

***Le traitement avancé des occurrences : le projet ABEL***  
***Analyse du texte fantastique***

Séminaire d'Analyse du texte fantastique :  
*Littérature et informatique*

Suzanne Van Weddingen  
Université de Haute-Alsace, Mulhouse

## ***Sommaire***

### Introduction

#### 1. Naissance du projet

##### 1.1. Le Traitement des occurrences

##### 1.2. VICAP (Violent Criminal Apprehension Program) et CAIN (Crime Artificial Intelligent Network)

##### 1.3. Une application possible à la littérature ?

#### 2. Présentation du Projet Abel

##### 2.1. Principe de fonctionnement

##### 2.2. Limites du logiciel

#### 3. En pratique...

#### 4. Vers un logiciel idéal

### Conclusion

### Bibliographie

### Annexe

## **Introduction**

À l'heure de l'informatique et de l'Internet, les recherches littéraires ont pris une nouvelle dimension. Il est de plus en plus facile d'avoir accès aux textes tombés dans le domaine public, et de nombreux sites se spécialisent dans la numérisation de textes antiques, classiques ou même modernes. À côté de ce développement, les outils du chercheur se sont également modernisés. Le compte des mots manuel a trouvé son équivalent sous la forme d'un logiciel, diminuant ainsi les risques d'erreur.

Simplifier le travail de recherche au profit du travail de réflexion se fait de plus en plus, mais les logiciels destinés à la littérature ne sont pas courant et se réduisent la plupart du temps à la description mathématique d'une œuvre. Les possibilités de l'informatique permettent pourtant de rêver un autre type de programme, susceptible de simplifier la recherche, en intégrant des outils déjà existants. ABEL mêle le concordeur classique avec un moteur de recherche qui propose des éléments bibliographiques susceptibles d'être utilisés dans le travail d'interprétation.

Afin de décrire au mieux ABEL, nous nous intéressons aux origines de ce programme, afin de voir comment diverses inspirations peuvent amener à la composition d'un outil destiné en priorité à la littérature.

### **1. Naissance du projet**

À la base du projet Abel, se trouve une volonté de dépasser le simple concordeur pour le transformer en un outil de recherche plus complet, à l'aide des technologies modernes et de quelques programmes existants.

#### **1.1. Le Traitement des occurrences**

Les outils informatiques mis à la disposition des chercheurs permettent à l'heure actuelle de gagner du temps dans le compte des occurrences d'un mot dans un texte. Alors qu'autrefois, il fallait faire ce compte manuellement, avec les risques d'erreurs que cela pouvait comporter, il est devenu facile de confier à un ordinateur cette tâche. Comme l'explique le concepteur d'un logiciel de ce type appelé Le Concordeur :

*Le Concordeur* est un logiciel Macintosh permettant de créer, éditer, trier et imprimer des concordances tirées de textes en français, anglais, russe et autres langues (un octet par caractère), à l'aide de plusieurs alphabets. Un éditeur de texte efficace, muni de diverses fonctions de recherche et de remplacement, y est incorporé. Les textes peuvent être entrés directement, ou importés de documents externes. La définition du "mot" est sous le contrôle de l'utilisateur de par son choix d'alphabet. Outre les alphabets prédéterminés fournis, le logiciel permet à l'utilisateur de définir ses propres alphabets. Une concordance est réalisée sous la forme d'une liste de mots, ou "catalogue", établie à partir du texte. L'utilisateur peut éditer ce catalogue afin d'y insérer d'autres informations (la forme générique, des traductions, etc.), le trier de diverses façons, et en extraire des sous-listes selon des critères de sélection. Le catalogue peut être imprimé sous divers formats, soit comme une simple liste de mots, soit sous forme de concordance où chaque mot est accompagné de tous ses contextes. Il est possible d'établir une concordance à partir d'un seul fichier-texte ou d'un ensemble de fichiers-texte. Des commandes macro permettent d'automatiser les procédures importantes. Plusieurs polices de caractères matricielles sont incluses.

Outre le fait que ce type de logiciel nécessite souvent un matériel bien particulier – dans le cas cité, il ne peut être utilisé que sur un Macintosh alors que d'autres logiciels plus ou moins efficaces fonctionnent exclusivement sur PC – il s'agit avant tout d'un outil de comptage destiné à faciliter l'interprétation et l'explication de textes. Son principal point positif réside dans la possibilité qu'a l'utilisateur de juxtaposer une série de résultats en provenance de plusieurs textes, ceci afin de permettre de comparer les statistiques et l'utilisation des occurrences.

## 1.2. VICAP (Violent Criminal Apprehension Program) et CAIN (Crime Artificial Intelligent Network)

VICAP est un système informatique dont la fonction est de traiter et répertorier toutes les caractéristiques significatives d'affaires policières, principalement des meurtres. Les rapports de police sont systématiquement entrés dans la base de données, afin que les bureaux reliés au système puissent y avoir accès. Les enquêteurs peuvent, lors d'une enquête, interroger VICAP pour trouver des similitudes avec d'autres affaires, ou pister un éventuel meurtrier, par exemple.

C'est dans les années cinquante à la suite de recherches fastidieuses pour trouver un cas similaire à celui sur lequel il enquêtait, que le détective Brooks a conçu ce système pour faciliter le travail de recherche des enquêteurs, ainsi que pour permettre la communication et l'échange d'informations entre les polices des différents états.

Un second logiciel complète VICAP. S'il n'est pas reconnu officiellement par le FBI, il n'est pas non plus une invention de Patricia Cornwell qui décrit son fonctionnement dans ses romans. Seul le concept est retenu dans cette étude. CAIN est « un programme d'intelligence artificielle, une sorte de réseau qui permettait d'emmagasiner toutes les informations concernant les criminels »<sup>1</sup>. Si son principe ressemble à la description de VICAP, il est cependant plus perfectionné. La recherche s'effectue selon d'autres critères, puisqu'il s'agit d'un logiciel « plus *déclaratif* que *navigational*. En d'autres termes, cela signifie que l'utilisateur doit préciser à *quoi* il veut avoir accès plutôt que *comment* il veut y accéder »<sup>2</sup>. Dans le cas de VICAP, l'enquêteur fait une recherche qui se rapproche d'une recherche sur Internet, alors que CAIN offre une dimension supplémentaire. Si les deux logiciels fonctionnent sur le même principe, puisque les informations recueillies sont identiques, CAIN en revanche est « un programme qui modélise le monde réel de façon aussi proche et vraisemblable que possible »<sup>3</sup>. C'est-à-dire que le logiciel peut en théorie se substituer aux enquêteurs, dans la mesure, où il accomplit leur travail de recherche. L'enquêteur amène la matière que le programme traite et trie. Il réagit ensuite en fonction de la demande.

CAIN est aussi un système interactif. Si CAIN a une hypothèse ou qu'il a besoin d'informations complémentaires, il pourra communiquer directement avec le policier soit par modem soit par l'intermédiaire d'une messagerie électronique, ou même par serveur vocal<sup>4</sup>.

Ces deux programmes pourraient être adaptés à d'autres domaines, en particulier à la littérature, et proposer une nouvelle approche des outils informatiques déjà disponibles, ou en cours de développement.

---

<sup>1</sup> CORNWELL, P., *The Body Farm*. Pour la traduction française : *La Séquence des corps*, Editions du Masque, 1995, p.44.

<sup>2</sup> *Ibid.*, pp. 45 et 46, souligné par l'auteur.

<sup>3</sup> *Ibid.*, p.44.

<sup>4</sup> *Ibid.*, p.44.

### 1.3. Une application possible à la littérature ?

Lors de l'étude d'un thème, pour prendre un exemple concret, le concordeur facilite certes le travail, mais il reste toutefois sommaire, dans la mesure où seuls les textes constituant le corpus de l'utilisateur sont utilisés.

En ne retenant que le concept de chacun des logiciels – le concordeur, VICAP et CAIN – il est possible de penser un système fonctionnant sur le même principe, mais appliqué à la littérature. Les informations incluses dans les rapports de police pourraient trouver leur équivalence dans un domaine littéraire, sous la forme de textes numérisés, accompagnés de d'indications bibliographiques. À partir de cette base de données, le logiciel pourrait proposer une recherche annexe, liée au texte initial, ainsi qu'à l'utilisateur, mais en ouvrant le champ à d'autres critères de sélection.

On peut également lier un concordeur, tel celui décrit plus haut, à la base de données, susceptible de réagir comme un moteur de recherche pour amener d'autres éléments de comparaison.

Le but du logiciel est donc, si l'on se réfère aux logiciels des enquêteurs, de prendre un texte initial et de confronter ses caractéristiques principales à la base de données, afin d'obtenir de nouvelles pistes de recherches.

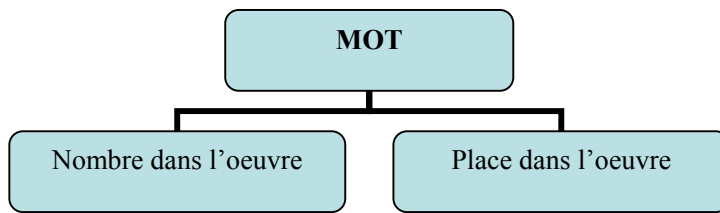
## **2. Présentation du Projet Abel**

### 2.1. Principe de fonctionnement

Pour comprendre le principe du logiciel, chaque étape est décomposée, et chaque outil qui compose ABEL est détaillé pour aboutir ensuite à une vision d'ensemble.

#### **- PHASE 1 : Le concordeur**

Fondé sur le principe du concordeur, la première phase du logiciel se comporte de la même façon ; il propose de rechercher les occurrences d'un mot, d'en donner le compte, et de localiser les résultats dans l'œuvre utilisée. D'une manière simplifiée, l'équation du concordeur peut être :



### - PHASE 2 : Le renvoi à des références

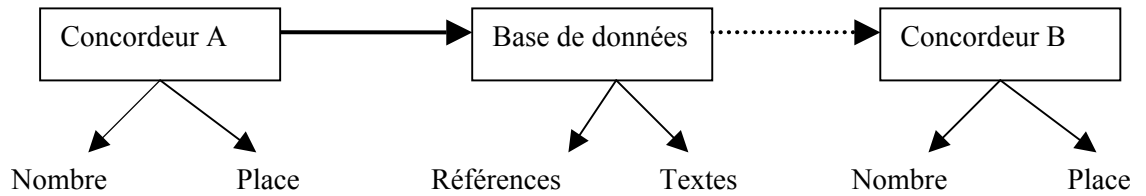
Placée à la portée du concordeur, une base de données peut apporter un éclairage supplémentaire à l'interprétation. Au compte des occurrences s'ajoute une dimension comparatiste supplémentaire, puisqu'un mot précis, voire une citation se retrouve la plupart du temps dans une autre œuvre<sup>5</sup>. À ce niveau, un premier choix s'offre à l'utilisateur : utiliser ABEL comme un concordeur, ou lui suggérer de rechercher un complément d'information dans une base de données pré existante. Dans l'idéal, ABEL devrait également permettre la recherche sur des textes dans une langue différente, ce qui permettrait d'élargir la vision d'un mot et de son contexte à son équivalent dans une autre langue et offrir la possibilité d'une analyse comparée avec le même logiciel, sans manipulations supplémentaires.

### - PHASE 3 : La comparaison des références

Lorsque l'utilisateur utilise la base de données, outre les références qui lui sont fournies, il a la possibilité de rechercher les occurrences du mot initial dans les textes proposés par la base de données. Le principe redevient celui du concordeur, mais hiérarchiquement, il s'agit de la troisième étape, ou du dernier choix :

---

<sup>5</sup> Il s'agit de partir du postulat selon lequel le mot recherché justifie une recherche approfondie. Il est clair que la recherche comparée d'articles n'a nul besoin de s'appuyer sur Abel, puisqu'elle n'apportera aucune information digne d'intérêt.

**ABEL :***Schéma complet**Schéma final*

Concordeur *A* → Base de données → Concordeur *B*

Le concordeur nommé *A*, renvoie à la première phase, c'est-à-dire à la recherche initiale de l'occurrence d'un mot. La flèche en pointillés symbolise le choix laissé à l'utilisateur. À partir des références fournies par la base de données le concordeur *B* peut intervenir sur les textes proposés dans lesquels se retrouve la forme initiale recherchée.

À cette description, il convient d'ajouter quelques facteurs. La base de données reste à définir. D'une manière simplifiée, la base est évidemment l'ensemble des textes numérisés, mis à la disposition d'ABEL, puisque celui-ci va agir comme un moteur de recherche et trier les résultats avec pertinence, rappelant le fonctionnement de VICAP. Cependant, elle comporte des informations bibliographiques capables de guider le chercheur. ABEL doit pouvoir être capable de spécifier le type de texte, s'agit-il d'un article, d'un roman, d'une nouvelle... Mais également de renseignements bibliographiques tels que l'auteur, le titre, la date de parution, le traducteur...

À partir de ces renseignements, on peut penser ABEL en termes de véritable moteur de recherche, dans la mesure où à partir de la bibliographie, il ne se contente plus de lister tous les textes avec l'occurrence demandée, mais permet aussi de faire un tri selon des critères spécifiques : une période donnée, un auteur, voire une collection ou un éditeur...



## 2.2. Limites du logiciel

Comme tout programme, il a cependant ses limites, à commencer par la création du corpus constituant la bibliographie. Tous les textes ne sont pas numérisés, et les rassembler tous est irréalisable. Cependant, en travaillant sur une thématique précise, on peut arriver à créer un corpus spécifique, qui peut apporter un éclairage sur l'évolution d'un mot dans son contexte précis.

La seconde difficulté visible dès la création, et qui se retrouve dans les concordeurs classiques est la langue. En effet, ABEL travaillant sur les occurrences, ne peut faire une recherche que sur une seule langue à la fois. Pour utiliser des textes en langue étrangère, il convient de s'appuyer sur une base de données comportant les textes spécifiques, dans une thématique précise.

Enfin, tout comme pour VICAP, ABEL n'est pas à l'abri d'un « oubli ». Un enquêteur est libre de ne pas entrer son rapport dans la base de données VICAP, tout comme ABEL ne peut forcément posséder tous les textes sur un même sujet. Le logiciel reste exposé à des manques.

## 3. En pratique...

Afin d'illustrer le fonctionnement d'ABEL, nous pouvons simuler une recherche précise. Nous utilisons comme textes de départ *L'Homme voilé* et *Le Train 081* de Marcel Schwob, ainsi que *La Peur* de Maupassant. À partir de ces trois textes, nous désirons constituer un corpus autour du vocabulaire ferroviaire. Nous commençons par déterminer une liste de termes (*Figure 1*) sur lesquels peut se porter la recherche. Les résultats (*Figure 1*) montrent des termes communs, tels que « portière(s) » ou « wagon(s) », entrés dans la langue. En effet, le vocabulaire ferroviaire s'est construit petit à petit, d'abord à l'aide d'emprunts à l'anglais qui avait été précurseur du développement ferroviaire, comme c'est le cas par exemple dans *Arria Marcella*, où Gautier utilise le mot « railway » ou « station ». Le terme qui revient le plus souvent dans les trois textes est « wagon », au singulier et au pluriel. C'est sur celui-ci que se porte notre recherche.

*L'homme Voilé*

|            |   |
|------------|---|
| BANQUETTE  | 3 |
| BANQUETTES | 1 |
| BILLET     | 1 |
| CHEMIN     | 2 |
| FER        | 2 |
| JONCTIONS  | 1 |
| L'ESSIEU   | 1 |
| LAMPE      | 4 |
| PORTIÈRE   | 1 |
| RAILS      | 2 |
| ROUES      | 1 |
| VOITURES   | 1 |
| WAGON      | 3 |
| WAGONS     | 1 |

*Le Train 081*

|                  |   |
|------------------|---|
| CHAUFFE          | 1 |
| CHAUFFEUR        | 3 |
| COMPAGNIE        | 1 |
| DÉPÔT            | 1 |
| EMBRANCHEMEN...  | 1 |
| ENTRECROISEME... | 1 |
| FOUR             | 1 |
| FREIN            | 2 |
| GARE             | 4 |
| LIGNE            | 1 |
| LOCOMOTIVE       | 1 |
| MACHINE          | 6 |
| MACHINES         | 4 |
| MARCHEPIED       | 1 |
| MÉCANICIEN       | 2 |
| MÉCANICIENS      | 1 |
| PORTIÈRE         | 1 |
| PORTIÈRES        | 1 |
| RAPIDE           | 2 |
| SIGNAUX          | 1 |
| TIMBRE           | 1 |
| TRAIN            | 9 |
| VAPÉUR           | 4 |
| VITRES           | 1 |
| VOIE             | 4 |
| VOITURER         | 1 |
| WAGON            | 4 |
| WAGONS           | 2 |

*La Peur*

|          |   |
|----------|---|
| PORTIÈRE | 1 |
| TRAIN    | 2 |
| VAPÉUR   | 1 |
| WAGON    | 1 |
| WAGONS   | 1 |

*Figure 1*

Dans une seconde phase, le terme retenu est recherché dans une base de données. Pour cette simulation, nous retenons *ABU, la bibliothèque universelle*<sup>6</sup>, puisqu'elle offre l'avantage d'avoir intégré un concordeur à son site, ce qui permet de chercher les occurrences d'un mot dans chaque texte sélectionné, ou dans toute la base. Nous notons cependant deux inconvénients. D'une part le mélange des textes qui ne sont pas répertoriés selon un classement spécifique. D'autre part, le moteur ne donne que les titres d'ouvrages sans autre forme de précision. Il convient donc d'ajouter les éléments manquants manuellement. Sur les deux cent quatre-vingt-huit textes du site, nous cherchons les occurrences du mot « wagon », au singulier et au pluriel.

Au singulier, on obtient 62 occurrences réparties dans 11 œuvres.

Joris Karl Huysmans, *A Rebours* (1884)

1

Jules Verne, *Les Cinq cents millions de la Begum* (1879)

4

Guy de Maupassant, *Bel Ami* (1885)

4

Arthur de Gobineau, *La Chasse au Caribou* (1872)

1

Léon Bloy, *Le Désespéré* (1885)

1

Raymond Radiguet, *Le Diable au corps* (1923)

5

Frères Goncourt, *Germinie Lacerteux* (1865)

1

Paul Ginisty, *Anthologie du journalisme* (1922)

1

Henri Poincaré, *La Science et l'hypothèse* (1902)

1

Jules Verne, *Le Tour du monde en quatre-vingt jours* (1873)

41

---

<sup>6</sup> <http://abu.cnam.fr/index.html>

Jules Verne, *De la Terre à la Lune* (1865)

2

Le classement est celui donné par le moteur, et n'est pas forcément le plus clair. Un classement chronologique, par auteur ou par occurrences serait bienvenu. De plus, toutes les catégories du discours sont prises en compte et *L'Anthologie du Journalisme* côtoie des œuvres de fictions. Un classement chronologique permettrait de replacer les trois nouvelles de départ dans leur contexte et ainsi de composer un corpus.

Frères Goncourt, *Germinie Lacerteux* (1865)

1

Jules Verne, *De la Terre à la Lune* (1865)

2

Arthur de Gobineau, *La Chasse au Caribou* (1872)

1

Jules Verne, *Le Tour du monde en quatre-vingt jours* (1873)

41

Jules Verne, *Les Cinq cents millions de la Begum* (1879)

4

Joris Karl Huysmans, *A Rebours* (1884)

1

Guy de Maupassant, *Bel Ami* (1885)

4

Léon Bloy, *Le Désespéré* (1885)

1

Henri Poincaré, *La Science et l'hypothèse* (1902)

1

Paul Ginisty, *Anthologie du journalisme* (1922)

1

Raymond Radiguet, *Le Diable au corps* (1923)

5

*L'Homme voilé* et *Le Train 081* sont placés après Léon Bloy, alors que *La Peur* se trouve entre les *Contes de la Bécasse* et *Bel Ami*. La suppression des textes de Henri Poincaré, Paul Ginisty et Raymond Radiguet (nous ne retenons que les romans de fiction, et en ce qui concerne Radiguet, l'intervalle de temps est trop important), on obtient le résultat suivant :

Frères Goncourt, *Germinie Lacerteux* (1865)

1

Jules Verne, *De la Terre à la Lune* (1865)

2

Arthur de Gobineau, *La Chasse au Caribou* (1872)

1

Jules Verne, *Le Tour du monde en quatre-vingt jours* (1873)

41

Jules Verne, *Les Cinq cents millions de la Begum* (1879)

4

Joris Karl Huysmans, *A Rebours* (1884)

1

Guy de Maupassant, *La Peur* (1884)

1

Guy de Maupassant, *Bel Ami* (1885)

4

Léon Bloy, *Le Désespéré* (1885)

1

Marcel Schwob, *Le Train 081* (1891)

3

Marcel Schwob, *L'Homme Voilé* (1891)

4

Le champ se restreint, mais d'autres œuvres sont susceptibles de rejoindre le corpus étudié, comme *Bel-Ami* et *Les Cinq cents millions de la Begum* en ne retenant que les textes où les occurrences de « wagon » sont plus nombreuses, ou *Le Désespéré* et *A Rebours*, si les occurrences uniques sont acceptées. En élargissant la recherche au

pluriel de « wagon », d'autres résultats apparaissent, qui complètent le corpus (on ne retiendra ici que les titres différents).

Victor Hugo, *Les contemplations*, tome premier (1858)

1

Emile Zola, *Germinal* (1885)

2

Jules Verne, *Robur le Conquérant* (1886)

1

Alphonse Daudet, *Souvenirs d'un Homme de Lettres* (1888)

1

Pierre Louÿs, *La Femme et le pantin* (1898)

2

La poésie et les mémoires ne faisant pas partie du domaine qui nous intéresse, nous ne retenons que Emile Zola, Jules Verne et Pierre Louÿs, que nous intégrons au corpus formé plus haut.

Jules Verne, *Les Cinq cents millions de la Begum* (1879)

4

Joris Karl Huysmans, *A Rebours* (1884)

1

Guy de Maupassant, *La Peur* (1884)

1

Guy de Maupassant, *Bel Ami* (1885)

4

Léon Bloy, *Le Désespéré* (1885)

1

Emile Zola, *Germinal* (1885)

2

Jules Verne, *Robur le Conquérant* (1886)

1

Marcel Schwob, *Le Train 081* (1891)

3

Marcel Schwob, *L'Homme Voilé* (1891)

4

Pierre Louÿs, *La Femme et le pantin* (1898)

2

Avec ABEL, nous pourrions déjà combiner la recherche au singulier et au pluriel afin d'obtenir un corpus provisoire. Les informations bibliographiques éviteraient un mélange entre œuvres de fiction, poésie, documentaire pour permettre un regroupement des œuvres du même genre. Enfin, un classement chronologique direct permettrait de situer plus précisément les textes par rapport au corpus de base.

Le nouveau corpus ainsi obtenu montre donc l'intérêt d'ABEL pour la recherche avancée d'occurrences. Il permet d'ouvrir le champ à d'autres textes qui peuvent faire l'objet d'une recherche. ABU permet un travail quasi similaire, mais comme nous le notions précédemment, certains éléments bibliographiques sont manquants, et le classement n'est pas assez pertinent. Enfin, la dernière phase d'ABEL est présente chez ABU, puisque le site permet de voir toutes les citations dans lesquelles figure l'occurrence recherchée.

#### 4. Vers un logiciel idéal

Aux limites du logiciel que nous avons évoqué plus haut s'ajoute la question du choix. En effet, si nous prenons par exemple pour sujet l'étude des textes fantastiques à dimension ferroviaire, nous nous heurtons à un problème d'identification du texte fantastique. La recherche d'occurrence est insuffisante pour permettre de dire quel texte appartient ou non au fantastique. Les indications bibliographiques d'ABEL peuvent certes aiguiller la recherche, mais sans précision, il faut alors prendre les textes proposés un à un afin d'éviter les risques d'erreurs liés aux réponses générales obtenues. En revanche, en appliquant le principe de CAIN, c'est-à-dire une comparaison des textes avec une série de critères précis, le corpus de base peut être complété et utilisé avec de meilleurs résultats.

Reprenons par exemple *Le Train 081* et *La Peur*. Certains éléments présents dans les textes les font appartenir de droit au fantastique. Le sujet, d'une part. Dans *Le Train 081*, le train est dédoublé, comme s'il se trouvait face à un miroir :

Nous avons sur une ardoise le numéro du train, marqué à la craie : 180. – Vis-à-vis de nous, à la même place, un grand tableau blanc s'étalait, avec ces chiffres en noir : 081. [...] Sur la brume rouge, je vis ainsi se détacher l'ombre de Graslepoix<sup>7</sup>.

Maupassant, lui, suggère d'emblée que sa nouvelle soit fantastique :

Ce fut tout à coup comme une apparition fantastique. [...] Certes, ce fut une apparition fort étrange ! [...]

« Il est juste minuit, Monsieur, nous venons de voir une singulière chose »<sup>8</sup>.

D'autre part, les termes choisis pour la description identifient le texte ferroviaire fantastique, dans la majorité des cas. Les couleurs, et surtout le blanc (la vapeur), le rouge et le noir sont fréquemment utilisées dans ces textes et Marcel Schwob et Guy de Maupassant ne font pas exception à cette règle.

La recherche d'occurrences de ces couleurs montre chez Maupassant l'importance du noir avec « nuit » repris six fois, « noir » quatre fois et « ténèbres » deux fois. De même, chez Marcel Schwob nous pouvons voir ce même phénomène : « nuit » est repris sept fois, « noir(s) » six fois. Alors que dans *La Peur*, c'est le « feu » qui revient à trois reprises, dans *Le Train 081* c'est la « vapeur » l'élément important (repris quatre fois), alors que « feu » et « blanc » sont équivalents avec deux occurrences. En plus des occurrences ferroviaires (incluant le vocabulaire « technique ») nous pouvons décider de compléter notre recherche à l'aide de ces termes. Notre liste d'occurrences est donc : Noir – Ténèbres – Nuit – Feu – Noirs – Vapeur – Blanc(s) ; puisque ce sont les mots les plus fréquemment rencontrés dans nos deux exemples. Si nous appliquons cette liste à *Arria Marcella* par exemple, nous obtenons les résultats suivants :

|  |                                     |
|--|-------------------------------------|
|  | <i>Arria Marcella -<br/>Gautier</i> |
|--|-------------------------------------|

<sup>7</sup> SCHWOB, M., « Le Train 081 », in *Le Cœur Double*, Paris, Edition 10/18, 1979, pp. 77-78.

<sup>8</sup> Guy de Maupassant, *La Peur*, Le Figaro, 25 juillet 1884.



|          |    |
|----------|----|
| noir     | 6  |
| noirs    | 2  |
| ténèbres |    |
| nuît     | 12 |
| feu      | 2  |
| vapeur   | 3  |
| blanc    | 4  |
| blancs   | 4  |

La constitution d'un corpus par le biais d'un logiciel se décompose donc en plusieurs phases qui peuvent pourtant être simultanées. Dans nos exemples, seules les indications bibliographiques nous ont fait défaut. Celles-ci auraient permis de replacer les œuvres dans leur contexte, et d'ouvrir éventuellement le champ à une recherche plus approfondie<sup>9</sup>.

Le processus d'ABEL de la première phase à la dernière a pu être (en partie) reproduit ici, en utilisant des outils existants, à savoir une base de donnée et un concordeur, et qui mettent en valeur les avantages d'un travail de recherche bibliographique combiné à la recherche d'occurrence.

L'outil informatique redéfini pour la littérature permet donc de compléter une recherche. À l'aide des nouvelles technologies, un logiciel comme ABEL devrait pouvoir être intégré à l'Internet plus précisément à des sites dont la vocation est de proposer des textes numérisés comme outil de travail, comme c'est déjà le cas pour ABU. Il est évident que ce genre de programme ne peut être fonctionnel immédiatement, en partie à cause de la conception de la base de données. De plus, la thématique est laissée sous la responsabilité de l'utilisateur. Si dans le cas d'Internet un système de mots clefs peut être imaginé, l'utilisateur est libre d'intégrer ses propres textes dans la base. Les possibilités d'un tel programme sont donc à l'image de la recherche littéraire : quasi infinies.

---

<sup>9</sup> Par exemple, nous aurions pu apprendre qu'un texte absent de la base de donnée ABU écrit par Jules Verne, *La Maison à vapeur*, est une nouvelle ferroviaire dont s'est inspiré Paul D'Ivoi pour son *Docteur Mystère...* Cette référence nous aurait permis d'ajouter potentiellement deux textes supplémentaires à notre corpus.

### ***Bibliographie***

- CORNWELL, Patricia, *The Body Farm*. Pour la traduction française : *La Séquence des corps*, Editions du Masque, 1995.
- MAUPASSANT, Guy de, *La Peur*, in « Le Figaro », 25 juillet 1884.
- SCHWOB, Marcel, « Le Train 081 », in *Le Cœur Double*, Paris, Edition 10/18, 1979.