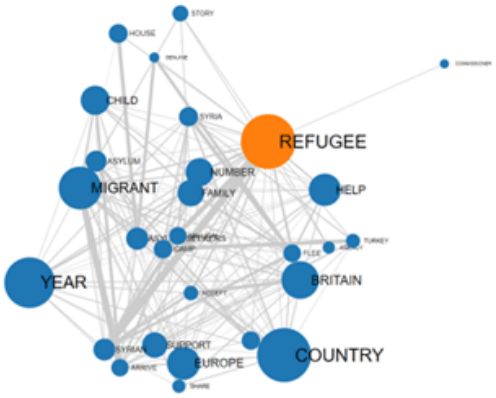
Cet outil **T-LAB** tient compte des **positions** des différentes unités lexicales à l'intérieur des phrases et il nous permet de représenter et d'explorer n' importe quel texte comme un **réseau de relations**.

Ceci signifie, après avoir exécuté ce type d'analyse, que l'utilisateur peut vérifier les relations entre les nœuds du réseau (c'est-à-dire les mots-clés) à plusieurs niveaux: a) en relations du type un-à-un; b) à l'intérieur d' «ego network»; c) à l'intérieur des «communautés» auxquelles ils appartiennent; d) à l'intérieur du réseau entier constitué par le texte en analyse.

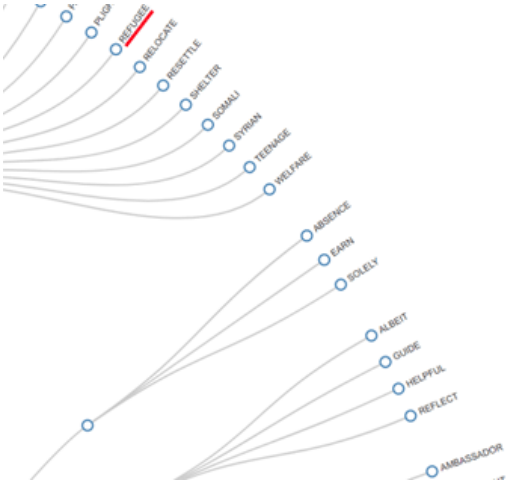
RELATIONS DU TYPE UN-A-UN



EGO-NETWORK

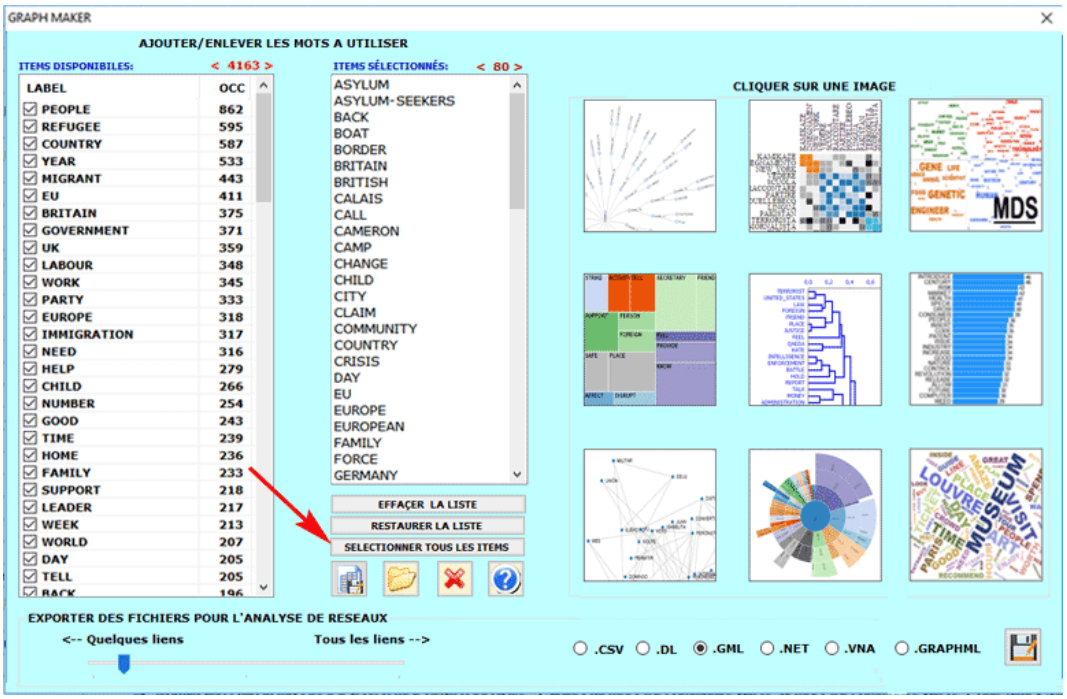


COMMUNAUTES



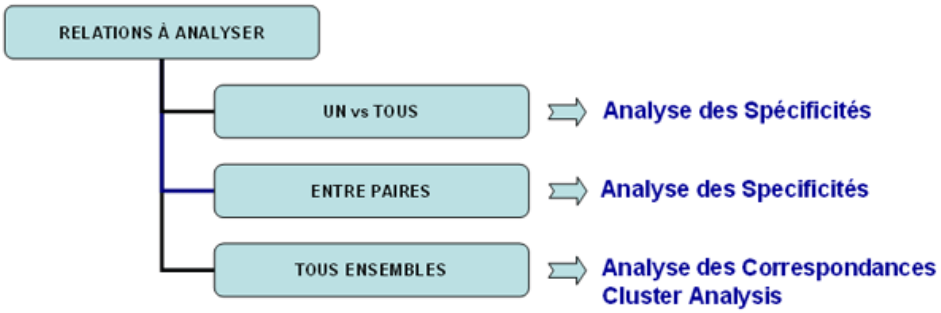
RESEAU ENTIER

De plus, à n'importe quel moment, en faisant clic sur l'option GRAPH MAKER, l'utilisateur peut créer des différents types de graphiques en utilisant des listes personnalisées de mots-clés, (voir ci-dessous).

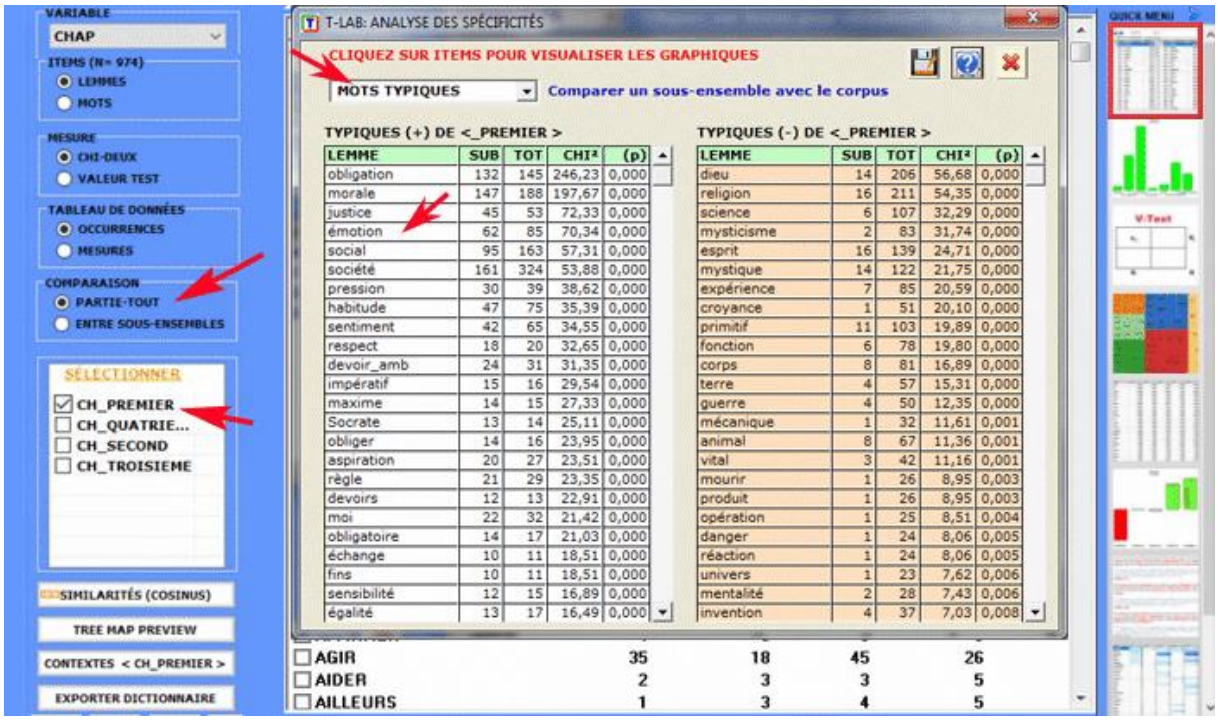


B : OUTILS POUR LES ANALYSES COMPARATIVES

Ces outils nous permettent d'analyser différentes typologies de relations entre les unités de contexte.

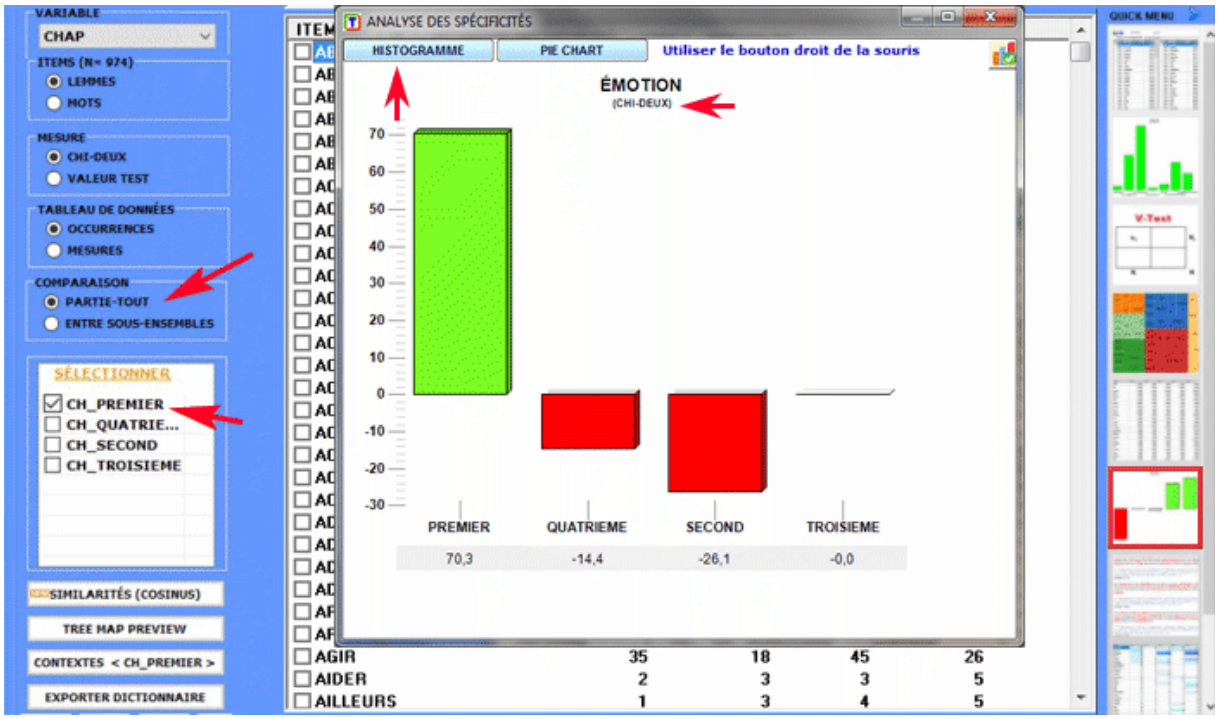


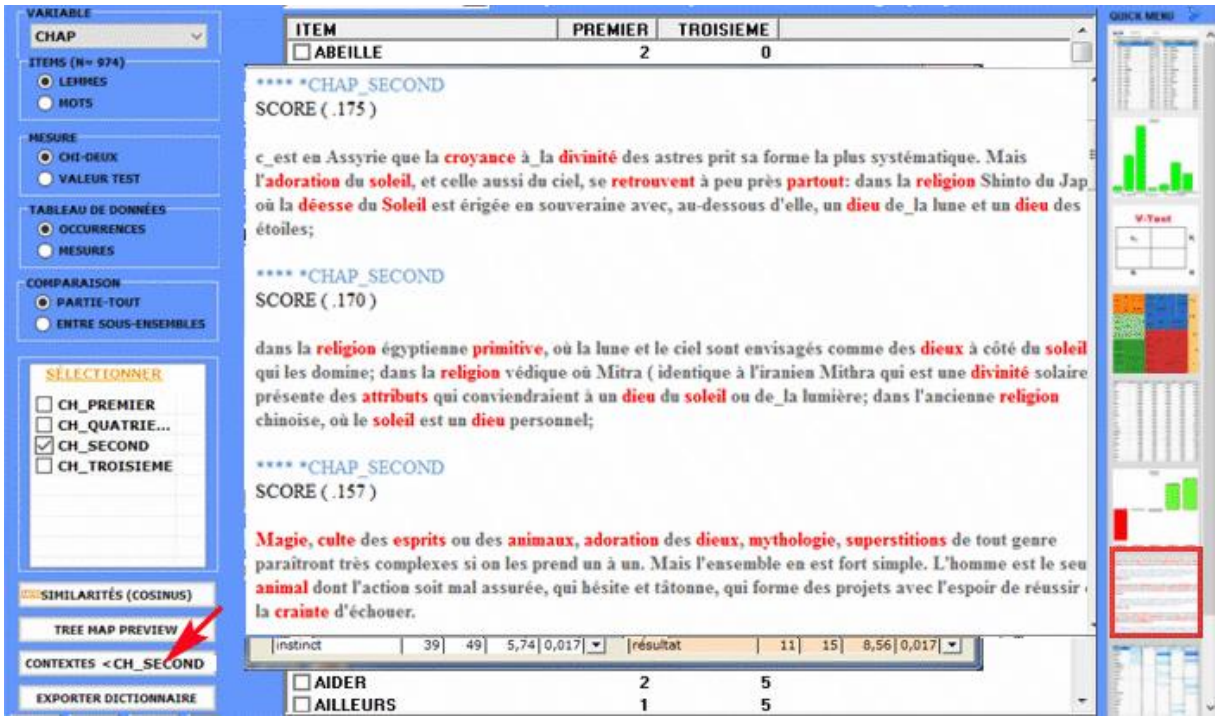
L'**Analyse des Spécificités** permet de vérifier quels mots sont “typiques” ou “exclusifs” de chaque sous-ensemble du corpus. En outre il nous permet d'extraire les contextes typiques, c'est-à-dire les contextes élémentaires caractéristiques, de chacun des sous-ensembles analysés (par exemple, les phrases "typiques" utilisées par certains leaders politiques).



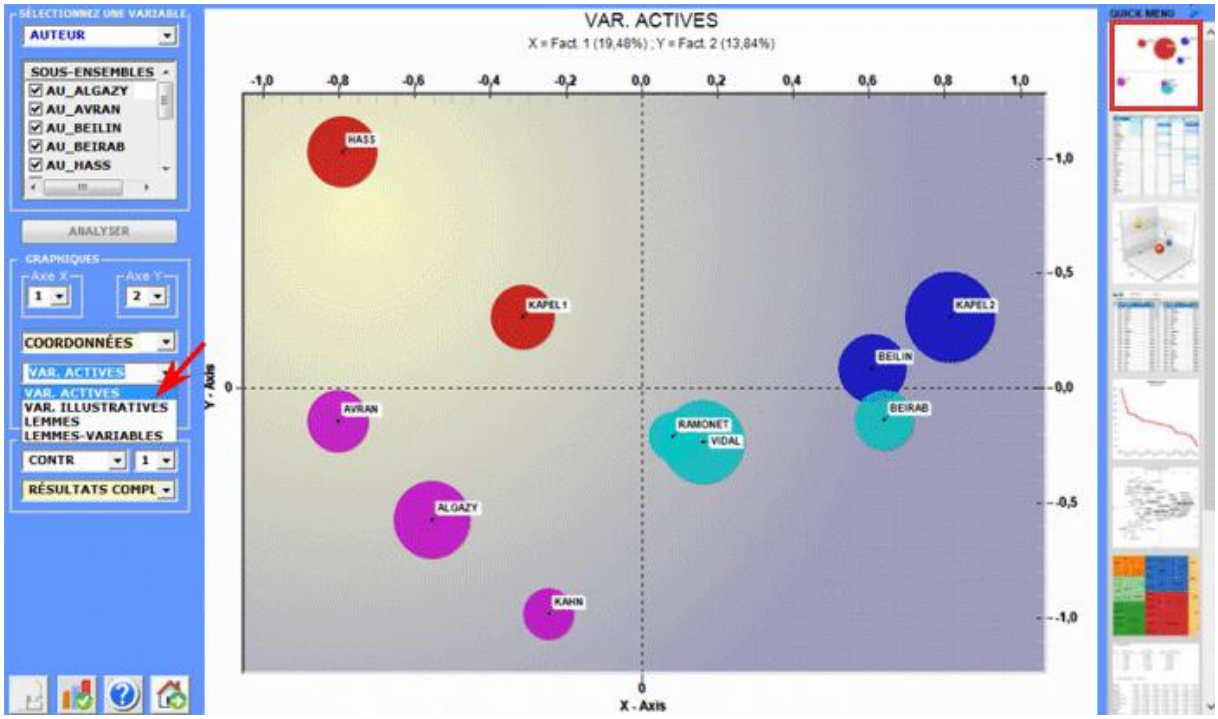
LEMME	SUB	TOT	CHI²	(p)
obligation	132	145	246,23	0,000
morale	147	188	197,67	0,000
justice	45	53	72,33	0,000
émotion	62	85	70,34	0,000
social	95	163	57,31	0,000
société	161	324	53,88	0,000
pression	30	39	38,62	0,000
habitude	47	75	35,39	0,000
sentiment	42	65	34,55	0,000
respect	18	20	32,65	0,000
devoir_amb	24	31	31,35	0,000
impératif	15	16	29,54	0,000
maxime	14	15	27,33	0,000
Socrate	13	14	25,11	0,000
obliger	14	16	23,95	0,000
aspiration	20	27	23,51	0,000
règle	21	29	23,35	0,000
devoirs	12	13	22,91	0,000
moi	22	32	21,42	0,000
obligatoire	14	17	21,03	0,000
échange	10	11	18,51	0,000
fins	10	11	18,51	0,000
sensibilité	12	15	16,89	0,000
égalité	13	17	16,49	0,000

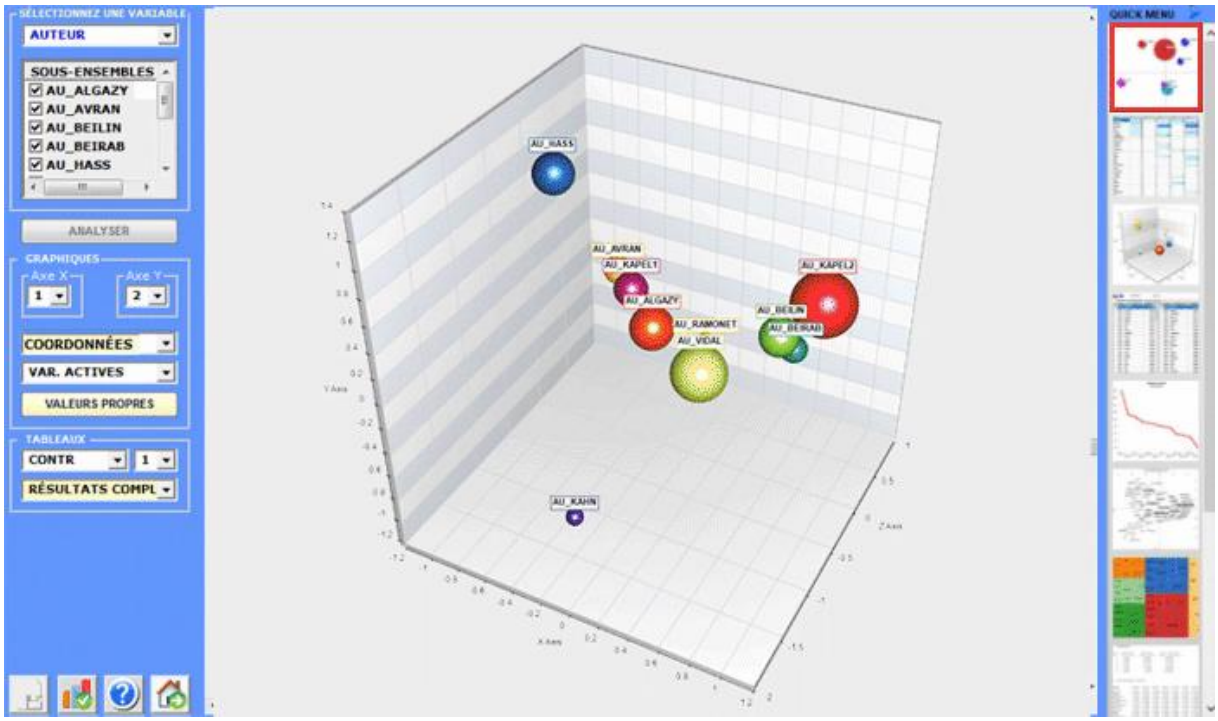
LEMME	SUB	TOT	CHI²	(p)
dieu	14	206	56,68	0,000
religion	16	211	54,35	0,000
science	6	107	32,29	0,000
mysticisme	2	83	31,74	0,000
esprit	16	139	24,71	0,000
mystique	14	122	21,75	0,000
expérience	7	85	20,59	0,000
croyance	1	51	20,10	0,000
primitif	11	103	19,89	0,000
fonction	6	78	19,80	0,000
corps	8	81	16,89	0,000
terre	4	57	15,31	0,000
guerre	4	50	12,35	0,000
mécanique	1	32	11,61	0,001
animal	8	67	11,36	0,001
vital	3	42	11,16	0,001
mourir	1	26	8,95	0,003
produit	1	26	8,95	0,003
opération	1	25	8,51	0,004
danger	1	24	8,06	0,005
réaction	1	24	8,06	0,005
univers	1	23	7,62	0,006
mentalité	2	28	7,43	0,006
invention	4	37	7,03	0,008



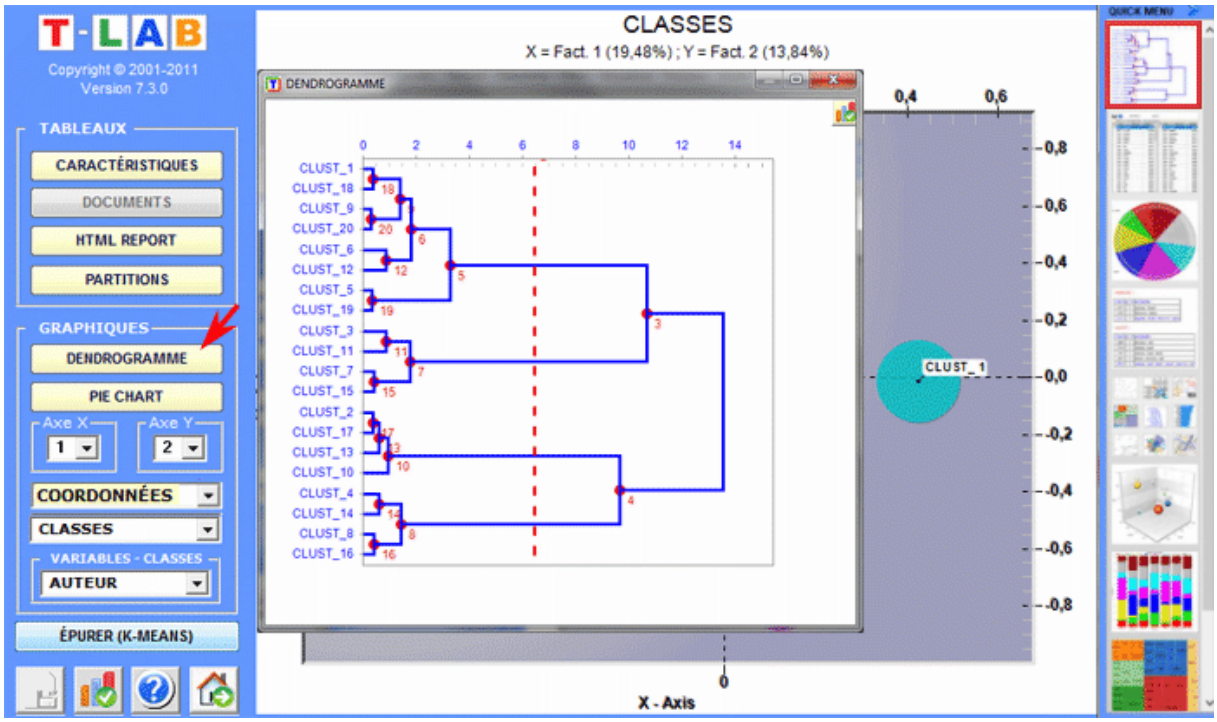


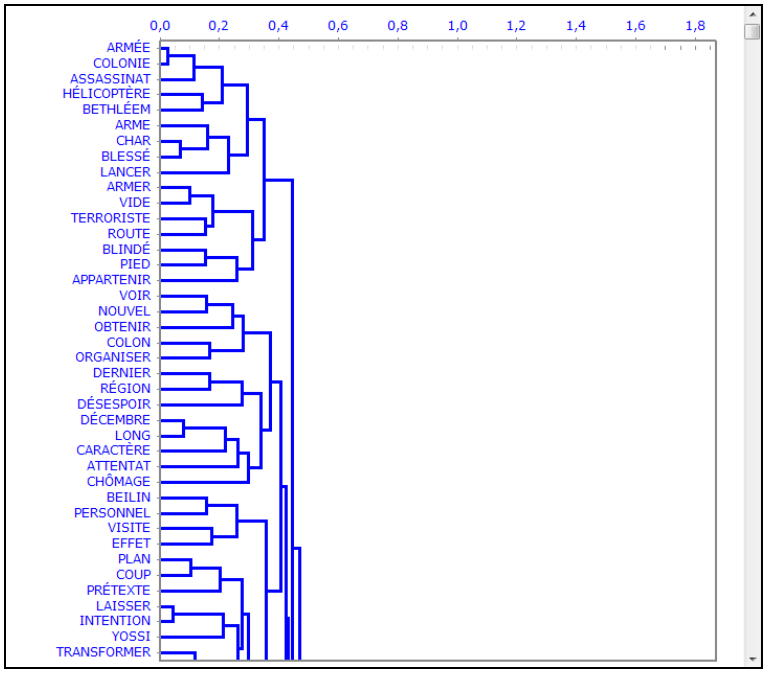
L'**Analyse des Correspondances** permet d'explorer différentes typologies de relations (différences et ressemblances) entre les unités de contexte.





La **Cluster Analysis**, qui peut être réalisée avec différentes techniques, permet d'identifier des groupes d'unités textuelles qui aient deux caractéristiques complémentaires: maximum d'homogénéité dans leur interne et maximum d'hétérogénéité entre eux deux et les autres clusters.





C : OUTILS POUR LES ANALYSES THÉMATIQUES

Ces outils permettent de repérer, examiner et cartographier les "thèmes" présents dans les textes analysés.

Puisque "thème" est un mot polysémique, dans ce cas il est utile se référer à des définitions opérationnelles. En fait, dans ces outils de **T-LAB**, le mot "thème" est un label utilisé pour indiquer quatre entités différentes :

1- un **cluster thématique** d'unité de contexte caractérisé par les mêmes patterns de mots-clés (voir **Analyse Thématique des Contextes Elémentaires** et **Classification thématique des Documents)**;

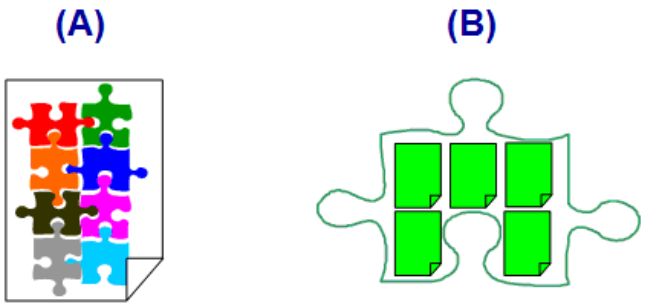
2- un **groupe thématique de mots-clés** classés comme appartenant à la même catégorie (voir l'outil **Classification basée sur des Dictionnaires**);

3- un **élément d'un modèle probabiliste** qui représente chaque unité de contexte (soit un contexte élémentaire, soit un document), comme généré par un mélange de "thèmes" ou "topics" (voir les outils **Modélisation des Thèmes émergentes** et **Textes et Discours comme Systèmes Dynamiques**);

4- un **mot-clé** ("thématique") **spécifique** utilisé pour extraire un ensemble de contextes élémentaires dans lesquels ce mot est associé à un groupe de mots spécifique présélectionnés par l'utilisateur (voir **Contextes-Clé de Mots Thématiques**).

Par exemple, selon le type d'outil que nous sommes en train d'utiliser, un document spécifique peut être analysé comme étant composé de différents « thèmes » (voir «A» ci-dessous) ou bien comme appartenant à un ensemble de documents concernant le même «thème» (voir « B » ci-dessous). En effet, dans le cas «A» chaque thème peut correspondre à

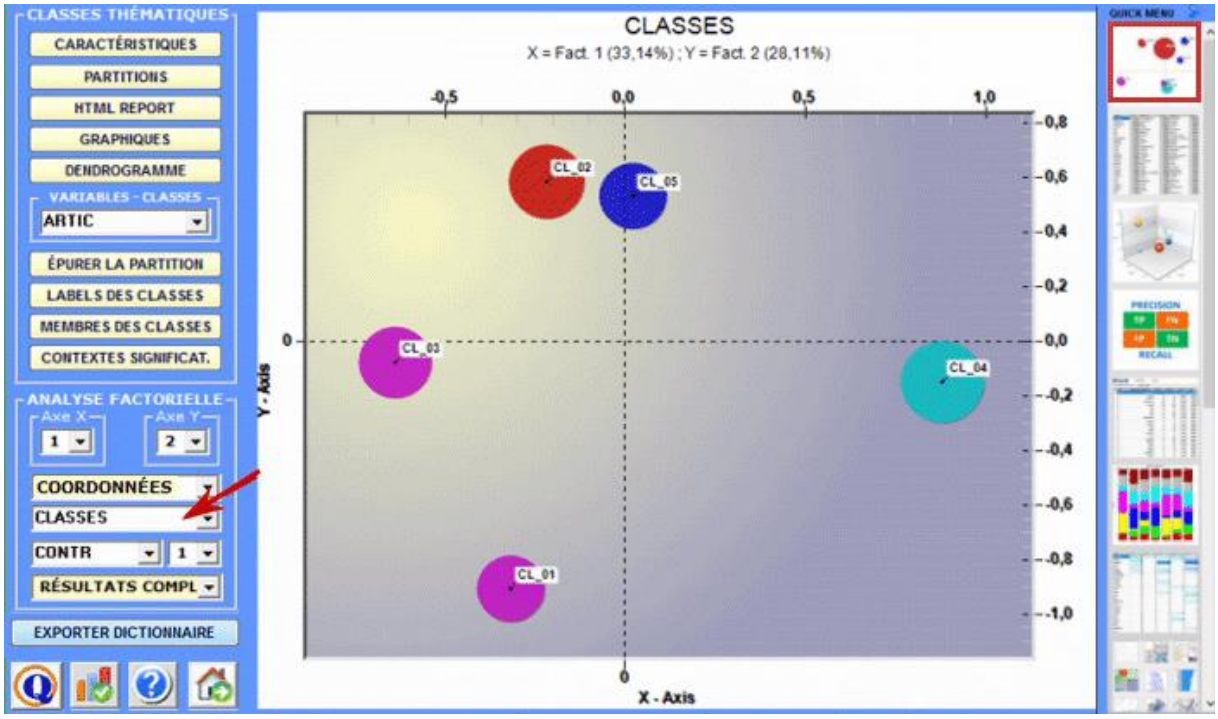
un mot ou bien à une phrase, tandis que dans le cas «B» un thème peut être une étiquette attribuée à un groupe de documents caractérisés par les mêmes patterns de mots-clés.

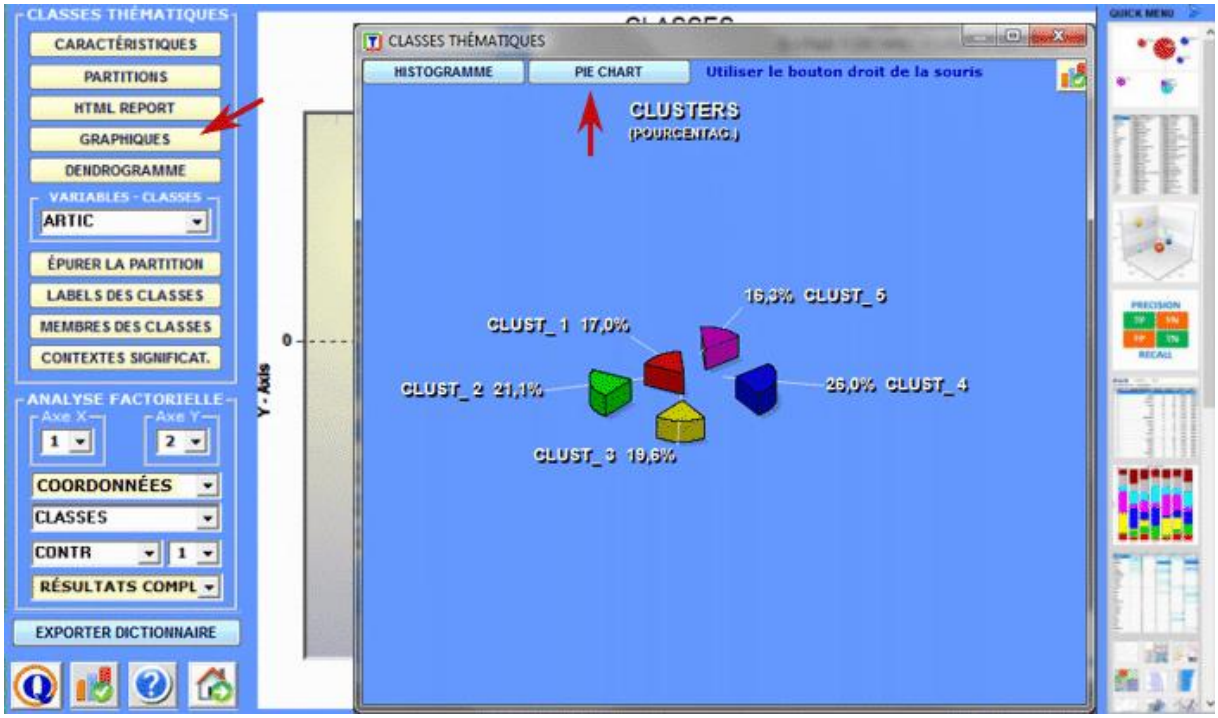


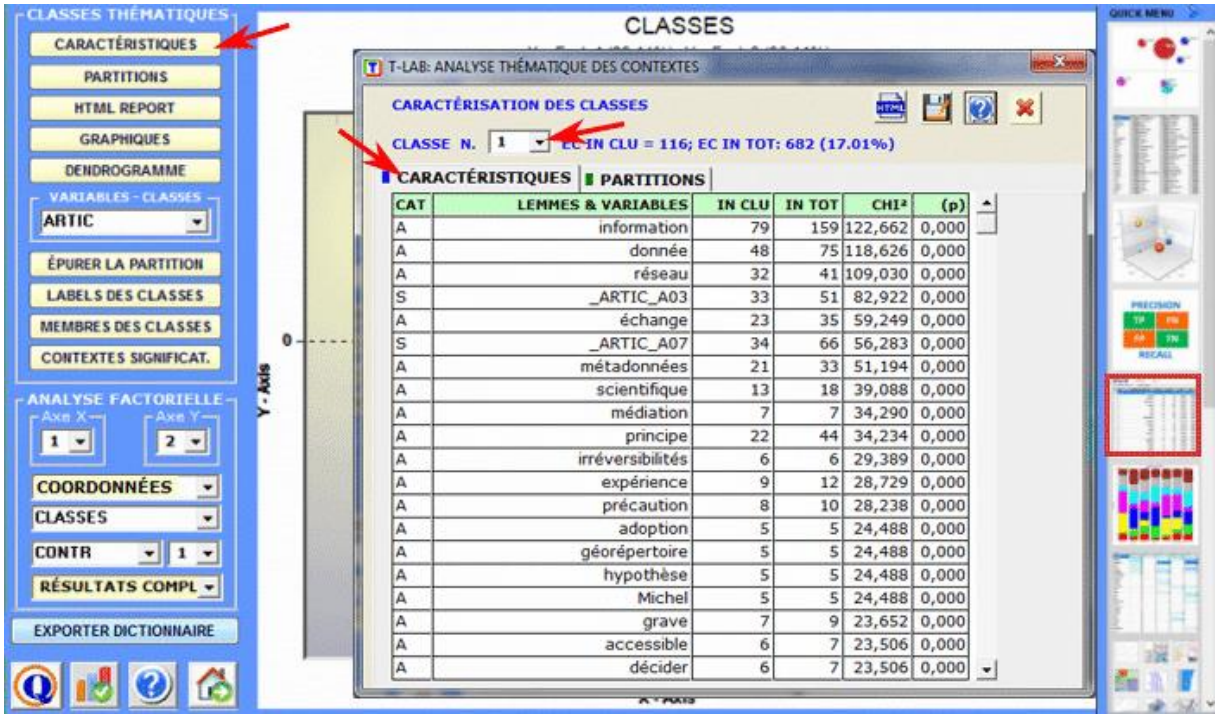
En détail, les façons dont **T-LAB** extrait les thèmes sont les suivantes:

1 - soit l'outil **Analyse Thématiques des Contextes Elémentaires**, soit l'outil **Classification Thématique des Documents** fonctionnent de manière suivante :

a- ils réalisent une **analyse des cooccurrences** pour identifier les classes thématiques de unités de contexte;
b- ils réalisent une **analyse comparative** pour confronter les profils des différentes classes;
c- ils produisent différents types de graphiques et de tableaux (voir ci-après);
d- ils permettent d'archiver les **nouvelles variables** obtenues (classes thématiques) et de les utiliser dans d'autres analyses.

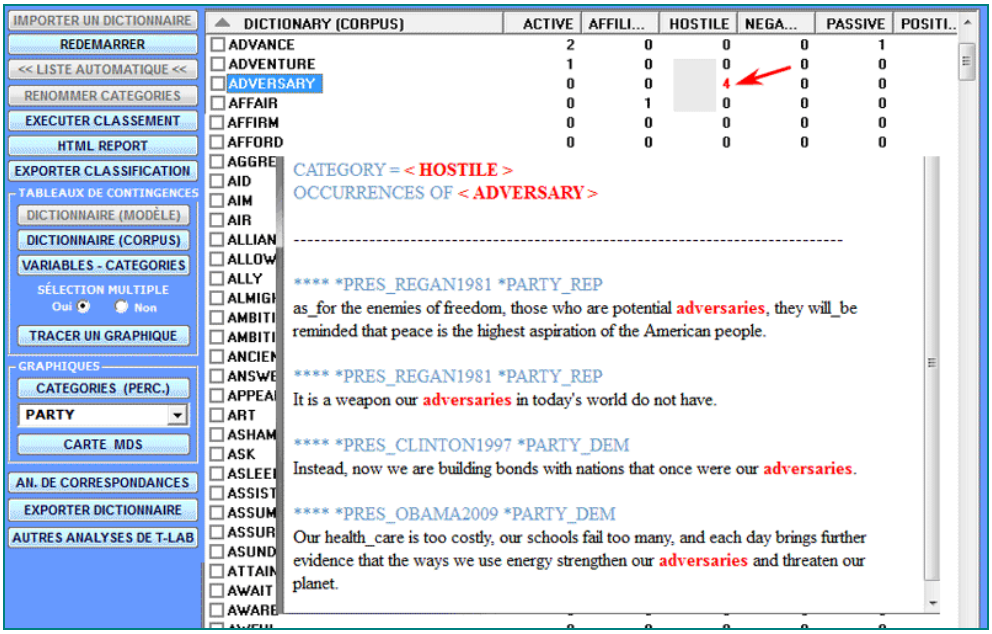


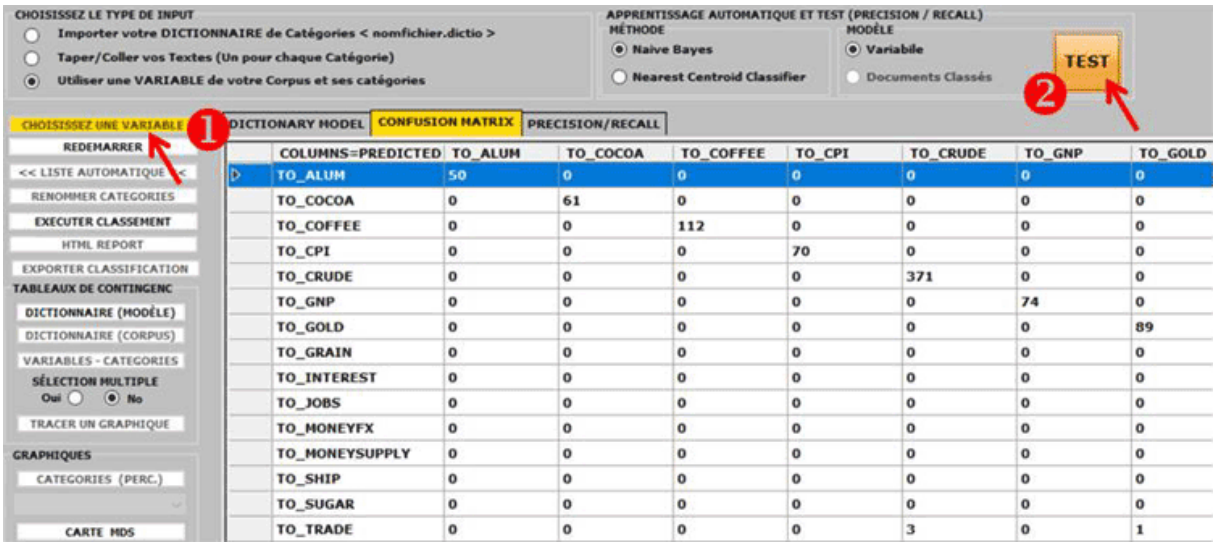




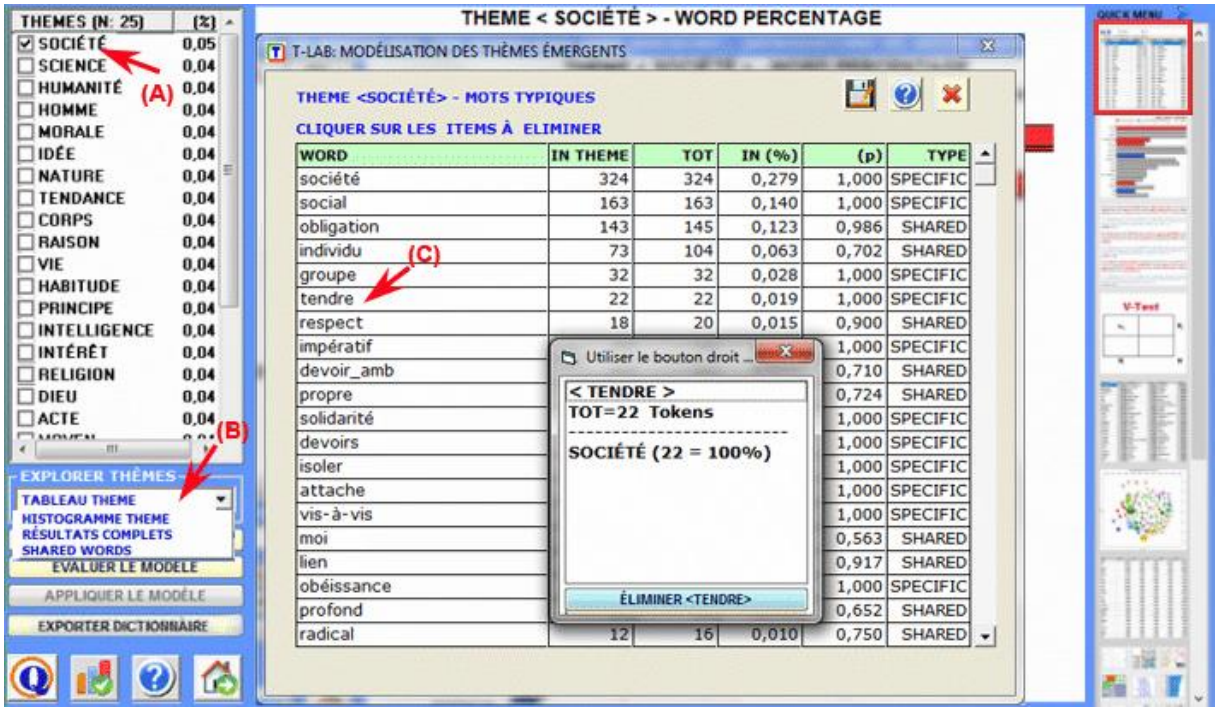
CAT	LEMMES & VARIABLES	IN CLU	IN TOT	CHI²	(p)
A	information	79	159	122,662	0,000
A	donnée	48	75	118,626	0,000
A	réseau	32	41	109,030	0,000
S	_ARTIC_A03	33	51	82,922	0,000
A	échange	23	35	59,249	0,000
S	_ARTIC_A07	34	66	56,283	0,000
A	métadonnées	21	33	51,194	0,000
A	scientifique	13	18	39,088	0,000
A	médiation	7	7	34,290	0,000
A	principe	22	44	34,234	0,000
A	irréversibilités	6	6	29,389	0,000
A	expérience	9	12	28,729	0,000
A	précaution	8	10	28,238	0,000
A	adoption	5	5	24,488	0,000
A	géorépertoire	5	5	24,488	0,000
A	hypothèse	5	5	24,488	0,000
A	Michel	5	5	24,488	0,000
A	grave	7	9	23,652	0,000
A	accessible	6	7	23,506	0,000
A	décider	6	7	23,506	0,000

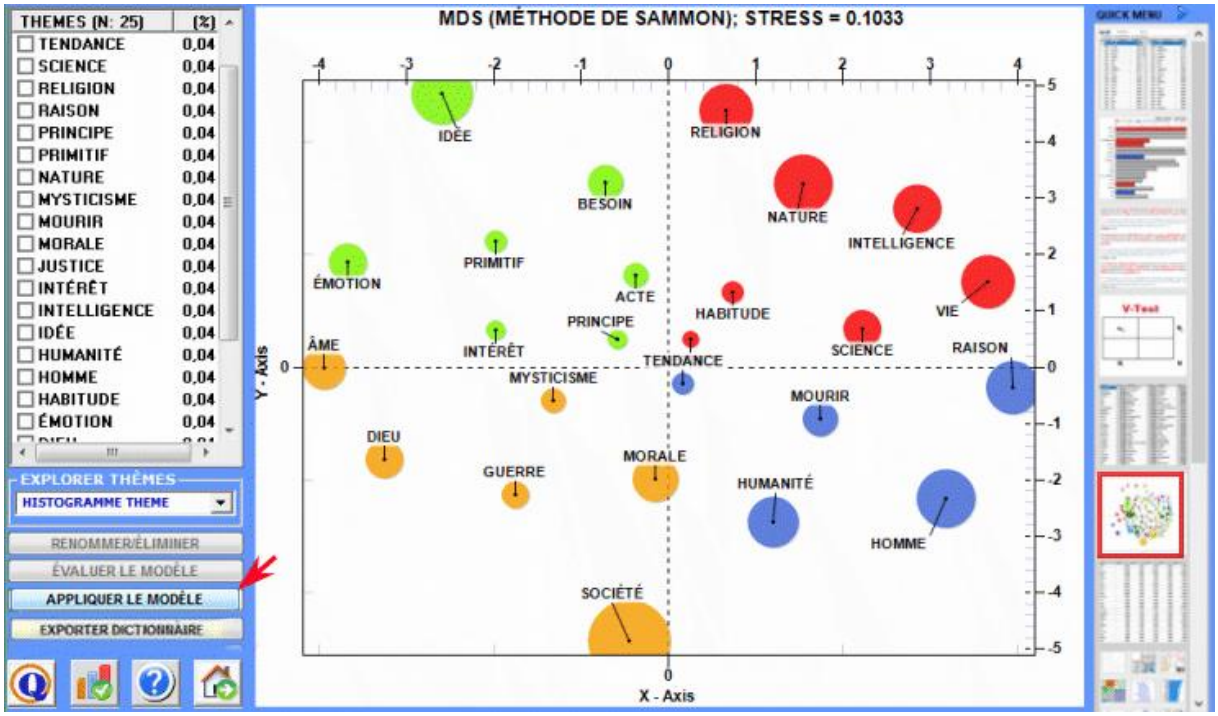
2 - à l'aide de l'outil **Classification Basée sur des Dictionnaires** nous pouvons facilement construire / tester / appliquer des modèles (par ex. des dictionnaires de catégories) soit pour l'analyse classique du contenu soit pour la sentiment analysis. En effet cet outil nous permet d'effectuer une classification automatique de type top-down aussi bien des unités lexicales (c'est-à-dire mots et lemmes) que des unités de contexte (c'est-à-dire phrases, paragraphes et documents courts).



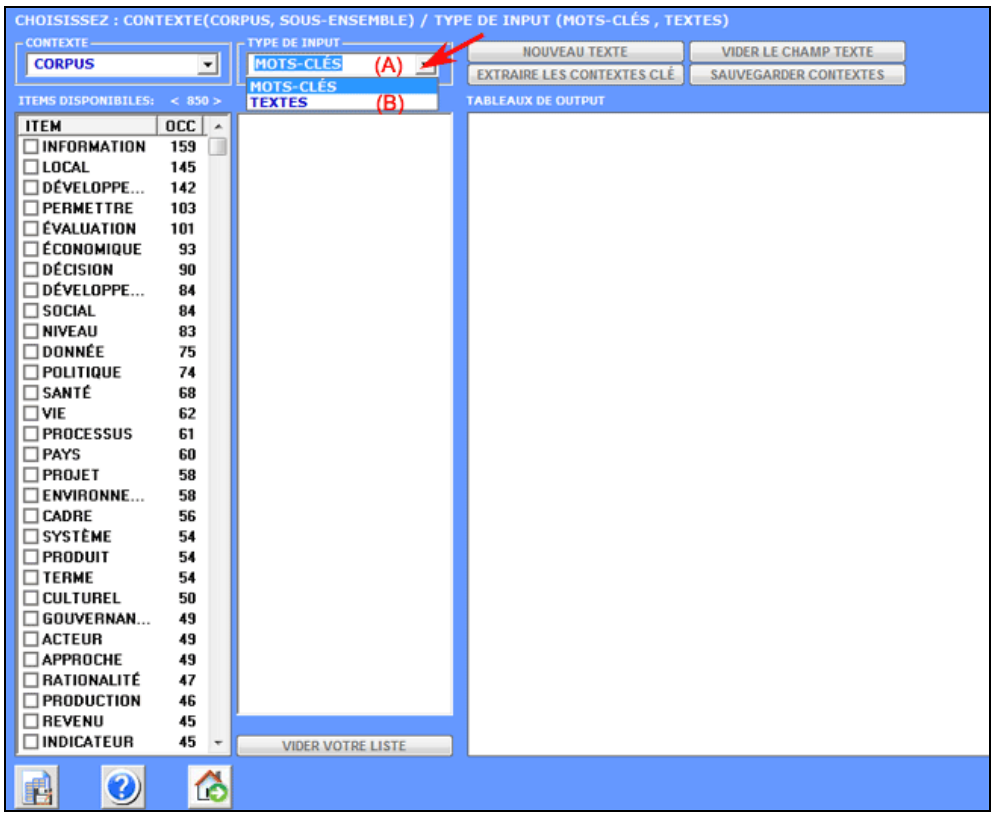


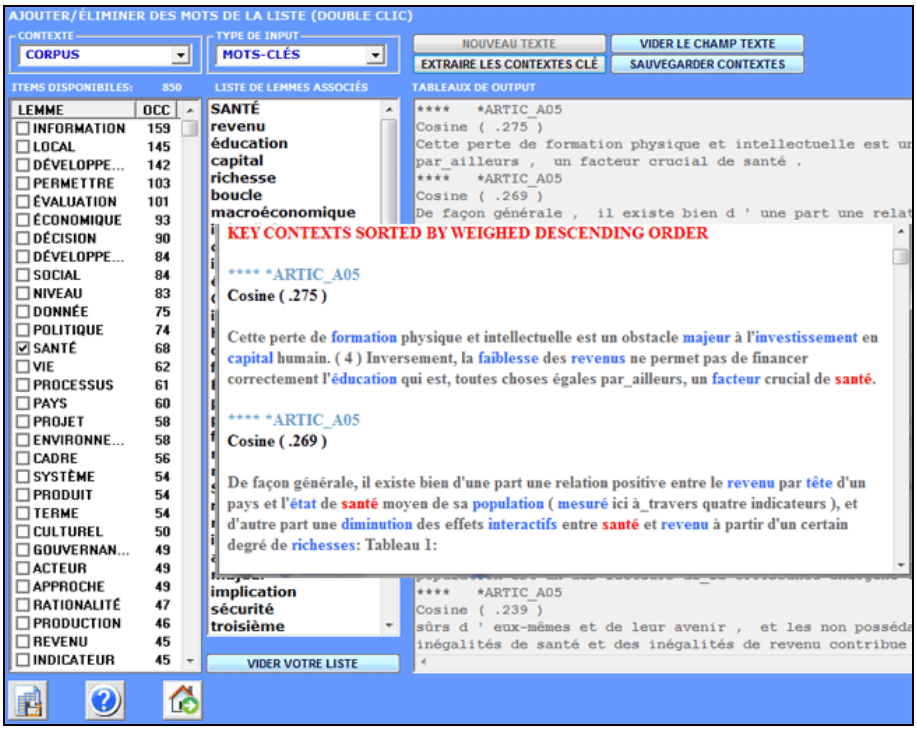
3 - grâce à l'outil **Modélisation des Thèmes Émergents** (voir ci-dessous) les composants du «mélange» thématique peuvent être décrits par leur vocabulaire caractéristique et peuvent être utilisés pour la construction de grilles pour l'analyse qualitative et / ou pour la classification automatique des unités de contexte (c'est-à-dire contextes élémentaires ou documents).





4 - l'outil **Contextes Clé des Mots Thématiques** (voir ci-dessous) peut être utilisé pour deux buts différents: (a) extraire des listes d'unités de contexte (c'est-à-dire contextes élémentaires) qui permettent d'approfondir la valeur thématique de **mots-clés** spécifiques, (b) extraire des groupes d'unités de contexte qui sont semblables à n'importe quel **texte** « exemple » choisi par l'utilisateur.

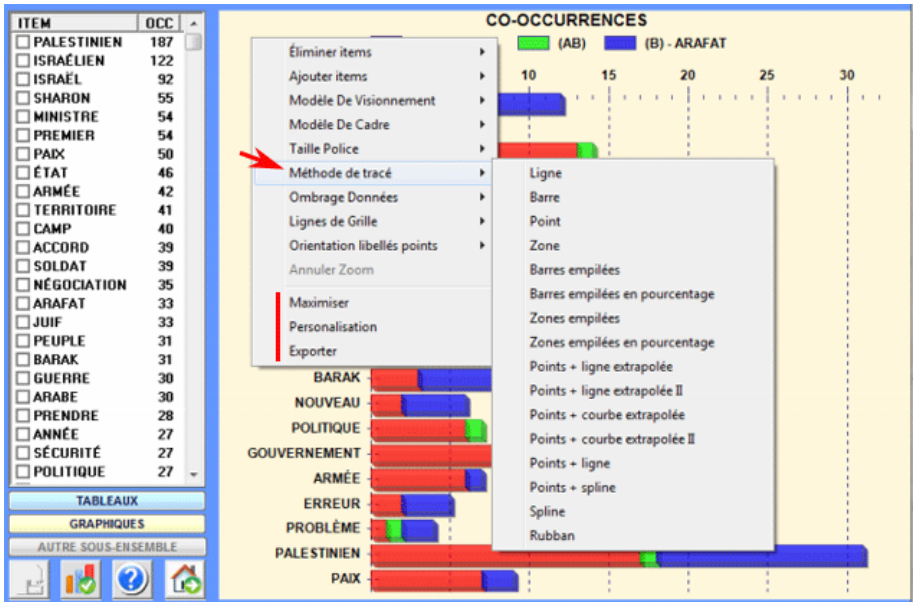


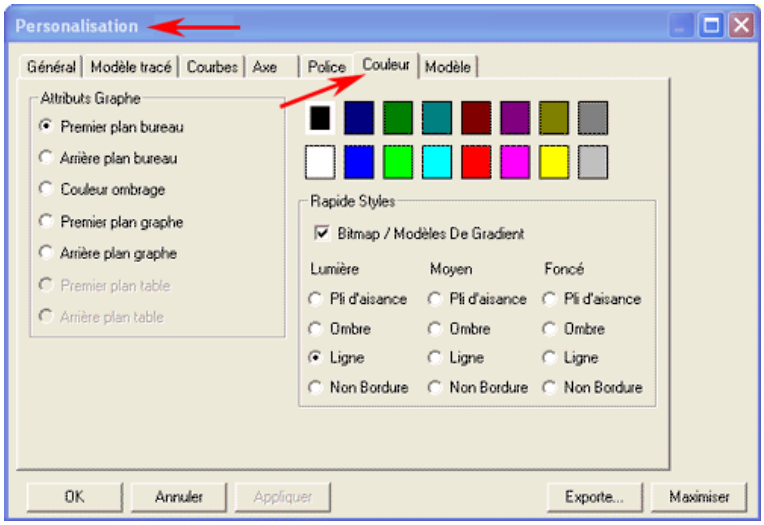


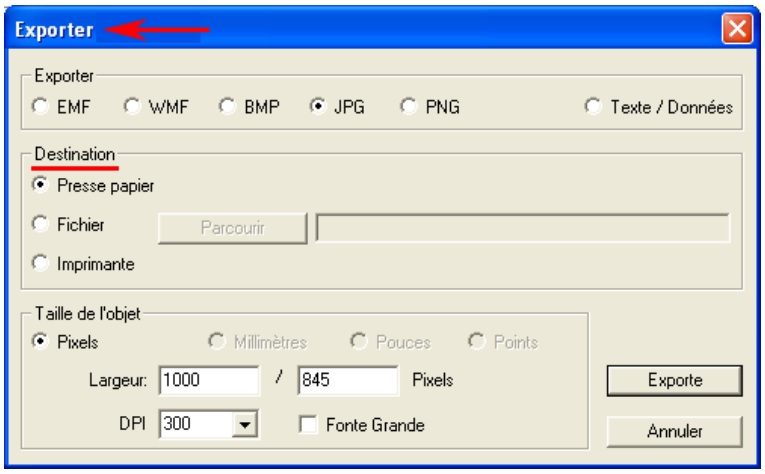
6 - L' INTERPRÉTATION DES OUTPUTS consiste en la consultation des tableaux et des graphiques produits par **T-LAB**, en l'éventuelle personnalisation de leur format et dans le fait de faire des inférences sur la signification des relations représentées.

Dans le cas des **tableaux**, selon les cas, **T-LAB** permet de les exporter dans des fichiers avec les extensions suivantes: **.DAT**, **.TXT**, **.CSV**, **.XLXS**, **.HTML**. Ceci signifie que, en se servant de n'importe quel éditeur de textes et/ou d'un applicatif de la suite Microsoft Office, l'utilisateur peut facilement les importer et les réélaborer.

Dans le cas des **graphiques**, les sous-menus appropriés activés avec le clic droit de la souris permettent d'effectuer plusieurs opérations: zoom (clic gauche et glisser), maximisation, personnalisation et exportation des outputs en plusieurs formats.





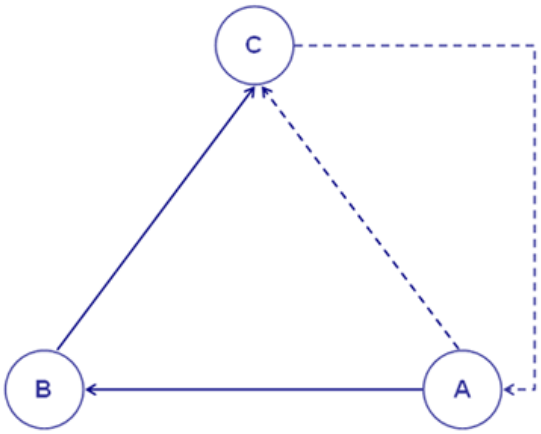


Certains critères généraux pour l'interprétation des outputs **T-LAB** sont illustrés dans un papier cité dans la **Bibliographie** et disponible sur le site https://www.tlab.it (Lancia F.: 2007). Dans ce dernier on propose l'hypothèse que les outputs des élaborations statistiques (tableaux et graphiques) sont un type particulier de textes, c'est-à-dire des objets multi-sémiotiques caractérisés par le fait que les relations entre les signes et les symboles sont ordonnées par des mesures qui renvoient à des **codes** spécifiques.

Dans d'autres termes, aussi bien dans le cas des textes écrits dans le langage naturel que dans ceux écrits dans le langage de la statistique, la possibilité de faire des inférences sur les relations qui organisent les **formes du contenu** est garantie par le fait que les relations entre les **formes de l'expression** ne sont pas casuelles (random); en effet, dans le premier cas (langage naturel) les unités signifiantes se succèdent ordonnées de façon linéaire (l'une après l'autre dans le chaîne du discours), alors que dans le second cas (tableaux et graphiques) les principes d'ordonnance sont constitués par les mesures qui déterminent l'organisation des **espaces sémantiques** multidimensionnels.

Même si les espaces sémantiques représentés dans les cartes **T-LAB** sont très variés, et chacun d'eux requiert des procédures interprétatives spécifiques, nous pouvons faire l'hypothèse que - en général - la logique du processus inférentiel est la suivante:

A - relever une relation significative entre les unités "présentes" sur le plan de l'expression (par ex. entre "données" des tableaux et/ou entre "labels" des graphiques);
B - explorer et confronter les traits sémantiques des mêmes unités et les contextes auxquels elles sont mentalement et culturellement associées (plan du contenu);
C - construire une hypothèse ou une catégorie d'analyse qui, dans le contexte défini par le corpus, rendent raison des relations entre formes de l'expression et formes du contenu.



Actuellement les options de **T-LAB** ont les **limitations** suivantes:

- dimension du corpus: max 90Mo correspondant à environ 55.000 pages de format .txt;
- documents primaires : max 30.000 (N.B.: Lorsque aucun des textes dépasse 2.000 caractères, la limite est de 99.999 documents);
- variables catégorielles: maximum 50, chacune avec un maximum de 150 modalités;
- modélisation des thèmes émergents : max 5.000 unités lexicales (*) pour 5.000.000 occurrences;
- analyse thématique des contextes élémentaires: max 300.000 lignes (unités de contexte) par 5.000 colonnes (unités lexicales);
- classification thématique des documents: max 99.999 lignes (unités de contexte) par 5.000 colonnes (unités lexicales);
- analyse des spécificités (unités lexicales x modalités d'une variable): max 10.000 lignes x 150 colonnes;
- analyse des correspondances (unités lexicales x modalités d'une variable): max 10.000 lignes x 150 colonnes;
- analyse des correspondances (unités de contexte x unités lexicales): max 10.000 lignes x 5.000 colonnes;
- analyse des correspondances multiples (contextes élémentaires x modalités des variables): max 150.000 lignes x 250 colonnes;
- décomposition en valeurs Singulières (SVD) : max 300.000 lignes par 5.000 colonnes;
- classification (cluster analysis) qui emploie les résultats d'une précédente analyse des correspondances (ou SVD): max10.000 lignes (unités lexicales ou contextes élémentaires);
- association des mots, comparaisons entre paires de mots-clés: max 5.000 unités lexicales;
- analyse des mots associés et cartes conceptuelles: max 5.000 unités lexicales;
- analyse de séquences: maximum 5.000 unités lexicales (ou catégories) pour 3.000.000 occurrences.

(*) Dans **T-LAB**, les 'unités lexicales' sont mots, multi-mots, lemmes et catégories sémantiques. Ainsi, lorsque la lemmatisation automatique est appliquée, 5.000 unités lexicales correspondent à environ 12.000 mots.