

## **Mode d'emploi du logiciel TROPES**

(source : <http://www.acetic.fr/semantique.htm>)

### **L'analyse de textes**

Le fonctionnement de Tropes (et de Zoom) s'appuie sur une logique d'Intelligence Artificielle capable de résoudre les ambiguïtés lexicales et sémantiques des langues. Il intègre des lexiques et des réseaux sémantiques (dictionnaires de classification).

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

- découpage des phrases et des propositions,
- levée d'ambiguïté des mots du texte,
- identification des classes d'équivalents sémantiques,
- statistiques, détection des rafales et des épisodes,
- détection des propositions remarquables (contraction du texte),
- mise en forme et affichage du résultat.

Les trois premières étapes sont exploitées par les processus d'indexation de nos deux moteurs de recherche sémantiques Zoom et Index.

Les mots sont regroupés en plusieurs grandes catégories, 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,
- les pronoms personnels,
- les substantifs et les noms propres.

Pour effectuer une analyse sémantique, Tropes 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, rafales, épisodes, propositions remarquables).

### **Découpage propositionnel**

Afin d'en simplifier l'analyse, Tropes 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 (appelées "mises en relations"), puisqu'il n'est pas possible que deux mots se trouvent dans la même proposition grammaticale s'ils ne sont pas étroitement liés.

### **La levée d'ambiguïté sémantique**

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 un mot sur quatre est ambigu. Une des fonctions importantes de Tropes est de résoudre ces ambiguïtés en utilisant plusieurs algorithmes d'intelligence artificielle. Il lui est impossible d'effectuer ce travail de façon parfaite, mais son taux d'erreur est suffisamment faible pour permettre une analyse correcte de votre texte

### Les (méta)catégories de mots

Tropes effectue une classification sémantique des mots-outils, qui est très utile pour faire de l'analyse qualitative ou de l'analyse du discours. Six catégories grammaticales (et sémantiques) sont analysées automatiquement : les verbes, les connecteurs (conjonctions), les modalisations (adverbes), les pronoms personnels et les adjectifs.

#### 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 où", "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 qualificatifs :

- 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", ...)

Les autres catégories de mots contiennent certains pronoms, des articles, des prépositions, et des adjectifs non qualificatifs. Elles ne sont utilisées que pour le traitement de levée d'ambiguïté du logiciel.

### Analyses statistiques, probabilistes et cognitives

**Tropes** effectue automatiquement plusieurs traitements 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 méta catégories de mots,
- une analyse probabiliste des mots arrivant en rafales et une analyse géométrique des rafales délimitant les épisodes,
- une analyse cognitivo-discursive (A.C.D.) permettant de détecter les propositions remarquables (contraction du texte).

Les statistiques sont utilisées, entre autres, pour construire les graphes et pour mettre en forme les résultats. **Tropes** identifie des Catégories de mots fréquentes et le Style général du texte en comparant la répartition des fréquences d'apparition des catégories sémantiques 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.

### Episodes et rafales

Ces deux notions permettent d'étudier la chronologie d'un discours :

- Une *rafale* regroupe des occurrences de mots (contenus dans une classe d'équivalents sémantiques ou une (méta)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 *épisode* correspond à une partie du texte où un certain nombre de rafales se sont formées et terminées. Ce sont de grands blocs d'argumentation, représentatifs de la structure du discours observé.

### Classes d'équivalents sémantiques

Les classes d'équivalents regroupent les mots (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 à **Tropes** d'élaborer une représentation du contexte.

Ainsi, le dictionnaire des équivalents sémantiques de **Tropes** est construit selon trois 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	instance politique	gouvernement	gouvernement

Ces classifications sont exploitées, par l'intermédiaire d'un Scénario, dans les processus d'indexation de nos deux moteurs de recherche sémantiques Zoom et Index.

### Propositions remarquables

Les propositions remarquables caractérisent le texte, sans se répéter. 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 pour les extraire. Pour simplifier, disons que chaque proposition du texte se voit attribuée un score calculé en fonction de son poids relatif, de l'ordre d'arrivée et de son rôle argumentatif. Les propositions sont ensuite triées, puis filtrées en fonction de leur score. Afin de vous permettre de jouer sur le taux de contraction du texte, et pour garantir que le résultat obtenu reflète bien le texte analysé, **Tropes** permet de régler la quantité de propositions affichées.

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 résumé du texte (qui nécessiterait de le réécrire).