

ANALYSEUR DE TEXTE

(Programme Gratuit)

L'Analyseur de textes a pour vocation de donner divers renseignements sur un (ou un ensemble de plusieurs) texte(s) français de style général. C'est, par définition, un produit imparfait. Les points qu'il signale ne sont pas toujours des erreurs et, réciproquement, des points qui pourraient être des erreurs ne sont pas signalés... Il lui arrive quand même parfois de signaler des passages critiques, du moins espérons-le. Des tests vous permettront de savoir dans quelle proportion cet outil peut vous être utile.

PRINCIPE

L'Analyseur a en gros deux types de fonctions :

1- Des fonctions statistiques. Il compte les mots d'un texte, les classe par catégorie (mots de longueur 1, 2, 3, 4, 5, 6, 7 et plus), et édite la liste des quinze ou vingt mots les plus utilisés par catégorie, avec leur nombre respectif d'occurrences. Ce n'est pas nécessairement enthousiasmant pour un texte de 4 000 signes mais se révèle parfois plein d'enseignements pour un roman...

Remarque :

Les majuscules sont ignorées par le programme. **CE = ce = Ce**

Une particularité d'Analyseur est de tenir compte, dans une large mesure, des pluriels et féminins. Ainsi, un adjectif comme **soumis** est pris en compte de façon identique sous toutes ses formes, i.e **soumise, soumis, soumises**. Par contre, il ignore que **vitrail** est le singulier de **vitraux** (moi-même je ne l'ai appris que récemment). Ce n'est pas un système expert.

De la même façon, la plupart des verbes (en particulier les verbes de longueur supérieure ou égale à 7) sont regroupés sous la forme de l'infinitif. Les mots **déterminent, déterminant détermination, déterminisme et déterminera** sont regroupés sous la racine **déterminer**.

Ceci entraîne que des noms (une **marche** par exemple) sont considérés comme des formes verbales (ici le présent du verbe **marcher**). Analyseur signalera donc une répétition du verbe **marcher** dans la phrase : Il saute de **marche** en **marche**.

Enfin, deux mots dont les sept premiers caractères sont identiques sont considérés comme égaux par le programme.

2- Des fonctions d'analyse **locale**, facultatives (l'utilisateur a moyen d'en mettre une ou plusieurs hors-service comme nous le verrons plus loin). Elles sont de trois ordres, la recherche d'adverbes, de participes présents, et de répétitions.

- Les adverbes

L'Analyseur signale les paragraphes avec trop d'adverbes, **trop** étant un nombre entre 1 et ce que vous voulez (à définir au départ dans la fenêtre **paramètres** du programme).

Est considéré comme adverbe tout mot de sept lettres ou plus terminé par **ment** (hormis **comment**). Le programme assimile donc les substantifs comme **grognement** à des adverbes (ce qui n'est pas toujours

gênant, l'effet de lourdeur généré par un adverbe ou un substantif étant du même ordre), mais aussi des verbes conjugués au présent comme **assomment**. Il peut donc signaler "**trop d'adverbes**" à mauvais escient.

- Les **participes présents**

De même, l'Analyseur signale les paragraphes avec trop de participes présents, **trop** étant un nombre entre 1 et ce que vous voulez (à définir aussi dans la fenêtre **paramètres** du programme).

Est considéré comme participe présent tout mot de sept lettres ou plus terminé par **ant**. Les restrictions du paragraphe ci-dessus s'appliquent aussi à celui-ci.

- Les formes du **verbe FAIRE**

L'emploi trop fréquent de "Faire" est souvent considéré comme une maladresse, ou un facteur d'imprécision. On peut fixer le nombre maximum de formes de ce verbe (et éventuellement, du verbe "Refaire") que le programme "accepte sans sourciller".

- Les **répétitions**

Le programme ne signale les répétitions que pour les mots de longueur 5 et plus, et en suivant les règles ci-dessous :

Pour les mots de longueur 5 et 6, il y a répétition à partir de **trois fois** le même mot.

Pour les mots de longueur 7 et plus, il y a répétition à partir de **deux fois** le même mot (y compris les mots dérivés).

Les répétitions peuvent concerner un seul paragraphe ou plusieurs consécutifs (de façon paramétrable par l'utilisateur). Dans le premier cas, le paragraphe contenant la répétition est édité. Lorsque le terme trop souvent répété est réparti entre plusieurs paragraphes consécutifs, seul le dernier d'entre eux est édité.

Remarque :

Ces règles vont évoluer au gré des demandes des premiers utilisateurs. Dans une prochaine version du programme, elles seront paramétrables par l'utilisateur.

.....

Remarque

Lorsque l'Analyseur signale un problème local, il édite en même temps le paragraphe où se situe le problème. Il n'est malheureusement pas possible d'indiquer le numéro de la page où se trouve le paragraphe.

MODE D'EMPLOI

Dans le dossier **ANALYSEUR** se trouvent le programme et les deux tables de mots nécessaires au fonctionnement du programme (MOT1-6 et MOT7).

Il existe deux versions du programme : **Analyseur** pour Mac + et Mac SE (avec processeur 68000), et **Analyseur II** pour Mac II ou Mac SE/30 (avec processeur 68020 ou 68030). Les deux versions ont des fonctionnalités rigoureusement identiques, la version II a simplement été compilée de façon optimisée pour le 68020.

On peut utiliser la version de base pour Mac + sur un Mac II, sans aucun problème (c'est plus lent). Par contre, on ne peut pas utiliser la version II sur un Mac + ou un Mac SE, le message "**Hardware configuration error**" apparaît immédiatement !

ATTENTION

Ne déplacez pas les tables hors de portée du programme Analyseur sous peine d'erreur lors du chargement. Pour cette raison, il vaut mieux laisser le dossier (Programme + tables) tel qu'il est. Il peut se placer n'importe où, sur un disque dur ou une disquette.

De même, rien ne vous empêche de copier l'ensemble du dossier et d'en faire des sauvegardes. Rien n'est protégé. Faites donc circuler le programme !

1- Les textes à analyser (maximum 50 à la fois et il est conseillé de travailler sur des copies) doivent être sous la forme **texte seul, Word ou Write Now** (avec quelques imperfections dans ces deux derniers cas). Analyseur ne fonctionne vraiment de façon optimale qu'avec des documents au format **texte seul**. Il n'est pas encore interfacé Mac Write (j'y travaille, mais c'est long).

2- On clique sur l'icône Analyseur. Après quelques instants apparaît une fenêtre de paramètres.

PARAMETRES

Nombre d'adverbes : Sortie sur écran Sortie sur fichier

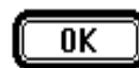
Verbe FAIRE : Analyse indépendante

Formes en - ANT- : Ecart Répétition :

Analyse des mots de longueur au moins égale à :



SAVE



Les cases **nombre d'adverbes**, **Verbe FAIRE** et **formes en -ANT -** servent à définir à partir de combien d'adverbes, de formes du verbe Faire ou de formes en - ANT - dans un même paragraphe le programme signale qu'il y en a trop... Pour supprimer ces fonctions, mettre la valeur **0** dans chaque case.

La case **écart répétition** (par défaut à 1), indique le nombre de paragraphes consécutifs pris en compte pour la recherche de répétitions. Pour supprimer la recherche de répétition, mettre la valeur **0** dans cette case.

Le bouton **sortie sur fichier** (par défaut à oui) , sert à éditer les résultats sur **UN FICHER TEXTE SEUL du nom d'ANALYSE + heure (pouvant être par la suite consulté sous Editeur, ou Mac WRITE)** Ceci est vivement conseillé si le ou les textes sont longs. Le programme ne demandant pas d'intervention de l'opérateur, on peut le lancer et aller faire ses courses en le laissant tourner (il faut près d'une heure pour analyser un roman entier de 400 000 signes...).

Sinon, lorsque le bouton **sortie sur fichier** est désélectionné (et devient **sortie sur écran**), l'analyse a lieu sur écran, sans sauvegarde des résultats. Ceci peut convenir pour un texte court que l'on corrige au fur et à mesure...

Le bouton **Analyse indépendante** (par défaut à oui) permet d'analyser plusieurs textes successifs de façon indépendante (ou au contraire de comptabiliser les fréquences de mot pour plusieurs textes à la fois).

La case **Analyse des mots de longueur au moins égale à** permet de ne s'occuper que des mots les plus longs. L'analyse va plus vite... D'après mes essais, 3 ou 4 semblent de bonnes moyennes.

Remarque : On passe d'une case à l'autre au moyen de la touche de Tabulation...

Le bouton **Save** permet de sauver l'état des paramètres pour toute utilisation ultérieure.

Une fois les paramètres établis, cliquer OK.

.....

L'Analyseur affiche une fenêtre d'ouverture de fichier classique, avec une petite case en bas à droite indiquant combien de fichiers sont déjà sélectionnés. Cliquer sur un nom de fichier à analyser : la fenêtre s'affiche à nouveau. Lorsque tous vos fichiers ont été entrés (maximum 50), cliquer sur **annuler** dans la fenêtre. Après un certain temps de lecture des tables, l'exécution commence. Les fichiers seront traités l'un après l'autre, de façon indépendante ou non suivant le choix effectué.

SORTIES

Le programme édite :

- Le nom du fichier-texte en cours de traitement -

(éventuellement) Des messages concernant un paragraphe suivis du paragraphe en question
Le nombre d'adverbes, de formes du verbe FAIRE et/ou de formes en ANT dans le texte

A LA FIN

Analyseur édite le nombre de mots d'une longueur donnée rencontrés dans le texte, avant analyse. En effet, le mot "neuves" est de longueur 6 mais sera assimilé au mot "neuf" (donc de longueur 4) après analyse...

Pour chaque longueur de mot analysé, édition de la table des occurrences maximales, classées de façon décroissante. Lorsque la sortie se fait sur écran, des pauses sont incorporées dans le défilement (signalées par un **bouton "Suite"**). Cliquez dessus pour relancer.

PROBLEMES DIVERS

Problèmes de mémoire (ERROR 7)

Analyseur requiert 800 K... C'est beaucoup, quand on a un Mac avec un seul Méga de mémoire. Il peut n'en utiliser que 760 (version limite) sous Finder, mais ne descend pas plus bas. Donc, si le message **Error 7** apparaît (qui veut dire Mémoire insuffisante), il faut reconfigurer son système en éliminant des Inits et/ou de la mémoire cache, de façon à libérer de la place. Ne pas oublier de redémarrer avant de relancer Analyseur...

Au fur et à mesure de l'apparition de mots nouveaux inconnus de lui, l'Analyseur enrichit ses tables en mémoire. C'est une opération coûteuse en temps. **IL EST DÉCONSEILLÉ D'UTILISER ANALYSEUR AVEC AUTRE CHOSE QUE DES TEXTES FRANCAIS SOUS PEINE DE VOIR PEINER DURANT DES HEURES. Inutile aussi de lui faire analyser des fautes d'orthographe, il vaut mieux passer un coup de correcteur orthographique avant...**

Certains INITS ou CDEV font planter Analyseur... C'est en particulier vrai de "WINDOW", de "AltCDEF" (il suffit de mettre Analyseur dans la liste des applications à exclure, à partir du tableau de bord) et de "Disk Express II", même si celui-ci est mis à OFF dans le tableau de bord (en effet, Disk Express a tendance à modifier les paramètres d'un fichier lorsque celui-ci est utilisé, et les modifications entrent en conflit avec le programme Analyseur)... De même, les INITS qui altèrent les ressources WDEF du système entrent souvent en conflit avec l'Analyseur.

En cas de problème, essayer de nouveau en supprimant tous les INITS...

Problèmes avec des documents Word

Le programme Word sauve ses textes en gardant les "versions précédentes"... Donc, si l'Analyseur signale un problème et que ce problème est ensuite corrigé dans le document, Analyseur signalera de nouveau le même problème la fois suivante, parce qu'il analysera aussi la version précédente sauvegardée avec le reste du document ! Désagréable, et difficile à corriger.

Une solution (bâtarde, je sais) consiste à sauver son texte depuis Word en écrasant la version précédente (option **Save as** ou **Sauver sous**), en gardant le même nom.

Tout autre problème est probablement dû à une Bogue. Me consulter ou m'insulter, au choix. J'essaierai de réparer. (**Nouveau Téléphone : 61 93 69 72, heures de Bureau**).

Remarques

Le programme est écrit en BASIC Microsoft, pour des raisons de portabilité, et compilé par Microsoft Quick Basic 1.0. Il est assez lent. Si quelqu'un a eu la bonne idée d'écrire un sous-programme assembleur de recherche d'un mot dans une table, ou un sous-programme permettant de lire des programmes Mac Write et d'en extraire les mots, je suis **très** intéressé ! De même, je suis prêt à fournir le programme source Basic original à qui s'y intéresse. Toute suggestion, remarque, critique, sera la

bienvenue.

COPYRIGHT ET DISTRIBUTION

Le programme Analyseur est © J. C. DUNYACH 1989. Il peut être distribué et copié librement mais ne peut être vendu, seul ou inclus dans un ensemble de logiciels, sans l'autorisation écrite de l'auteur.