

Projet PERL (rapport)

Corpus :

Le choix de mon corpus s'est porté sur mon sujet de mémoire qui s'intitule : La vie intérieure dans *Thérèse Desqueyroux* de François Mauriac. Le texte a été copié sur le site http://www.alyon.org/litterature/livres/XX/monades/therese_desqueyroux_de_francois_mauriac.html . Ensuite a été collé dans un traitement de texte puis converti en format txt, pour pouvoir l'utiliser dans Atom. Mon document comporte 35,000 mots.

Cible :

Tournant autour de plusieurs thèmes tels que le monologue intérieur, la métaphore de la cage, la solitude et l'incommunicabilité, il nous a semblé pertinent de mettre en relief certains verbes et mots qui montrent une certaine incommunicabilité entre le personnage principal et son entourage. Pour compléter cette approche, nous souhaitons compter le nombre d'occurrence qui reflète les thèmes cités.

Objectif :

Dans ce roman, comme le dialogue est dissous par un monologue, et que le personnage principal (Thérèse) ne cesse de songer intérieurement. L'objectif de notre projet, est d'abord de pouvoir mettre en évidence dans le texte le verbe « songer » (conjugué soit au présent ou à l'imparfait voire au passé simple) puis le mot « silence » à travers un concordancier qui pourra récupérer les mots et les mettant dans leur contexte et de pouvoir en faire une sortie HTML.

Étapes :

La première étape, a d'abord été de pouvoir repérer dans mon texte le verbe « songer » (conjuguer soit au présent ou à l'imparfait voire au passé simple) ainsi que le mot « silence » et enfin compter le nombre d'occurrences trouvées. Ce qui nous a conduit à créer un script s'appelant *reperage.pl* qui récupérera les lignes de mon texte où se retrouvent les différentes conjugaisons du verbe « songer » et du mot « silence » et nous faire un affichage dans le terminal.

- Nous avons d'abord créé une variable (pour chaque cible) qui nous permettra de pouvoir compter le nombre le nombre d'occurrence avec un compteur de 0 : `my $nb_occurrences = 0;` `my $nb_occurrences_silence = 0;` et aussi, nous avons aussi créé une variable, pour le comptage de nombre de ligne : `my $nb_lignes = 0;`

- Ensuite la création d'un test a été créé avec un opérateur de matching au sein duquel nous avons une expression régulière permettant de voir si dans la ligne nous avons le verbe « songer » et ses déclinaisons et aussi pour le mot « silence ».

Après avoir repéré les occurrences, nous avons créé un fichier *songer_silence.pl* qui nous a permis de faire un concordancier en HTML.

Dans le déroulé des étapes, nous nous focaliserons sur les éléments qui les plus importants, à savoir la création d'un tableau, la récupération des occurrences et la mise en contextes.

- dans l'élément <body>, nous avons créé un élément <table> qui nous permettra d'avoir un tableau pour chaque concordancier.
- comme dans *reperage.pl*, nous avons créé des variables pour chaque tableau représentatif du concordancier :
 - my \$nb_occurrences = 0; my \$nb_occurrences_silence = 0; pour le nombre d'occurrence.
 - my \$nb_lignes = 0; pour le nombre de ligne.
- création d'un test, pour chaque tableau, afin de voir si dans les lignes nous avons le verbe « songer » et ses déclinaisons, et aussi le mot « silence » :
 - if (\$ligne =~ /^(.*\W)(songe\w*)(.*)\$/i).
 - if (\$ligne =~ /^(.*\W)(silence)(.*)\$/i).

À l'intérieur de chaque if nous avons différentes instructions :

- pour pouvoir récupérer les mises en contexte, nous avons créé des variables :
 - my \$gauche = \$1; qui représente le contexte de gauche.
 - my \$cible = \$2; pour la cible, c'est à dire le mot ou le verbe qui sera en contexte.
 - my \$droite = \$3 pour le contexte de droite.
- pour pouvoir récupérer les lignes ou les caractères des différents contextes que nous aimerions afficher dans notre concordancier, il a fallu qu'on utilise un opérateur de caractères *substr* qui permet de récupérer ou d'extraire des caractères. À l'intérieur de cet opérateur, nous avons des chiffres qui sont comme des pourcentages, permettant de réguler le nombre de caractères que l'on souhaite avoir.
- Nous avons aussi créé une balise <tr> qui définit une rangée dans le tableau au sein duquel nous avons des balise <td> (pour la variable \$nb_lignes, \$gauche, \$cible, my \$droite) et qui définissent chaque cellule. Pour chaque balise, nous avons mis des print HTML qui nous permettront d'avoir un affichage dans la sortie HTML.

Le résultat final se présente sur une page HTML, mis en forme par une CSS (*concordanciers_occurences.css*) permettant une belle apparence et une meilleure valorisation et rendu du résultat :

- Mot ou verbe recherché en noir avec un fond jaune dans chaque cellule.
- Contexte de gauche en couleur vert.
- contexte de droite en couleur rouge.
- nombre de ligne, en violet.
- Pour chaque concordancier, nous avons créé une balise <h1> pour des titres.
- À la fin de chaque concordancier, nous avons le nombre d'occurrences trouvées.

NB : Il est à noter qu'un fichier *pl* (*concordancier_songer_silence.html*), sans une mise en relation avec une css, a été créé. Les éléments css y ont été intégrés.

Problèmes rencontrés :

Le premier problème rencontré se retrouve dans la récupération de certains verbes suivis du pronom personnel, tels que : « songeait-il ou songe-t'elle ». Il a été difficile à trouver une expression régulière qui pourrait récupérer le verbe ainsi que le pronom personnel. Lorsque nous utilisons l'expression régulière suivante : `if ($ligne =~ /^(*W)(songe\w*)(.*)$/i)`, nous arrivons à récupérer le verbe conjugué seulement. Si nous ajoutons l'expression suivante : `if ($ligne =~ /^(*W)(songe\w*)(.*)$/i)`, nous récupérons le caractère qui le suit (-) mais pas le pronom.

La seconde difficulté qui a été rencontrée, est dans la mise en page de notre concordancier en sortie HTML. La difficulté a été de mettre côte à côte les deux tableaux, avec la création d'un <div> pour chaque tableau <table>.

Améliorations/points à résoudre :

Plusieurs problèmes cités plus haut n'ont pas trouvé de réponses et seraient amenés à être retravaillés.

Par exemple la récupération des verbes suivis de leurs pronoms personnels.

Une amélioration est à voir aussi au niveau de la mise en page avec un affichage avec le texte complet tout en haut et en bas les deux concordanciers côte à côte.

CONCLUSION :

Grâce aux résultats obtenus à partir de nos scripts, nous pouvons dire que l'objectif du projet a été atteint dans l'ensemble. Nous avons pu avoir une mise en contexte dans un concordancier les différents mots que nous voulions représenter (*songer* et ses déclinaisons, *silence*). Le fait de ressortir les différents passages où nous retrouvons ces mots, nous permettra de mieux nous repérer dans notre corpus et de mieux savoir dans quel contexte ils ont été employés et de mieux faire ressortir les différents thèmes que l'on retrouve dans *Thérèse Desqueyroux*. Le nombre

d'occurrences trouvées : 32 pour le verbe « songer » et 22 pour « silence » viennent corroborer nos propos sur les différents thèmes que nous avons dans ce roman, à savoir incommunicabilité entre le personnage principal et son entourage.