Utilisation d'un concordancier

Une concordance, dans sa forme la plus simple, est une liste alphabétique des mots d'un texte, donnés avec les contextes dans lesquels ils apparaissent. Le plus souvent on représente les mots sous la forme d'un index KWIC (*KeyWord In Context*), dans lequel chaque mot est centré dans une fenêtre de taille fixe (cf. figure 1).

```
1320       taste it is that such  poor cattle always have in their mouths
948              of sparing the  poor child the inheritance of any part of
778         small property of my  poor father, whom I never saw--so long
1870        desolate, while your  poor heart pined away, weep for it
947                Miss, if the  poor lady had suffered so intensely
1884              the love of my  poor mother hid his torture from me
1615      stockings, and all his  poor tatters of clothes, had, in a long
1577           faded away into a  poor weak stain.  So sunken and
1001         on your way to the  poor wronged gentleman, and, with a
1036        detachment from the  poor young lady, by laying a brawny hand
```

FIG. 1 – *Concordance de poor dans A Tale of Two Cities (Dickens)*

Un **concordancier** est donc un programme permettant de réaliser de telles listes sur un corpus de son choix. En général, les concordanciers donnent aussi diverses informations statistiques sur le corpus traité.

Nous utiliserons pour ces manipulations *Concordance* (version 1.1.3), un concordancier développé par R. J. C. Watt, et disponible à l'adresse `http://www.rjcw.freeserve.co.uk/`.

Manipulations

Les manipulations suivantes (1 à 5) doivent être réalisées par chaque groupe, et les résultats remis à l'enseignant à l'issue de la séance.

a. Lancer le concordancier « Démarrer/Programmes/TAL/Concordance ». Vérifier que l'alphabet comprend bien les lettres accentuées, et pas l'apostrophe.[1]

b. En utilisant WordPad (« Démarrer/Accessoires/WordPad »), créer un petit fichier de test à partir du fichier fourni en exemple : `H:corpusLeMonde.txt`. Sauver ce fichier sur le bureau. Lancer le calcul (« File/Make Full Concordance ») (F5) sur ce fichier. Essayer les différentes fonctionnalités : affichage des concordances, tris par fréquence des mots, affichage des collocations (« Headwords/Collocations »)...

1. Lancer « Internet Explorer », et choisir un texte parmi ceux qui sont proposés sur le site `http://cedric.cnam.fr/ABU/BIB/index.html`. (Sauvegarde en « texte seul »). Lancer le calcul de concordances sur ce texte.
2. Donner le nombre d'occurrences (en pourcentage), du mot *en* suivi par un participe présent.
3. Donner les 10 mots « pleins » les plus fréquents.
4. Repérer les collocations pertinentes pour 5 de ces mots.
5. Donner le pourcentage de cas où le mot *le* est un article, et où il est un pronom. Quel est le contexte privilégié dans ce dernier cas?

1. Menu « Text/Alphabet ». Si certains symboles manquent, cliquer sur « Languages », et couper/coller les symboles nécessaires. Il faut aussi ajouter l'apostrophe dans les séparateurs de mot.