

Tropes[©]

Version 8.1

Manuel de référence

Logiciel développé par Pierre Molette et Agnès Landré
sur la base des travaux de Rodolphe Ghiglione

Sommaire

Introduction	5
CHAPITRE 1 - Analyser des textes	6
● Enregistrer un document en format texte.	7
● Lancer l'analyse d'un texte.	7
● Examiner le texte.	8
● Afficher le contexte.	9
● Styles et mises en scène	10
● Propositions remarquables.	11
● Univers de référence : le contexte global.	11
● Références utilisées : de quoi est-il question ?	12
● Actants et actés (analyse des acteurs).	12
● Relations : quels éléments sont fréquemment associés ?	13
● Remarques sur les classes d'équivalents.	13
● Afficher toutes les catégories de mots.	14
● Catégories de mots fréquemment utilisées	14
● Listes de verbes et d'adjectifs.	15
● Episodes et rafales.	16
● Options d'analyse.	17
● Options générales.	19
● Imprimer les résultats.	20
● Graphes des classes d'équivalents et des relations.	21
● Graphes des acteurs.	23
● Graphes par répartition.	24
● Fonctionnement général des graphes.	26
● Copie des graphes.	27
● Copie du texte sélectionné.	27
● Copie des résultats.	27
● Propriétés des graphes.	27
● Modification des couleurs.	28
● Rechercher un mot.	29
● Générer un rapport	30

CHAPITRE 2 - Scénarios sémantiques - ontologies	31
● Introduction.	32
● Création de Scénarios.	33
● Association d'une classe d'équivalents à un groupe sémantique.	34
● Application du Scénario sur le texte.	36
● Méthodologie de conception de Scénarios.	37
● Gestion des Scénarios.	38
● Extraction terminologique.	39
CHAPITRE 3 - Délimiteurs	43
● Segmentation de textes.	44
● Création de délimiteurs.	45
● Fichiers de délimiteurs.	46
CHAPITRE 4 - Notions d'analyse de texte	47
● Le travail de Tropes.	48
● Le découpage propositionnel.	49
● La levée d'ambiguïté.	49
● Catégories de mots.	50
● Analyses statistiques, probabilistes et cognitives.	52
● Classes d'équivalents et relations entre équivalents.	52
● Considérations relatives à la taille du texte.	53
● Analyse de propos hétérogènes : questions ouvertes, dépêches, etc.	54
● Traitement de très gros corpus textuels.	55
● Analyse de discours et d'entretiens.	56
● Etudes littéraires.	57
● Comparer deux textes.	57
Annexes	58
● Conversions des formats de fichiers.	59

Licence

Tropes peut être utilisé selon les conditions décrites dans son contrat de licence. Ce n'est pas un logiciel du domaine public, parce qu'il appartient à ses auteurs. Il est protégé par les réglementations en vigueur concernant les droits d'auteurs et la propriété intellectuelle.

Overtex, Tropes, Microsoft, Windows, Excel, Word, Outlook, PowerPoint, Internet Explorer, Adobe, Java, Apple, MacIntosh sont des marques déposées par leurs propriétaires respectifs.

Conventions typographiques

Les commandes Windows, accessibles via les menus ou les boîtes de dialogue, ainsi que le texte correspondant aux boutons sont présentés en italiques et entre crochets [], par exemple :

- *[Fichier][Ouvrir]* correspond à la commande [Ouvrir] du menu [Fichier],
- *[Annuler]* correspond au bouton [Annuler] d'une boîte de dialogue Windows.

Les mots spécifiques au logiciel sont généralement en majuscules et présentés en italiques (par exemple : *Classes d'équivalents*).

Remarques

Bien qu'il soit possible d'installer le logiciel dans un autre dossier que celui proposé par défaut, ce manuel considère que 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...*].

Les ressources linguistiques (dictionnaires, réseaux sémantiques et Scénarios) de ce logiciel évoluent constamment. Les résultats présentés ici à titre d'exemple peuvent donc varier en fonction de la version que vous utilisez.

Ce manuel doit être complété par les informations du site Internet de Tropes (www.tropes.fr).

Edition du 4 novembre 2011

Introduction

Qu'est-ce que le contenu d'un texte ? Ou plus précisément qu'est-ce qui est essentiel pour comprendre le sens d'un texte ?

Chaque texte, qu'il s'agisse d'un article, d'un ouvrage, d'un discours, etc. contient quelques phrases clés correspondant aux idées qui constituent son ossature, son squelette. Le problème dès lors est - hors de toute interprétation préalable - d'atteindre ce(s) noyau(x) du texte qui contiennent l'essentiel du sens.

Disons qu'un texte est composé d'Univers dans lesquels des acteurs font, sont ou disent avec d'autres acteurs des choses diverses. Disons encore que ces Univers - qui ont tous un format propositionnel - n'ont pas la même importance dans la structuration du texte. Disons enfin que certains de ces Univers - en nombre faible - sont des fondements du texte, au sens où si on les retirait, « l'édifice textuel » s'écroulerait, et donc le sens serait perdu.

L'analyse de contenu est dès lors un ensemble de moyens pour tenter de savoir :

- Quels sont les principaux acteurs à l'oeuvre dans le texte ?
- Quelle est la structure des relations qui les lient ?
- Quelle est la hiérarchie de ces relations et leur évolution ?

En un mot l'analyse de contenu consiste à faire apparaître l'ossature du texte, c'est-à-dire son sens. Ceci suppose un ensemble théorique qui rende compte à la fois de l'économie textuelle des énoncés et de l'économie cognitive des énonciateurs. Ceci suppose également un outil qui - issu de la conception théorique - mette la subjectivité du codeur à distance... au moins durant le temps de l'analyse. C'est cet outil que nous vous présentons : Tropes.

Contributeurs

Tropes a été développé par Pierre Molette et Agnès Landré, sur la base des travaux de Rodolphe Ghiglione.

Durant son évolution, le logiciel s'est appuyé sur d'autres travaux scientifiques que ceux des premiers auteurs. En particulier les *Rafales* (adaptées à partir de la thèse de Mathieu Brugidou), le *Scénario* (inspiré de la linguistique anglo-saxonne et partiellement des travaux de John Lyons) ainsi que le *Style général du texte* (adapté à partir de la *Grammaire du sens et de l'expression* de Patrick Charaudeau).

Il faut aussi remercier les nombreuses personnes qui ont contribué à Tropes, directement ou indirectement. La liste est trop longue pour figurer ici.

Tropes

CHAPITRE 1

-

Analyser des textes

● Enregistrer un document en format texte.

Les documents doivent être sauvegardés au format texte ANSI (Windows), HTML (pages Web), Microsoft Word®, RTF, etc. Ils doivent comporter l'extension d'un fichier texte (Monfich.txt, PageWeb.htm, ...) Des restrictions existent pour certains de ces formats. Consultez les annexes de ce manuel pour en savoir plus.

Nous vous conseillons d'enregistrer vos textes au format *[Page Web (HTML)]* ou *[Texte seul]*. Pour effectuer une conversion de fichier, vous utiliserez généralement l'option *[Texte seul]* ou *[Page Web]* de la fonction *[Enregistrer sous]* de votre traitement de texte. Vous pouvez enregistrer des pages Web sur votre disque dur au moyen de votre navigateur Internet, ou d'un « aspirateur de site » .

Le texte peut contenir des signes de ponctuation, des majuscules et des caractères spéciaux (comme des parenthèses, des nombres, des pourcentages, etc.) Seules les lettres de l'alphabet et les caractères de ponctuation seront utilisés durant l'analyse. Si vous voulez qu'une suite de mots soit considérée comme un mot unique, vous devrez les lier avec le caractère souligné "_" (par exemple : Cyrano_de_Bergerac).

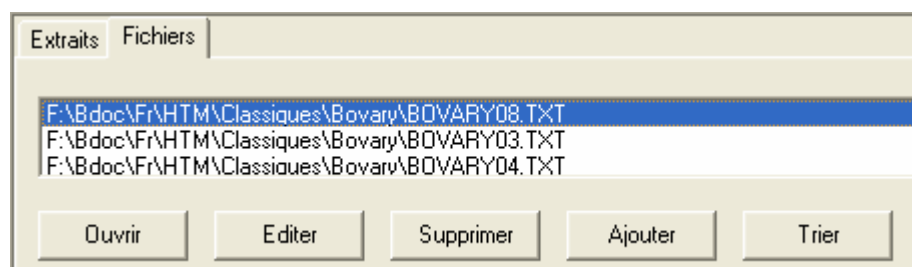
Pour de meilleurs résultats, vous devez utiliser de la ponctuation et respecter les conventions typographiques du français. Le correcteur orthographique de certains traitements de texte permet de s'assurer que ces conventions sont respectées.

● Lancer l'analyse d'un texte.

Pour analyser un texte, lancez *Tropes*, utilisez le menu *[Fichier][Ouvrir]*, sélectionnez un fichier, puis appuyez sur *[Ouvrir]* .

Vous pouvez ouvrir plusieurs fichiers simultanément en utilisant les touches [Ctrl] et [Maj] de votre clavier lorsque vous sélectionnez les fichiers. Vous pouvez aussi utiliser les fonctions de glisser-lâcher (*drag & drop*) de l'*Explorateur* de Windows : pour cela, prenez avec la souris un texte de votre choix et lâchez-le sur la fenêtre principale de Tropes.

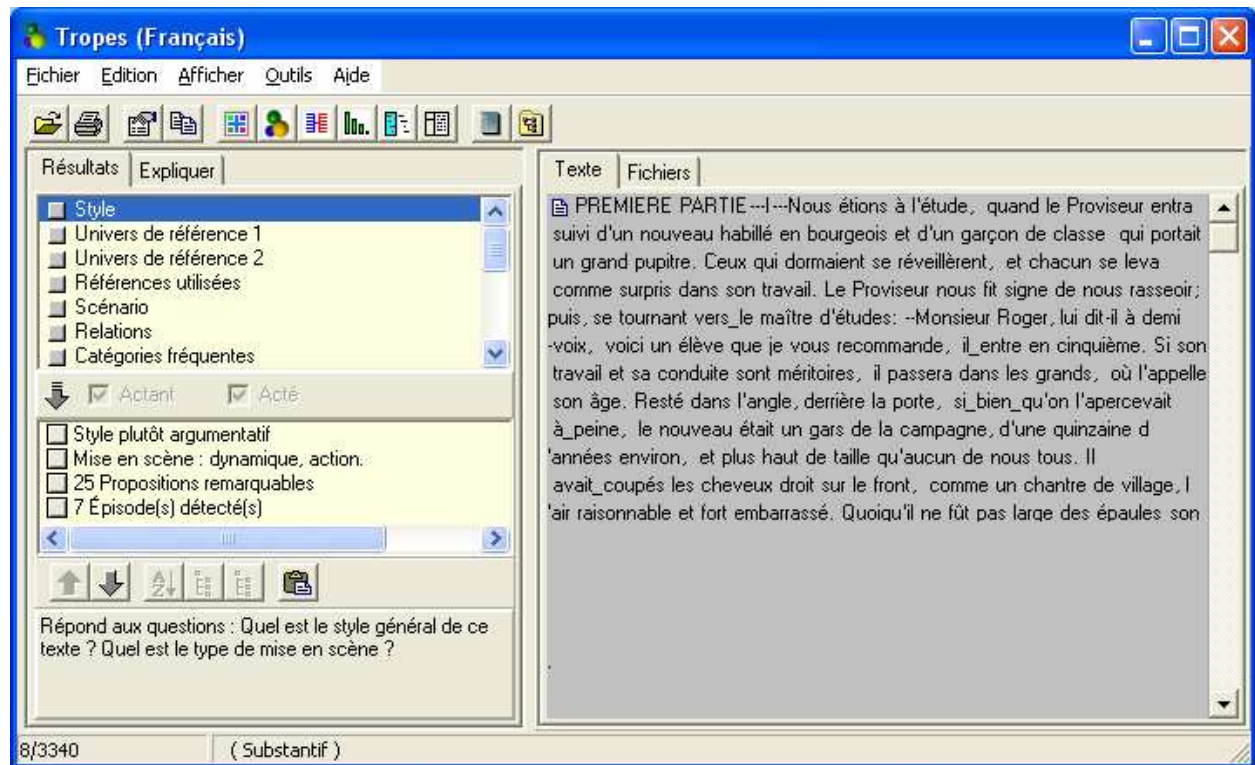
Après avoir lancé l'analyse, vous pouvez aussi utiliser l'onglet [Fichier] pour ajouter ou supprimer des documents, les visualiser, les ouvrir, les trier, etc.



● Examiner le texte.

Pour une première approche, nous vous proposons d'ouvrir le fichier d'exemple *Bovary1.txt*.

Lorsqu'un document est ouvert, et dès que l'analyse est terminée, vous voyez apparaître le texte correspondant dans la fenêtre principale de Tropes :



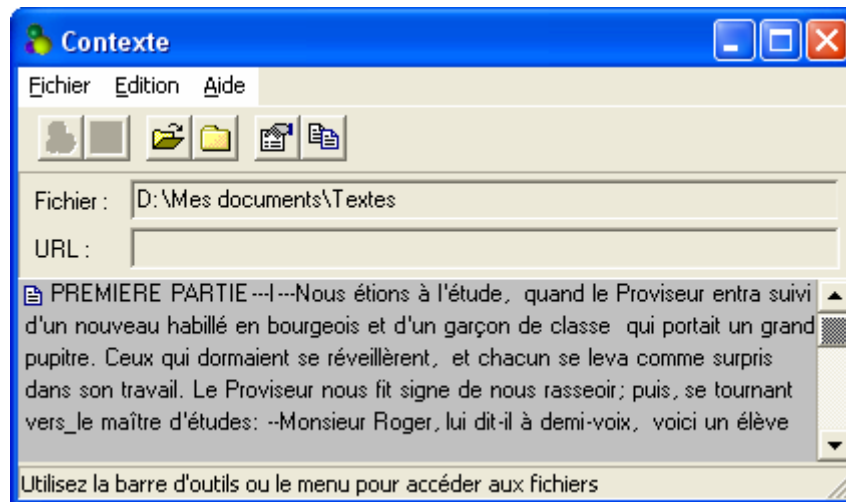
Les résultats d'analyse sont présentés à gauche, le texte et les graphes sont à droite.

Cette fenêtre "réagit" en fonction de ce que vous faites :

- Lorsque le **curseur de la souris** est positionné sur un mot, la barre de message, fait apparaître la catégorie du mot pointé (dans l'exemple ce mot est un « substantif »). Le nombre indiqué dans la barre de message indique le **numéro de séquence** de l'occurrence pointée par la souris suivi du nombre total de mots du texte.
- Quand l'affichage se fait sur un sous-ensemble du texte, seules les catégories concernées sont en couleurs. Pour afficher tout le texte utilisez le menu *[Afficher][Tout le texte]* ou le bouton *[Tout]* de la barre d'outils.
- Un **message d'aide**, en bas du cadre de gauche, vous explique pas à pas les différents résultats d'analyse et ce que vous pouvez en faire.

● Afficher le contexte.

Pour afficher le contexte d'une proposition, il suffit de cliquer sur l'icône qui est affichée dans la marge droite de la fenêtre contenant le texte. Le dialogue suivant apparaît :



L'affichage du contexte est effectué de la même manière que pour la fenêtre principale (coloration syntaxique des mots, etc.), mais de façon plus compacte, parce qu'il n'y a pas de retour à la ligne à la fin de chaque proposition. Le texte est donc plus facile à lire.

Le haut du dialogue affiche le nom du fichier en cours et éventuellement son adresse URL sur le Web (si ce fichier est associé à un raccourci Internet [.url]).

Le menu et la barre d'outils offrent des fonctions permettant :

- d'**ouvrir** le document, avec l'application définie par défaut sous Windows,
- d'**analyser** avec Tropes le document sélectionné,
- d'**ouvrir** le document à son emplacement d'origine **sur Internet**,
- de **localiser** le document (ouvrir l'Explorateur Windows sur le dossier où il se trouve),
- de **copier** des parties de textes et de **chercher** un mot dans le texte.

Ce dialogue est utile lorsque l'affichage porte sur une partie du texte (lorsque toutes les propositions ne sont pas affichées), mais vous pouvez aussi l'utiliser pour stocker provisoirement un passage du texte que vous trouvez intéressant, ou pour lire les documents.

● Styles et mises en scène

Cliquez sur la ligne [Style], ou utilisez le menu [Afficher][Style].

Le logiciel effectue un diagnostic du **style** général du texte et de sa **mise en scène** verbale, en fonction des indicateurs statistiques récupérés au cours de l'analyse.

Les **styles** possibles sont les suivants :

Style :	Explication :
Argumentatif	le sujet s'engage, argumente, explique ou critique pour essayer de persuader l'interlocuteur
Narratif	un narrateur expose une succession d'événements, qui se déroulent à un moment donné, en un certain lieu
Enonciatif	le locuteur et l'interlocuteur établissent un rapport d'influence, révèlent leurs points de vue
Descriptif	un narrateur décrit, identifie ou classe quelque chose ou quelqu'un

Les **mises en scène** verbales possibles sont les suivantes :

Mise en scène :	La mise en scène s'exprime à travers :
Dynamique, action	des verbes d'action
Ancrée dans le réel	des verbes de la famille d'être et avoir
Prise en charge par le narrateur	des verbes qui permettent de réaliser une déclaration sur un état, une action, ...
Prise en charge à l'aide du « Je »	de nombreux pronoms à la première personne du singulier (« je », « moi », « me », ...)

Vous verrez aussi immédiatement si des *notions (adverbes) de doute* ont été détectées.

Pour obtenir une explication du style affiché (ou de la mise en scène), cliquez sur la ligne concernée : la fenêtre principale affichera en couleur tous les mots dont les catégories ont été prises en compte pour effectuer le diagnostic.

L'étude du *style général* et de la *mise en scène* d'un texte dont vous êtes l'auteur (directement ou indirectement) est particulièrement intéressante si le logiciel a émis un diagnostic contraire à vos objectifs. Par exemple, vous chercherez probablement à savoir pourquoi le style détecté est [*argumentatif*] alors que vous n'aviez pas cette intention. De la même manière, vous essayerez sans doute d'éviter au maximum les [*prises en charge à l'aide du « je »*] si votre texte est rédigé en nom collectif, ou bien vous chercherez à éliminer toutes les [*notions de doute*] d'une proposition financière, d'un contrat, etc.

Important : les analyses de Tropes étant statistiques, les textes doivent être assez longs pour que les résultats soient significatifs.

● Propositions remarquables.

Pour afficher les *Propositions remarquables*, cliquez sur la ligne [*nn propositions remarquables*] du [*Style*], dans les résultats.

Les *Propositions remarquables* sont obtenues par contraction du texte. Ce sont « des propositions qui introduisent des thèmes ou des personnages principaux, qui expriment des événements nécessaires à la progression de l'histoire (attributions causales, des conséquences, des résultats, des buts) ».

Tropes effectue un traitement complexe d'Analyse cognitivo-discursive (ACD) pour les extraire. Pour simplifier, disons que chaque proposition du texte se voit attribuée un score calculé en fonction de son poids relatif, de son ordre d'arrivée et de son rôle argumentatif. Les propositions sont ensuite filtrées en fonction de leur score. Afin de permettre de contrôler la quantité de propositions affichées et pour garantir que le résultat obtenu reflète bien le texte analysé, Tropes permet de régler le taux de contraction du texte (Cf. infra, *Options d'analyse*).

Nous insistons sur le fait que les *Propositions remarquables* ne présentent un intérêt réel que sur un discours monolithique et structuré, pas trop long. Elles ne constituent en aucun cas un véritable résumé du texte (ce qui nécessiterait sa réécriture).

Ces indicateurs n'ayant une signification que pour des documents de taille raisonnable, les *Propositions remarquables* ne sont pas affichées pour des textes très longs.

● Univers de référence : le contexte global.

Cliquez sur [*Univers de référence 1*], ou utilisez le menu [*Afficher*][*Univers de référence 1*].

Cette fonction affiche, par fréquence décroissante, les *Univers de référence* des mots du texte. Chaque ligne comporte un *Univers*, précédé d'un compteur indiquant le nombre de mots (occurrences) qu'il contient. Seuls les *Univers* significatifs sont affichés.

Les *Univers de référence* représentent le contexte. Ils regroupent, dans des *classes d'équivalents*, les principaux substantifs du texte que vous analysez. Le logiciel détecte les *Univers de référence* en utilisant deux niveaux de représentation du contexte (*Univers de référence 1 et 2*).

Pour voir le contenu d'un *Univers*, sélectionnez celui qui vous intéresse, la fenêtre principale affichera tous les mots qu'il contient (ceux-ci seront colorés en bleu sur l'écran).

Remarque : pour détecter ces *Univers de référence*, le logiciel utilise un dictionnaire des *équivalents sémantiques* dans lequel tous les mots du français ne sont pas classés (Cf. infra - *Remarques sur les classes d'équivalents*); vous ne verrez apparaître ici que les substantifs très significatifs de votre texte et certains noms propres.

● Références utilisées : de quoi est-il question ?

Cliquez sur *[Références utilisées]*, ou utilisez le menu *[Afficher][Références utilisées]*.

Cette fonction affiche, regroupés par classes d'équivalents et triés par en fréquence décroissante, les substantifs (*références*) utilisés dans le texte. Chaque ligne comporte une *référence*, précédée d'un compteur indiquant le nombre de mots (occurrences) qu'elle contient. Seules les *références* significatives sont affichées (cf. *Options d'analyse*, ci-après).

Les *Références utilisées* regroupent dans des *classes d'équivalents* les noms communs et noms propres ayant un sens voisin (par exemple : « père » et « mère » sont regroupés dans la classe « famille »).

Pour voir le contenu d'une *classe*, cliquez sur celle qui vous intéresse, la fenêtre principale affichera les mots qu'elle contient (ceux-ci seront colorés en bleu sur l'écran).

● Actants et actés (analyse des acteurs).

Lorsque vous affichez des *classes d'équivalents* (*Univers*, *Références* ou *Groupes du Scénario*), les deux cases à cocher *[Actant]* et *[Acté]* sont activées dans le cadre de gauche. Cet outil permet d'analyser les *acteurs*, en distinguant les *classes d'équivalents* qui sont généralement placées en position :

- d'*actant*, c'est-à-dire avant le verbe (et souvent sujet de ce dernier),
- d'*acté*, c'est-à-dire après le verbe (et rarement sujet de ce dernier).

Quand vous sélectionnez une seule de ces cases, la liste indique le nombre de fois (pourcentage entre parenthèses) où la *classe* correspondante s'est trouvée placée en position d'*actant*, pour le texte considéré. Vous retrouvez le même compteur lorsque vous imprimez la liste des *classes*.

Pour revenir à un affichage normal, cochez à la fois les cases *[Actant]* et *[Acté]*.

L'identification des *actants* et des *actés* constitue une des étapes fondamentales de l'analyse de texte. En effet, lorsqu'un *Univers de référence* significatif (ou une *Référence utilisée*) se trouve fortement placé en position d'*actant* (taux supérieur à 60 %) on peut généralement considérer qu'il *effectue l'action*. Dans le cas contraire, lorsqu'un *Univers* (ou une *Référence utilisée*) significatif se trouve fortement placé en position d'*acté*, on peut généralement considérer qu'il reflète *subit l'action*.

● Relations : quels éléments sont fréquemment associés ?

Cliquez sur [*Relations*], ou bien utilisez le menu [*Afficher*][*Relations*].

Cette fonction affiche, triées par fréquence décroissante, les *relations* de cooccurrence entre les *classes d'équivalents*. Chaque *relation* est précédée d'un compteur indiquant sa fréquence d'apparition dans le texte.

Les *relations* indiquent quelles *classes d'équivalents* sont fréquemment reliées (rencontrées à l'intérieur d'une même proposition), dans le texte analysé. Ces *relations* sont orientées suivant l'ordre d'apparition des mots qui les composent (généralement en allant des *actants* vers les *actés*, ou, plus simplement, dans le sens de lecture). Pour afficher le texte correspondant, cliquez sur la ligne qui vous intéresse.

Par défaut, les *relations* sont construites à partir des *Références utilisées*. Vous pouvez paramétrer le *niveau de construction des relations* (Cf. *Options d'analyse*, ci-après).

Lorsque vous construisez les *relations* sur les *groupes sémantiques* du *Scénario*, notez bien quelles sont calculées uniquement à partir du bas de l'arborescence, et donc pas sur toutes les branches (si c'était le cas, cela impliquerait de compter un même mot dans plusieurs relations sur des groupes de noms différents et fausserait les résultats d'analyse).

L'affichage des *relations* ne laisse pas beaucoup de place au hasard. En effet, il est assez peu probable que deux *classes d'équivalents* se retrouvent plusieurs fois, dans le même ordre, à l'intérieur d'un même texte. Dans ce cas, cela signifie que ces deux *classes* sont fortement liées, cela montre les notions sur lesquelles l'auteur du texte a insisté (mais pas forcément ce qu'il a voulu y mettre), mais aussi des mots composés ou des associations triviales (comme les prénoms qui précèdent les noms, par exemple « Amadeus Mozart »).

L'affichage des *Univers de référence*, des *Références utilisées* et de leurs *Relations* vous conduit au coeur du discours : vous verrez apparaître, par importance décroissante, tous les acteurs, objets, choses et concepts mis en scène dans le texte.

● Remarques sur les classes d'équivalents.

Dans ce manuel, on utilise indifféremment l'appellation *classes d'équivalents* pour parler des *Univers de référence*, *Références utilisées*, ou des *Groupes sémantiques* du *Scénario*.

Pour plus d'information sur les *classes d'équivalents*, consultez le chapitre 4 : « *Notions d'analyse de texte* ».

Remarques concernant les dictionnaires :

- Etant donné qu'il n'est ni possible, ni pertinent de classer tous les substantifs (noms, prénoms et noms propres) de toutes les langues, le logiciel génère automatiquement des *classes d'équivalents* pour tous les mots non référencés dans ses dictionnaires. Ces *classes générées* contiennent aussi des mots identifiés par le logiciel mais qui ne peuvent pas être regroupés par équivalents (par exemple, « Charles »).
- Ces *classes générées* sont visibles uniquement dans les *Références utilisées*, et précédées par un carré noir, alors que les *classes d'équivalents* identifiées par le logiciel sont précédées par un carré bleu.
- Pour regrouper des *classes générées* avec des *classes d'équivalents*, et créer votre propre classification, vous devez utiliser un *Scénario* (Cf. chapitre correspondant).

● Afficher toutes les catégories de mots.

Cliquez sur [*Toutes catégories de mots*], ou utilisez le menu [*Afficher*][*Toutes catégories de mots*].

Cette fonction affiche toutes les (méta) *catégories de mots* du texte analysé. Chaque ligne comprend une catégorie, sa répartition dans la sous-catégorie concernée (pourcentage) et le nombre d'occurrences trouvées.

Quand vous cliquez sur une ligne, Tropes vous donne une explication de la signification de la catégorie concernée dans la bulle d'aide. Pour plus d'information sur *ces catégories de mots*, consultez le chapitre 4 de ce manuel : « *Notions d'analyse de texte* ».

Nous vous conseillons d'utiliser un *graphe de type étoilé* (Cf. infra, *Propriétés des graphes*) pour visualiser les références rencontrées autour de ces *catégories*.

Notez bien que, pour des raisons méthodologiques (parce que les fréquences d'occurrence des *catégories* et des *classes d'équivalents* ne sont pas comparables), les *graphes en aires* ne sont pas disponibles sur les *catégories de mots*.

Pour imprimer ces *catégories de mots*, utilisez le menu [*Fichier*][*Imprimer*] et cochez l'option [*Style et catégories*].

● Catégories de mots fréquemment utilisées

Cliquez sur [*Catégories fréquentes*], ou utilisez le menu [*Afficher*][*Catégories fréquentes*].

Cette fonction donne un résultat semblable à la précédente, excepté que seules les *catégories de mots* les plus significatives sont affichées.

Une *catégorie de mots* est considérée comme significative lorsque sa fréquence d'apparition est nettement supérieure à la moyenne. Ces résultats sont construits en comparant les statistiques (de répartition des catégories) du discours analysé avec des tables internes au logiciel (ces tables ont été élaborées en effectuant une analyse de variance sur un grand nombre de textes différents).

Si vous sélectionnez une catégorie, la fenêtre principale fait apparaître les propositions concernées.

● Listes de verbes et d'adjectifs.

Cliquez sur *[Verbes]* ou *[Adjectifs]*, ou bien utilisez le menu *[Afficher][Verbes]* ou *[Afficher][Adjectifs]*.

Ces deux fonctions affichent, par fréquence décroissante, des verbes et adjectifs rencontrés dans le texte analysé. Chaque ligne est précédée d'un compteur indiquant le nombre de mots (occurrences) trouvés.

Ces mots sont réduits à leur forme canonique (lemme) :

- les verbes sont réduits à l'infinitif ;
- les adjectifs sont présentés au masculin singulier.

Vous pouvez ajouter ces verbes et adjectifs directement dans l'*Outil Scénario* et les déplaçant directement avec la souris.

Ces verbes et adjectifs peuvent être utiles pour compléter certaines analyses (en particulier si vous analysez une suite de messages très courts, où la moindre information doit être récupérée), pour définir et tester vos propres classifications « qualitatives », ou pour effectuer des études linguistiques.

Pour imprimer ces résultats, utilisez le menu *[Fichier][Imprimer]* et cochez les options *[Verbes]* ou *[Adjectifs]*.

● Episodes et rafales.

Cliquez sur la ligne [*nnn épisode(s) détecté(s)*] du [*Style*], ou bien utilisez le menu [*Afficher*][*Episodes*].

Cette fonction permet d'étudier la chronologie d'un discours. Elle se fonde sur deux notions, les *Rafales* et les *Episodes* :

- Une *Rafale* regroupe des occurrences de mots (contenus dans une *classe d'équivalents* ou une *catégorie de mots*) ayant tendance à arriver avec une concentration remarquable dans une partie limitée du texte (quelque part au début, au milieu ou à la fin du texte, mais jamais de façon uniforme sur l'intégralité de celui-ci).
- Un *Episode* correspond à une partie du texte dans lequel un certain nombre de rafales se sont formées et terminées. Ce sont de grands blocs d'argumentation, représentatifs de la structure du discours observé.

Les *Episodes* sont affichés les uns à la suite des autres, et numérotés en fonction de leur ordre d'arrivée dans le texte.

Dans chaque *Episode*, les *Rafales* sont triées en fonction de leur adresse (moyenne de la position des mots) et préfixées par la fréquence d'occurrence des mots qui la composent.

Par exemple, le texte [*Bovary1.txt*] commence par un premier *Episode* comportant des petites *Rafales* portant sur la référence « étude », des *connecteurs* et des *modalisations*, ... Puis un deuxième *Episode* comportant une *Rafale* longue sur « chapeau » (comportant 11 mots), etc. Le texte se termine par des *Rafales* sur « Charles », « Bovary », « femme » et « amour ». Cette analyse met en relief la construction du discours par le narrateur (ici Flaubert), qui a commencé son récit (ici le chapitre d'un livre) en situant l'action à l'école, a raconté ensuite une anecdote concernant un *chapeau*, et l'a terminé en évoquant les premiers amours d'un de ses personnages principaux, *Charles Bovary*.

Important : Des *Rafales* portant sur une même référence peuvent apparaître à plusieurs endroits du texte. Par exemple, « Charles » fait l'objet de deux *Rafales*, au milieu et à la fin du texte [*bovary1.txt*]. Nous vous conseillons d'utiliser le *graphe des épisodes* (Cf. infra) pour visualiser ces résultats.

Vous pouvez paramétrer le niveau de construction des *Rafales* (sur des *Univers*, *Références utilisées*, *Scénario*) en modifiant la *base de construction des relations* (Cf. *Options d'analyse*). L'utilisation d'*Univers de référence* à la place des *Références utilisées* a généralement pour conséquence de réduire le nombre d'*Episodes* détectés dans un texte (le logiciel utilise alors un nombre réduit des concepts génériques ayant des fréquences d'occurrence élevées, ce qui a pour effet de regrouper certains propos et de réduire la finesse de l'analyse).

Ces indicateurs n'ayant une signification que pour des documents de taille raisonnable, les *épisodes* et les *rafales* ne sont pas affichés pour des textes très longs.

● Options d'analyse.

Utilisez le menu [*Outils*][*Options d'analyse*].

Ce dialogue permet d'agir sur le moteur d'analyse du logiciel.

Le *seuil de détection des classes* permet d'indiquer au logiciel le niveau de significativité des *classes d'équivalents* :

- Si ce seuil est calculé en fonction d'un *nombre minimum de mots*, alors toutes les *classes d'équivalents* ayant une fréquence d'apparition inférieure à ce seuil seront ignorées (i.e. elles n'apparaîtront plus).
- Si ce seuil est calculé en fonction d'un *facteur de pertinence*, alors toutes les *classes* ayant un *facteur de pertinence* inférieur à ce seuil seront ignorées.

Le *facteur de pertinence* est calculé en dix-millièmes de nombre de mots du texte, c'est-à-dire qu'un *facteur de pertinence* de 10 correspond, par exemple, à une fréquence d'occurrence minimum de 3 mots, pour un texte de 3000 mots.

Vous pouvez modifier le *seuil de détection des classes d'équivalents* si vous voulez que le logiciel ne traite que les *classes* très fréquentes ou, au contraire, si vous voulez qu'il prenne en compte les *classes* peu fréquentes.

Important : Plus vous augmentez ces seuils, plus vous perdez de l'information. Et réciproquement : si vous réduisez ces seuils, vous augmentez la quantité d'information prise en compte par les *classes d'équivalents*.

Le dialogue des options d'analyse permet aussi de modifier la *base de construction des relations*, (encadré [*construire les relations sur*]), c'est-à-dire le niveau de *classes d'équivalents* devant être utilisé pour construire, afficher et imprimer les *relations*, les *épisodes* et les *rafales*.

Par exemple, vous pouvez construire les *relations* à partir du contenu du *Scénario courant* (Cf. chapitre 2 - *Scénarios sémantiques*).

Vous pouvez aussi agir sur le taux de contraction du texte en utilisant le curseur permettant de régler la [*quantité de propositions remarquables*] affichées.

Parce qu'il n'est pas certain que le logiciel soit capable de détecter correctement les propositions essentielles d'un corpus, nous vous conseillons d'utiliser un seuil relativement bas (c'est-à-dire d'afficher beaucoup de propositions remarquables), dans un premier temps, puis de l'augmenter progressivement jusqu'à ce que vous ayez trouvé un bon compromis entre la quantité de propositions affichées et la pertinence du résultat obtenu.

L'option [*utiliser le Scénario sur toutes catégories de mots*] permet de demander au logiciel de transformer en substantifs (en références) tous les mots entrés dans le *Scénario sémantique*. Lorsque cette case est cochée, vous pouvez entrer et afficher dans le *Scénario* des mots appartenant à d'autres catégories que les substantifs (sans être obligé de faire appel aux lemmes de verbes et d'adjectifs). Vous pourrez utiliser cette option pour forcer le logiciel à considérer comme substantifs des mots qui ne sont pas classés dans cette catégorie (par exemple, si vous analysez un texte comportant deux personnages, « Monsieur Blanc » et « Madame Rouge », vous pourrez rentrer « Blanc » et « Rouge » dans votre *Scénario* afin de les comptabiliser comme des références).

Pour valider votre choix, appuyez sur le bouton [*Accepter*]. Dans le cas contraire, si vous ne voulez pas modifier vos options d'analyse, appuyez sur [*Annuler*].

Attention : à partir du moment où vous cochez la case [*Utiliser le Scénario sur toutes catégories de mots*], tous les mots que vous entrez dans le Scénario seront transformés en substantifs (en références) à la fin de l'analyse automatique du texte. Ceci a plusieurs conséquences sur le fonctionnement du logiciel :

1 - Le *Scénario* devient prioritaire sur les autres classifications : tous les mots entrés dans le *Scénario*, qui ne sont pas des substantifs, vont disparaître de leurs catégories d'origine (vous pouvez éviter cela en utilisant des lemmes de verbes et d'adjectifs). Par exemple, si vous entrez le mot « vers » dans le *Scénario*, l'adverbe de lieu correspondant ne sera plus affiché dans la catégorie *modalisation* (où il sera toutefois comptabilisé).

2 - Le *Scénario* devient préemptif sur l'analyse syntaxique du texte : si, dans un même texte, on trouve un mot ambigu (utilisé simultanément dans plusieurs catégories grammaticales différentes) et si ce mot est placé dans le *Scénario*, alors toutes ses formes seront comptabilisées dans le *Scénario*. Par exemple, si vous avez rentré le mot « livre » dans un *Scénario* et si vous l'utilisez pour analyser un texte qui contient les phrases suivantes « On nous *livre* une caisse de limonade. » et « Nous avons reçu des *livres*. », alors les deux occurrences de « livre » (le verbe et le nom commun) seront comptabilisées par le *Scénario*.

3 - Les remarques qui précèdent ne s'appliquent que lorsque vous entrez des mots dans le *Scénario* : si votre *Scénario* est élaboré à partir de *classes d'équivalents*, elles ne poseront pas de problème de classification (sans compter que les *classes d'équivalents* bénéficient aussi d'une levée d'ambiguïté lexicale et sémantique).

L'onglet [*Conversions*] permet de paramétrer les filtres de lecture de certains formats de fichiers. Il y a deux groupes d'options :

1 - **L'analyse des documents Word**, qui peut s'effectuer en mode **binaire** (le logiciel extrait le texte sans tenir compte du format du fichier, ce qui peut conduire à l'affichage de certains caractères parasites) ou **approfondie** (qui donne de meilleurs résultats, mais peut poser bloquer le logiciel si vos fichiers sont corrompus).

2 - **L'analyse des documents Adobe (PDF)**, qui doit être activée en utilisant l'option correspondante. Ceci nécessite l'installation d'un pilote de périphérique (driver) spécifique sur votre ordinateur (cf. *Conversion de formats de fichiers*, dans ce même manuel).

● Options générales.

Utilisez le menu [*Outils*][*Options générales*].

Vous utiliserez ce dialogue pour personnaliser l'affichage et le comportement général de certains outils.

L'onglet [*Résultats*] permet de :

- désactiver l'affichage des bulles d'aide dans la fenêtre principale ;
- ne pas sauvegarder la liste des derniers fichiers ouverts ;
- doubler la taille des polices de caractères, si vous êtes mal-voyant ou si vous affichez à très haute résolution ;
- d'afficher les extraits de texte par propositions grammaticales (conforme aux théories d'analyse de contenu) ou par fragments de contexte (ce qui améliore leur lisibilité) ;
- régler le seuil à partir duquel un message d'avertissement vous demande de confirmer si vous voulez continuer l'analyse, quand vous analysez une grande quantité de texte.

L'onglet [*Scénario*] permet :

- de désactiver le message d'avertissement lors des suppressions ;
- d'afficher l'assistant automatiquement lorsque le *Scénario* est vide ;
- d'ouvrir automatiquement l'*Outil Scénario* lorsque vous lancez Tropes ;
- de mettre en majuscule (en option) la première lettre des noms des groupes ajoutés dans le *Scénario* ;
- d'ajouter automatiquement les formes non-accentuées (par exemple, si vous insérez le mot "bénéfice" dans l'*Outil Scénario*, le logiciel lui associera directement le mot "benefice", sans accents) ;
- de choisir le critère de tri du *Scénario* (cette fonction est aussi directement accessible via un menu contextuel sur la liste de résultats).

L'onglet [*Installation*] permet de changer le type d'installation (Cf. *Guide d'installation, Installation en mode Multi-Utilisateurs*).

● Imprimer les résultats.

Utilisez le menu [*Fichier*][*Imprimer*], pour imprimer vos résultats.

Le dialogue affiche une liste de résultats, que vous devez cocher avant de lancer l'impression.

Les états imprimés des *classes d'équivalents* et des *Scénarios* font apparaître un *taux d'utilisation*, exprimé en pourcentage, qui correspond au nombre de mots de chaque *classe* divisé par le nombre total de mots du texte.

L'état imprimé des *relations* fait apparaître une information supplémentaire, qui n'est pas affichée dans la fenêtre principale : le *taux de liaison*. Ce *taux* est calculé en divisant le nombre de *relations* observées par le nombre maximum de *relations* possible. Un taux de liaison de 100% indique que l'un des deux termes de la *relation* est toujours présenté avec l'autre. Un taux de liaison proche de zéro indique que les deux termes ne sont pratiquement jamais présentés ensemble.

Dans les options d'impression, des boutons permettent de modifier la configuration de l'imprimante et de choisir la police de caractère devant être utilisée pour l'impression.

La case à cocher [*Couleurs*], permet d'effectuer des impressions en couleur, si votre imprimante en est capable.

L'impression est lancée quand vous appuyez sur [*Imprimer*], sinon appuyez sur [*Annuler*].

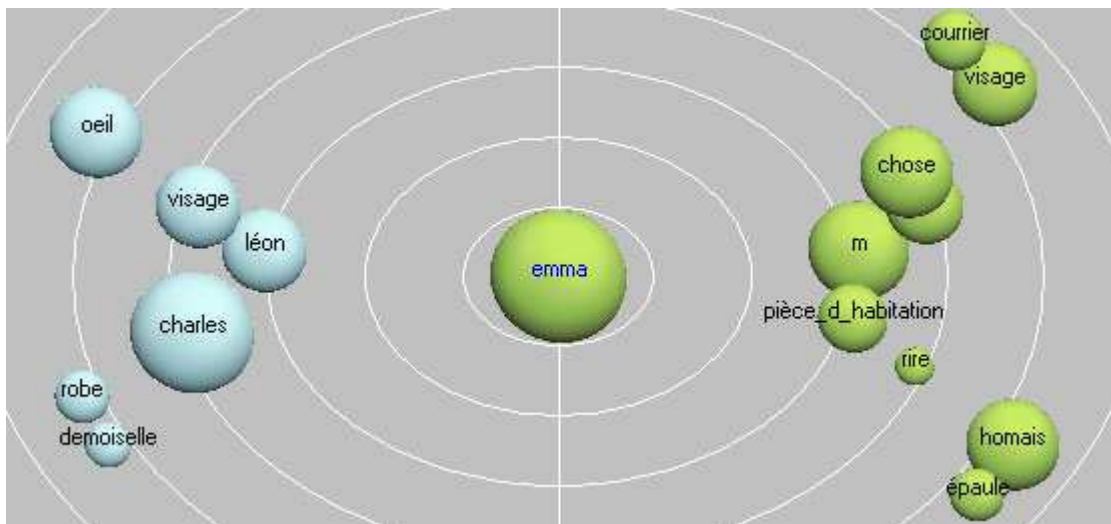
Notez que l'impression des graphes peut poser des problèmes si votre imprimante n'est pas capable de travailler en mode graphique (ce qui est assez rare).

● Graphes des classes d'équivalents et des relations.

Vous pouvez affecter l'affichage du graphe sur les modes *[acteurs]*, *[aires]*, *[étoilé]*, *[répartition]* ou *[épisodes]* en utilisant la barre d'outils ou le menu contextuel qui apparaît lorsque vous pressez le bouton droit de la souris sur le graphe.

Les trois premiers graphes indiquent les *relations* entre les *classes d'équivalents*. Les deux autres graphes concernent la chronologie du texte (nous en parlerons plus tard).

Commençons par le graphe en **aires** :



Chaque référence est représentée par une sphère dont la surface est proportionnelle au nombre de mots qu'elle contient. Dans l'exemple ci-dessus¹, on voit qu'Emma et Charles ont des fréquences d'occurrence comparables. Alors que « robe » et « demoiselle » ont une fréquence beaucoup plus faible.

La distance entre la classe centrale et les autres classes est proportionnelle au nombre de relations qui les lient : autrement dit, lorsque deux classes sont proches du centre elles possèdent beaucoup de relations en commun. Dans l'exemple ci-dessus, on voit que Léon est un peu plus souvent associé à Emma que Charles.

Les graphes en aires et en étoile sont orientés : les éléments affichés à gauche de la classe centrale sont ses prédécesseurs, ceux qui sont placés à droite sont ses successeurs. Dans l'exemple ci-dessus, on voit qu'on parle de Charles avant Emma.

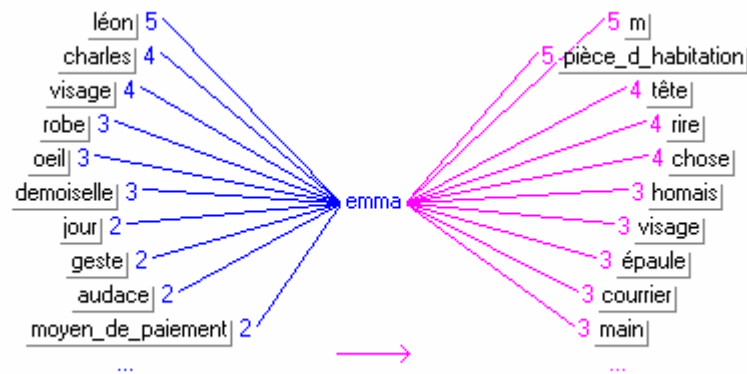
Pour reprendre la métaphore de système planétaire utilisée pour ce graphe, disons qu'autour d'une planète centrale (classe d'équivalent) gravitent d'autres planètes (classes d'équivalent qui ont des relations avec la classe centrale) qui sont plus ou moins proche du centre (fréquemment utilisées ensemble) et sont plus ou moins grosses (en fonction du nombre d'occurrences de mots qu'elles contiennent).

Remarques:

- 1 - le recouvrement de deux sphères n'a pas de signification particulière ;
- 2 - le graphe en aire ne peut pas être affiché sur les *catégories de mots* ;
- 3 - utilisez le curseur à droite du graphe pour modifier le nombre d'éléments affichés.
- 4 - Les objets du graphe se repositionnent quand la souris passe dessus. Utilisez la touche [Maj]/[Shift] pour éviter que les objets disparaissent lorsqu'on bouge la souris.

¹ Cet exemple est issu de l'analyse du roman « Madame Bovary » de Flaubert.

Le **graphe étoilé** affiche les relations entre *classes d'équivalents*, ou entre une *catégorie de mots* et des *classes d'équivalents* :

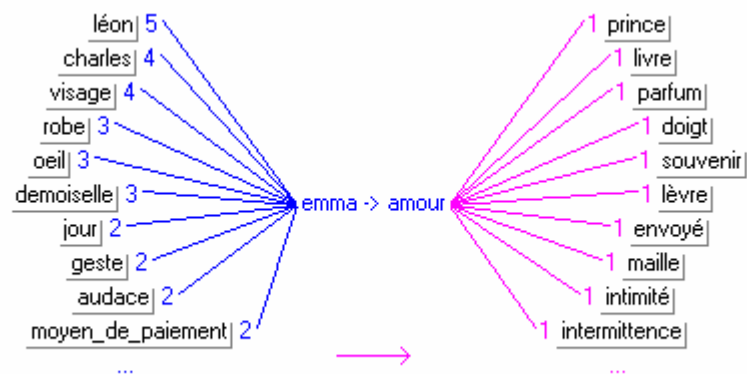


Les nombres qui apparaissent sur le graphe indiquent la quantité de *relations* (fréquence de cooccurrence) existant entre les éléments affichés.

Dans l'exemple ci-dessus, on voit qu'on mentionne souvent « Léon » (5 fois), « Charles » et « visage » (4 fois), avant de parler d'Emma (variable centrale). Alors que « M » (pour monsieur) et « Pièce d'habitation » ont été trouvés cinq fois après Emma.

Vous pouvez suivre les *relations* sur le graphe en cliquant directement sur les éléments qui vous intéressent. Cette fonction permet de naviguer à l'intérieur d'un texte en analysant les relations entre références.

Lorsque vous affichez les *relations*, le graphe comporte deux classes centrales :



Notez qu'une même référence peut se retrouver placée avant et après l'élément central, en particulier lorsqu'elle est à la fois en position de sujet et d'objet.

● Graphes des acteurs.

Le graphe des acteurs représente la concentration de relations entre les principaux acteurs (actants/actés) sur la totalité du texte. Il permet de faire une comparaison visuelle du "poids" des relations entre les principales références (ou bien entre les groupes du Scénario).

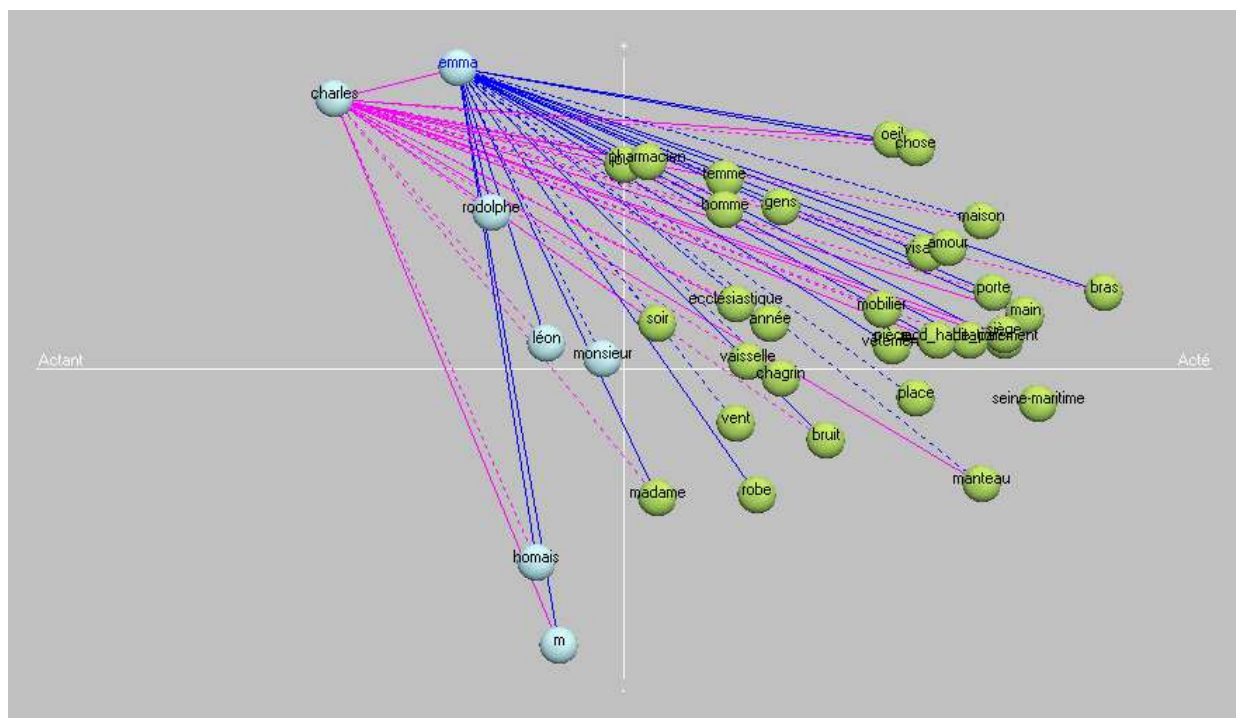
Les références sont présentées selon deux axes :

- L'axe des X (horizontal) indique le taux actant/acté (de gauche à droite).
- L'axe des Y (vertical) indique la concentration de relations pour chaque référence affichée (forte en haut du graphe, et faible en bas).

La concentration de relations est calculée en divisant, pour chaque référence, le nombre total de relations (R) par le nombre de relations différentes (D). Cette valeur est pondérée².

Les traits indiquent les relations entre la référence sélectionnée et les autres références affichées. Un trait en pointillé indique une relation peu fréquente. Un trait plein indique une relation fréquente. En survolant une référence avec la souris on affiche un deuxième réseau de relations. Un clic sur le fond du graphe affiche toutes les relations.

Dans l'exemple ci-dessous, on a sélectionné Charles et Emma :



Le logiciel a placé les principaux personnages (Emma, Charles, etc.) en haut à gauche (actants ayant une forte concentration de relations), alors que les personnages secondaires se trouvent soit en bas à gauche (actants avec une faible concentration de relations) ou à droite (actés). La majorité des autres références sont actés (à droite).

La fréquence de relations et le nombre de relations différentes sont des indicateurs de centralité du texte analysé. Si l'auteur, ou les auteurs, du texte ont associé une référence avec

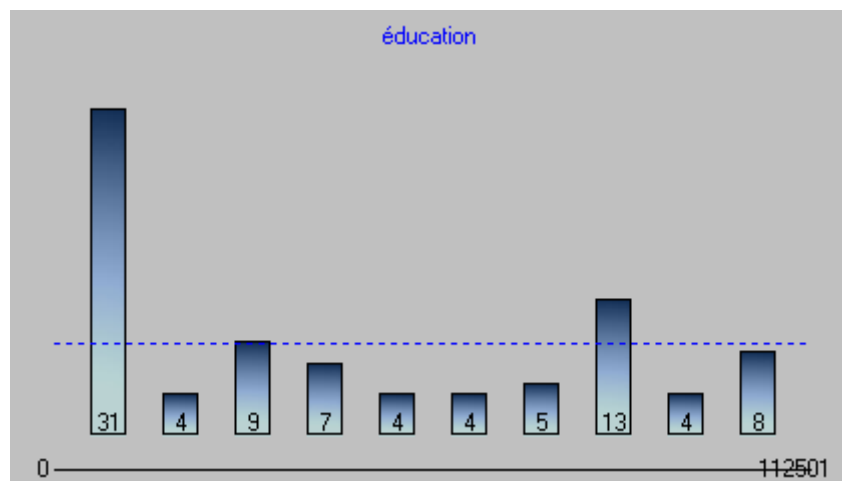
² Depuis la version 8.1 de Tropes, la concentration des relations du graphe des acteurs est pondérée par le logarithme de la fréquence d'occurrence, ce qui donne un peu plus de poids aux classes fréquentes. La formule de calcul est : $(R / D) * \text{Log}(F / 2)$. Où F est la fréquence d'occurrence.

de nombreuses autres références, alors on peut penser que cette référence est très importante, ou en tout cas plus importante que d'autres. Par ailleurs, cette représentation graphique présente un graphe global, qui ne dépend pas uniquement de la référence sélectionnée (ce qui est le cas avec les graphes en *étoile* et en *aires*).

Notez que seules les références présentant un grand nombre de relations sont représentées sur le graphe. Mais vous pouvez faire varier le nombre d'objets affichés en utilisant le curseur à la droite du graphe.

● Graphes par répartition.

Les graphes par répartition permettent d'afficher un histogramme de répartition (chronologique) d'une *classe d'équivalents*, d'une *mise en relation* (i.e. de deux *classes d'équivalents*), ou d'une *catégorie de mots* :



Dans l'exemple ci-dessus, on voit que la *classe d'équivalents* « Education » apparaît beaucoup plus au début (partie gauche) qu'à la fin du texte (partie droite).

Ce graphe est construit en divisant le texte en secteurs contenant un nombre égal de mots et en calculant la fréquence d'apparition de la *classe d'équivalents* ou la *catégorie de mots* sélectionnée à l'intérieur de chaque secteur. Les barres de l'histogramme affichent chaque secteur dans l'ordre chronologique, de gauche (début du texte), à droite (fin du texte). La ligne en pointillés indique la taille moyenne des barres de l'histogramme.

Si vous cliquez sur une barre, l'affichage se positionnera automatiquement sur les propositions apparaissant à partir de cet endroit du texte. Lorsque vous pointez une partie de l'histogramme, la ligne de message (au bas de l'écran) affiche, pour le secteur sélectionné, le nombre de mots qu'il contient et sa position (début et fin du secteur, calculés en nombre de mots, depuis le début du texte).

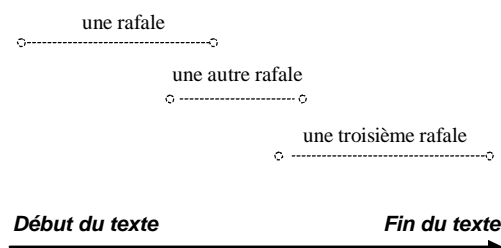
Le nombre de barres de l'histogramme et le nombre de mots contenus dans chaque barre sont déterminés automatiquement par le logiciel en fonction du nombre total de mots du texte et de la taille de la fenêtre principale.

Lorsque le graphe de répartition porte sur une *relation*, l'histogramme représente les fréquences d'occurrence cumulées des *classes* contenues dans la *relation*.

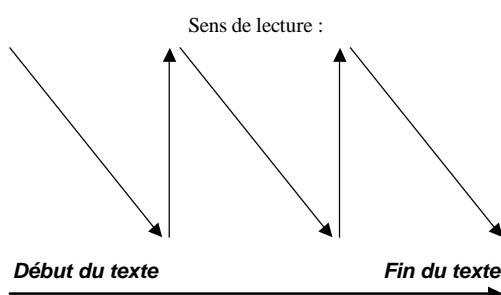
Utilisez le curseur à droite du graphe pour faire varier le nombre de barres de l'histogramme.

● Graphes des rafales et des épisodes.

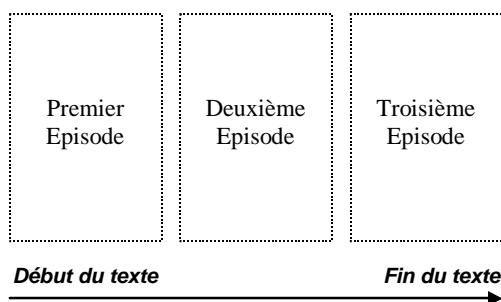
Le graphe des *rafales* et des *épisodes* est affiché quand vous visualisez un *épisode*. Chaque *rafale* est affichée sous la forme d'une ligne pointillée horizontale indiquant son étendue (longueur de la *rafale*) et sa position par rapport au début du texte. Comme pour le graphe de *répartition*, l'ordre chronologique est représenté sur l'axe horizontal du début (à gauche) à la fin du texte (à droite) :



Les *rafales* sont affichées de haut en bas et de gauche à droite, en fonction de leur ordre d'arrivée dans le texte. Lorsque le bas de l'écran est atteint, le logiciel effectue un balayage en zigzag, pour afficher le plus d'information possible :



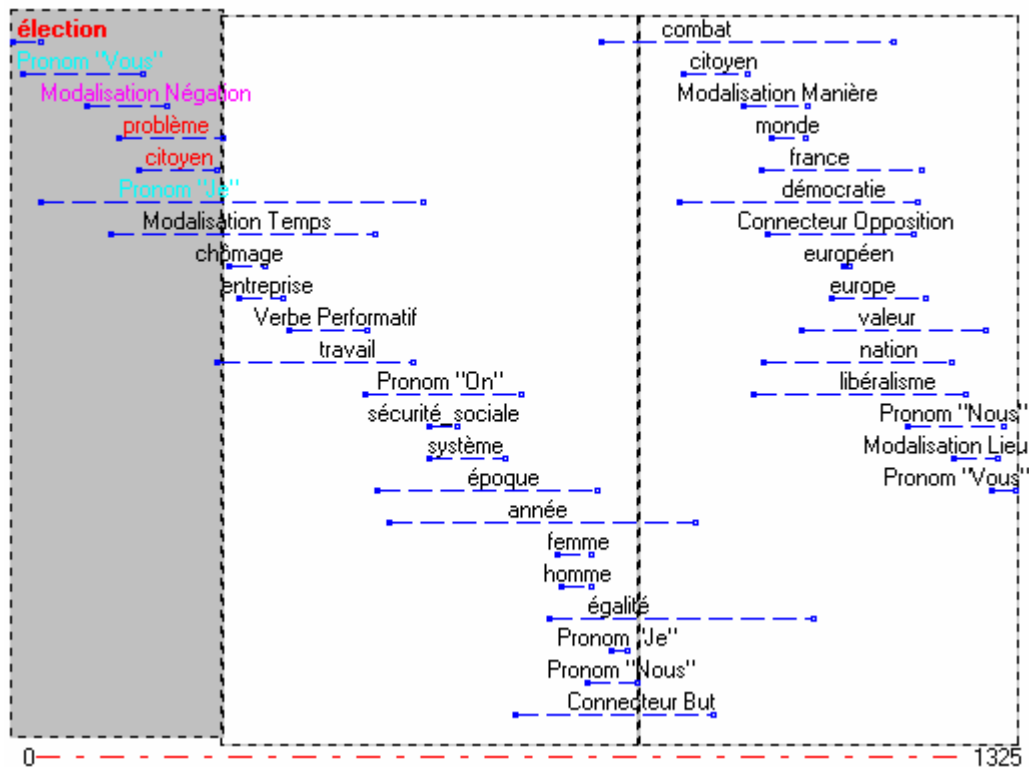
Les *épisodes* sont affichés sur le même graphe, de gauche à droite, dans l'ordre chronologique, sous la forme de grands cadres en pointillés :



Seules les *rafales* contenues dans l'*épisode* sélectionné (représenté par un cadre hachuré) sont affichées en couleurs.

Lorsque vous cliquez sur une *rafale*, la partie haute de l'écran affiche les propositions correspondantes, et son *épisode* d'appartenance est automatiquement sélectionné. Si d'autres *rafales* de même type ont été détectées, elles seront affichées en subbrillance.

Ci-dessous, on voit que le texte (exemple sur *[Jospin.txt]*) commence par des *rafales* portant sur « *élection* », « *Vous* », « *Je* », etc. et qu'il se termine par des *rafales* sur « *Europe* », « *libéralisme* », « *nation* », « *famille* », « *Nous* », « *Vous* », etc.



S'il n'y a pas assez de place sur l'écran pour afficher toutes les *rafales*, Tropes essaiera de supprimer les moins significatives (pour simplifier, disons que ce sont celles qui contiennent le moins de mots). Vous pouvez éviter la saturation de ce graphe en le construisant sur un *Scénario* (Cf. *Options d'analyse, Base de construction des relations*), ou en utilisant un *Seuil de détection des classes* élevé (Cf. *Options d'analyse, Seuils*) et/ou en supprimant l'affichage des (méta)catégories de mots (Cf. *Propriété des graphes*).

● Fonctionnement général des graphes.

Pour désactiver l'affichage des graphes, utilisez la commande *[Cacher]* du menu contextuel qui apparaît lorsque vous pressez le bouton droit de la souris sur un objet du graphe. Pour afficher à nouveau le graphe, cliquez sur un type de graphe dans la barre d'outils.

Quand on affiche une branche du scénario, Tropes évalue les relations du graphe en comptabilisant les groupes contenus dans toute la branche sélectionnée. Ce n'est qu'une approximation sur des variables qui ne sont pas indépendantes. Ce mécanisme évite d'afficher un graphe vide et apporte des informations utiles sur cette branche de l'ontologie.

Pour modifier la taille de la fenêtre allouée aux graphes (ou celle du texte), utilisez la souris pour déplacer les barres séparatrices horizontale et verticale situées au centre de l'application.

Vous pouvez changer rapidement de type de graphe en utilisant la barre d'outils de la fenêtre principale, ou le menu contextuel.

● Copie des graphes.

Lorsqu'un graphe est affiché, vous pouvez utiliser le menu *[Edition][Copier][Graphe]* pour transférer le graphe dans le *Presse-papiers* de *Windows*. Vous pourrez alors utiliser le menu *[Edition][Coller]* de votre traitement de texte ou de votre logiciel de dessin, pour récupérer ce graphe et, par exemple, l'inclure dans un rapport.

● Copie du texte sélectionné.

Lorsque l'affichage porte sur un sous-ensemble du texte (par exemple, quand vous visualisez une *classe d'équivalents*), vous pouvez utiliser le menu *[Edition][Copier][Texte sélectionné]* pour transférer les propositions sélectionnées dans le *Presse-papiers* de *Windows*. Vous pourrez alors, par exemple, utiliser le menu *[Edition][Coller]* de votre traitement de texte pour récupérer ces propositions.

● Copie des résultats.

Lorsque vous affichez un résultat d'analyse (Scénario, Catégories, etc.), vous pouvez utiliser le menu *[Edition][Copier][Résultats]* pour copier directement les résultats dans le *Presse-papiers* de *Windows*. Vous pourrez alors utiliser le menu *[Edition][Coller]* de votre logiciel de statistiques ou de votre tableur pour constituer un rapport.

● Propriétés des graphes.

Utilisez le menu *[Edition][Graphes]*, ou la commande *[Propriétés]* du menu qui apparaît quand vous pressez le bouton droit de la souris sur le graphe.

Ce dialogue permet de paramétrer les graphes. Il est possible de modifier :

- Le type de graphe affiché *[étoilé]*, *[aires]*, *[répartition]*, *[épisodes]*.
- Le graphe des rafales peut être construit sur *[toutes les catégories de mots]* (incluant connecteurs, modalisations, etc.) ou si cette case n'est pas cochée, uniquement sur les classes d'équivalents, ce qui le rend plus lisible pour des textes longs.
- Le nombre maximum d'objets affichés sur le graphe en aires.
- Vous pouvez aussi choisir la texture des aires qui peut être transparente (rayée) ou opaque et en trois dimensions (3D).

Les modifications sont prises en compte immédiatement. Pour les mémoriser définitivement cliquez sur *[Ok]*, dans le cas contraire cliquez sur *[Annuler]*.

● Modification des couleurs.

Utilisez le menu *[Edition][Couleurs]* ou la commande *[Couleurs]* du menu qui apparaît quand vous pressez le bouton droit de la souris sur la fenêtre principale.

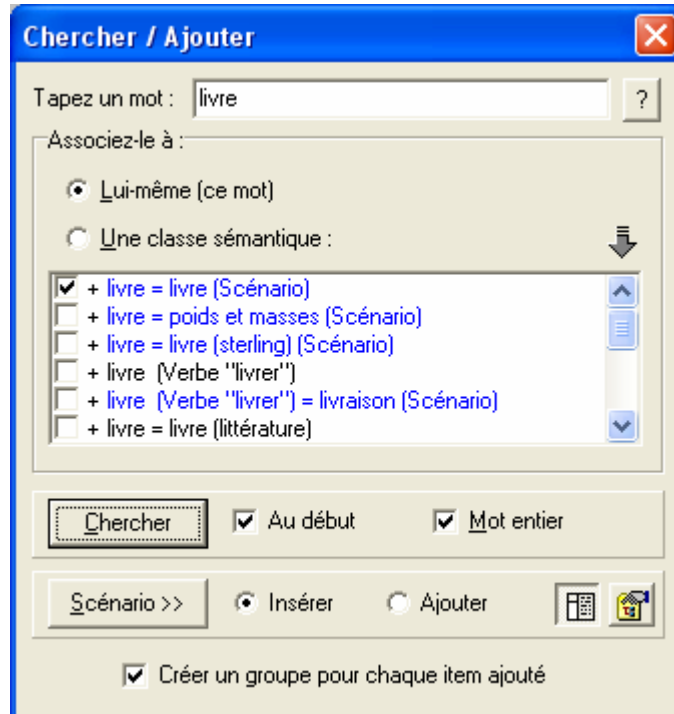
Ce dialogue permet de modifier les couleurs :

- De chaque *catégorie de mots* affichée.
- Du contour des classes affichées *avant* et *après* la classe centrale du graphe.
- Du *fond* des aires (tracées sur le graphe).
- Vous pouvez aussi choisir entre une coloration *sélective* (seul les éléments affichés sont en couleurs) ou *syntaxique* (toutes les catégories de mots sont colorées).

Le bouton *[Rétablir]* annule toutes les modifications effectuées sur les couleurs depuis la première installation du logiciel.

● Rechercher un mot.

Utilisez le menu *[Edition][Chercher]* ou bien la commande *[Chercher]* du menu contextuel qui apparaît quand vous pressez le bouton droit de la souris sur un mot du texte (ainsi que dans de nombreux outils du logiciel) :



Pour utiliser ce dialogue, vous devez taper un mot (ou un mot composé) puis appuyer sur la touche *[Entrée]* ou sur le bouton *[Chercher]*.

Si le mot recherché est trouvé, les propositions contenant ce mot sont affichées dans la fenêtre principale. Dans le cas contraire, vous entendrez un bip sonore.

Suivant l'option choisie, la recherche s'effectuera dans *[le texte]*, ou bien dans les *[Classes d'équivalents]* ou les *groupes sémantiques* du *[Scénario]* (en bleu), dans ces deux derniers cas il faut que la *classe d'équivalents correspondante* existe dans le texte et que les dernières modifications du *Scénario* aient été enregistrées.

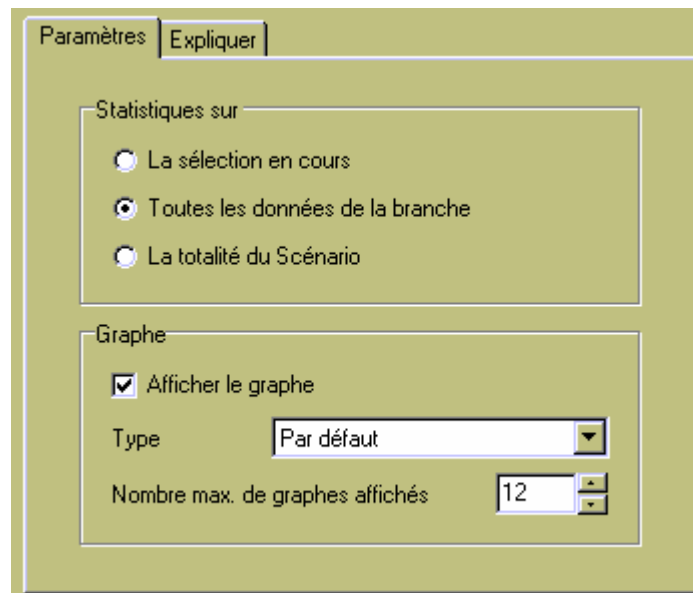
Si la case *[Commencer au début]* est cochée, la recherche s'effectuera depuis le début du texte. Dans le cas contraire, elle commencera à partir du dernier mot recherché.

Si l'*Outil Scénario* est ouvert, vous pouvez ajouter (créer un nouveau groupe) ou insérer (dans un groupe existant) directement un mot (ou un mot composé) ou une classe d'équivalents dans le *Scénario* en appuyant sur le bouton *[Scénario]* qui est au bas de ce dialogue. Une petite barre d'outils (en bas à droite) contient une option pour repositionner automatiquement le *Scénario* sur les résultats lors de la recherche.

● Générer un rapport

Un générateur d'état permet de construire automatiquement des rapports (contenant des tableaux statistiques et des graphiques) dans Microsoft Excel[®], en utilisant les résultats obtenus avec le Scénario.

Pour l'utiliser vous devez sélectionner une branche du Scénario puis utiliser le menu [Outils][Générer un rapport] :



Vous pouvez produire des tableaux statistiques :

- sur l'**élément sélectionné** (i.e. les groupes sémantiques du même niveau),
- **toutes les données de la branche** contenant l'élément sélectionné,
- la **totalité du Scénario** (ce qui peut être très volumineux).

Vous pouvez choisir d'**afficher des graphes** pour chaque tableau statistique généré, choisir leur **type** (histogramme, etc.) et contrôler le **nombre maximum de graphes affichés** (sachant qu'il est parfois difficile d'en exploiter un très grand nombre et que votre tableur ne peut pas en gérer une quantité illimitée).

Ce module nécessite qu'une version récente de Microsoft Excel[®] soit installée sur l'ordinateur. Les résultats sont exprimés en nombre d'occurrences de mots.

Notez que la modification du type de graphique par défaut dans Microsoft Excel[®] permet de contrôler la mise en page du rapport.

Tropes

CHAPITRE 2

-

Scénarios sémantiques - ontologies

● Introduction.

Les *Scénarios* sont conçus pour enrichir et filtrer les *classes d'équivalents* en fonction d'une stratégie d'analyse. Ce sont des ontologies spécifiques, qui permettent de :

- définir vos propres classifications ;
- modifier ou restructurer les dictionnaires du logiciel ;
- remplacer un thesaurus et personnaliser vos fonctions de recherche d'informations ;
- définir une grille d'analyse pour générer automatiquement un rapport (cf. *Générer un rapport*, dans ce manuel).

Pourquoi faut-il utiliser des *Scénarios* alors qu'on pourrait être tenté de modifier directement les dictionnaires de Tropes ? Parce que vous devez utiliser des classifications adaptées à vos objectifs d'analyse, qui dépendent de ce que vous voulez faire avec vos textes. Par exemple, il est parfaitement correct de classer les pastèques, concombres, melons, courges et potirons dans la famille des cucurbitacées, si vous voulez faire des analyses en botanique. Toutefois, si vous vous intéressez au domaine alimentaire, il serait plus correct de classer les pastèques et melons dans les fruits, puis les concombres, courges et potirons dans les légumes. Dans d'autres contextes, vous voudrez différencier les grandes entreprises et les PME, les substances nocives et les médicaments, certaines personnalités politiques et leurs adversaires, ... Ou bien ce qui vous intéresse, et ce que vous considérez comme secondaire.

L'utilisation d'un *Scénario* est indispensable pour effectuer une analyse correcte d'un texte. Dans la majorité des cas vous devrez retravailler ces classifications, corriger d'éventuels contresens et personnaliser vos analyses.

Par ailleurs les dictionnaires de Tropes évoluent constamment et vous pouvez en télécharger de nouvelles versions sur le site de l'éditeur. Alors que vous conservez vos *Scénarios* personnalisé, même quand vous mettez à jour votre logiciel. Lorsque vous installez une nouvelle version, l'analyse s'affine, vous bénéficiez de nouvelles classifications, mais vous ne perdez pas votre travail (qui est conservé dans vos *Scénarios*).

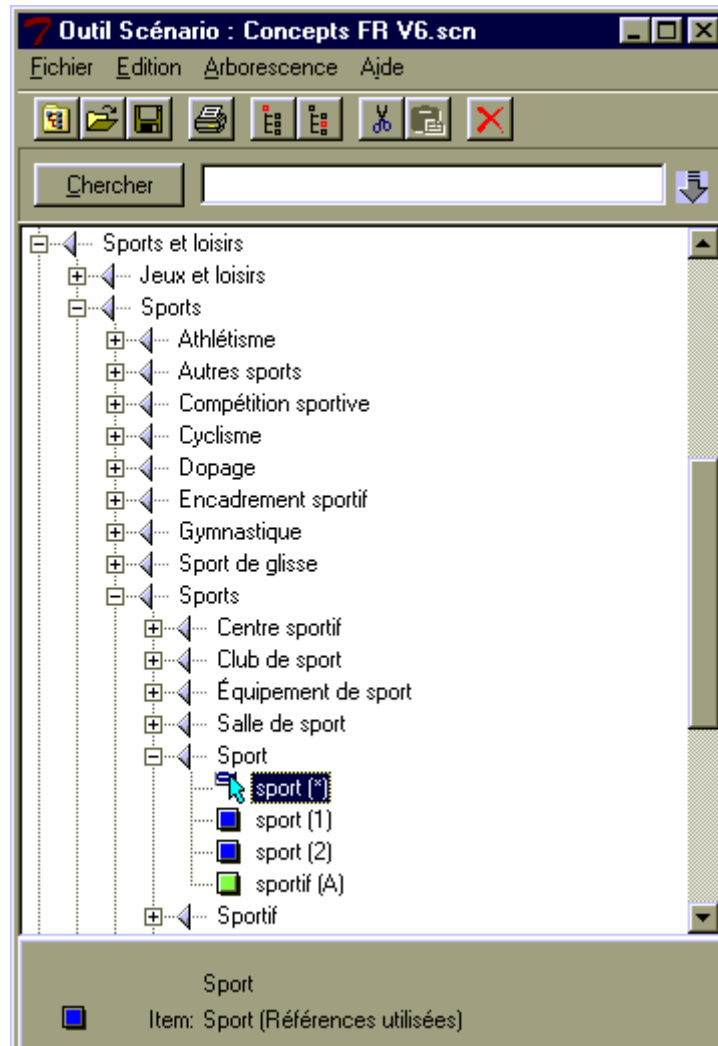
Tropes est livré avec plusieurs *Scénarios* prédéfinis qui correspondent à différentes approches de classification de vos documents :

- 1 – globale (avec le *Scénario Concept Fr V7* qui regroupe les références par grands thèmes, à la manière d'un thesaurus généraliste grand public) ;
- 2 – détaillée (avec le *Scénario Concept Fr V7 détaillés* qui regroupe les références dans un plus grand nombre de thèmes, à la manière d'une encyclopédie) ;
- 3 – très spécialisée (avec les autres *Scénarios*, ou ceux disponibles sur demande).

Notez que le mot scénario est aujourd'hui intégré à la langue française, on écrit un scénario et des scénarios, même si le pluriel italien ("scenarii") est parfois utilisé. On écrit aussi "un spaghetti" et "des spaghettis" ; on ne dit pas "un spaghetti".

● Création de Scénarios.

Pour créer un *Scénario*, actionnez le menu [Outils][Scénario] :



Tropes est livré avec des *Scénarios* par défaut, riches en classifications, qui peuvent servir de base pour constituer rapidement un nouveau plan de classement. Nous vous conseillons de conserver précieusement les *Scénarios* livrés avec le logiciel, et de les enregistrer sous un autre nom, si vous voulez les modifier.

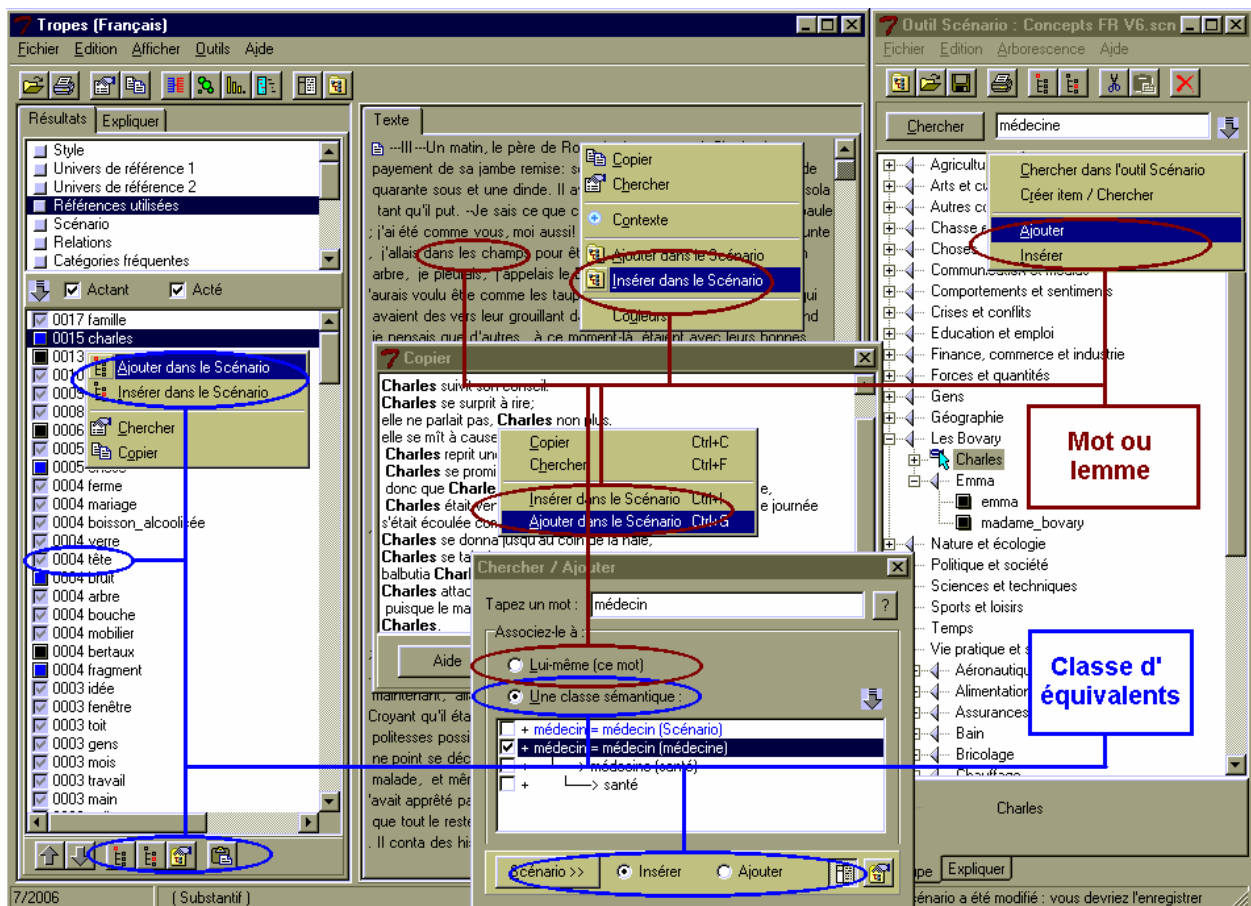
Un *Scénario* est constitué d'un certain nombre de *groupes sémantiques*, c'est-à-dire de regroupements de mots et/ou de *classes d'équivalents*, qui peuvent être hiérarchisés sur neuf niveaux de profondeur. La plupart des regroupements peuvent être réalisés à la souris, à l'intérieur de l'*Outil Scénario* ou en interaction avec la fenêtre principale de Tropes. Il suffit de créer un groupe et de mettre quelque chose dedans.

Lorsque vous visualisez une *classe d'équivalents*, à partir des résultats, vous pouvez l'ajouter directement au *Scénario* en la prenant avec la souris et en la déposant sur l'*Outil Scénario*. Depuis la version 6 de Tropes, vous pouvez aussi ajouter des lemmes de verbes et d'adjectifs directement dans le *Scénario*. Une autre approche consiste à prendre un mot dans la fenêtre [Texte] de Tropes et le faire glisser sur l'*Outil Scénario*. Les touches [Ctrl] et [Maj] peuvent être utilisées pour contrôler l'ajout ou l'insertion durant le déplacement.

Le menu contextuel (clic + bouton droit de la souris) des résultats comporte aussi deux fonctions liées au *Scénario*. La commande [Ajouter classe dans Scénario] crée une entrée dont le nom de groupe est le même que celui de la classe d'équivalents sélectionnée. La commande [Insérer classe dans Scénario] crée une entrée en utilisant le nom du groupe actuellement affiché dans le *Scénario*.

Enfin, vous pouvez créer manuellement une nouvelle entrée dans le *Scénario* à partir du dialogue [Chercher/Ajouter]. Il suffit de taper un mot, de sélectionner un groupe, et d'appuyer sur le bouton [Ajouter]. Vous pouvez choisir d'ajouter dans le *Scénario* un mot ou une *classe d'équivalents*. Une case à cocher, au bas du dialogue [Chercher/Ajouter], permet de basculer de la fonction [Insérer] à [Ajouter], et réciproquement.

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* :



Le menu [Enregistrer] de l'*Outil Scénario* permet de prendre en compte les modifications effectuées, de le sauvegarder et de mettre à jour les résultats, sans quitter ce dialogue.

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.

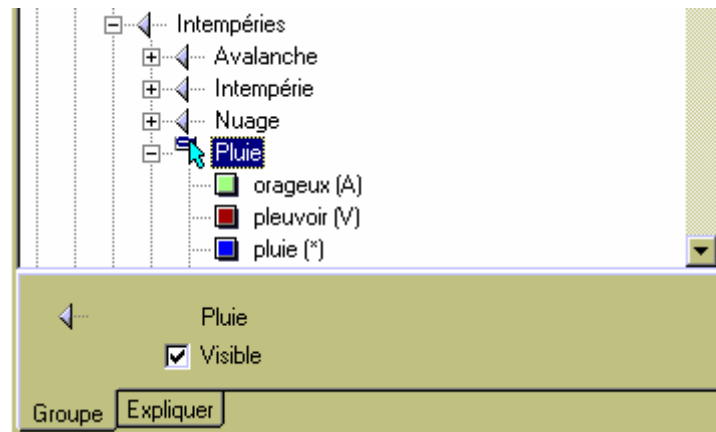
● Association d'une classe d'équivalents à un groupe sémantique.

Lorsqu'une entrée d'un groupe sémantique fait référence à une *classe d'équivalents*, l'*Outil Scénario* affiche la ligne correspondante avec une note entre parenthèses qui contient :

- une étoile (*), si l'*item* contient une *Référence utilisée*;

- un nombre (1 ou 2), si l'*item* contient un *Univers de référence* (1 ou 2);
- un (V), si vous avez ajouté un lemme de verbe;
- un (A), si vous avez ajouté un lemme d'adjectif.

Dans l'exemple ci-dessous, la classe d'équivalent "pluie" (qui fait partie des *Références utilisées*) a été regroupée avec le verbe "pleuvoir" et l'adjectif "orageux".



Les *groupes sémantiques* peuvent avoir l'attribut [Visible] ou [Invisible]. Si cette case n'est pas cochée (groupe invisible), tous les mots et *classes d'équivalents* entrés dans l'arborescence en dessous du groupe invisible ne seront pas affichés dans les résultats. Ceci permet de masquer temporairement une branche sans être obligé de la supprimer.

Le *Scénario* a été conçu pour résoudre manuellement des ambiguïtés, c'est-à-dire que si vous ajoutez un mot directement dans un *groupe sémantique*, Tropes n'effectuera plus de levée d'ambiguïté. Par exemple, si vous ajoutez le mot "culture" dans le groupe [Agriculture] du Scénario concept, vous bloquez le mécanisme de levée d'ambiguïté ("culture" sera donc toujours associé à [Agriculture], quel que soit le texte analysé). Pour bénéficier de la levée d'ambiguïté, utilisez des *classes d'équivalents*. En vertu de ces principes, vous ne devez pas mettre une *classe d'équivalents* ou un mot dans deux groupes différents (si vous le faites, l'anomalie sera signalée lorsque vous enregistrerez le *Scénario*, et la seconde entrée ne sera pas prise en compte).

Important :

- Si vous entrez un substantif au singulier le logiciel prendra automatiquement en compte ce mot à la fois au singulier et au pluriel. Dans le cas contraire, si vous entrez un mot au pluriel, le logiciel ne tiendra pas compte du mot au singulier. Par exemple si vous ajoutez le mot "libertés" (au pluriel) dans un *groupe sémantique*, le mot "liberté" (au singulier) ne sera pas pris en compte.
- Si vous entrez dans un *Scénario* un mot composé qui n'est pas reconnu comme tel, il sera ajouté au dictionnaire du logiciel. Ceci implique que les mots qui le composent ne seront plus pris en compte séparément dans les *classes d'équivalents*. Par exemple, si vous entrez "Cyrano de Bergerac" dans le *Scénario*, le logiciel n'identifiera plus "Bergerac" comme une ville de Dordogne; par contre le mot composé "Cyrano_de_Bergerac" apparaîtra dans les *Références utilisées*.

- Les *équivalents* du plus bas niveau sont prioritaires sur ceux des niveaux supérieurs. C'est-à-dire que si vous créez un *groupe sémantique* contenant un simple mot, il ne sera pas compté dans un autre *groupe sémantique* incluant une *classe d'équivalents* contenant ce même mot. Par exemple, si vous affectez le mot "Médecin" au *groupe* "Charles", il sera exclu de l'*Univers* "Médecine", même si cet *Univers* est affecté à un autre *groupe sémantique*.

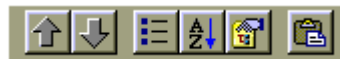
Pour réaliser un *Scénario* pertinent pour un ensemble de documents, nous vous conseillons de regrouper, à l'intérieur d'un même fichier, un échantillon de textes représentatif de ce vous voulez analyser et de construire votre *Scénario* à partir de cet échantillon. Vous pouvez utiliser ouvrir simultanément plusieurs fichiers pour constituer dynamiquement votre échantillon.

● Application du Scénario sur le texte.

Quand vous avez créé les *groupes sémantiques* de votre *Scénario*, et dès qu'il a été enregistré, vous pouvez utiliser le *Scénario* dans la fenêtre principale de Tropes.

A ce stade, vous pouvez utiliser les *groupes sémantiques* exactement comme s'il s'agissait de *classes d'équivalents* : vous pouvez visualiser les différents types de graphes, les *groupes* placés en position d'*actants* ou d'*actés*, etc.

Lorsque vous affichez ces résultats, la barre d'outils au bas de la fenêtre Tropes et un menu contextuel (clic avec le bouton droit de la souris) permettent de naviguer dans l'arborescence en rétractant ou dépliant les branches (commandes [*Tout réduire*] / [*Tout développer*]) et de changer l'ordre de tri du *Scénario*.



Les classes sémantiques présentes dans le *Scénario* sont indiquées par une case cochée dans les listes de résultats. Ceci permet de voir tout de suite si une classe est bien prise en compte dans votre classification. Dans l'exemple ci-dessous, on voit que "famille", "saison", ... sont classés dans le *Scénario*, ce qui n'est pas le cas de "Charles", "Rouault" et "Emma" :



● Méthodologie de conception de Scénarios.

Plusieurs approches sont possibles pour concevoir votre *Scénario*. Vous pouvez par exemple utiliser un *Scénario* existant (comme ceux qui sont livrés avec le logiciel), le compléter et l'affiner en fonction de vos hypothèses de travail. Vous pouvez aussi partir d'un *Scénario* vide que vous allez construire progressivement à partir d'un texte.

Nous vous proposons la méthodologie suivante, qui donne rapidement de bons résultats :

1. ouvrez l'*Outil Scénario* (menu [Outils][Scénario] de la fenêtre principale) et chargez un des *Scénarios* existants (menu [Fichier][Ouvrir] de l'*Outil Scénario*) que vous enregistrerez sous un autre nom (menu [Fichier][Enregistrer sous] de l'*Outil Scénario*) ;
2. analysez une collection de documents représentative du sujet que vous voulez traiter ;
3. affichez les Références utilisées ;
4. complétez le *Scénario* en rajoutant tout les mots qui ne sont pas classés (c'est-à-dire qui ne sont pas cochés dans la liste des Références utilisées, ou les expressions proposées par l'extracteur terminologique) et qui sont pertinents par rapport à votre problématique d'analyse ;
5. affichez les Relations et utilisez-les pour classer des mots composés pertinents dans le *Scénario* (par exemple, si vous étudiez les fichiers Bovary... dans les textes d'exemple, vous devez regrouper "Madame Bovary" avec "Emma" pour effectuer une analyse correcte de ce classique littéraire) ;
6. étudiez éventuellement les listes de verbes et d'adjectifs, dont certaines entrées permettent de renforcer le *Scénario* ;
7. appliquez votre *Scénario* sur le texte (menu [Fichier][Enregistrer] de l'*Outil Scénario*) ;
8. placez-vous sur le *Scénario* dans la fenêtre résultats de Tropes (si besoin en utilisant le menu [Afficher][Scénario] de la fenêtre principale) ;
9. examinez le résultat obtenu en descendant la liste des groupes du *Scénario* ;
10. rectifiez les classifications erronées, en ajoutant des directement les mots problématiques dans l'*Outil Scénario* (par exemple, si un "Monsieur Blanc" est classé dans les couleurs, vous devrez déplacer ce mot dans un autre groupe sémantique) ;
11. supprimez les branches anecdotiques (qui ne sont pas utiles pour l'analyse) de votre *Scénario* (par exemple, les jours de la semaine ne présentent généralement pas d'intérêt quand vous analysez des articles de journaux) ;
12. interprétez le résultat et retournez à l'étape 7, si nécessaire ;
13. quand vous avez tout vérifié, alors votre analyse est terminée.

Parce que les *Références utilisées* sont triées par fréquences d'occurrence décroissantes, vous avez la garantie d'avoir classé les références principales de vos textes quand vous utilisez cette méthode, même si vous n'intégrez pas toutes les classes sémantiques dans le *Scénario*. Par exemple, si vous arrêtez la construction du *Scénario* lorsque les *Références utilisées* atteignent une fréquence de deux occurrences (donc généralement peu pertinentes), on ne pourra pas vous accuser d'avoir bâclé l'analyse. Vous vous serez simplement focalisé sur l'essentiel, en arrêtant l'analyse à un certain moment.

● Gestion des Scénarios.

Le menu [Fichier] de l'*Outil Scénario* permet de :

- créer un nouveau *Scénario*,
- ouvrir un autre *Scénario* (qui doit déjà exister),
- enregistrer le *Scénario* courant sous un autre nom,
- imprimer la définition du *Scénario*,
- retrouver le dossier contenant vos fichiers *Scénarios*.

Lorsque vous créez un nouveau *Scénario* (menu [Fichier][Nouveau...]), un assistant vous aidera à le générer automatiquement à partir du texte ou à le fabriquer à partir d'un (ou plusieurs) *Scénario(s)* existants :



Si vous ne voulez pas utiliser de *Scénario* existant, il suffit de créer un *Scénario* vide, que vous complétez. Il est toutefois fortement déconseillé de travailler sans *Scénario*.

Notez que le menu [Edition][Annuler toutes les modifications] permet d'annuler toutes les modifications effectuées depuis le dernier enregistrement du *Scénario*.

Important :

- 1 - Les noms des fichiers contenant les *Scénarios* doivent obligatoirement être suffixés par [.SCN] (par exemple : *Bovary.scn*).
- 2 - Les fichiers des *Scénarios* doivent impérativement être stockés dans le sous-dossier [*Scenario*] du dossier d'installation du logiciel (*c:\Program files\Tropes\Tropes...\scenario*), ou dans le dossier que vous avez sélectionné lors de l'installation du logiciel, si vous avez choisi le mode multi-utilisateurs.
- 3 - Vous devez sauvegarder régulièrement le contenu de ce dossier. Utilisez le menu [Fichier][Dossier] de l'*Outil Scénario* pour copier vos fichiers de *Scénario*.

● Extraction terminologique.

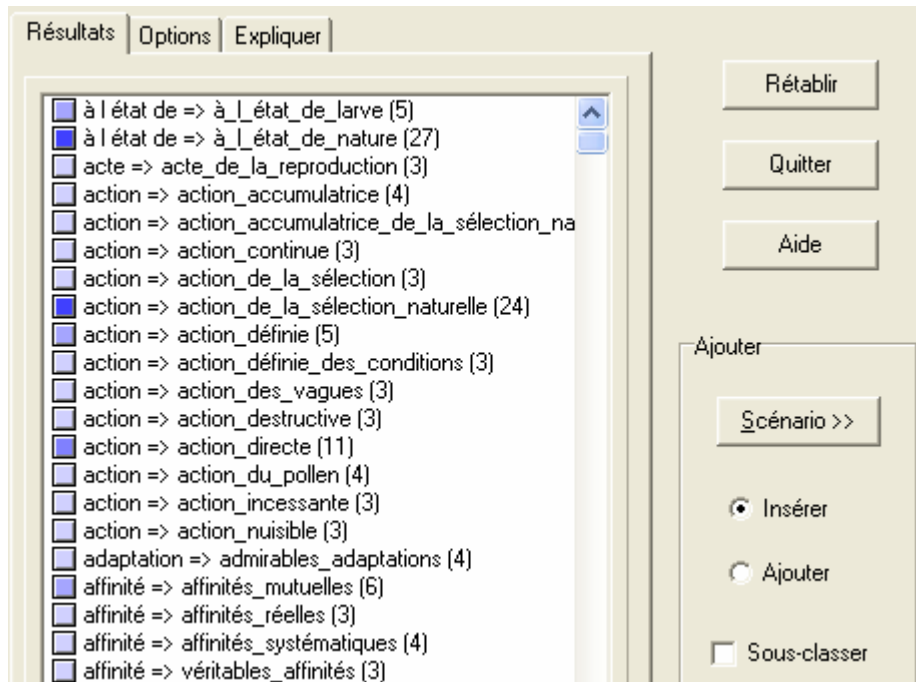
L'extracteur terminologique est capable de repérer automatiquement la majorité des expressions et mots composés significatifs, ainsi que tous les substantifs non classés dans le Scénario existant. Cet outil permet, à la fois, d'enrichir rapidement les Scénarios du logiciel (en regroupant, par exemple, tous les sigles avec les expressions qui y correspondent) et d'obtenir une classification plus précise (en proposant, par exemple, de câbler les termes qui posent des problèmes d'ambiguïté et/ou qui peuvent « parasiter » la liste des Relations).

L'extracteur terminologique répond à un triple objectif :

- 1 – Extraire automatiquement du texte tous les mots composés (i.e. suite de termes répétés contenant au moins un substantif et cohérents d'un point de vue linguistique) qui peuvent présenter un intérêt pour l'analyse.
- 2 – Proposer une liste de références pour compléter le Scénario et/ou dresser la liste de tout ce qui n'est pas encore classé.
- 3 – Accélérer la construction du Scénario.

L'extracteur terminologique est couplé sémantiquement au Scénario, c'est-à-dire que lorsque vous sélectionnez d'une famille de termes, Tropes essaiera automatiquement de repositionner l'Outil Scénario sur le groupe qui paraît le plus pertinent pour accueillir l'expression choisie. Le logiciel est aussi capable de rechercher automatiquement une famille d'expression dans le texte. Ces deux fonctions de couplage sémantique sont désactivables (cf. options "Localiser" en bas et à droite du dialogue). Les termes extraits sont précédés par un code de couleur qui permet de voir quels sont les termes ou expressions les plus fréquents (sombre = fréquent, clair = peu fréquent).

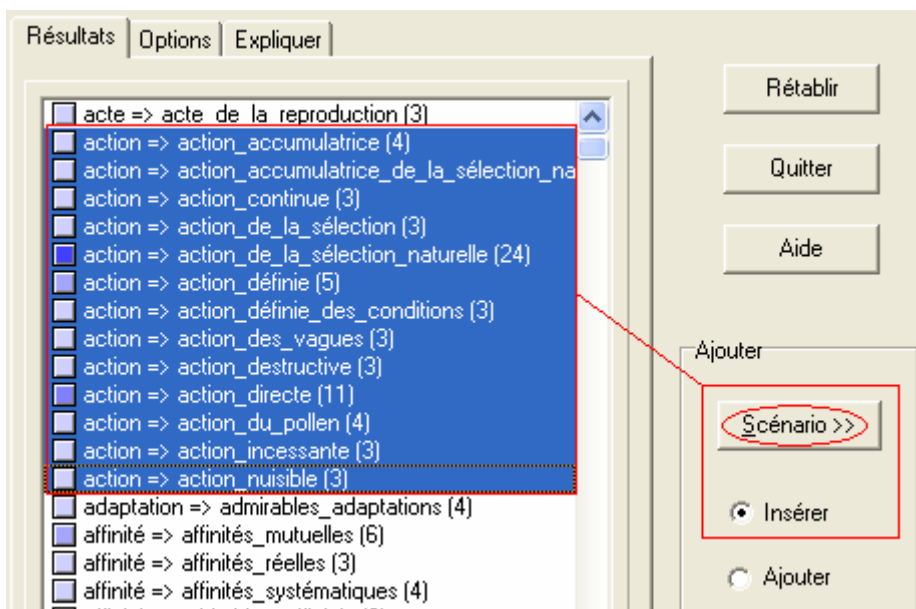
L'extracteur se lance via le menu [Outils] de Tropes, quand vous avez analysé un texte ou (de préférence) plusieurs textes, à partir desquels vous voulez bâtir une classification. L'exemple suivant affiche le résultat d'analyse terminologique de *L'origine des espèces* de Darwin (212.000 occurrences) :



Chaque terme (ou expression) extrait est affiché en ligne, avec son groupe d'appartenance. Il est précédé d'un carré de couleur indiquant sa fréquence d'occurrence.

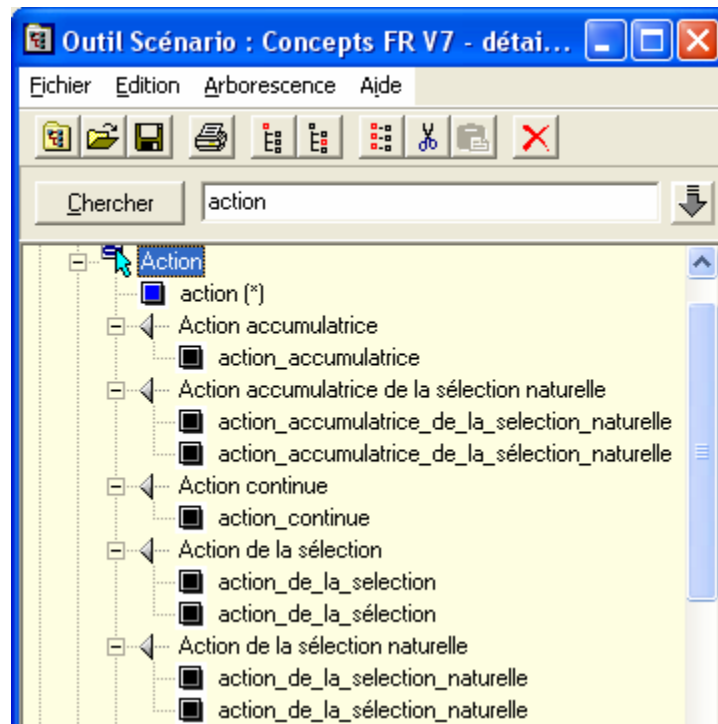
Pour enrichir votre classification, vous devez ouvrir l'outil Scénario et sélectionner les termes qui vous intéressent dans l'extracteur terminologique.

Dans l'exemple qui suit, on a sélectionné un certain nombre d'expressions qui nous paraissent intéressantes et qui ont été regroupées par le logiciel dans la catégorie "action" :



Puis nous allons cocher la case *Insérer*, et appuyer sur le bouton Scénario (à droite), pour les transférer automatiquement dans le Scénario.

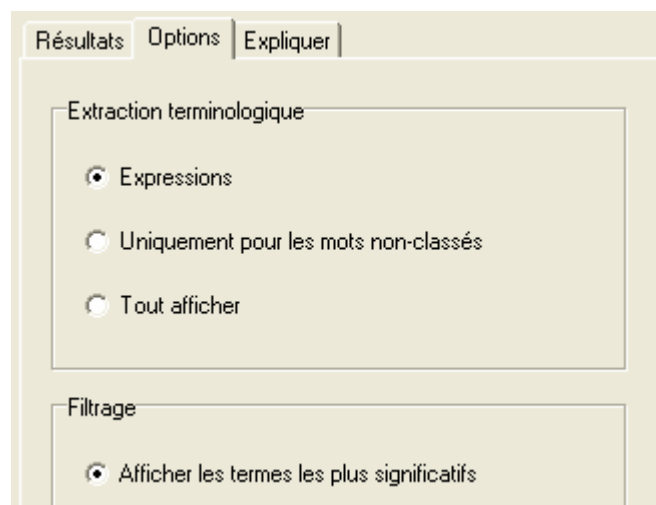
Le logiciel va automatiquement créer toutes les entrées et les étiquettes de groupe correspondantes dans l'outil Scénario.



Vous n'aurez ensuite plus qu'à enregistrer votre Scénario (menu [Fichier][Enregistrer] de l'outil Scénario) pour voir le résultat dans le texte analysé.

Lorsque que vous sélectionnez un terme, et si l'outil Scénario est ouvert, le logiciel va essayer de positionner automatiquement le Scénario sur la catégorie sémantique la plus proche. Ce comportement peut être désactivé en enlevant la case Scénario du cadre [Localiser], à droite du dialogue.

Pour finir, l'extracteur terminologique est doté d'un onglet [Options] qui vous permet d'ajuster finement le niveau d'extraction.



A vous de jouer !

Tropes

CHAPITRE 3

-

Délimiteurs

● Segmentation de textes.

L'outil *Délimiteur* a été conçu pour effectuer une segmentation automatique de documents. Autrement dit, à l'intérieur d'un même texte, les *délimiteurs* peuvent être utilisés pour isoler automatiquement de multiples acteurs, un interviewer et un interviewé, des populations, les chapitres d'un livre, etc.

L'utilisation de *délimiteurs* peut nécessiter un codage préalable des documents (vous devrez souvent insérer des codes spécifiques pour que cela fonctionne).

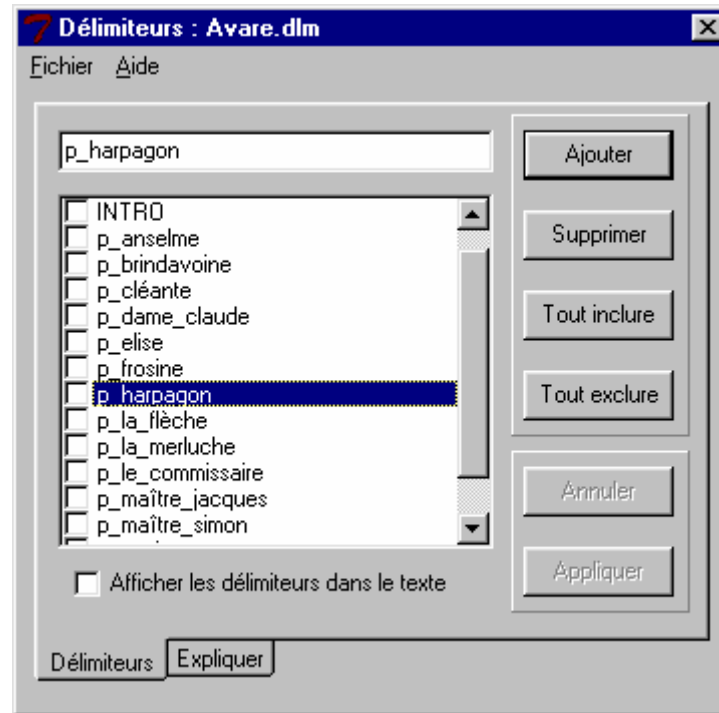
Un *délimiteur* commence à partir de l'endroit où il a été trouvé dans le texte. Il s'arrête lorsque le *délimiteur* suivant a été trouvé. Par exemple, si « p_harpagon » et « p_maitre_jacques » sont deux *délimiteurs*, la séquence suivante délimite quatre parties d'un texte :

	<i>Partie</i>
P_HARPAGON.-- Allons, venez çà tous, que je vous distribue mes ordres pour tantôt et règle à chacun son emploi. Approchez, dame Claude. Commençons par vous. (Elle tient un balai.) Bon, vous voilà les armes à la main. Je vous commets au soin de nettoyer partout, et surtout prenez garde de ne point froter les meubles trop fort, de peur de les user. Outre cela, je vous constitue, pendant le souper, au gouvernement des bouteilles ; et, s'il s'en écarte quelqu'une et qu'il se casse quel-que chose, je m'en prendrai à vous et le rabattrai sur vos gages.	1
P_MAITRE_JACQUES, à part.-- Châtiment politique.	2
P_HARPAGON.-- Allez... Vous, Brindavoine, et vous, La Merluche, je vous établis dans la charge de rincer les verres et de donner à boire, mais seulement lorsque l'on aura soif, et non pas selon la coutume de certains impertinents de laquais qui viennent provoquer les gens et les faire aviser de boire lorsqu' on n'y songe pas. Attendez qu'on vous en demande plus d'une fois, et vous ressouvenez de porter toujours beaucoup d'eau.	3
P_MAITRE_JACQUES, à part.-- Oui ; le vin pur monte à la tête.	4

Ces quatre parties correspondent à une discussion entre deux personnages, Harpagon et Maître Jacques (voir le fichier Avare.txt, dans les textes d'exemple livrés avec le logiciel). Si nous voulons isoler automatiquement le discours d'Harpagon (parties 1 et 3) de celui de Maître Jacques (parties 2 et 4), alors nous devons utiliser des *délimiteurs*.

● Création de délimiteurs.

Pour créer un fichier de *délimiteurs*, utilisez le menu [Outils][Délimiteurs].



Pour créer une nouvelle entrée dans les *délimiteurs*, vous devez inscrire un mot, dans le champ en haut du dialogue, puis appuyer sur le bouton [Ajouter]. Ce mot doit être représentatif d'une séquence du texte analysé. Dans l'exemple, les codes « DESC » et « INTRO » ont été utilisés pour identifier les parties descriptives (Desc) et l'introduction (Intro) d'une pièce de théâtre, tandis que les différents tours de paroles des personnages de la pièce ont été codés en utilisant le code « p_ » avant chacun des noms des personnages.

Pour effacer un *délimiteur* existant, sélectionnez- le, puis appuyez sur [Supprimer].

Lorsque tous vos *délimiteurs* sont créés, vous pouvez choisir les parties de texte que vous désirez ne pas prendre en compte en cochant les *délimiteurs* correspondants : si votre *délimiteur* est coché tout le texte qui suit ne sera pas pris en compte, et, a contrario, s'il n'est pas coché, le reste du texte sera visible.

Pour désactiver tous les *délimiteurs*, utilisez le bouton [Tout inclure]. Pour exclure tout le texte pris en compte par les délimiteurs, utilisez le bouton [Tout exclure]. Ceci permet, par exemple, de voir si l'intégralité du texte est bien délimitée.

Lorsque vous avez terminé, utilisez le bouton [Appliquer] pour relancer l'analyse de votre texte en excluant les parties que vous avez voulu lui enlever (si vous utilisez des *délimiteurs* pour la première fois, le logiciel vous demandera de donner un nom de fichier pour les sauvegarder).

Si l'option [Afficher les délimiteurs dans le texte] est cochée, Tropes comptabilisera tous vos *délimiteurs* comme des mots du texte, et les affichera. Dans le cas contraire, si cette option n'est pas cochée, Tropes n'affichera pas les codes correspondants, qui ne seront pas comptabilisés comme des mots.

Remarque : si vos *délimiteurs* sont des mots composés non reconnus par le logiciel, vous devez relier les mots qui les composent avec le caractère souligné (par exemple « p_harpagon »). Evitez d'utiliser des *délimiteurs* ambigus avec les mots du texte (par exemple,

« *Harpagon* » n'est pas un bon délimiteur pour analyser l'Avare de Molière, parce que ce mot est utilisé à la fois pour introduire le personnage et dans les dialogues entre les personnages).

● **Fichiers de délimiteurs.**

Utilisez le menu [Fichier], qui permet de :

- créer un **nouveau** fichier de *délimiteurs* (vide),
- **ouvrir** un fichier de *délimiteurs* (qui doit déjà exister),
- **enregistrer** et appliquer les modifications.
- **enregistrer** le fichier de *délimiteurs* sous un autre nom.

Notez que les *délimiteurs* sont enregistrés automatiquement chaque fois que vous utilisez le bouton [Appliquer].

Les *délimiteurs* sont stockés dans le même dossier que les *Scénarios*, et doivent faire l'objet des mêmes précautions d'emploi (cf. chapitre 2, *Gestion des Scénarios*).

Tropes

CHAPITRE 4

-

Notions d'analyse de texte

● Le travail de Tropes.

Pour traiter un texte, le logiciel travaille en 6 étapes :

- ❶ - découpage des phrases et des propositions,
- ❷ - levée d'ambiguïté des mots du texte,
- ❸ - identification des *classes d'équivalents*,
- ❹ - statistiques, détection des *rafales* et des *épisodes*,
- ❺ - détection des *propositions remarquables*,
- ❻ - mise en forme et affichage du résultat.

Les mots sont regroupés en plusieurs grandes catégories lexicales, dont six nous intéressent vraiment :

- les verbes,
- les connecteurs (conjonctions de coordination et subordination, locutions conjonctives),
- les modalisations (adverbes ou locutions adverbiales),
- les adjectifs qualificatifs et les nombres,
- les pronoms personnels,
- les substantifs et les noms propres (i.e. les références).

Pour effectuer une analyse, le logiciel effectue un traitement complexe visant à affecter tous les mots significatifs dans ces catégories, à analyser leur répartition en sous-catégories (*catégories de mots, classes d'équivalents*), à étudier 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, rafaes, épisodes, propositions remarquables*).

● Le découpage propositionnel.

Afin d'en simplifier l'analyse, le logiciel découpe le texte en propositions (phrases simples). Cette première étape s'appuie sur un examen de la ponctuation ainsi que sur des fonctions complexes d'analyse de syntaxe qui ne seront pas détaillées ici.

Ceci permet d'obtenir des statistiques de co-occurrence très fiables (*relations*), puisqu'il n'est pas possible que deux mots se trouvent dans la même proposition grammaticale s'ils ne sont pas étroitement liés.

Le découpage propositionnel comporte inévitablement des erreurs (propositions trop courtes ou trop longues), mais cela ne fausse pas les résultats.

● La levée d'ambiguïté.

L'interprétation automatique des mots du français (et de toute langue vivante écrite ou parlée) nécessite de résoudre de nombreuses ambiguïtés :

- grammaticales et syntaxiques (par exemple le mot « livre » correspond à un *nom* dans « une pile de livre » et à un *verbe* dans « on nous livre du poisson »),
- sémantiques (par exemple le mot « livre » peut renvoyer à une *œuvre littéraire*, à une *monnaie* ou à un *poids*).

On considère qu'en moyenne au moins un mot sur quatre est ambigu.

Une des fonctions importantes de ce logiciel est de résoudre ces ambiguïtés en utilisant plusieurs algorithmes de résolution de problèmes. Il lui est impossible d'effectuer ce travail très complexe de façon parfaite, mais son taux d'erreur est suffisamment faible pour permettre une analyse correcte de votre texte.

● Catégories de mots.

Les **verbes** :

- ♦ *factifs* expriment des actions (« travailler », « marcher »,...)
- ♦ *statifs* expriment des états ou des notions de possession (« être », « rester »,...)
- ♦ *déclaratifs* expriment une déclaration sur un état, un être, un objet, (« dire », « croire »,...)
- ♦ *performatifs* expriment un acte par et dans le langage (« promettre », « exiger »,...).

Les **connecteurs** (conjonctions de coordination et de subordination, locutions conjonctives) relient des parties de discours par des notions de :

- ♦ *condition* (« si », « dans l'hypothèse », « au cas où », ...)
- ♦ *cause* (« parce que », « puisque », « car », « donc », ..),
- ♦ *but* (« pour que », « afin de », ...)
- ♦ *addition* (« et », « ensuite », « puis »...),
- ♦ *disjonction* (« ou ... ou », « soit ... soit », ...),
- ♦ *opposition* (« mais », « cependant », « toutefois », ...),
- ♦ *comparaison* (« comme », « tel que », « ainsi que », ...),
- ♦ *temps* (« quand », « lorsque », « avant que », ...),
- ♦ *lieu* (« où », « jusqu'où », ...).

Les **pronoms personnels** sont affichés en genre (« Je », « Tu », « Il », ..) et en nombre (« Ils », « Elles », ...).

Les **modalisations** (adverbes ou locutions adverbiales) permettent à celui qui parle de s'impliquer dans ce qu'il dit, ou de situer ce qu'il dit dans le temps et dans l'espace, par des notions de :

- ♦ *temps* (« maintenant », « hier », « demain », ...),
- ♦ *lieu* (« là-bas », « en haut », « ici », ...),
- ♦ *manière* (« directement », « ensemble », ...),
- ♦ *affirmation* (« tout à fait », « certainement », ...),
- ♦ *doute* (« peut-être », « probablement », ...),
- ♦ *négation* (« ne...pas », « ne...guère », « ne...jamais », ..),
- ♦ *intensité* (« très », « beaucoup », « fortement », ...).

Les adjectifs :

- ♦ *objectifs* permettent de caractériser des êtres ou des objets, indépendamment du point de vue du locuteur (par exemple les adjectifs de couleur),
- ♦ *subjectifs* indiquent une appréciation sur quelque chose ou quelqu'un, ils permettent d'exprimer le point de vue du locuteur (« intéressant », « gentil », « agréable », ...)
- ♦ *numériques* regroupent les nombres (en lettres ou en chiffres) et les adjectifs ordinaux et cardinaux.

Les **autres catégories** de mots contiennent des *pronoms*, des *articles*, des *prépositions*, et des *adjectifs non qualificatifs*. Nous vous demandons de ne pas tenir compte de ces catégories, car elles ne sont utilisées que pour le traitement de levée d'ambiguïté du logiciel.

D'un point de vue général, on peut dire que :

- ♦ les *connecteurs* et *modalisations* de *temps* et de *lieu* permettent de situer l'action,
- ♦ les *modalisations d'intensité* et de *négation* permettent de dramatiser le discours,
- ♦ les *connecteurs* de *cause* et de *condition* permettent de construire un raisonnement,
- ♦ les *connecteurs d'addition* permettent d'énumérer des faits ou des caractéristiques,
- ♦ plus particulièrement, les *connecteurs d'opposition* permettent à la fois d'argumenter, de relativiser et de présenter des points de vue opposés.

Dans les versions précédentes de Tropes la catégorie « Connecteur » s'appelait « Joncteur ». Ce mot a été modifié pour être en conformité avec les appellations en usage actuellement, mais le contenu de cette catégorie est inchangé.

● Analyses statistiques, probabilistes et cognitives.

Tropes effectue plusieurs types d'analyse du texte :

- des statistiques sur la fréquence globale d'apparition des grandes *catégories de mots*, et de leurs sous-catégories,
- des statistiques sur la cooccurrence et le taux de liaison des *classes d'équivalents* et des *catégories de mots*,
- une analyse probabiliste des mots arrivant en *rafales* et une analyse géométrique des *rafales* délimitant les *épisodes*,
- une analyse cognitivo-discursive (ACD) permettant de détecter les *propositions remarquables*.

Les statistiques sont utilisées, entre autres, pour construire les graphes et pour mettre en forme les résultats.

Les *Catégories de mots fréquentes* et le *Style général du texte* sont obtenus en comparant la répartition des fréquences d'apparition des catégories observées dans le texte avec des normes de production langagière. Ces normes ont été élaborées en étudiant un grand nombre de textes différents et sont stockées dans des tables internes au logiciel.

● Classes d'équivalents et relations entre équivalents.

Les *classes d'équivalents* regroupent les références (noms communs ou noms propres) qui apparaissent fréquemment dans le texte et qui possèdent une signification voisine. Par exemple : « père » et « mère » seront regroupés dans la classe « famille » par le logiciel.

Les *Univers de référence* regroupent les mots contenus dans les *classes d'équivalents* afin de permettre au logiciel d'élaborer une représentation du contexte. Ainsi, le dictionnaire sémantique de Tropes est construit selon 3 niveaux de classification. Au niveau le plus bas se situent les *Références utilisées*, elles-mêmes regroupées de façon plus large dans les *Univers de référence 2*, qui à leur tour sont regroupés dans les *Univers de référence 1*.

Dans l'exemple ci-dessous, le mot « garde des sceaux » fait partie de la *référence* « ministre », incluse dans l'*Univers 2* « homme politique », et dans l'*Univers 1* « politique ». L'*Univers 1* « politique » regroupe des concepts très larges comme les « doctrines politiques », les « instances politiques », etc :

Univers 1	Univers 2	Classes	Mots
politique	doctrine politique	communisme	communisme
politique	doctrine politique	communisme	marxisme
politique	doctrine politique	libéralisme	capitalisme
politique	doctrine politique	libéralisme	libéralisme
politique	homme politique	chef d'état	chef d'état
politique	homme politique	chef d'état	président de la république
politique	homme politique	ministre	garde des sceaux
politique	homme politique	ministre	ministre
politique	homme politique	parlementaire	député
politique	homme politique	parlementaire	sénateur
politique	instance politique	gouvernement	gouvernement

● Considérations relatives à la taille du texte.

Parce que la plupart des analyses effectuées par Tropes sont statistiques, le logiciel est sensible à la quantité de texte traitée.

Si votre corpus est constitué de fichiers contenant des textes **très courts** (moins d'une page), les résultats obtenus ne seront probablement pas significatifs, et le logiciel peut avoir des difficultés à résoudre les ambiguïtés sémantiques (parce qu'il ne disposera pas de suffisamment de mots pour résoudre les implicites). Si vous disposez d'un nombre raisonnable de textes courts, nous vous conseillons de les regrouper dans un seul fichier. Si vous disposez d'un grand nombre de petits textes, mettez-les dans plusieurs fichiers, en les regroupant, par exemple, par ordre chronologique.

Dans le cas contraire, si votre corpus est constitué de fichiers contenant des textes **très longs** (plus de cent pages) et/ou regroupant des propos qui font référence à des contextes radicalement différents, la pertinence des résultats risque d'en être affectée : les discours sont mélangés. Essayer, autant que possible, de diviser vos gros corpus en fichiers ne dépassant pas quelques centaines pages de texte. Vous pouvez utiliser un *Scénario sémantique* pour fusionner les résultats obtenus en analysant plusieurs fichiers (Cf. Manuel de Zoom).

Certains « symptômes » permettent de détecter que vous analysez un texte trop gros :

- Tropes met plus d'une minute pour traiter votre document.
- Les *classes d'équivalents* sont difficilement exploitables : elles regroupent souvent des mots ayant des sens différents.
- Tropes n'a détecté aucun actant significatif, et ceci ne peut être expliqué par l'utilisation massive de pronoms personnels.

Important : à cause de limitations techniques internes au logiciel, il est possible que certaines *classes d'équivalents* peu fréquentes ne s'affichent pas, si le texte est très long.

● Analyse de propos hétérogènes : questions ouvertes, dépêches, etc.

Les remarques qui suivent s'appliquent à tous les corpus hétérogènes, obtenus en regroupant dans un même fichier des propos émanant de multiples individus, sans cohérence linéaire (i.e. il ne s'agit pas du discours d'un même narrateur, ou d'une interaction entre plusieurs interlocuteurs, respectant une suite logique ou un ordre chronologique strict).

Considérant que le découpage propositionnel de Tropes s'appuie exclusivement sur des règles de grammaire, vous devrez utiliser un signe de ponctuation non-ambigu (point d'interrogation ou d'exclamation) pour forcer le logiciel à séparer les différents propos (vous mettrez, par exemple, un point d'exclamation à la fin de chaque réponse à une question ouverte, pour la séparer de la suivante).

Si le corpus comprend des réponses à plusieurs questions, vous devrez utiliser des *délimiteurs* pour regrouper les réponses et/ou séparer les réponses des questions, et analyser chaque réponse séparément.

Si vous disposez d'un indicateur permettant de constituer votre corpus en fonction d'une variable externe (zone géographique, type de population, période, etc.) vous pouvez l'utiliser pour répartir les propos à l'intérieur de plusieurs fichiers (chaque fichier contenant, par exemple, les propos correspondant à une seule variable), que vous pourrez alors analyser séparément et comparer. Vous pouvez aussi coder ces variables dans les textes et les utiliser comme *délimiteurs*.

Si le corpus n'a aucune cohérence linéaire (par exemple, si à l'intérieur de chaque fichier, les propos ont été regroupés en vrac, sans respecter aucune logique particulière), les résultats qui dépendent des analyses chronologiques de Tropes (*propositions remarquables, rafales, épisodes, graphes de répartition*) ne seront pas significatifs : vous ne devrez donc pas essayer de les interpréter.

● Traitement de très gros corpus textuels.

Cette technique a été conçue pour extraire rapidement de l'information à partir de gros corpus de presse, effectuer certains traitements d'Intelligence économique, constituer des revues de presse automatisées, effectuer des études historiques ou sociologiques, et/ou générer des mots-clés à partir d'un thésaurus.

Elle n'est exploitable que si vous utilisez Tropes Zoom.

La méthode proposée est la suivante :

- 1 - Récupérez sur votre disque dur le plus grand nombre possible de documents traitant du même thème (abonnez-vous à des services télématiques, récupérez l'information sur Internet, exportez le contenu de cédéroms bibliographiques, ...) Si possible, regroupez vos très petits fichiers (brèves, dépêches) dans des fichiers de taille moyenne en prenant soin de respecter la cohérence dans chaque document (Cf. supra, *Analyse de propos hétérogènes*).
- 2 - Créez une base documentaire avec Zoom.
- 3 - Effectuez une recherche documentaire en utilisant un critère suffisamment large pour garantir l'exhaustivité du résultat, sélectionnez tous les fichiers trouvés (si c'est possible, ou un échantillon conséquent, s'il y a beaucoup trop de fichiers), puis lancez l'analyse avec Tropes.
- 4 - Construisez un *Scénario* regroupant les thèmes les plus représentatifs de ce que vous voulez analyser.
- 5 - Relancez l'indexation de votre base documentaire, en cochant éventuellement l'option [*Cumul de Scénario*].
- 6 - Analysez le *Scénario* avec le générateur de rapports (cf. chapitre correspondant dans ce même manuel), ou bien traitez le fichier cumul de *Scénario* avec un tableur, un gestionnaire de base de données ou un logiciel de statistiques (Cf. Manuel de Zoom, Annexes, Exemple de fichier cumul de *Scénario*).
- 7 - Repérez des irrégularités dans les données obtenues et analysez les textes correspondants avec Tropes. Si nécessaire, retournez à l'étape 2, perfectionnez votre *Scénario* et recommencez le processus d'analyse jusqu'à ce que vous ayez trouvé les documents ayant la meilleure valeur explicative du corpus observé.

● Analyse de discours et d'entretiens.

Si vous analysez un fichier texte contenant la retranscription du discours de plusieurs individus, effectuez d'abord une analyse globale de l'ensemble de vos documents, puis utilisez éventuellement des *délimiteurs* pour traiter séparément les propos des différents personnages, et comparez les différents résultats obtenus, en vous posant (entre autres) les questions suivantes :

- Est-ce que tous les participants/narrateurs ont parlé de la même chose ?
- Qu'est-ce qui les différencie (comparez les références) ?
- Est-ce qu'ils associent les mêmes références (exploitez les relations, les graphes en aires et en étoile) ? Si ce n'est pas le cas, pourquoi ?
- Comment en parlent-ils (exploitez les catégories de mots) ?
- Donnent-ils la même importance à toutes les références (comparez l'ordre d'arrivée des classes d'équivalents et les actants) ? Si ce n'est pas le cas, pourquoi ?
- A-t-on particulièrement insisté sur certaines références ? Pourquoi ?
- Que s'attendait-on à trouver dans les résultats (vous pouvez en établir la liste *a priori* avant de commencer l'analyse) ? Quels sont les références inattendues ? Pourquoi ?
- Dans quel ordre chronologique sont présentés les références (exploitez les rafales et les épisodes, graphes de répartition) ? Essayez d'expliquer pourquoi (rappelez-vous qu'on mémorise plus facilement les introductions et conclusions) ?
- A-t-on refusé de répondre à certaines questions ? Pourquoi ?
- A-t-on cherché à convaincre quelqu'un ? Pour quelle raison ? Est-ce réussi ?
- etc.

Si vous avez le temps de le faire, vous pouvez résoudre manuellement les anaphores, c'est-à-dire remplacer chaque pronom personnel par ce à quoi il renvoie. Prenons un exemple. Supposons que nous voulions analyser le discours de deux individus, Pierre et Philippe, qui ont parlé de trois autres personnes, Alain, Marie et Sandrine, et que le texte comporte de nombreux pronoms personnels (« je », « tu », « il », « elle », ...). Vous pouvez alors remplacer certains de ces pronoms (en utilisant la commande *[rechercher/remplacer]* de votre traitement de texte) de la façon suivante : « Je », « Tu », .. par « Pierre » ou « Philippe », « Elle(s) », ... par « Marie » et/ou « Sandrine », « Il », ... par «Alain », etc. Ceci permettra de compter très précisément le nombre de fois où on a parlé de telle ou telle personne, de voir si ce sont des *actants* ou des *actés*, etc.

Lorsque vous effectuez des retranscriptions à partir de l'oral, vous devez impérativement inclure de la ponctuation dans le texte, sinon le logiciel ne sera pas capable d'effectuer correctement le découpage propositionnel et le fonctionnement de l'analyse sera perturbé.

● Etudes littéraires.

Pour analyser une pièce de théâtre, vous pouvez utiliser la méthode décrite ci-dessus (l'analyse des propos de multiples acteurs est quasiment équivalente à l'analyse d'un entretien entre plusieurs interlocuteurs).

Lorsque vous étudiez des textes de grande taille (un livre entier, par exemple) nous vous conseillons d'analyser tout d'abord chaque chapitre, puis éventuellement d'effectuer une synthèse en traitant ensuite la totalité du texte (Cf. supra, *Considérations relatives à la taille du texte*).

● Comparer deux textes.

La comparaison de deux textes revient à la fois à faire une analyse des contenus (i.e. des *classes d'équivalents*) et de la mise en scène (i.e. des *catégories de mots*).

On pourra, par exemple, comparer :

- les poids respectifs (*taux d'utilisation pondérés*) et les positions (*actants/actés*) des *classes d'équivalents*,
- la chronologie d'apparition des thèmes principaux (*rafales, épisodes et graphes de répartition*),
- la fréquence de co-occurrence des références (*graphes en étoile et en aires, relation, taux de liaison des relations, Scénarios*),
- les types d'actes à travers l'analyse des catégories verbales (*catégories de mots, styles et mises en scène*),
- le type de logique développée et de prise en charge à travers l'analyse des *connecteurs* et des *modalisations* (*catégories de mots, rafaes, épisodes*),
- et, d'une façon plus générale, une synthèse des propositions centrales permettant d'introduire les personnages et les thèmes principaux (*propositions remarquables*).

Tropes

Annexes

● Conversions des formats de fichiers.

Observations concernant la conversion des fichiers :

Format	Extension	Description
Texte	*.TXT	Aucune conversion n'est effectuée sur ce format, considéré en ANSI ISO-8859-1 (ISO-Latin1, aussi appelé ANSI Windows) ou au format Unicode, s'ils respectent la norme BOM (Byte Order Mark).
HTML	*.HTM	Même remarques pour le (HyperText Markup Language) que pour les formats XML et SGML (voir ci-dessous).
Microsoft Word	*.DOC *.DOCX	Le logiciel effectue une conversion native des formats Microsoft Word®. Par ailleurs, il peut extraire le texte par analyse binaire, pour les formats Word antérieurs à Office 2007.
Microsoft PowerPoint	*.PPT *.PPTX	Le logiciel effectue une conversion native des formats Microsoft Word®. L'installation de Microsoft Filter Pack peut être nécessaire pour convertir ces fichiers.
OpenOffice	*.ODT *.ODP	Le logiciel effectue une conversion native des formats OpenDocument Text ou Presentation. L'installation de Microsoft Filter Pack ou des filtres correspondants peut être nécessaire pour convertir ces fichiers.
PDF	*.PDF	Le logiciel utilise un pilote (driver) de périphérique Adobe® (IFilter) pour extraire le texte pertinent des fichiers PDF® (Portable Document Format). Si l'installation d'Adobe Reader® n'est pas toujours nécessaire, il faut parfois installer le driver Adobe® IFilter pour bénéficier de ce format. L'usage d'un logiciel de reconnaissance de caractères (OCR) externe peut être nécessaire pour certains fichiers image (ceux qui sont issus d'une numérisation par scanner).
RTF	*.RTF	Le logiciel interprète les formats RTF (Rich Text File) contenant des caractères ANSI ISO-8859-1, Apple Macintosh, ou ASCII IBM, codés sur 7 ou 8 bits. Le format Unicode n'est pas accepté. Parce que des disparités existent entre les normes RTF, il est possible que des caractères parasites apparaissent dans certains fichiers.
SGML	*.SGM	Le logiciel interprète les balises des fichiers SGML (Standard Generalized Markup Language, ISO 8879) et effectue une conversion des variables HTML spécifiques aux langues européennes (comme les caractères accentués). Le logiciel n'interprète pas les DTD. Le format UTF-8 (UCS Transformation Format-8, ISO 10646/RFC 2279) est automatiquement converti en ANSI ISO-8859-1. Les autres formats Unicode (Universal Character Set, ISO 10646) ne sont pas toujours directement acceptés.
ASCII	*.ASC	Le logiciel utilise l'API Windows OemToChar pour convertir ce format en ANSI ISO-8859-1.
Macintosh	*.MAC	Les jeux caractères Apple Macintosh® Latin sont convertis en caractères ANSI ISO-8859-1. L'extension .MAC est arbitraire et peu prête à confusion avec certaines macros : préférez le format TXT.
XML	*.XML	Le logiciel utilise son interpréteur SGML (voir ci-dessus) pour lire le format XML (Extensible Markup Language). Les mêmes restrictions sont donc applicables à ce format. Le moteur n'interprète pas les scripts ni les feuilles de style XSL (Extensible Stylesheet Language).

L'analyse de certains formats de fichiers nécessite que des pilotes IFilter 32 bits soient installés sur votre ordinateur. Si ce n'est pas le cas, alors vous pouvez installer Microsoft Office Filter Pack 32 bits pour activer la conversion de ces fichiers (téléchargement gratuit sur le site de Microsoft).

Le logiciel utilise les logiciels installés dans votre système ou bien un composant interne pour convertir les documents Microsoft Word® (conversion approfondie). En cas de problème (fichier corrompu, etc.) la conversion est basculée automatiquement sur une analyse heuristique, qui va tenter de récupérer le texte par extraction binaire.

Lorsque le logiciel effectue une analyse binaire (heuristique) pour extraire le texte, il est possible que des caractères parasites apparaissent à l'affichage. Par définition, cette méthode d'extraction du texte ne peut pas être parfaite, parce que le logiciel ne tient pas compte du format natif du fichier.

Dans tous les cas, les fichiers cryptés ou, plus généralement, protégés par mots de passe, ne peuvent pas être lus et convertis directement par le logiciel. Vous devez convertir ces fichiers manuellement pour les analyser.

Pour plus d'information sur les formats de fichiers supportés, consulter :

- <http://www.apple.com>,
- <http://www.adobe.com>,
- <http://www.microsoft.com>,
- <http://www.unicode.org>,
- <http://www.utf-8.com>,
- <http://www.w3.org>,
- <http://www.w3.org/MarkUp>,
- <http://www.w3.org/MarkUp/SGML>.