

Le traitement informatique du discours juridique dans la presse française avec le logiciel *Tropes*

Yacoub Ghérissi

gh.yac.2013@gmail.com

Résumé

Notre article est une application du Traitement automatique du langage (TAL) en vue d'extraire, à partir d'un corpus journalistique, le vocabulaire spécifique au domaine juridique. Il s'agit d'utiliser les outils de fouille des données pour d'éventuelles applications notamment en traduction automatique. Nous utiliserons à cet effet le logiciel *Tropes*, téléchargeable gratuitement sur Internet.

Mots clés : Traitement Automatique du langage, fouille de données, Logiciel *Tropes*, droit, analyse textuelle, presse.

1. Introduction

Le présent article se propose de présenter le logiciel *Tropes*¹ qui est capable de traiter différents types de fichier : *.doc, *.pdf, *.txt, *.html, etc., et ce dans trois langues, le français, l'anglais et l'espagnol. C'est un logiciel automatique, rapide, doté d'une interface graphique simple et moderne.

Nous proposons une méthode qui scrute les textes de tout bord pour en déceler les traces du langage juridique, qui n'est pas, bien évidemment, l'apanage des textes strictement juridiques.

Le corpus que nous allons utiliser est composé de textes journalistiques issus des livraisons quotidiennes de journaux français, *Le Monde* et *L'Est-Républicain*.²

Nous expliquons dans un premier temps les soubassements théoriques de l'analyse de contenu appliquée par *Tropes* et les spécificités du discours juridique. Dans un second temps, nous montrons comment configurer le logiciel pour fouiller dans des fichiers tels que des journaux ou autres supports numériques écrits.

Grâce à son menu *Scénario*, utilisé comme tel et modulable à souhait, le logiciel permet de relever tous les énoncés d'un texte qui comportent des termes juridiques.

En ce qui concerne notre propos, l'application que nous allons présenter intéresse aussi bien les avocats que les journalistes ou les simples citoyens qui lisent les versions numériques des journaux et qui aimeraient bien retrouver (rapidement) les articles qui traitent des affaires judiciaires.

La simple question à laquelle nous voudrions répondre ici est de savoir dans quels articles d'une livraison il est question de droit ou de justice.

¹ *Tropes* est téléchargeable gratuitement sur internet à l'adresse <https://www.tropes.fr/download.htm>. Nous utilisons la dernière version 8.5.0.

² Les livraisons de ces journaux sont facilement téléchargeables sur Internet moyennant un abonnement pour les livraisons anciennes même d'un jour.

2. Analyse de contenu et lexicométrie

Le départ de conception de *Tropes* est une tentative de dépassement de l'analyse thématique qui définit arbitrairement les unités de codage et le choix subjectif des indicateurs utilisés pour l'interprétation des textes. A cet effet, Rodolphe Ghiglione et le Groupe de Recherche sur la Parole (GRP, Université Paris 8) ont élaboré successivement deux théories d'analyse de contenu : l'Analyse Propositionnelle du Discours (APD) puis l'Analyse Cognitive Discursive (ACD).

En appliquant une analyse lexicométrique, *Tropes* découpe le texte en propositions (phrases simples appelées *propositions remarquables*). Cette première étape est réalisée grâce à un simple examen de la ponctuation. Ceci permet également d'obtenir des statistiques de co-occurrence très fiables (appelées "mises en relations"), puisque deux mots ne peuvent pas se trouver dans la même proposition grammaticale s'ils ne sont pas étroitement liés.

Ces théories se fondent également, comme le souligne Pierre Molette³ « sur la catégorisation sémantique des mots outils, l'identification de classes paradigmatiques de substantifs [...] et la modélisation des propositions sous un formalisme simplifié (appelé "modèle argumentatif" dans l'APD ou "noyau générateur" et "structure fondamentale de la signification" dans l'ACD). [...] Il s'agit d'extraire du texte une série de variables qui [font] l'objet d'un traitement statistique permettant de révéler des résultats d'analyse objectifs, qui ne seraient pas forcément identifiés après une lecture approfondie du texte ».

Avec la huitième version qui a nécessité une quinzaine d'années d'expérimentation et de vérifications, *Tropes* est devenu une plate-forme d'analyse sémantique de contenu très performante.

3. Le discours juridique

Le discours juridique utilise un langage difficilement accessible aux non-initiés. L.-M. Raymondis et M. Le Guern ont mené, dans les années 1970, une enquête sur le langage de la justice pénale française afin d'en déceler les ambiguïtés et les « dysfonctionnements observés lors de la communication entre les utilisateurs du service public de justice et ceux qui ont pour charge de la faire fonctionner »⁴.

Par ailleurs, nous savons que chaque type de texte présente des spécificités lexicales et grammaticales telles que des suites de mots qui existent dans l'un sont absentes dans les autres. Une de ces spécificités consiste donc dans la fréquence assez élevée dans le corpus de structures ou suites de mots propres au domaine considéré⁵ ; ce qui laisse supposer que ces structures appartiennent assez largement à la pratique linguistique d'un groupe d'utilisateurs de la langue, et pour le cas qui nous intéresse cette pratique est celle des hommes de loi. Si dans le langage

³ Pierre Molette, « De l'APD à Tropes : comment un outil d'analyse de contenu peut évoluer en logiciel de classification sémantique généraliste », 2009.

⁴ Raymondis et Le Guern, *Le langage de la justice pénale*, p.2.

⁵ Une des limites des statistiques lexicales est l'absence de résolution de l'anaphore pronominale. Une étape réalisable à l'utilisation de ce type de logiciel consisterait à remplacer les termes anaphoriques par les termes qu'ils anaphorisent. Ainsi les comptes seront bons.

courant les substantifs sont, par exemple, le plus souvent actualisés par des déterminants, dans le langage juridique, nous relevons des structures sans détermination comme :

- décerner Ø mandat d'arrêt*
- déposer Ø copie*
- il résulte Ø charges suffisantes*
- obtenir Ø paiement*

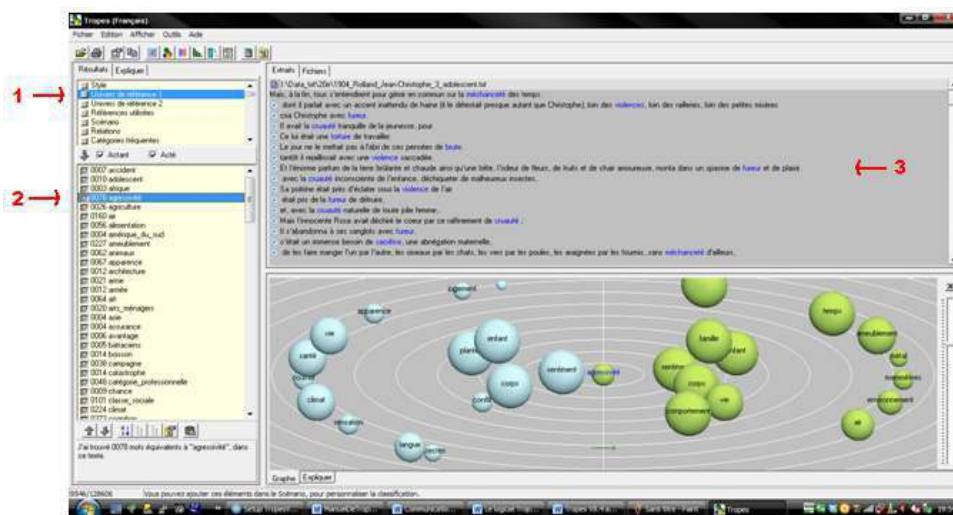
Le langage juridique, qui présente à plusieurs égards des réminiscences des états anciens du français, a aussi tendance à omettre l'article défini ou indéfini devant les noms qui suivent les prépositions :

- selon reconnaissance de Ø dettes*
- sur renvoi d' Ø information*
- s'en remettre à Ø justice*

pour ne citer que ces exemples.

4. L'analyse avec Tropes

Tropes propose quelques types d'analyse très intéressantes. Ces analyses, appliquées à des données numériques écrites, se présentent sous forme de menus dès la fenêtre principale⁶. Un clic sur l'un de ces menus ouvre instantanément les sous-menus respectifs dans la fenêtre en dessous. Si on clique, par exemple, sur le deuxième menu « Univers de référence 1 » (1), on obtient une liste de termes qui correspondent aux grands thèmes classés par ordre de fréquence décroissante. En cliquant sur l'un des items du sous-menu (2), instantanément, le résultat s'affiche dans la fenêtre supérieure droite⁷ (3).



Les menus proposés sont :

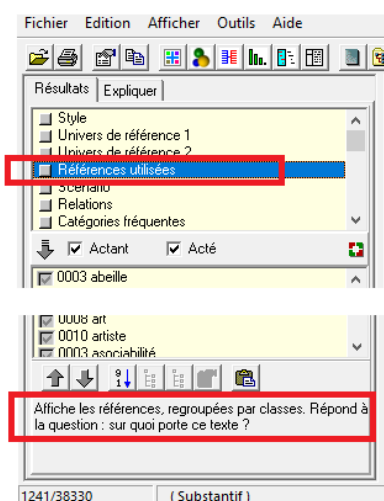
- *Style* : affiche des indications sur le style général du texte et sur le type de mise en scène
- *Univers de référence 1* : affiche le contexte général avec les grands thèmes classés par ordre de fréquence décroissante

⁶ Au lancement du logiciel, la fenêtre est encore vide et il faut fournir au programme le ou les fichiers à traiter.

⁷ Les couleurs dans les résultats peuvent être changées avec le Menu *Edition* > *Couleurs*.

- *Univers de référence 2* : affiche le contexte détaillé avec les thèmes classés par ordre de fréquence décroissante
- *Références utilisées* : affiche les références groupées par classe afin de répondre à la question : “sur quoi porte le texte ?”
- *Scénario* : affiche, par fréquence décroissante, le contenu du scénario sous forme d’une arborescence
- *Relations* : affiche les résultats des relations entre les mots trouvés dans le même ordre et la même proposition
- *Catégories fréquentes* : indique les statistiques des catégories de mots les plus fréquents
- *Toutes catégories de mots* : affiche les statistiques de toutes les catégories de mots, quelle que soit leur fréquence
- *Verbes* : affiche les lemmes de tous les verbes du texte
- *Adjectifs* : affiche tous les adjectifs
- *Substantifs* : affiche tous les substantifs
- *Épisode 1, 2, 3, n.* : segmente le texte proposé en séquences plus ou moins longues pour aérer la présentation en rafales⁸.

Chaque menu, sous-menu ou résultat est expliqué brièvement en bas à gauche de la fenêtre.



Excepté *Style*, chaque menu est développé en plusieurs sous-menus qui apparaissent dans la fenêtre en dessous.

En outre, le logiciel analyse le(s) texte(s) qu’on lui soumet en 6 étapes :

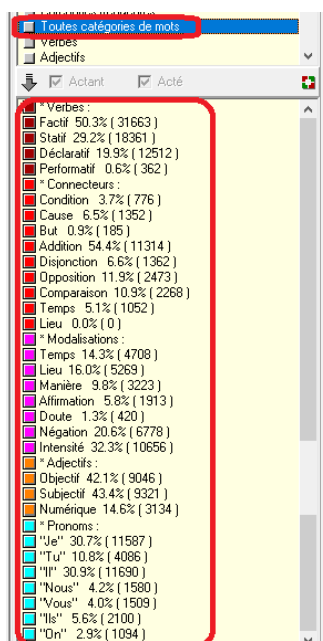
1. Il procède par un découpage en phrases et en propositions ;
2. Il permet de lever les ambiguïtés des mots du texte ;
3. Il identifie les classes d’équivalents sémantiques ;
4. Il donne les statistiques des classes de mots pleins (mots lexicaux) et des mots grammaticaux ;
5. Il détecte les propositions remarquables en vue d’une contraction du texte ;
6. Enfin, il met en forme et affiche les résultats.

Le fonctionnement de *Tropes* s’appuie sur une logique d’Intelligence Artificielle capable de résoudre les ambiguïtés lexicales et sémantiques des langues. Il intègre dans son moteur de recherche des dictionnaires et des grammaires qui permettent de classifier le vocabulaire d’un domaine dominant dans un texte pour en découvrir les réseaux sémantiques.

⁸ Voir le manuel du logiciel à l’adresse : <https://tropes.fr/ManuelDeTropesV840.pdf>.

Le logiciel commence par le processus d'indexation des mots du texte. L'utilisateur obtient ainsi les statistiques de cinq catégories, les plus significatives en langue, à savoir :

1. les verbes ;
2. les connecteurs (conjonctions de coordination et de subordination, locutions conjonctives) ;
3. les modalisations (adverbes ou locutions adverbiales) ;
4. les adjectifs qualificatifs ;
5. les pronoms personnels.



Mais ce qui nous intéresse présentement, c'est le menu *Scénario* qui propose une analyse très fine des textes.

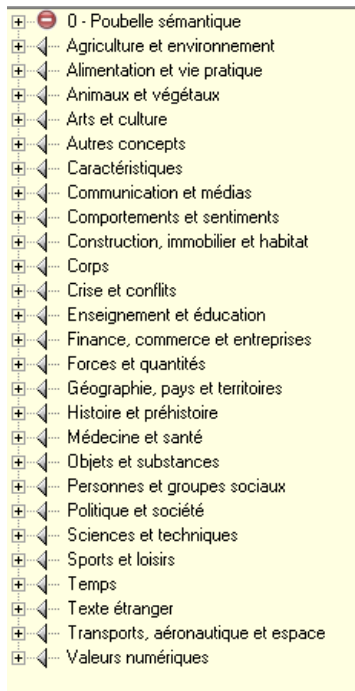
5. Le Scénario

Pour les mots pleins (Verbe, Nom, Adjectif et Adverbe), le logiciel intègre un scénario préconstruit où les verbes, les adjectifs et les adverbes sont ramenés aux seules formes nominales correspondantes. Inspirée de la linguistique anglo-saxonne et partiellement des travaux de John Lyons, la méthode se base sur un processus récursif d'analyse. L'opération consiste à retenir pour chaque mot son hypéronyme (le terme superordonné ou générique)⁹ et l'hypéronyme de cet hyperonyme, et ainsi de suite jusqu'à atteindre une profondeur de 6 niveaux. Au total, il y a 26 hyperclasses qui s'affichent selon l'ordre de fréquence du texte proposé.

Il s'agit donc d'une ontologie, sous forme d'arborescence qui fonctionne comme un dictionnaire indexant les mots par paradigmes allant des classes hypéronymiques aux hyponymes. Ce scénario est intégré à la base du logiciel et fournit à l'utilisateur la possibilité de voir les mots du texte classés selon l'arborescence de ce dictionnaire interne.

Voici une comparaison entre le scénario du logiciel et l'affichage du résultat obtenu lors du traitement d'un texte :

⁹ Il s'agit généralement de la classe supérieure qui correspond approximativement au début de la définition lexicographique.

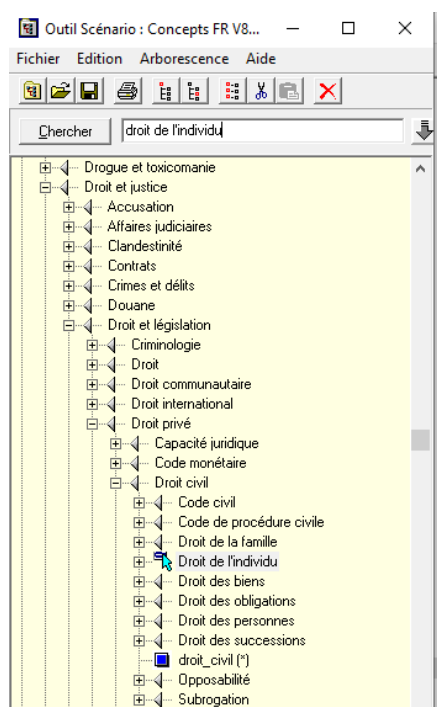


Le scénario préconstruit



Le scénario d'un texte donné

Le scénario de *Tropes* est visible depuis le menu *Outils > Outils scénario* du logiciel :



Pour avoir une idée sur ce scénario préconstruit, on peut forcer l'ouverture du fichier *Concept FR V8.scn*, placé lors de l'installation du logiciel dans le répertoire d'accueil, avec un éditeur de texte brut (*Bloc-notes* de Windows ou *Notepad++*)¹⁰.

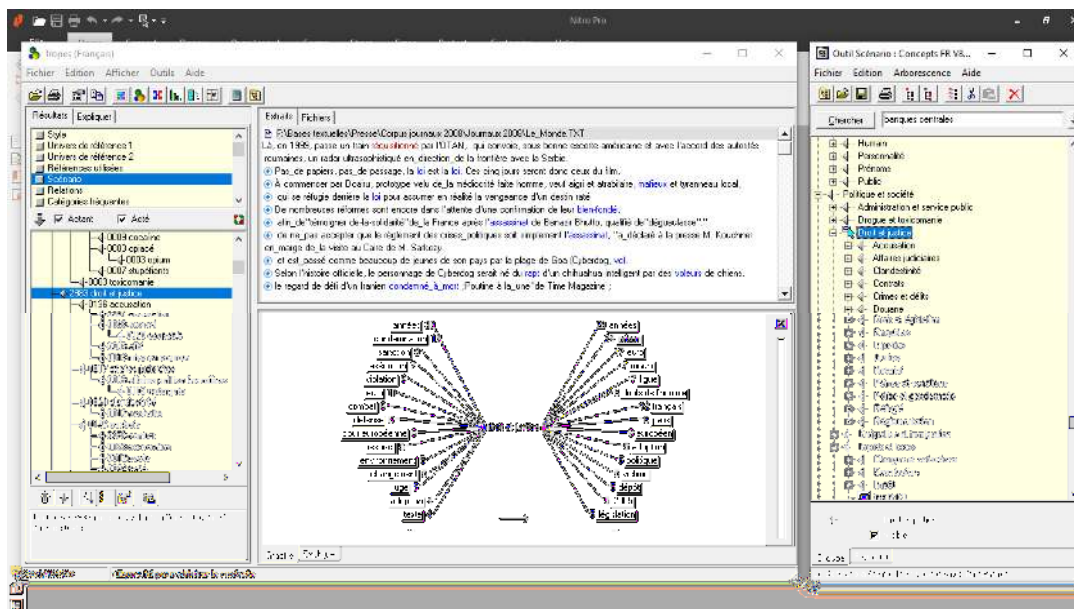
¹⁰ Bien qu'il soit possible d'installer le logiciel dans un autre dossier que celui proposé par défaut, le répertoire d'accueil de *Tropes* s'appelle « Tropes » et est placé dans le disque dur C:, c'est-à-dire [C:\Program files\Tropes\Tropes...].

```

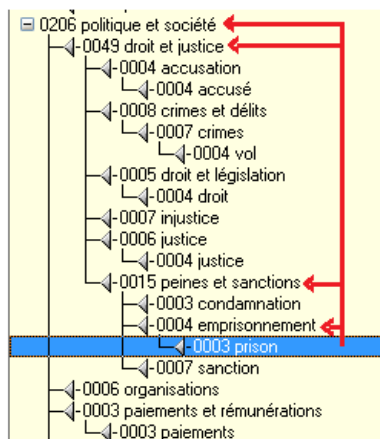
48764 1#6# → subrogation>SubrogationCRIFB
48765 A#6# → subrogé>SubrogationCRIFB
48766 V#6# → subroger → SubrogationCRIFB
48767 G#4V → 26888 → Droit de la consommationCRIFB
48768 1#5# → code de la consommation>Droit de la consommationCRIFB
48769 G#4V → 26889 → Droit des affairesCRIFB
48770 G#5V → 26890 → Code de commerceCRIFB
48771 S#6# → code du commerce → Code de commerceCRIFB
48772 G#5V → 26891 → Droit bancaireCRIFB
48773 1#6# → droit bancaire → Droit bancaireCRIFB
48774 G#5V → 26892 → Droit cambialCRIFB
48775 S#6# → droit cambiaire → Droit cambialCRIFB
48776 S#6# → droit cambial → Droit cambialCRIFB
48777 G#5V → 26893 → Droit de la concurrenceCRIFB
48778 1#6# → droit de la concurrence>Droit de la concurrenceCRIFB
48779 G#5V → 26894 → Droit des affairesCRIFB
48780 S#6# → droit des affaires → Droit des affairesCRIFB
48781 G#5V → 26895 → Droit des brevetsCRIFB
48782 1#6# → droit des brevets → Droit des brevetsCRIFB
48783 G#5V → 26896 → Droit des sociétésCRIFB
48784 1#6# → droit des sociétés → Droit des sociétésCRIFB
48785 1#5# → droit commercial → Droit des affairesCRIFB
48786 G#5V → 26897 → FiducieCRIFB
48787 1#6# → fiducie>FiducieCRIFB
48788 G#4V → 26898 → Droit des transportsCRIFB
48789 G#5V → 26899 → Code de la routeCRIFB

```

Une fois le texte proposé est géré par le programme, un clic sur le signe + d'une case développe l'arborescence de la classe correspondante et affiche les mots de cette classe précédés de leurs fréquences. Les énoncés qui comportent les mots en question sont affichés immédiatement dans la fenêtre à droite.



Le substantif *prison*, par exemple, est classé, de bas en haut, d'abord, dans la sous-classe *emprisonnement*. Cette sous-classe est, elle-même, classée dans la sous-classe supérieure *peines et sanctions* puis dans celle de *droit et justice* et enfin dans *politique et société*. C'est ce que les concepteurs appellent les *Classes d'équivalents*.



Le Scénario n'impose aucune contrainte. Pour les exemples du substantif « syndicaliste », du verbe « syndicaliser » et de l'adjectif « syndical », ils seront regroupés sous la référence commune [Syndicat], même si ce dernier mot n'existe pas dans le texte, c'est seulement le point de vue linguistique qui est pris en considération. Mais ces mots seront relevés dans les énoncés de la fenêtre des résultats.

Les classes d'équivalence sont ensuite présentées selon leur ordre d'arrivée à la fois à l'intérieur des propositions (relations, actants et actés), et sur l'intégralité du texte (graphe de répartition, rafales, épisodes, propositions remarquables).

A ce stade, le résultat obtenu est réalisé selon le scénario préétabli intégré dans le moteur de *Tropes*. Le logiciel offre également à l'utilisateur la possibilité de créer son propre scénario en fonction du type de son texte ou de sa base de données.

Pour éviter la procédure fastidieuse d'entrer manuellement chaque mot, l'un après l'autre, l'opération consiste, dans un premier temps, dans l'élaboration d'une ontologie constituée des classes d'équivalence et des sous-classes à partir de textes et de manuels spécialisés. Il reste ensuite à faire entrer « massivement » cette ontologie dans le moteur du logiciel.

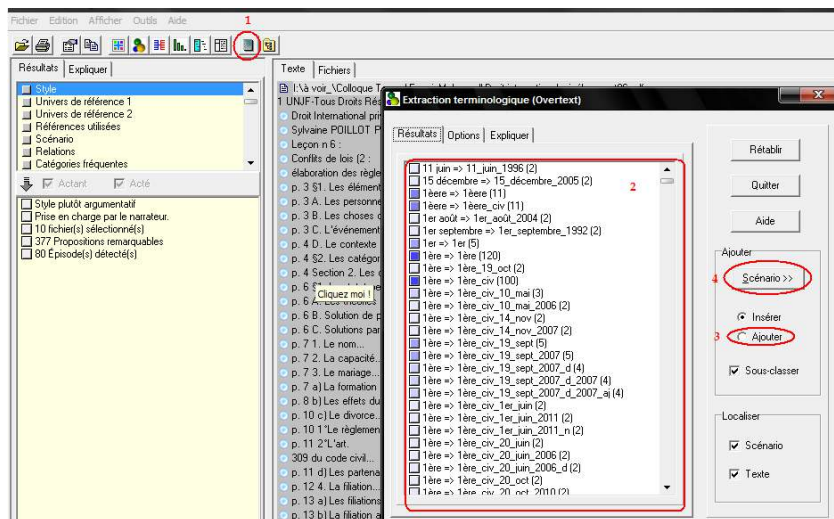
5.1. La création d'un scénario personnalisé

Le transfert d'une ontologie dans un scénario de *Tropes* est une opération complexe. Cependant, les concepteurs de *Tropes* proposent pour les utilisateurs qui n'ont aucune formation en programmation un déroulement en quatre étapes avec, malheureusement, la perte de la structure hiérarchique. Dans notre exemple, nous avons travaillé, comme indiqué plus haut, sur des livraisons des journaux *Le Monde* et *L'Est Républicain*.

A cette fin, on procède comme suit :

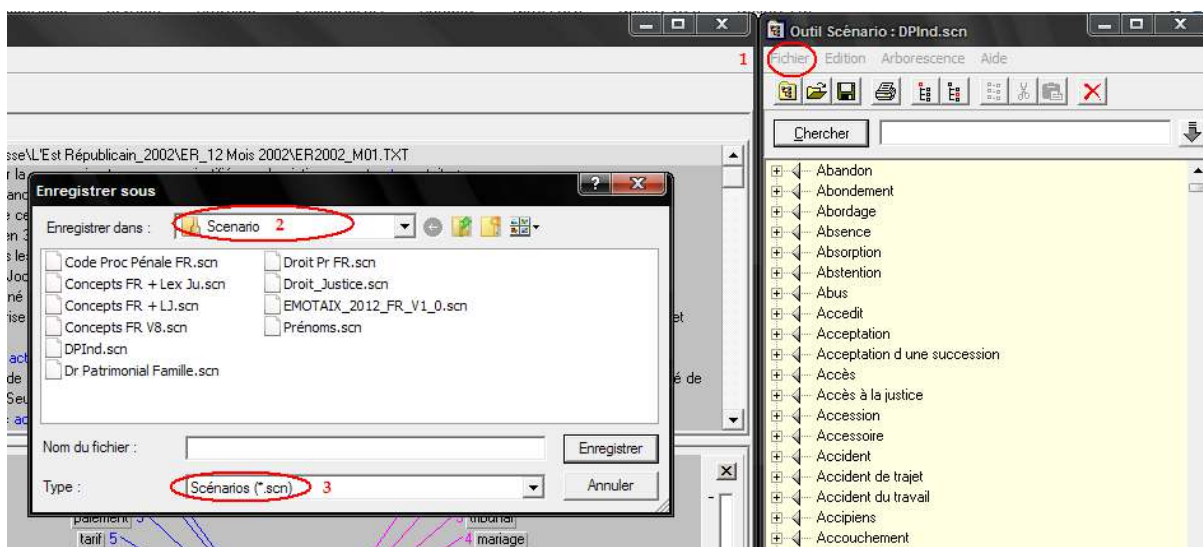
1- On indique dans une première étape l'adresse d'un ou de plusieurs fichiers présents dans l'ordinateur. Ces fichiers doivent obligatoirement être des textes théoriques relatifs au domaine à étudier, ici des manuels de droit¹¹.

¹¹ Nous avons utilisé les documents suivants à partir du site <https://www.legifrance.gouv.fr/> : Charges Locatives, Code Civil, Code de Déontologie Médicale, Code de Justice Administrative, Code de la Consommation, Code de la Construction et de l'Habitation, Code de la Famille et de l'Aide Sociale, Code de la Mutualité, Code de la Propriété Intellectuelle, Code de la Route, Code de l'Aviation Civile, Code de l'Education, Code de l'Expropriation Pour Cause d'Utilité Publique, Code de l'Urbanisme, Code des Assurances, Code des Collectivités Territoriales, Code des Communes, Code des Pensions Civiles et Militaires de Retraite, Code des Postes et Télécommunications, Code du Commerce, Code du Travail, Code Electoral, Code Pénal, Conseil 408 J224- Régler Litige Locatif, Constitution de 1958, Decret 87-713 Charges Locatives, Droit Français, Guide pratique des charges recuperables, Le Loyer et les Charges Locatives, Loi-89-462. Pour d'autres domaines, on procède de la même manière, par exemple, avec des thèses de médecine, des études historiques, des livres de philosophie, des manuels de linguistique, etc.



2- On utilise l'extracteur terminologique pour extraire les libellés. On clique sur le bouton *Extraction terminologique* (1) pour visualiser la liste des termes et expression récurrents dans les textes de référence (2). On choisit ensuite d'ajouter (3) l'ontologie obtenue au Scénario (4).

3- Il suffit, à ce stade, d'enregistrer (1) le scénario qui vient d'être créé dans le dossier où *Tropes* a installé déjà les scénarii existants (2). Ce fichier aura l'extension *.scn¹² (3)

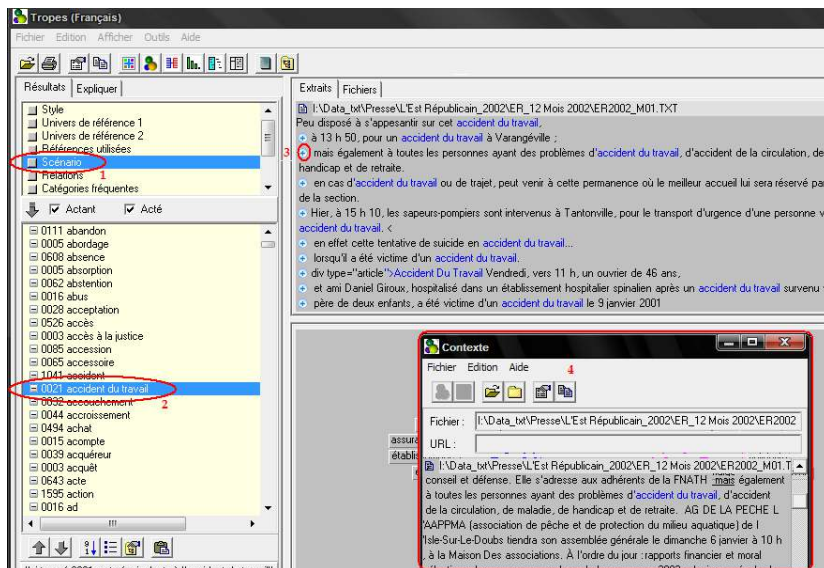


4- On procède ensuite à la fermeture des fichiers de référence et à l'ouverture du corpus de journaux à analyser comme pour la première étape.

5- On indique au logiciel d'utiliser le nouveau scénario en appelant ce dernier grâce au menu *Outils > Outils scénario*. Une nouvelle fenêtre apparaît à droite de l'écran où il faut donner l'adresse du nouveau scénario en utilisant le menu *Fichier* de la nouvelle fenêtre puis *Ouvrir*, et enfin sélectionner le fichier *.scn qui convient.

6- Le logiciel se positionne immédiatement sur le menu *Scénario* (1). Si on clique sur un mot ou une structure (2), les énoncés qui contiennent ces mots s'affichent à droite. En cliquant sur la petite croix (3) à gauche d'un exemple, une fenêtre volante (4) affiche l'exemple dans un contexte plus large.

¹² Pour *scénario*.

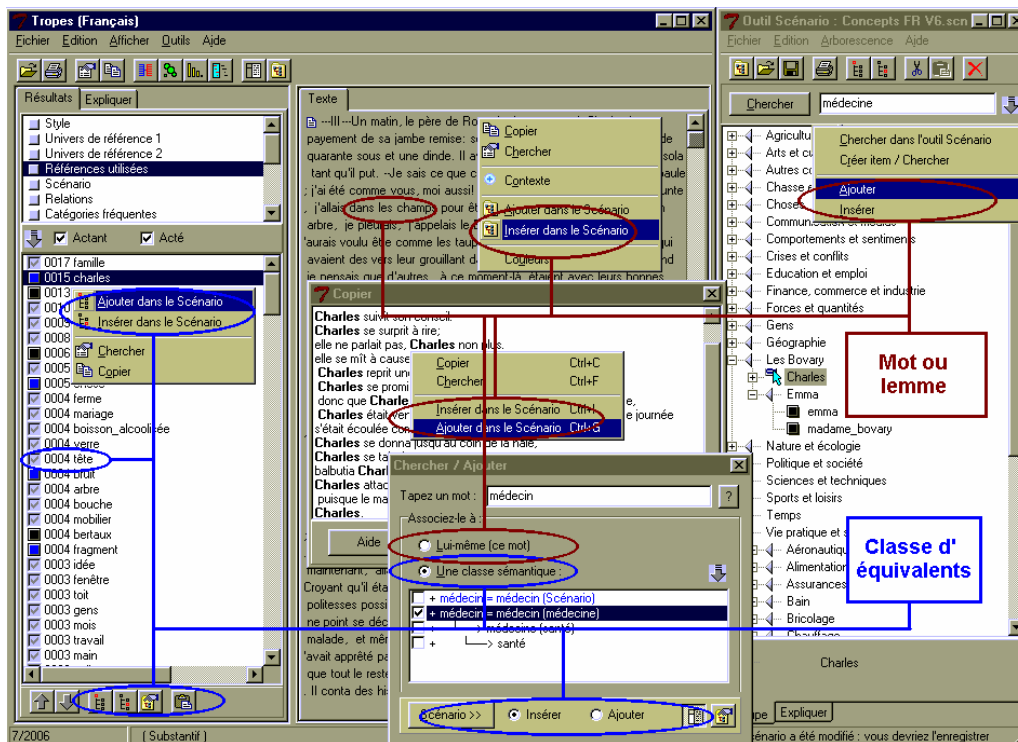


5.2. Enrichissement manuel d'un scénario

Comme le suggère le manuel, « tous les éléments de l'interface de Tropes contiennent des menus contextuels qui permettent d'ajouter facilement des mots ou des classes d'équivalents dans le Scénario ».

Quand un mot ou une expression du texte apparaissent dans les menus des verbes, des adjectifs ou des substantifs, mais pas dans le menu Scénario, il suffit d'ouvrir l'*Outil Scénario* et de créer un nouveau groupe ou juste d'insérer, dans un groupe déjà présent, directement un mot (ou un mot composé) ou une classe d'équivalents dans le Scénario. Il faut sélectionner le mot puis appuyer sur le bouton [Scénario] qui est au bas de ce dialogue. Une petite barre d'outils, en bas à droite, ([Chercher/Ajouter], permet de basculer de la fonction [Insérer] à [Ajouter], et réciproquement) contient une option pour repositionner automatiquement le Scénario sur les résultats lors de la recherche.

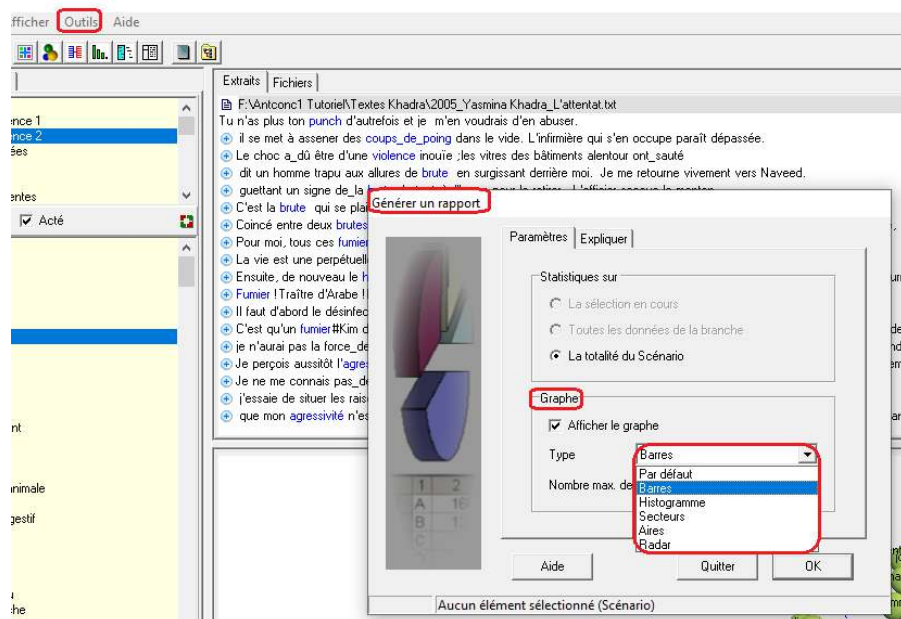
Voici un condensé des opérations possibles selon le manuel :



La plupart des fonctions de navigation dans l'arborescence du Scénario sont similaires à celles de l'Explorateur de fichiers de Windows. Par exemple, pour renommer un groupe sémantique, il suffit de le sélectionner pendant quelques secondes pour avoir la possibilité de changer son nom.

5.3. Exportation des résultats du scénario

Dans le menu *Outils* > *Gérer un rapport*, il est possible de sauvegarder les résultats du Scénario sous forme d'arborescence, avec l'option d'enregistrer du même coup l'une des visualisations proposées.



6. Conclusion

Nous avons vu comment un juge ou un avocat peuvent trouver facilement les textes qui intéressent leur domaine en fouillant dans n'importe quelle base de données ou n'importe quel document numérique en utilisant l'un des modules du logiciel *Tropes*, l'outil *Scénario*. Il suffit de créer son propre scénario avec les notions du domaine juridique. La méthode est très rapide et efficace, fait gagner beaucoup de temps et permet de trouver les thématiques d'un texte en quelques clics. La même procédure peut être appliquée à n'importe quel domaine.

Références

- CHARNOCK (R.), « Les mots du droit », in Actes de la 12ème Journée scientifique de la CRL
- 17 novembre 2007 - Université Paris-Diderot, textes réunis sous la direction de D.
De Vecchi & C. Martinot, pp.41-56.
- CORNU (G.), *Linguistique juridique*, Paris, Montchrestien, 2005.
- GÉMAR Jean-Claude, KASIRER Nicholas (éds.), *Jurilinguistique : entre langues et droits*,
Bruxelles, Bruylant, 2005.

GHIGLIONE (R.), KEKENBOSCH (C.), LANDRE (A.), *L'analyse cognitivo-discursive*, Grenoble, Presses Universitaires de Grenoble, 1995.

GHIGLIONE (R.), LANDRE (A.), BROMBERG (M.), MOLETTE (P.), *L'analyse automatique des contenus*. Paris, Dunod, 1998.

MOLETTE (P.), « De l'APD à Tropes : comment un outil d'analyse de contenu peut évoluer en logiciel de classification sémantique généraliste », Communication au colloque Psychologie et communication, Tarbes – Juin 2009.

PICOTTE (J.), *Juridictionnaire*, recueil des difficultés et des ressources du français juridique, Moncton, Université de Moncton, Faculté de droit, Centre de traduction et de terminologie juridiques, 2006.

RAYMONDIS (L.-M.) et LE GUERN (M.), *Le langage de la justice pénale*, Paris, Edit. CNRS, 1976.