
TROPES

Tropes est un programme d'analyse documentaire développé et distribué par la société ACETIC. La version que nous utilisons est une version de démo disponible sur le site web <http://www.acetic.fr>. Ce logiciel permet une analyse des formes du discours d'un texte. Il est basé sur la reconnaissance de termes (d'entrées lexicales), et utilise des classes d'équivalences sémantiques. A partir de la distance entre les termes repérés, de leur fréquence d'occurrence (ou de celle de termes notionnellement reliés), il propose, notamment, de calculer le « rythme » du texte (rafales et épisodes), les espaces notionnels, distingue les « propositions remarquables » (avec analyse thèse/rhème), ainsi que les situations énonciatives.

Manipulation

- Sous winnt z:\taln\lexical: lancer le raccourci tropes. Menu Fichier ⇒ Ouvrir : sélectionnez un texte de démo (.txt).
- Observez le résultat de l'étiquetage (catégorisation) représenté par différentes couleurs, et par un message en bas de la fenêtre si vous placez la souris sur un mot. Utilisez l'aide contextuelle du logiciel pour explorer les possibilités du programme.
- Exploitez les résultats : au moyen de la fenêtre de résultat et du menu contextuel (clic droit de la souris directement sur le texte) visualisez toutes les options :
 - le style général du texte,
 - les propositions remarquables,
 - les univers de référence,
 - les références utilisées,
 - les éléments du scénario courant,
 - les mises en relation,
 - les catégories de mots utilisées,
 - les épisodes et les rafales.

(Voir aussi : Comparaison de textes, Graphes des classes et des relations, Graphes des rafales et des épisodes, Afficher le contexte d'un mot, Historique des classes affichées, Création de scénarios, Options d'analyse).

- Repérez les différentes étapes de l'analyse d'un texte (ouvrez d'autres textes de démo) :

- 1 - découpage des phrases en propositions,
- 2 - levée d'ambiguïté des mots du texte,
- 3 - identification des classes d'équivalents et de leurs relations,
- 4 - élaboration de statistiques, détection des rafales et des épisodes,
- 5 - détection des propositions remarquables,
- 6 - mise en forme et affichage du résultat.

Les mots sont regroupés en plusieurs catégories, dont 6 sont pertinentes pour l'analyse :

- 1 - les verbes,
- 2 - les joncteurs (conjonctions de coordination et de subordination),
- 3 - les modalisations (adverbes ou locutions adverbiales),
- 4 - les adjectifs qualificatifs,
- 5 - les pronoms personnels,
- 6 - les classes d'équivalents.

le logiciel effectue un calcul 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).

- Quelles critiques pouvez-vous formuler sur la justesse des analyses ?

LEXICO

Lexico est un outil puissant pour le traitement lexicométrique de grands corpus basé sur l'analyse statistique des séquences. Ce programme est développé par l'équipe LEXICO à l'université de la Sorbonne nouvelle (Paris 3) dirigée par André Salem (<http://www.cavi.univ-paris3.fr/Ilpga/ilpga/tal/lexicoWWW/>). Il permet le calcul de fréquences, du pourcentage (taux) de réalisation des formes, donne une classification hiérarchique des termes (par fréquence, ordre alphabétique, rang).

Le répertoire où se trouvent les exécutables est sur winnt : z:\taln\lexical\lexico2. Utilisez l'aide en ligne ainsi que le fichier fbrochure.html.

On dispose de quatre outils :

1. la segmentation
2. la recherche contextuelle
3. la comparaison par parties
4. la comparaison chronologique

Manipulations

1. Préparation du texte

Choisissez un petit texte .txt ; ouvrez-le à partir de démarrer⇒programmes⇒accessoires : word pad.

L'analyse statistique d'un texte se base sur l'unité textuelle qu'est la forme pour estimer ses occurrences, délimiter les séquences de mots où elle apparaît et comparer ses fréquences dans une partie donnée du corpus par rapport à une ou plusieurs autres. Pour rendre possible et fiable cette segmentation il est nécessaire de définir des normes de saisie qui assurent la cohérence du texte et de son codage.

Le texte ne doit comporter que des minuscules ; pour désambigüiser les noms propres et noms communs homographes on accolera au nom propre un signe de distinction (par exemple : ~barre). Toutefois, ce signe ne doit alors pas figurer dans la liste des délimiteurs. **Ponctuations, délimiteurs** : afin que le programme d'analyse LEXICO2 puisse reconnaître les formes du corpus que l'on segmente, on affecte un statut à chaque signe présent dans le texte. On distinguera d'une part les délimiteurs : -- _ : ; / . , ? & iquest ; ! & iexcl ; * \$ " + = () { } ainsi que l'espace. (Cette liste est donnée à titre indicative et peut être modifiée). D'autre part, sont appelés non-délimiteurs tous les caractères restants contenus dans la police : a, b, c, ... Toute suite de caractères non-délimiteurs bornée à ses deux extrémités par des délimiteurs est considérée comme une occurrence, une forme à relever et à classer. On pourra également insérer des **clés** et des marques de **paragraphes** si nécessaire.

Enregistrer les modifications opérées et fermer le texte.

2. Segmentation

Lancer l'exécutable segmentation.exe ; ouvrir le texte que vous venez de modifier. Vérifier les délimiteurs. lancer le traitement.

Pour un rapport du traitement effectué ouvrir via word pad le fichier atrace.txt

Les fichiers .dic et .par ont été créés dans votre répertoire ; ouvrir le fichier .dic (résultat du calcul de fréquence des formes) via word pad.

3. Recherches contextuelles

Lancer le programme de concordance context.exe. Explorer les nombreuses options proposées : les tris possibles, les multiples requêtes, etc. utilisez le menu contextuel de la souris en cliquant sur les mots trouvés.

4. Comparaison par parties

Lancer le programme parties.exe et utiliser le fichier démo DUTCH.par

Quels résultats obtenez-vous ?

5. Comparaison chronologique

Lancer chrono.exe et utiliser le fichier démo DUTCH.par. Quelles fenêtres apparaissent, quels fichiers sont créés ? à quoi servent-ils ?