

INSTITUT NATIONAL DES LANGUES ET
CIVILISATIONS ORIENTALES

MASTER TRAITEMENT AUTOMATIQUE DES LANGUES,
SPÉCIALITÉ INGÉNIERIE MULTILINGUE

Définition d'un modèle d'émotions pour la catégorisation de textes narratifs courts

Lucie GIANOLA

21 octobre 2014



The logo for 'shortédition' is a dark grey rectangle. The word 'shortédition' is written in a white, lowercase, sans-serif font. Below it, the tagline 'AVEC VOUS, LE TALENT COURT' is written in a smaller, orange, uppercase, sans-serif font.

Table des matières

| | |
|--|-----------|
| Table des figures | 4 |
| Liste des tableaux | 5 |
| Introduction | 7 |
| 1 État de l’art | 8 |
| 2 Présentation de Short Edition | 13 |
| 2.1 Short Edition, “Éditeur communautaire de littérature courte” | 13 |
| 2.2 Projets de recherche | 14 |
| 2.2.1 Prédiction de la qualité littéraire | 14 |
| 2.2.2 Classification automatique selon les genres littéraires | 14 |
| 2.2.3 Classification automatique selon les émotions | 15 |
| 3 Données | 16 |
| 3.1 Format | 16 |
| 3.2 Pré-traitements et nettoyage | 17 |
| 3.3 Sous-corpus | 17 |
| 3.3.1 Publication | 17 |
| 3.3.2 Catégories | 18 |
| 3.3.3 Genre littéraire | 18 |

| | | |
|----------|--|-----------|
| 3.4 | Données quantitatives | 19 |
| 4 | Construction du modèle | 23 |
| 4.1 | Focalisation | 24 |
| 4.2 | Émotions retenues | 25 |
| 4.3 | Établissement des critères de définition | 27 |
| 5 | Phase d’annotation | 30 |
| 5.1 | Rédaction des documents à destination des annotateurs | 30 |
| 5.2 | Interface d’annotation | 32 |
| 6 | Validation du modèle par une étude sémantique | 35 |
| 6.1 | Short Edition, réservoir de littérature stéréotypée | 36 |
| 6.1.1 | Stéréotypes liés à l’âge | 36 |
| 6.1.2 | Blonde <i>vs</i> brune, la construction des stéréotypes | 37 |
| 6.2 | Corpus annoté selon les émotions | 39 |
| 6.2.1 | Extraction et traitement du corpus | 39 |
| 6.2.2 | Description du corpus | 40 |
| 6.2.3 | Validation du modèle par les sous-corpus | 41 |
| 6.2.4 | Contraste sémantique entre spleen et tristesse | 44 |
| 7 | Discussion | 47 |
| 7.1 | Limites de l’expérience | 47 |
| 7.2 | Premier retour d’expérimentation : modifications du modèle | 48 |
| 7.2.1 | Attribution de deux émotions avec relation de hiérarchie | 49 |
| 7.2.2 | Refonte des émotions | 49 |
| 7.2.3 | Modifications de l’option de ne pas annoter | 52 |
| | Conclusion | 55 |

| | |
|-----------------------------------|-----------|
| Bibliographie | 57 |
| Annexes | 59 |
| .1 Matrice de contraste | 60 |
| .2 Guide d'annotation | 62 |
| .3 Fiche récapitulative | 73 |
| .4 Arbre de décision | 76 |
| .5 Scripts | 79 |
| .5.1 extract.R | 79 |
| .5.2 normalization.R | 80 |
| .5.3 cleaning.sh | 81 |
| .5.4 sorting.pl | 82 |
| .5.5 make_corpus.pl | 86 |

Table des figures

| | | |
|-----|---|----|
| 1.1 | Roue de Plutchik | 9 |
| 1.2 | <i>The Hourglass of emotions</i> | 10 |
| 3.1 | Composition du corpus des textes soumis selon les catégories (nombre de fichiers) | 21 |
| 3.2 | Composition du corpus des textes soumis selon les catégories (nombre de mots) | 21 |
| 3.3 | Composition du sous-corpus des textes publiés en nombre de fichiers | 21 |
| 3.4 | Composition du sous-corpus des textes publiés en nombre de mots . | 22 |
| 5.1 | Champs de notation et d'appréciation | 32 |
| 5.2 | Interface d'annotation des émotions implémentée dans le comité éditorial | 33 |
| 5.3 | Visualisation des autres appréciations | 34 |
| 6.1 | Répartition des sous-corpus (nombre de mots) | 41 |
| 7.1 | Nouvelle interface d'annotation | 54 |

Liste des tableaux

| | | |
|-----|--|----|
| 3.1 | Rubriques | 19 |
| 3.2 | Répartition du corpus en fichiers et nombre de mots | 20 |
| 3.3 | Taux d'acceptation des textes suivant les catégories | 20 |
| 6.1 | Comparaison des rangs des occurrences | 37 |
| 6.2 | Cooccurents de blonde, brune, blond, brun | 37 |
| 6.3 | Volume des sous-corpus annotés selon les émotions en nombre d'oc- currences | 41 |

Remerciements

Je remercie tous ceux qui ont contribué à l'élaboration de ce mémoire, et plus particulièrement Patrick Paroubek et Mathieu Valette, pour leurs nombreux conseils, leur disponibilité, leur confiance et leur soutien à toute épreuve.

Je tiens également à remercier mes camarades de master, mes collègues du LIMSI et l'équipe de Short Edition.

Et enfin, mes parents pour m'avoir soutenue tout au long de ces années.

Introduction

Les plates-formes communautaires relatives à l'industrie culturelle sont aujourd'hui largement répandues. Si les plus connues d'entre elles concernent la musique ou les vidéos, certaines innovent et cherchent à exploiter le domaine de la littérature. Ainsi, Short Edition se présente comme l'éditeur communautaire de la littérature courte¹, le "Youtube de la nouvelle"². Bien que le domaine soit novateur et les acteurs présents peu nombreux, la concurrence est déjà forte et la course à l'innovation fait rage. Dans ce contexte, Short Edition cherche à explorer de nouvelles voies, comme celle qui nous intéresse : la classification automatique des textes du site selon les émotions. Afin d'atteindre cet objectif, il est nécessaire de mettre en place un processus d'annotation manuelle des textes d'après un modèle original et encadré par des directives spécialement conçues. Ce mémoire présente le travail réalisé dans le cadre de ce projet. Après avoir réalisé un état de l'art, puis nous présentons Short Edition et les données d'étude. Nous définissons ensuite un nouveau modèle et ses paramètres, ainsi que les documents et interfaces nécessaires au processus d'annotation. Enfin, nous procédons à la validation du modèle construit par une étude textométrique.

1. <http://short-edition.com/p/la-short-litterature>

2. Quentin Pleplé, <http://yakwala.fr/content/algorithmes-litterature-media-quentin-pleple-short-edition>

Chapitre 1

État de l’art

Afin de placer notre travail dans le champ des travaux déjà réalisés dans ce domaine, nous avons réalisé un état de l’art regroupant des références relevant de divers domaines (psychologie, traitement automatique des langues, linguistique) touchant à notre problématique.

Nous avons considéré dans un premier temps le modèle de formalisation des émotions de Robert Plutchik, notamment sa conceptualisation des émotions en une roue (figure 1.1). Cette roue reprend en son centre huit émotions que Plutchik considère comme basiques, répondant chacune à un besoin particulier relatif à l’instinct de survie. Ces huit émotions, organisées au centre de la roue en quatre paires opposées, sont également déclinées en degré d’intensité : plus on s’éloigne du centre, moins l’émotion est forte. Enfin, les émotions peuvent s’articuler deux à deux, pour former de nouvelles émotions dites secondaires, que Plutchik désigne sous le nom de dyades. D’autre part, dans son article *The Nature of Emotions* (Plutchik, 2001), Plutchik présente sa conception du processus émotionnel : “An emotion is not simply a feeling state. Emotion is a complex chain of loosely connected events that begins *with a stimulus* and includes feelings, psychological

changes, impulses to action and specific, goal-directed behavior.” L’idée selon laquelle une émotion est toujours déclenchée par un évènement en particulier nous a paru particulièrement pertinente, et nous avons pris cette conception en compte pour l’établissement de notre propre modèle.

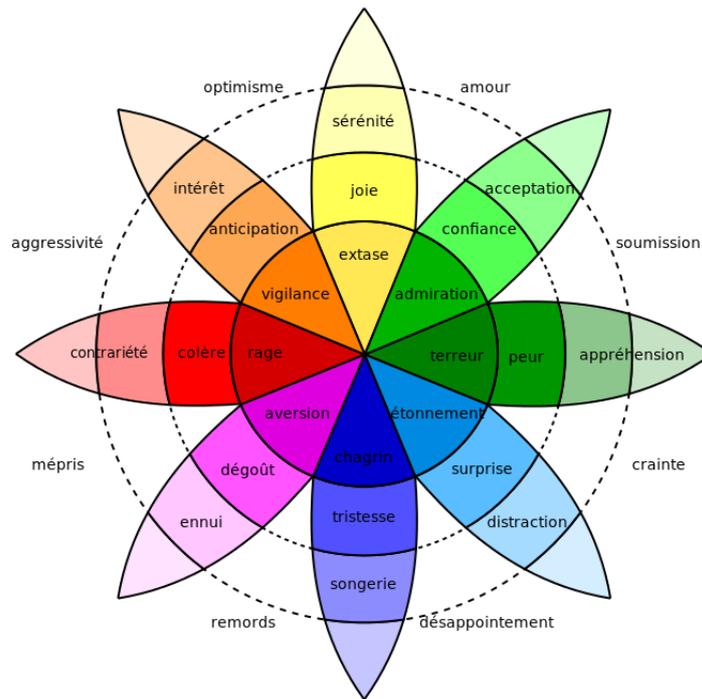


FIGURE 1.1 – Roue de Plutchik

Nous avons également considéré les travaux de (Cambria *et al.*, 2011), qui proposent une nouvelle représentation graphique du modèle de Plutchik en sablier (figure 1.2), mettant en évidence de façon plus prononcée le paramètre d’intensité de l’émotion. Cependant, cette nouvelle configuration ne se distingue pas suffisamment de celle dont elle s’inspire, puisque le sablier comporte strictement les mêmes émotions que la roue.

Nous avons également pris en compte les travaux de (Kappas, 2013), notamment lorsqu’il considère le processus émotionnel comme une boucle, composée des

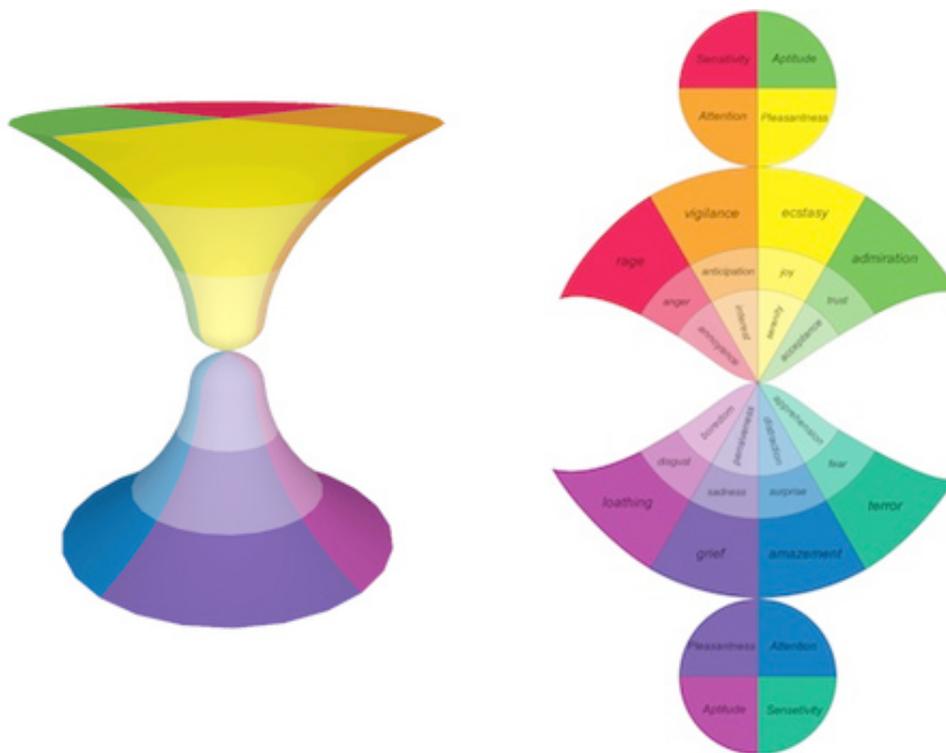


FIGURE 1.2 – *The Hourglass of emotions*

étapes suivantes :

- déclencheur extérieur,
- perception de ce déclencheur,
- engendrement de réaction et de réponses,
- et retour à la perception des changements.

Dans un second temps, nous avons considéré des travaux à visée plus applicatives et moins théoriques que ceux que nous venons de citer.

Concernant les modèles cognitifs de représentations des émotions, nous avons examiné avec intérêt le modèle *Beliefs-Desire-Intentions*, tel que présenté par (Sadek, 1992), utilisés dans le domaine de l'intelligence artificielle. Ce modèle considère trois composantes relatives aux aspirations du sujet :

- les croyances (*Beliefs*), correspondant aux informations dont le sujet dispose à propos de son environnement,
- les désirs (*Desire*), correspondant aux aspirations que le sujet veut voir se réaliser,
- et les intentions (*Intentions*), correspondant à ce que l’agent fait ou a l’intention de faire pour satisfaire ses désirs.

L’une des problématiques principales relatives à la modélisation des émotions tient au fait qu’à l’heure actuelle, aucun consensus n’a pu être trouvé en ce qui concerne l’établissement d’une définition (par exemple, la question de la différence entre un sentiment et une émotion est encore largement ouverte). Par conséquent, le modèle BDI fait office de support formel pour la rationalisation, l’explicitation et la distinction des émotions, et permet également d’uniformiser la comparaison des modèles disponibles.

Nous avons pu constater que ces modèles étaient repris dans un certain nombre de travaux relatifs à l’extraction d’information et à la classification automatique selon les émotions (Pak, 2012). Les travaux de (Fraisie et Paroubek, 2014) présentent l’intérêt de proposer une formalisation des émotions selon des critères très précis (temporels, causaux, etc) sous forme de formules logiques, pour un nombre important d’émotions.

Néanmoins, aucun de ces modèles n’a été conçu spécifiquement pour être appliqués à des textes relevant du genre littéraire. Au mieux, ils sont destinés à la classification de produits textuels propres au web 2.0 de type avis consommateurs ou posts de blogs. Par conséquent, nous avons également souhaité nous appuyer sur des travaux portant plus précisément sur la caractérisation des émotions dans des corpus littéraires, tels que (Gorcy, 1995) et (Bourion, 1995), qui proposent une

telle caractérisation au cours d'une étude du corpus FRANTEXT¹ Ces travaux nous ont permis de ne pas perdre de vue l'aspect singulier des textes traités dans notre étude. D'autre part, dans la perspective de la validation de nos partis pris, nous nous sommes appuyés sur les travaux de (Ensoo et Valette, 2013), visant à la caractérisation de traits relatifs à l'expression des sentiments par le biais d'une étude textométrique.

En somme, l'enjeu majeur de notre travail se trouve dans la pluridisciplinarité de la problématique. Nous avons fait en sorte de ne négliger aucun de ces aspects. Dans la mesure où aucune de ces références ne s'applique parfaitement à nos objectifs et afin de produire le modèle le plus précis et pertinent, nous avons pris la décision d'innover et de créer nos propres critères et référentiels, inspirés des références précédemment citées.

1. <http://www.frantext.fr/>

Chapitre 2

Présentation de Short Edition

2.1 Short Edition, “Éditeur communautaire de littérature courte”

Short Edition est une startup fondée en 2011 et basée à Grenoble qui se définit comme “l’éditeur communautaire de la littérature courte”. Le principe est de mettre en ligne, diffuser et éditer des textes soumis par des auteurs amateurs pouvant se lire en moins de 20 minutes : textes, poésies, bandes dessinées. Les œuvres sont diffusées grâce à plusieurs plate-formes, le site “classique” www.short-edition.com et les applications mobiles (Apple et Android), qui permettent une lecture nomade auquel le format court se prête bien. Soumises à l’appréciation d’un comité éditorial composé d’environ 80 “grands lecteurs”, les œuvres peuvent ensuite entrer en compétition pour le Grand Prix du Court, qui a lieu quatre fois par an. Celles qui sont retenues pour le concours sont alors soumises au vote de tous les utilisateurs, et les lauréats sont publiés dans la revue papier “Short”. Certaines œuvres sont également lues à haute voix par des comédiens pour la réalisation de podcasts.

Short Edition cherche donc à se positionner comme une maison d’édition no-

vatrice, aussi bien dans son fonctionnement que dans ses voies de diffusion.

2.2 Projets de recherche

Short Edition a développé un programme de recherche avec plusieurs laboratoires universitaires. Ce programme se décompose principalement en trois axes.

2.2.1 Prédiction de la qualité littéraire

Aujourd’hui, 1500 à 2000 textes environ sont soumis chaque mois à l’évaluation du comité éditorial. Un certain nombre de ces textes sont très largement en-dessous du niveau minimal requis pour être mis en ligne sur le site. L’objectif de ce programme de recherche est donc de développer un algorithme permettant d’émettre un premier avis sur un texte. Cette ambition découle de l’observation de points communs parmi les textes refusés comme de trop nombreuses fautes d’orthographe ou une utilisation anarchique de la ponctuation. Ainsi, un texte que l’algorithme aura supposé “mauvais” sera soumis à seulement trois évaluations humaines au lieu de dix.

2.2.2 Classification automatique selon les genres littéraires

Les textes diffusés par Short Edition sont classés en genres : romance, humour, policier, etc. Jusqu’à maintenant, les œuvres sont étiquetées (de une à trois étiquettes) manuellement par la direction éditoriale. Le deuxième volet de ce programme de recherche vise à développer un autre algorithme permettant d’opérer cette classification automatiquement.

2.2.3 Classification automatique selon les émotions

Le dernier volet du programme de recherche concerne également la classification automatique des textes, mais cherche cette fois à proposer un système de classification novateur relatif aux émotions exprimées dans le texte. Ce système permettrait de suggérer des lectures aux utilisateurs dépassant les simples critères de genre.

Chapitre 3

Données

Le corpus fourni par Short Edition en tant que base de travail rassemble plus de dix mille textes soumis au comité éditorial depuis plusieurs années.

3.1 Format

Le corpus a été fourni sous la forme d’un fichier XML présentant les métadonnées principales suivantes :

- numéro d’identification du texte,
- numéro d’identification de l’utilisateur,
- titre de l’œuvre,
- catégorie (nouvelles, poésie ou “très très court”),
- décision de publication (sous la forme d’un booléen),
- date de création de l’œuvre,
- date de publication,
- nombre de visites,
- nombre de votes,
- nombre d’avis issus du comité éditorial.

Nous avons réalisés plusieurs scripts R de traitement (extraction, nettoyage, etc) disponibles en annexe .5 exploitant ces meta-données, notamment afin d’obtenir des sous-corpus.

Short Edition a également fourni d’autres méta-données au format XML tels que les avis déposés sur une œuvre ainsi que les étiquettes relatives au genre littéraire qui lui ont été attribuées après acceptation.

3.2 Pré-traitements et nettoyage

3.3 Sous-corpus

3.3.1 Publication

Le comité éditorial, composé d’environ 80 utilisateurs du site considérés comme “grands lecteurs”, souvent eux-mêmes auteurs d’œuvres mises en ligne, lit ces œuvres par le biais d’une interface spéciale du site et attribue une note allant de 1 à 10. Les membres du comité sont invités par l’équipe éditoriale à attribuer une note entre 1 et 4 s’ils considèrent que le texte n’atteint pas un niveau suffisant pour être mis en ligne sur le site. À l’inverse, attribuer une note entre 5 et 10 revient à donner son assentissement pour la mise en ligne de l’œuvre. Enfin, les membres ont également la possibilité d’étiqueter un texte comme “Hors ligne éditoriale”. Ce choix est prévu pour les textes ne remplissant pas les conditions de mise en ligne soit pour une question de contenu, c’est-à-dire les textes relevant de l’actualité ou tenant des propos polémiques, racistes, homophobes, etc, soit pour une question de genre (pamphlets, essais, etc). Enfin, une zone “Commentaire” obligatoire est prévue afin de justifier l’attribution de la note.

Chaque œuvre est soumise à l’appréciation d’une dizaine de membres du comité. La décision de publication est ensuite prise par la Direction Éditoriale, en

s'appuyant sur les commentaires et les notes attribuées par les membres.

3.3.2 Catégories

Les auteurs peuvent soumettre leurs textes dans trois catégories différentes.

- Une catégorie *Nouvelles* pour les œuvres de fiction en prose de moins de 25 000 signes (entre 1 200 et 5 000 mots),
- Une catégorie *Très Très Court* de textes de moins de 6 000 signes (1 200 mots, moins de cinq minutes de lecture) correspondant au format émergent de “micro-nouvelle”,
- Une catégorie *Poésie* qui n’affiche pas ouvertement de contrainte de volume et regroupant divers formats du genre poétique (alexandrins, haiku, sonnets, slam, prose, etc).

Le site propose également une quatrième catégorie, les bandes dessinées courtes, qui n’entre pas dans le cadre de notre travail.

3.3.3 Genre littéraire

En dehors de la classification selon le format, le site propose actuellement une classification par genre, présentés dans une rubrique intitulée “Par rayon”, composée de plusieurs sous-catégories. Cette classification s’inspire des classements que l’on peut trouver dans une bibliothèque, mais reste perfectible.

La catégorie *Poésie* comporte également un “rayon” spécifique, relatif aux formes poétiques : Alexandrins, Chanson, Haikus, Slam, Sonnets, Vers libres, Prose, Comptine et Fable.

| Rubrique | Sous-rubrique |
|------------------|--|
| Chronique | Arts, Histoire, Nature, Spiritualité, Société, Voyage, Sport, Gastronomie, Sciences - Médical |
| Émotions | Amitié, Drame, Erotisme, Famille, Colère, Humour, Instant de vie, Enfance, Mélancolie - Solitude, Nostalgie - Souvenirs, Romance |
| Fantastique - SF | Esotérique, Fantasy, Merveilleux, Science-Fiction, Surnaturel |
| Jeunesse | Jeunesse |
| Noir | Aventure, Horreur, Policier, Thriller, Suspens |
| Autres | Allégorie, Conte, Fantaisiste, Autres, Lettre |

TABLE 3.1 – Rubriques

3.4 Données quantitatives

Les chiffres du tableau 3.2 nous permettent de calculer le taux d'acceptation des textes, présenté dans le tableau 3.3. Nous pouvons donc constater qu'en ce qui concerne le corpus dont nous disposons, les taux d'acceptation varient grandement d'une catégorie à l'autre, allant de 77,26% pour les nouvelles, à 20,90% pour la poésie, avec un taux global de 24,79%.

Les figures 3.1 et 3.2 fournissent un aperçu de la répartition des textes selon les catégories, ainsi que leurs volumes en termes de nombre de mots. Bien que la poésie soit la catégorie dominante, ce sont les nouvelles qui constituent le sous-corpus le plus important. Cette configuration est importante à prendre en compte dans le cadre d'une étude textométrique, nécessitant un volume de données suffisant pour être pertinente.

| Type de texte | Nombre de textes | Nombre de mots ¹ |
|-----------------------------|------------------|-----------------------------|
| Nouvelles publiées | 1 026 | 2 592 747 |
| Nouvelles refusées | 1 328 | 3 309 790 |
| Sous-total nouvelles | 2 354 | 5 902 537 |
| Très très courts publiés | 1 153 | 803 697 |
| Très très courts refusés | 3 300 | 1 725 846 |
| Sous total TTC | 4 453 | 2 529 543 |
| Poésie acceptées | 1 132 | 233 008 |
| Poésie refusées | 5 417 | 1 038 907 |
| Sous-total poésie | 6 549 | 1 271 915 |
| Sous-total acceptés | 3 311 | 3 629 452 |
| TOTAL | 13 356 | 6 394 205 |

¹ Chiffres obtenus à l'aide de la fonction "Description" de TXM.

TABLE 3.2 – Répartition du corpus en fichiers et nombre de mots

| | |
|------------------|--------|
| Nouvelles | 77,26% |
| Très très courts | 34,94% |
| Poésie | 20,90% |
| Taux global | 24,79% |

TABLE 3.3 – Taux d'acceptation des textes suivant les catégories

Considérant uniquement les textes publiés, la répartition est équitable en ce qui concerne le nombre de fichiers, ce qui peut refléter une volonté de la part de la rédaction de ne pas favoriser une catégorie par rapport à une autre et de conserver une ligne éditoriale variée. Pour ce qui est du volume, ce sont encore les nouvelles qui représentent la part la plus importante du corpus, avec 71,44%, ce qui est finalement logique, puisqu'il s'agit du format correspondant aux œuvres les plus longues, si l'on peut dire.

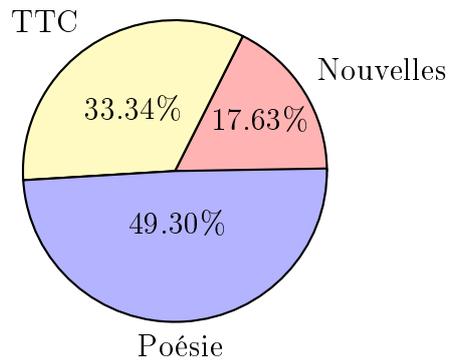


FIGURE 3.1 – Composition du corpus des textes soumis selon les catégories (nombre de fichiers)

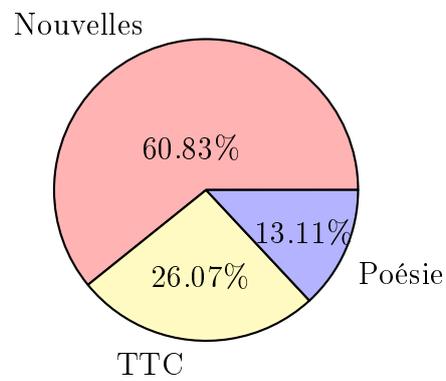


FIGURE 3.2 – Composition du corpus des textes soumis selon les catégories (nombre de mots)

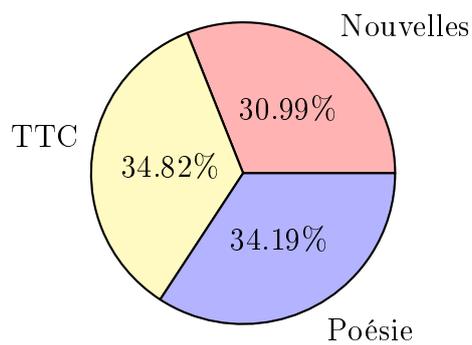


FIGURE 3.3 – Composition du sous-corpus des textes publiés en nombre de fichiers

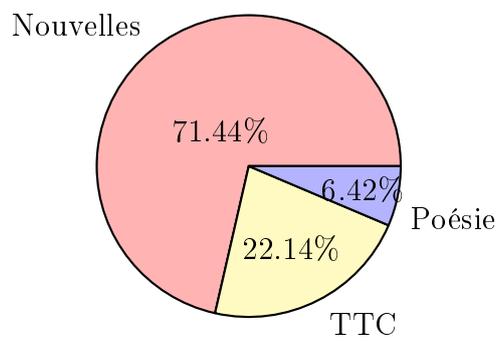


FIGURE 3.4 – Composition du sous-corpus des textes publiés en nombre de mots

Chapitre 4

Construction du modèle

Le choix des émotions a été effectué en s'inspirant des modèles présentés dans l'état de l'art, d'après un certain nombre de contraintes qu'il était déterminant de bien prendre en compte.

En premier lieu, parmi les modèles d'émotion que nous avons considérés dans l'état de l'art, nous n'avons trouvé aucune mention de leur utilisation dans le cadre d'un projet relatif à des textes littéraires, mais uniquement à des textes de type réseaux sociaux ou avis consommateurs. Or, il nous a paru primordial de considérer les textes sans perdre cette composante de vue. Il semble délicat de traiter des avis consommateurs ou des critiques de films de la même façon que des nouvelles, les émotions exprimées dans ces deux genres différant radicalement. Ainsi, parmi les émotions répertoriées dans ces modèles, certaines ne nous ont pas semblé pertinentes au premier abord. C'est le cas, entre autres, de la surprise, du dégoût ou de la colère (nous avons dû revenir sur ce postulat de départ après une première phase d'expérimentation).

D'autre part, l'une des demandes initiales de Short Edition était que le modèle

comporte un nombre relativement restreint d'émotions (de l'ordre de sept ou huit). Cette contrainte résulte du fait que le processus d'annotation a été dévolu aux membres du comité éditorial, qui sont bénévoles et non-experts. Cela impliquait en conséquence de produire un modèle convivial, facile à appréhender, et surtout ne risquant pas de lasser les membres du comité, puisqu'ils réalisent l'essentiel du travail de lecture et de critique des textes soumis au site.

4.1 Focalisation

Le premier paramètre à définir était celui de la focalisation des émotions. En effet, avec une telle problématique, de surcroît appliquée à un corpus de textes littéraires, nous pouvions envisager deux axes d'étude : celui des émotions générées par le texte chez le lecteur, et celle des émotions exprimées et éprouvées par les personnages dans le texte. La seconde piste s'est rapidement imposée. En effet, bien que la première piste soit tout à fait pertinente et intéressante, nous ne disposions pas des paramètres nécessaires à sa mise en œuvre, puisqu'il nous aurait fallu pouvoir étudier un panel de lecteurs, que nous aurions soumis au même texte et dont nous aurions évalué les réactions. En revanche, le corpus conséquent de textes issus de la plateforme nous permettait de nous orienter vers la seconde piste.

Nous avons donc décidé que le modèle évaluerait les émotions ressenties soit par le personnage principal du récit, soit par le narrateur. Ce point, déterminant pour la cohérence du modèle et des annotations, a été fortement souligné auprès des annotateurs.

4.2 Émotions retenues

Après une phase de réflexion et de concertation avec l'équipe de Short Edition, notre décision s'est arrêtée sur les six émotions présentées ci-dessous. Ces six émotions se partagent entre trois émotions "euphoriques" et trois émotions "dysphoriques", cependant cette répartition équilibrée n'est pas volontaire.

Joie La première émotion prise en compte est la joie. S'agissant d'une émotion de base mais a priori assez courante, il nous a paru pertinent de la prendre en compte. Le label "joie" vise à recouper toute notion de bonheur, félicité, etc.

Affection La deuxième émotion choisie a été intitulée "Affection". Nous souhaitons constituer un label regroupant les émotions de type "penchant" et "intérêt" destinées à des êtres vivants ou non (rassemblant ainsi des choses aussi diverses que l'amitié, la passion pour un loisir ou un sujet, le désir, l'amour filial, etc) sans avoir à l'agrémenter d'une composante d'intensité ou de degré, qu'il nous aurait fallu intégrer aux critères de définition des émotions, que nous voulions les plus génériques possibles (applicables à chacune des émotions de façon interchangeable). Nous avons donc eu recours au label "Affection", qui présentait, à première vue, l'avantage de la clarté et de la non-ambiguïté tout en traduisant relativement bien l'idée représentée. Ce choix d'un label consensuel s'est également opéré afin d'éluider des questions parfois épineuses, telle que celle de la frontière entre l'amour et l'amitié.

Sérénité Pour la dernière émotion du volet euphorique du modèle, notre choix s'est arrêté sur la sérénité. Ce choix s'est dessiné au fil de l'examen du corpus, où un certain nombre de textes comportaient un aspect "positif-calme" qu'il nous a semblé important de prendre en compte. Le concept de la sérénité repose sur une

idée d’accomplissement et de plénitude de la part du sujet.

Tristesse Dès le début de la conception du modèle, la tristesse s’est rapidement imposée comme incontournable. S’il s’agit bien entendu d’un thème classique et souvent traité en littérature, ce choix a également été pris d’après une information de Short Edition. En effet, les membres de l’équipe éditoriale, qui connaissent bien le corpus et les auteurs, depuis longtemps, a pu remarquer que la plateforme fait office pour certains d’exutoire, à la manière d’un journal intime.

Spleen Comme (Gorcy, 1995) l’a cité en reprenant une définition du TLF, le spleen se décrit comme un “état affectif, plus ou moins durable, de mélancolie sans cause apparente et pouvant aller de l’ennui, la tristesse vague au dégoût de l’existence”. L’idée de l’inclure dans le modèle nous est venue lorsque nous cherchions une émotion à mettre en parallèle de la tristesse, pour compléter son champ d’action à cet état négatif dont on ne parvient pas à identifier la cause. Après avoir envisagé un temps la mélancolie, le spleen est finalement apparu comme convenant mieux à ce que nous recherchions. La connotation littéraire de cette émotion n’y est bien sûr pas étrangère.

Peur Selon (Bourion, 1995), la peur se définit comme “un sentiment dysphorique, mêlé et intense”. La prise en compte de cette émotion était particulièrement pertinente dans le cadre de l’étude d’un corpus littéraire, regroupant une part importante de textes des genres policiers, horreur/épouvante, suspens, etc.

Suite à la description de ces émotions, nous souhaitons insister sur le fait qu’il faut absolument distinguer les émotions du texte (celles des personnages et/ou du narrateur) du genre littéraire duquel le texte relève. Par exemple, supposons un texte où un détective mène une enquête, tout en vivant une histoire d’amour. Le

texte, bien que relevant du genre “policier”, pourra être annoté “amour” (si cette émotion est suffisamment présente pour être prise en compte).

4.3 Établissement des critères de définition

Afin de permettre une description formelle et tangible des émotions du modèle et de les transmettre aux annotateurs sans équivoque, nous avons établi un ensemble de critères de définitions, pouvant s’appliquer de façons interchangeable à chacune d’entre elle et permettant ainsi de les contraster afin d’anticiper tout problème d’ambiguïté, de la même façon que (Mathieu, 2008) définit des propriétés sémantiques de verbes. Si la description subjective et informelle ne posent pas de problème particulier, nous avons pu constater qu’il est en autrement de la formalisation des émotions, qui ne répondent pas toujours parfaitement aux critères que l’on tente de leur appliquer.

Nous avons défini quatre critères de contraste, afin de rester le plus possible dans une configuration claire, rapidement et facilement appréhensible. Ces critères sont les suivants :

- La **cause** : En partant du principe qu’une émotion découle toujours d’un stimulus tel que (Plutchik, 2001) l’a spécifié, nous admettons l’existence d’une cause à chaque émotion, que nous désignons sous l’appellation *cause de l’émotion*. Après avoir établi cette fondation, nous avons désigné un ensemble de paramètres définissant la cause en elle-même : la conscience de la cause, son type (événements, actions, objet), et le désir ou non de cette cause par le sujet. En réalité, l’ensemble de ces paramètres est relatif à la perception de la cause par le sujet : a-t-il conscience de ce qui cause son émotion ? Est-il à l’origine de cette cause, ou la subit-il (ce qui est la

différence entre les évènements que l'on considère comme subis, les actions comme perpétrées par le sujet, et l'objet, vers qui l'émotion est dirigée)? La désire-t-il?

- La **temporalité** : Une fois la cause définie, le deuxième critère discriminant est d'ordre temporel. Il vise à déterminer dans quel temporalité la cause de l'émotion a eu lieu, par rapport au temps du ressenti de l'émotion. La temporalité peut donc être passée (la cause à eu lieu avant le ressenti de l'émotion), présente (la cause et l'émotion sont simultanées), ou future (la cause aura lieu après le ressenti de l'émotion).
- Les **intentions** : Nous avons également introduit un critère de satisfaction des intentions du sujet. Ce critère permet d'affiner le paramètre de désir de la cause. Il s'inspire directement du modèle *Beliefs Desire Intentions* (BDI) tel que Sadek cité par (Fraisie et Paroubek, 2014) l'a défini.
- La **valence** : La valence est finalement déduite des précédents paramètres établis. Nous considérons que par son caractère empirique, la valence d'une émotion ne peut être considérée comme un critère de définition majeur. C'est pourquoi nous la définissons en dernier, d'après d'autres critères établis précédemment.

Il est important de noter que ces critères doivent être définis *dans l'ordre dans lequel ils sont présentés*. En effet, dans le cas d'émotions dont la cause est inconnue du sujet (dans le cas de notre modèle, le spleen), il n'est pas possible de définir ensuite le type, le désir et la temporalité de la cause. Par conséquent, la définition de l'émotion se limite à cette composante : la cause est inconnue.

D'autre part, un certain nombre de critères envisagés n'ont pas été retenus, toujours dans un souci de simplicité du modèle. On compte notamment parmi eux le critère d'intensité, que nous avons déjà évoqué au cours de la présentation de l'af-

fection. Ce critère ne nous a finalement pas semblé crucial, car nous avons estimé que les émotions retenues se distinguaient suffisamment selon leur fond. L'intensité aurait éventuellement pu être pertinente pour distinguer la joie de la sérénité, mais n'aurait été applicable et réellement utile que dans ce cas-là. C'est donc notamment le manque d'adaptabilité de ce critère qui nous a fait l'écarter, couplé bien entendu à la "couche de complexité" supplémentaire qu'il apportait. En somme, le rapport bénéfice/complexité n'était pas satisfaisant.

À ce stade de l'expérimentation, nous espérions avoir produit un modèle certes et évidemment imparfait, mais suffisamment étayé pour satisfaire nos exigences, ménageant la chèvre et le chou, c'est-à-dire abordable pour les annotateurs, sans agir au détriment de la finesse des définitions.

Le modèle ainsi établi a été finalement représenté sous la forme d'une matrice à double-entrée disponible en annexe .1. Les émotions sont définies dans la diagonale et contrastées entre elles aux intersections (mise en évidence des critères qui diffèrent d'une émotion à une autre).

Chapitre 5

Phase d’annotation

Une fois le modèle d’émotions établi, la phase suivante était la mise en place des outils nécessaires au processus d’annotation, à savoir les documents à destination des annotateurs et l’interface d’annotation.

La nature des documents nécessaires a été décidée en concertation avec l’équipe de Short Edition. Nous avons également suggéré, pour la première expérimentation, la limitation de l’étendue de l’expérience aux textes du format *Très Très Court*, supposant que ceux-ci pouvait présenter une meilleure homogénéité dans les émotions exprimées.

5.1 Rédaction des documents à destination des annotateurs

Guide d’annotation Dans un premier temps, Short Edition a demandé la réalisation d’un guide d’annotation concis (deux pages maximum). Il s’est avéré que cet objectif était très délicat voir même impossible à atteindre. Nous avons en effet établi un plan du guide consistant essentiellement en une présentation des critères de définition et leurs différentes valeurs possibles, suivie de la définition formelle de

chaque émotion, une définition “résumée” simplifiée, et des exemples commentés tirés du corpus destinés à illustrer l’usage des émotions. Enfin, la dernière partie du guide rassemble des consignes d’ordre général et présente le dernier choix laissé aux annotateurs, à savoir celui de ne pas attribuer d’émotion à un texte. Dans ce cas de figure, nous demandons aux annotateurs de justifier ce choix de façon concise, ceci afin d’éviter qu’ils y aient recours en une sorte de “solution de facilité” et de pouvoir ensuite analyser ces commentaires pour en tirer des observations au sujet de ces textes “problématiques”. Finalement, le guide est constitué de cinq pages (annexe .2).

Arbre de décision Un arbre de décision constitué de questions fermées a également été réalisé à l’aide du logiciel XMIND. Le but de cet outil est de permettre aux annotateurs, en cas de doute, de “suivre un chemin” qui leur permette d’aboutir à l’émotion la plus pertinente pour le texte qu’ils évaluent (annexe .4).

Afin de lever une ambiguïté relative à la peur et à la tristesse, nous avons dû affiner le critère de type entre ces deux émotions, et déclarer la cause de la peur explicitement comme une menace ou un danger.

Fiche récapitulative Enfin, le dernier document produit est une fiche récapitulative d’une page structurée en deux parties principales. Le premier bloc reprend des définitions “informelles”, les définitions “formelles”, et est agrémenté d’icônes représentant chacune des émotions. Le second bloc quant à lui reprend quelques points synthétiques visant à orienter les annotateurs dans le processus de choix de l’émotion. La création graphique de cette fiche a été confiée à la graphiste de Short Edition, afin de produire un document coloré et convivial pouvant faire office de “guide d’annotation d’appoint” (annexe .3).

5.2 Interface d’annotation

Ensuite, nous avons contribué à concevoir l’interface d’annotation, en concertation avec l’équipe technique de Short Edition. Un module permettant de choisir une émotion a été ajouté aux champs destinés à recevoir la note et l’appréciation des membres du comité. Il se présente sous la forme d’une liste à puces avec les descriptions succinctes des émotions et les icônes de la fiche récapitulative. Le choix de ce format “liste à puce” permet aux annotateurs de conserver une vision globale de l’ensemble des émotions et de leurs définitions.



FIGURE 5.1 – Champs de notation et d’appréciation

La figure 5.1 présente l’interface “classique”, commune à tous les membres du comité éditorial. Le texte évalué se situe au-dessus.

La figure 5.2, quant à elle, présente l’interface réservée aux membres faisant partie de l’expérience émotions. Elle présente donc, en plus des icônes, la définition “informelle” de chaque émotion, rédigée initialement pour la fiche récapitulative.

Après avoir attribué une note et une émotion, l’annotateur débouche sur une page où il peut accéder aux notations et appréciations des membres ayant déjà évalué le texte. Les membres faisant partie de l’expérience émotions peuvent aussi voir l’émotion attribuée par les autres membres participants (ceux n’étant pas encore intégrés ne voient que l’interface habituelle telle qu’ils la connaissent déjà.)

Expérience Emotions
bêta

Quelle est l'émotion que ressent le personnage principal ?

 **Joie**
Le sujet ressent du bonheur. Il est content, joyeux, plein d'enthousiasme.

 **Peur**
Tout ce qui relève de la crainte, l'appréhension, la terreur.

 **Affection**
Vaut pour toutes les formes d'amour, d'intérêt, de désir (le sentiment amoureux, amour filial, attachement, etc), la passion pour une activité ou un sujet.

 **Tristesse**
Le désespoir, le chagrin, la peine, l'affliction.

 **Sérénité**
Le sujet a réalisé quelque chose qui le satisfait, il est apaisé, détendu, calme, serein, relaxé.

 **Spleen**
Le sujet est mal sans savoir précisément pourquoi. C'est le blues, la déprime, le moral dans les chaussettes, "l'angoisse d'exister" comme Baudelaire l'a défini.

Je ne peux pas choisir

Pourquoi ne pouvez-vous pas choisir d'émotion ici ?

Valider

FIGURE 5.2 – Interface d'annotation des émotions implémentée dans le comité éditorial

Pour des raisons de confidentialité, les pseudonymes et avatars ont été floutés sur la figure 5.3.

| | | |
|---|--|--|
|  | 5 / 10 Récit touchant. il y a 2 mois · répondre |  Affection |
|  | 2 / 10 Histoire bête et mal développée. A retravailler. il y a 2 mois · répondre |  Bonheur |
|  | 4 / 10 Le récit est bancal, l'écriture quelconque. il y a 2 mois · répondre |  Bonheur |

FIGURE 5.3 – Visualisation des autres appréciations

Chapitre 6

Validation du modèle par une étude sémantique

En parallèle de la construction du modèle et de la mise en marche du processus d’annotation, une étude textométrique des corpus a été réalisée, afin d’en dégager des traits caractéristiques, qui d’une part, corroboreraient les bases de notre modèle, et qui, d’autre part, pourraient être utiles dans la perspective de l’amélioration du modèle.

Nous insistons sur le fait que ces études sont présentées ici à un stade relativement peu avancé (notamment à cause de la progression assez lente du processus d’annotation), et qu’il serait nécessaire de les approfondir si l’on souhaitait obtenir une véritable vue d’ensemble des textes de Short Edition.

Nous avons eu recours, pour cette étude au logiciel de textométrie TXM ([Heiden Serge *et al.*, 2010](#)).

6.1 Short Edition, réservoir de littérature stéréotypée

Dans un premier temps, nous nous sommes intéressés à la construction des personnages. En effet, le programme de recherche de Short Edition visant à étudier la psychologie générale des œuvres du site, nous avons pensé que cette perspective s’inscrivait dans le courant de cette thématique de recherche, tout en restant liée aux questions relatives aux émotions.

Au fil de l’examen du corpus, nous avons affiné cette orientation, pour la recentrer sur l’étude des stéréotypes relatifs au genre.

Nous avons donc comparé les cooccurrents des termes “homme”, “femme”, “fille” et “garçon”, dans le premier corpus fourni par Short Edition.

6.1.1 Stéréotypes liés à l’âge

Considérées en parallèle, ces deux listes de cooccurrents nous permettent de remarquer quelques traits caractéristiques communs aux deux genres, notamment la présence importante de “jeune” et “vieux”/“vieille”. Cependant, il est important de remarquer que si “jeune” se classe en première place (pour “femme”) et deuxième place (pour “homme”) des cooccurrences, partageant donc apparemment un usage similaire, on ne peut pas en dire autant pour “vieux”, qui occupe le premier rang des cooccurrences les plus fréquentes du mot “homme”, alors que “vieille” se situe au rang 11. On repère également “mûr” du côté masculin, apparaissant au rang 16, alors que la version féminine “mûre” n’apparaît qu’au rang 64 des cooccurrences de “femme”. De surcroît, les noms du type “trentaine”, “quarantaine”, “cinquantaine”, etc, cooccurrent plus fortement avec “homme” qu’avec “femme” (tableau 6.1).

Si l’on explore alors les cooccurrences de “garçon” et de “fille”, on constate que “quinze” cooccur avec “garçon” et “dix-sept” cooccur avec “fille”, ce qui corres-

| | Homme | Femme |
|--------------|-------|--------|
| Trentaine | 21 | 40 |
| Quarantaine | 37 | 50 |
| Cinquantaine | 35 | 227 |
| Soixantaine | 30 | absent |

TABLE 6.1 – Comparaison des rangs des occurrences

pond à l’indication de l’âge des personnages. On peut dégager de ces observations un premier schéma stéréotypique selon lequel l’âge des protagonistes est relatif à leur genre. Dans le cas des adolescents, cette stéréotypie se manifeste à travers des âges précisément identifiés.

6.1.2 Blonde *vs* brune, la construction des stéréotypes

Afin de poursuivre l’exploration de cette hypothèse, nous avons cherché à mettre en évidence d’autres manifestations de la stéréotypie des personnages. Nous avons remarqué que les cooccurrents de “femme” présentaient plus souvent des termes relatifs à la description physique, lorsque les cooccurrents d’“homme” renvoyait plus souvent au statut social et professionnel (“costume”, “vêtu”, “complet”, “honnête”, “brave”). Parmi les cooccurrents de “femme”, “brune” et “blonde” se détachaient. Nous avons exploré plus précisément le cas de ces deux termes.

| | |
|--------|--|
| Blonde | chevelure, belle, femme, brune, jeune, pulpeuse, peroxydée, jolie, petite, mince, bleus, pétasse, fille, platine, perruque, grande, fausse, yeux, crinière |
| Brune | chevelure, jolie, femme, verts, belle, jeune, blond, grande, cheveux, yeux, courts, tignasse, peau, lunettes, chignon, petite, rousse |

TABLE 6.2 – Cooccurrents de blonde, brune, blond, brun

Le tableau 6.2 Ils nous révèlent une nouvelle dimension de stéréotypie, posée

entre différentes figures féminines. On y repère la construction de la blonde comme un personnage superficiel, fortement imprégnée d'une nuance de fausseté (explicitement citée comme telle, mais aussi bien à travers "peroxydée" ou "perruque") tout en restant fantasmée ("crinière", "pulpeuse").

Du côté de la brune, on remarque des traits communs à ceux de la blonde (notamment le fait que les deux termes se répondent et se citent l'un l'autre), mais il s'en dégage surtout une construction en opposition avec le personnage frivole de la blonde. La brune porte un chignon et des lunettes, créant un stéréotype de personnage plus sérieux et impliqué.

Quoiqu'il en soit, il est important de remarquer que les deux sont décrites physiquement de façon globale, dans une représentation commune : "jolie", "mince", "jeune", "grande" ou "petite". Les apparitions de "bleus" et "verts", soulignés par "yeux" dans les deux cas, montrent que la description de la chevelure des personnages s'accompagne de celle de la couleur de leurs yeux, et que celle-ci est propre à chaque couleur de cheveux (blonde aux yeux bleus et brune aux yeux verts). On pourrait donc envisager la description de la couleur des cheveux comme le signe de la construction d'un personnage physiquement à son avantage, sans quoi les auteurs s'attachent moins à cette description.

À travers ces deux parties, nous avons donc dégagé deux axes de stéréotypie relative aux genres dans la construction des personnages : un premier axe concernant l'âge des personnages masculins et féminins, et un second axe concernant plus particulièrement les stéréotypes dans la représentation des personnages féminins. Il est important de ne pas perdre de vue que le corpus que nous avons étudié, d'une part, est constitué de textes publiés et refusés, et qu'il est d'autre part produit entièrement par des auteurs amateurs et ne bénéficie d'aucune relecture ou remaniement. Cela peut donc, en partie, expliquer les résultats que nous venons

de décrire.

6.2 Corpus annoté selon les émotions

Environ un mois après le début de la mise en annotation, nous avons réalisé une extraction de la base de données, afin d'étudier le corpus annoté selon les émotions.

6.2.1 Extraction et traitement du corpus

Short Edition ayant fourni uniquement un accès à la base de données contenant les textes évalués selon les émotions par les annotateurs, nous avons développé et mis en place une chaîne pour extraire et traiter les textes annotés.

Cette chaîne se décompose ainsi :

- L'extraction est opérée à l'aide d'une requête SQL récupérant divers champs (numéro du post, titre, nombre d'évaluations, émotions attribuées, nombre de commentaires justifiant l'annotation "pas d'émotion") et produisant en sortie un fichier CSV regroupant les champs extraits.
- La deuxième phase est le nettoyage du texte, à l'aide d'un programme intitulé `cleaning.sh`, qui supprime les retours à la ligne, les balises, les espaces insécables, etc.
- Ensuite, le tri des annotations a été effectué à l'aide d'un programme intitulé `sorting.pl`. Ce programme commence par considérer le nombre d'évaluations qui ont été déposées sur un texte. En effet, le nombre de textes évalués étant relativement faible, nous avons fait en sorte de pouvoir en prendre le plus possible en compte. Ceci pose le problème de la catégorisation des textes : comment prendre la décision de catégorisation lorsque les annotateurs ne sont pas d'accord ? Le système est relativement sommaire et c'est l'émotion

la plus attribuée qui est choisie. Si cela ne pose pas forcément de problème pour des textes présentant un nombre suffisant d'annotations (de trois à six), cette tâche est plus problématique dans le cas des textes, assez nombreux, qui n'ont été évalués que par deux annotateurs. Dans ce cas de figure, nous avons ainsi fait le choix de ne considérer que les textes pour lesquels les deux annotateurs étaient d'accord, sans quoi le texte était écarté. Dans le cas des textes présentant plus de deux annotations, le programme incrémente des compteurs, et l'émotion obtenant la plus forte valeur est choisie. Ce système de tri des annotations très sommaire mériterait un développement plus poussé que nous n'avons pas eu le temps de mettre en place. En l'état, le programme nous a permis d'obtenir rapidement et à moindre effort un nouveau fichier CSV comportant les identifiants des textes, leur titre, l'émotion retenue, et le contenu textuel.

- Enfin, la dernière étape de la chaîne est le programme `make_corpus.pl` qui compile le corpus. Il prend en entrée le CSV généré par `sorting.pl`, redirige les contenus textuels dans des fichiers textes (un par émotion) et génère le fichier de méta-données nécessaire à l'importation du corpus dans le logiciel TXM.

La réalisation de ces étapes a été partiellement automatisée à l'aide d'un script qui appelle successivement les scripts suscités (disponibles en annexe [.5](#)).

6.2.2 Description du corpus

Le corpus est composé de sept sous-corpus : joie, affection, sérénité, tristesse, peur, spleen, pas d'émotion. Il est important de garder à l'esprit le fait que l'annotation a été réalisée sur les textes de format *Très Très Court*, ce qui explique partiellement le faible volume de certains sous-corpus.

Ces données montrent une répartition inégale des textes selon les émotions,

| Émotion | Nombre d'occurrences | Nombre de textes |
|---------------|----------------------|------------------|
| Joie | 7 046 | 12 |
| Affection | 42 270 | 60 |
| Sérénité | 10 872 | 16 |
| Tristesse | 21 247 | 42 |
| Spleen | 12 460 | 19 |
| Peur | 17 785 | 24 |
| Pas d'émotion | 39 358 | 72 |
| Total | 151 038 | 245 |

TABLE 6.3 – Volume des sous-corpus annotés selon les émotions en nombre d'occurrences

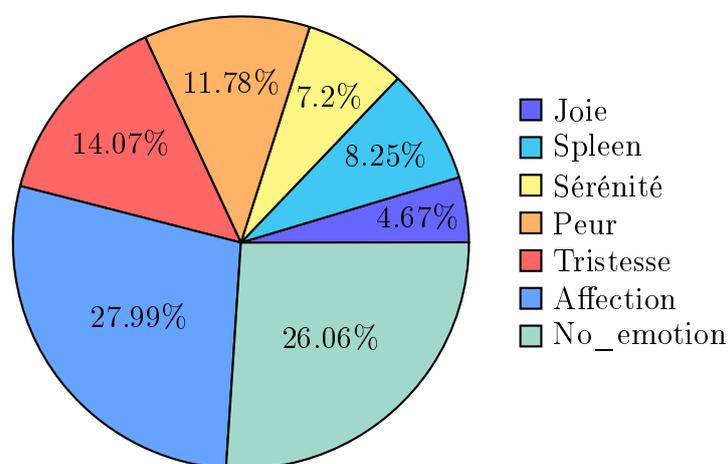


FIGURE 6.1 – Répartition des sous-corpus (nombre de mots)

ce qui peut s'expliquer par la "popularité" de certaines émotions de la part de certains auteurs de Short Edition, ou par la façon dont les annotateurs ont perçu les émotions du modèle.

6.2.3 Validation du modèle par les sous-corpus

Dans un premier temps, nous nous sommes attachés au repérage des caractéristiques de chaque sous-corpus relatifs à une émotion, afin de vérifier que chacun possède ses propres signes distinctifs. Nous considérerons pour cette partie les oc-

currences et les spécificités de chaque corpus.

Joie Malgré le faible nombre d’occurrences de ce corpus (de l’ordre de 7 000), nous remarquons certaines formes qui peuvent se rattacher en tant que motifs propres à la joie, comme par exemple “rire”. On remarque également plusieurs occurrences relatives au football : “but”, “joueur”, “ballon”, etc. Par conséquent, on peut supposer que la thématique du football relève de la joie. Cependant, cette observation est à prendre avec des précautions : le corpus de la joie comporte également un nombre important d’entités nommées (“Coluche”, “hommes-corbeau”, “Jean”, “Manuel”, “Rico...”), relatives à certains textes en particulier. Considérant le faible volume du corpus, nous en resterons là en ce qui concerne sa caractérisation.

Affection En ce qui concerne l’affection, bien qu’il s’agisse du sous-corpus le plus volumineux de l’ensemble, nous sommes d’emblée sceptiques concernant sa caractérisation, puisque nous avons pu constater que cette émotion posait particulièrement problème lors de la phase d’annotation. À l’examen du corpus, nous notons quand même la présence de mots tels que “bonheur”, “regard”, “amour”, “corps”, “cœur”, “cheveux”, qui peuvent contribuer en partie à la construction de l’émotion telle que nous l’avions conceptualisée. Si ces occurrences se distinguent, elles ne sont pas pour autant les plus fortes (les occurrences les plus fortes étant peu parlantes, et pas réellement propres au corpus de l’affection).

Sérénité Le corpus de la sérénité, avec seulement 10 872 occurrences, fait partie des corpus les moins fournis de l’ensemble. La recherche textométrique appliquée à son sujet est relativement peu parlante. Nous arrivons néanmoins à distinguer quelques formes qui pourraient être considérées comme caractéristiques de cette émotion, tels que les points de suspension, “monde”, “nouveau”, “humains”, qui pourraient être interprétés dans le sens de la représentation de cette émotion telle

que nous l'avons établie. Cependant, dans le cas de cette émotion, nous estimons que la validation textométrique n'est pas suffisamment parlante.

Tristesse En terme de volume, la tristesse rassemble plus de 21 000 occurrences, ce qui est suffisant pour espérer dégager des traits caractéristiques pertinents. On peut y repérer des motifs dans les fréquences fortes tels que “jamais”, “rien”, “temps”, “seule”, “larmes”, “amour”, “mort”, “hôpital”, etc, qui tracent les contours d'une représentation de la tristesse.

Spleen La consultation des spécificités du spleen révèle un grand nombre d'entités nommées : “Jay”, “Sylvestre”, “Tim”, “Môman”, “Mariona”, “Rouge-de-brique”, etc. Se démarquent quand même également les points de suspension, ainsi que le motif de la fenêtre, laissant ressortir une certaine thématique de l'introspection, sur laquelle nous reviendrons par la suite. La consultation du lexique, en revanche, fait ressortir les termes “toujours” et “jamais”, tous les deux avec des fréquences similaires, ce qui pourrait être interprété comme une présence importante du doute dans ce corpus.

Peur Enfin, le corpus de la peur comporte lui aussi un nombre important d'entités nommées (“Gaston”, “Marion”, “Hélène”, “Henri”, “Lucky Luke”, etc). Cependant, il s'en dégage également des termes caractéristiques de la construction d'une ambiance effrayante ou menaçante : “route”, “forêt”, “nuit”, “bruit”, “sang”, etc. Dans le cas de la construction de cette émotion, la thématique est donc explicite dans ses manifestations.

Ces observations, couplées à celles effectuées lors du processus d'annotation, ont permis de poser les bases des modifications nécessaires pour la production d'une seconde version du modèle.

6.2.4 Contraste sémantique entre spleen et tristesse

En second lieu, nous avons suivi la piste du contraste entre les émotions dysphoriques, plus particulièrement entre le spleen et la tristesse. Le parti pris du modèle étant, en ce qui concerne ces deux émotions, qu’elles se distinguent uniquement par la conscience ou non de leur cause, nous avons exploré le corpus à l’aide d’outils textométriques pour rechercher des manifestations lexicales et sémantiques de cette distinction, ce qui pourrait confirmer notre hypothèse de départ.

Nous avons ainsi eu recours de nouveau au logiciel TXM, en constituant un sous-corpus “spleen-tristesse”, permettant de considérer les spécificités propres à chaque émotion contrastivement l’une de l’autre.

Si ces deux émotions partagent a priori des points communs dans leurs manifestations, l’examen des spécificités nous a permis d’identifier des traits qui paraissent caractéristiques et distinctifs de la tristesse et du spleen.

Dans le cas de la tristesse, le corpus comporte un certain nombre de formes que nous avons identifiées comme des causes de tristesse, tel que nous avons défini ce concept pour l’établissement du modèle : “maladie”, “cancer”, “méningite”, “mort”, “hôpital”, “mensonges”, “souvenirs”, “amour” et ses nombreuses flexions “aimée”, “amant”, “amoureux”, “amours”. Il est également très intéressant de noter que le mot “cause” (et ses variantes, comme “fardeau”) sont caractéristiques de la tristesse. On trouve aussi un certain nombre de mots liés aux manifestations de la tristesse, tels que “hurler”, “hurlant”, “sanglots”, “larmes”, “souffrance”, “pensées”, etc. Il apparaît que les personnages et narrateurs du corpus de la tristesse identifient et désignent leur émotion et ses causes, comme l’illustre la présence de “nostalgique” et de “sens”, qui à l’examen des concordances ressort souvent dans “je me sens” (“seule”, “vidée”, “différente”, etc).

Dans le cas du spleen, le contraste fait ressortir de façon plus prononcée des formes relatives à l'introspection et à l'interrogation, présentes également dans le corpus de la tristesse : le point d'interrogation, "pourquoi", "est-ce", associée à "parce que", souvent utilisé en énumération, évoquant ainsi une recherche de cause. L'emploi de "quoi" semble s'inscrire également dans cette logique, lorsque l'on consulte les concordances : "je ne savais toujours pas quoi", "c'est quoi ... ?", "à quoi bon...". L'usage de "quoi" peut être contrasté avec, en regard dans le corpus de la tristesse, celui de "qui". "Qui" dans la tristesse est au contraire de "quoi" dans le spleen une forme qui désigne également une cause : "un feu qui les détruisait", "une autre personne qui hantait mon cerveau", "la personne qui les a rendu triste". "Oui" et "non" s'avèrent également caractéristiques du spleen. La consultation des concordances de ces deux mots révèle un usage en début ou fin de phrase, visant a priori à renforcer le propos : "Oui, il a pris son imper, ses godasses", "Parti, oui, enfin, disparu", "Mais non, t'es là, à gaspiller l'énergie électrique produite par les centrales nucléaires.". Cet usage (et celui de la deuxième personne du singulier) caractérise un monologue intérieur, confirmant l'hypothèse d'une introspection. En ce qui concerne l'identification de la cause, le démonstratif "ça", caractéristique du spleen, renforce l'idée de l'incapacité du sujet à identifier la cause de son mal-être, et ce qui pourrait l'aider à progresser : "C'est là que ça se déclenche", "Pourquoi est-ce que je pense à tout ça".

Nous n'avons relevé que peu d'occurrences de désignation formelle de l'état du sujet. Elles apparaissent sous la forme notamment des mots "angoisse" et "désarroi", que l'on peut rattacher entre autres aux occurrences de "dormir", "fenêtre" qui suggèrent doucement une thématique de la dépression et de l'apathie. Dans cette perspective, il est important de prendre en compte la prépondérance de pronoms personnels dans le corpus de la tristesse ("nous", "m", "J", "elle", "ses", "mes", etc).

La tristesse se dessine à travers ces motifs comme une émotion vécue en confrontation (présence de “monde” également, dans “face au monde”), au milieu de relations interpersonnelles, alors que le spleen se décrirait plutôt comme une émotion intime, vécue dans la solitude et la passivité, aussi bien en ce qui concerne ses causes que ses manifestations.

Le spleen et la tristesse présentent donc un fond dysphorique commun, mais on repère dans les corpus la distinction qui se fait sur la conscience de la cause de l’émotion, connue du sujet dans le cas de la tristesse, lorsque dans le cas du spleen, l’émotion est très liée à la réflexion à propos de la cause, que le sujet ne parvient pas à identifier. En ce sens, on peut émettre l’hypothèse selon laquelle la non-connaissance de cette cause serait en elle-même la cause ou une partie de la cause du mal-être du sujet. Quoiqu’il en soit, le spleen est une émotion vécue dans une intimité et sans recherche de confrontation. Il en résulte que le parti pris au début de notre expérimentation, à savoir la distinction entre spleen et tristesse basée sur la connaissance de la cause de l’émotion par le sujet, se trouve validée par l’observation des corpus générés d’après ce modèle.

Chapitre 7

Discussion

À l’issue de la première période d’expérimentation du modèle, nous sommes en mesure de produire un certain nombre de critiques à son encontre et d’envisager des pistes de modification et d’amélioration, dont certaines ont d’ailleurs été mises en place dans une deuxième version.

7.1 Limites de l’expérience

Dans un premier temps, nous estimons que l’expérience a été fortement limitée par les contraintes déjà longuement présentées dans les chapitres précédents.

Le principal obstacle nous semble être le recours aux annotateurs bénévoles. En effet, cette contrainte a limité drastiquement le nombre d’émotions prises en compte dans le modèle, comme nous l’avons déjà expliqué. Or, il ne nous semble pas infondé, dans ces circonstances, d’interroger la pertinence du modèle, surtout dans le cadre de son application à des “textes littéraires”.

Une autre conséquence collatérale de ce choix est la lenteur de l’annotation et son caractère aléatoire : dans la mesure où les annotateurs sont bénévoles et où la critique des textes sur la plateforme Short Edition représente pour eux un loisir, on

ne peut exiger de leur part un “investissement minimum”. Nous avons pu constater lors de la phase d’annotation la “disparition” de certains annotateurs, qui semblait coïncider avec la période de la rentrée des classes.

En somme, nous questionnons le recours aux annotateurs bénévoles dans ses conséquences relatives à la conception du modèle et à la production des annotations. Nous estimons ainsi que le modèle, les consignes d’annotation et le corpus finalement obtenu sont uniquement applicables aux textes de Short Edition, qui constituent en réalité un genre textuel à part entière¹, à mi-chemin entre les posts de blogs et l’exercice de style littéraire court, et très difficilement transposable à d’autres œuvres littéraires. On imagine mal se contenter de six ou sept émotions pour annoter les *Rougon-Macquart* de Zola, *L’Éducation Sentimentale* de Flaubert, ou même simplement, dans un style moins classique mais partageant l’usage du format court avec Short Edition, les œuvres d’Alice Munro.

7.2 Premier retour d’expérimentation : modifications du modèle

Les modifications apportées au modèle ont été prises après la première phase d’annotation (d’une durée d’environ trois semaines), principalement d’après les retours obtenus auprès des annotateurs et les observations que nous avons pu réaliser au cours de l’expérience.

1. Ce qui peut naturellement s’expliquer par l’effet communautaire du site.

7.2.1 Attribution de deux émotions avec relation de hiérarchie

L'un des paramètres sérieusement envisagé, et ce, depuis le début de la conception du modèle, était la mise en place d'un choix de deux émotions, impliquant une relation de hiérarchie entre les deux. En somme, les annotateurs auraient eu à choisir une émotion principale, et une seconde émotion moins présente dans le texte. Cette éventualité nous aurait permis de prendre en compte les textes difficiles à classer, comme c'est le cas par exemple des textes relatant des histoires d'amour qui se finissent mal. Dans le cas de ces textes, il est parfois difficile de distinguer quelle émotion de la tristesse ou de l'amour est prépondérante. Par convention, nous avons établi que dans ces cas l'amour primerait sur la tristesse. En effet, une histoire d'amour triste reste en premier lieu une histoire d'amour. La tristesse n'en est qu'une conséquence.

Cette possibilité a cependant été écartée, les annotateurs n'ayant pas manifesté un fort intérêt à son égard. De plus, nous avons constaté que lorsque deux émotions étaient de force similaire dans un texte, cela se ressentait fortement dans les annotations, la plupart du temps divisées assez équitablement entre elles. Nous avons donc estimé, une nouvelle fois, que le ratio bénéfice/complexité de cette mesure n'était pas suffisamment intéressant pour la mettre en place.

7.2.2 Refonte des émotions

Concernant le modèle en lui-même, nous avons effectué des modifications basées sur l'expérience et les retours des annotateurs et sur notre observation du processus d'annotation. Le modèle a donc subi des modifications relativement importantes.

Redéfinition de l'émotion "Affection"

En premier lieu, nous avons pu constater, aussi bien par le biais de l'interface d'annotation que par l'étude du corpus généré, que l'émotion "Affection" ne semblait pas toujours utilisée de façon pertinente par les annotateurs. Nous supposons que cela s'explique par le fait qu'ayant souhaité constituer une émotion "générique" rassemblant les idées de penchant et d'intérêt, l'idée même de l'émotion représentée en a pâti. Une définition trop vague, trop "fourre-tout", a laissé trop de places pour des interprétations trop différentes de la part des annotateurs. C'est ainsi que se sont trouvés annotés de la même façon des textes faisant le récit d'une histoire d'amour comme d'autres racontant une passion pour la musique ou la pêche à la ligne. Le corpus généré s'est ainsi révélé extrêmement hétérogène, et donc finalement peu exploitable.

En conséquence, nous avons scindé cette émotion en deux, la distinction s'opérant sur le caractère vivant ou non de l'objet de l'émotion. Nous avons d'un côté conservé l'appellation "Affection", destinée aux textes dont l'objet de l'émotion est vivant, et de l'autre, une nouvelle émotion intitulée "Intérêt" a été créée, destinée aux textes dont l'objet de l'émotion est non-vivant (objet matériel ou intellectuel). Ainsi, "Affection" permet d'annoter aussi bien les récits rapportant des histoires d'amour que des histoires d'amitié, en somme toutes sortes de rapport inter-personnels impliquant un attachement d'au moins une des parties.

Fusion des émotions "Sérénité" et "Joie"

À l'examen du corpus annoté selon les émotions, et tout en nous appuyant sur l'observation de l'interface d'annotation, nous avons constaté que la sérénité semblait mal perçue par les annotateurs, qui y avaient rarement recours, et majoritairement de façon isolée (souvent appliquée à des textes ne faisant pas consensus).

D'autre part, nous avons également remarqué que la joie, dans une moindre mesure, posait des problèmes similaires (bien moins prononcés en terme de consensus, cependant). Nous sommes alors revenus sur l'un de nos partis pris de départ, qui consistait à distinguer la joie et la sérénité d'après le critère de temporalité. Nous avons finalement estimé que ces deux émotions pouvaient être, cette fois, rassemblées en une seule. Dans un souci de généricité, et afin de ne pas troubler les annotateurs déjà impliqués dans le projet, cette nouvelle émotion a été intitulée "Bonheur", "Joie" nous semblant, de plus, impliquer une connotation d'intensité assez forte et d'activité qui ne correspondait pas toujours à ce que nous souhaitions².

Ajout de la Colère dans le modèle

Parmi les questions posées aux annotateurs au cours du bilan de la première phase d'annotation figurait celle de savoir s'ils avaient ressenti le manque d'une ou de plusieurs émotions en particulier. L'un d'entre eux nous a fait remarquer qu'à son sens, la colère manquait effectivement au modèle. Au début de la conception, nous avons écarté assez rapidement cette émotion, mais suite à cette remarque, cette décision a été reconsidérée. Il nous a semblé en effet que le corpus comportait suffisamment de textes évoquant la rage, la colère, ou une idée plus générale de violence pour intégrer cette émotion. Cela ne posait pas de problème en terme d'évolution du modèle, puisque nous restions ainsi dans les dimensions requises par Short Edition initialement, à savoir un modèle de sept à huit émotions. Nous avons donc établi une définition : il s'agit d'une émotion dont la cause est connue,

2. On pourrait nous faire remarquer que dans le cas de la modification de l'affection, nous n'avons pas pris la peine de proposer une appellation différente. Nous justifions ce choix par le fait qu'il s'agissait de retrancher une part du domaine de pertinence de l'affection, alors que dans le cas de la sérénité et de la joie, il s'agit de fusionner deux émotions ensemble. Il nous a donc paru important de renommer cette émotion, afin que cette modification ne suggère pas aux annotateurs que la sérénité était purement et simplement supprimée.

de type évènements (donc hors de la portée d'action du sujet), survenue dans le passé. Nous avons insisté dans la définition informelle, présente dans la fiche récapitulative sur le fait que nous souhaitions considérer toutes les manifestations de la colère, aussi bien rentrée, froide, que vive et exprimée.

Suite à ces modifications, nous avons repris chacun des documents d'annotation (guide, arbre de décision, fiche récapitulative), dont les nouvelles versions sont disponibles en annexe à la suite de la version initiale.

7.2.3 Modifications de l'option de ne pas annoter

L'un des reproches que nous avons également à l'encontre des annotateurs concerne le fait que certains, lorsqu'ils choisissaient de ne pas annoter le texte, négligeait la justification, alors que nous avons pourtant pris soin de leur expliquer son importance. Par exemple, il est arrivé que le champ soit rempli uniquement de points de suspension, ou encore d'un simple "Non", obscur au possible. Nous expliquons ce phénomène par une certaine lassitude, et par le fait que certains textes soumis, d'une qualité littéraire médiocre voir mauvaise, sont difficiles à apprécier et attaquent fortement le capital d'usure des annotateurs. Ceci étant, parmi les justifications exploitables que d'autres fournissaient, nous avons pu repérer des motifs revenant souvent. Par conséquent, afin de contrer ce phénomène et de faciliter la tâche des annotateurs, nous avons pris la décision de modifier l'interface et de prévoir un certain nombre de raisons prédéfinies de ne pas annoter. Nous avons également estimé que, dans l'optique d'un éventuel traitement automatique des justifications (par exemple, constitution de sous-corpus afin de dégager des traits caractéristiques), le tri des textes serait facilité puisque chaque raison implémentée se voit attribuer un numéro dans la base de données.

Les raisons pré-définies de ne pas annoter que nous avons implémentées sont au nombre de trois.

La première a été réclamée par plusieurs annotateurs : il s’agit du cas où le texte est un “simple constat” (expression reprise des annotateurs). Parmi les textes soumis, certains relèvent en effet purement de la description (description d’une ville, d’un paysage, d’un personnage, d’une œuvre d’art...). Si certaines émotions peuvent s’en dégager, elles sont en général ténues, mêlées, et bien entendu, rarement présentes dans le modèle. Face à la demande importante pour cette annotation, nous avons pris la décision de l’intégrer.

La deuxième a également été réclamée par les annotateurs, mais de façon moins “virulente” que la première. Il s’agit de prévoir le cas où un texte comporte un véritable mélange d’émotions de force équivalente, qu’il est difficile ou impossible de hiérarchiser. Nous avons donc ajouté un choix “Trop d’émotions”.

Enfin, la troisième et dernière possibilité est la plus simple : celle du cas où l’émotion du texte n’est pas comprise dans le modèle, et ne peut se rattacher à aucune autre (par exemple, la surprise).

Nous avons prévu un dernier choix, qui est celui où aucune des raisons proposées ne convient. Dans ce cas, l’annotateur trouve un champ dans lequel il est invité, de la même façon que dans le cadre de la première interface, à préciser une justification de façon concise.

Nouvelle interface d’annotation Suite à ces modifications, l’interface d’annotation a été reprise en conséquence, comme le montre la figure 7.1. De nouvelles

icônes ont été créés pour les émotions “Intérêt”, “Colère”, ainsi que pour les raisons prédéfinies de ne pas annoter. L’interface n’est cependant pas foncièrement remaniée.

Expérience Emotions bêta

Quelle est l'émotion que ressent le personnage principal ?

| | |
|---|--|
| <input type="radio"/> Bonheur Le sujet ressent du bonheur. Il est content, joyeux, plein d'enthousiasme. | <input type="radio"/> Spleen Le sujet est mal sans savoir précisément pourquoi. C'est le blues, la déprime, le moral dans les chaussettes, "l'angoisse d'exister" comme Baudelaire l'a défini. |
| <input type="radio"/> Affection Vaut pour toutes les formes d'amour, d'intérêt, de désir (le sentiment amoureux, amour filial, attachement, etc), la passion pour une activité ou un sujet. | <input type="radio"/> Tristesse Le désespoir, le chagrin, la peine, l'affliction. |
| <input type="radio"/> Intérêt Recoupe la passion, l'intérêt, l'attrait, la passion envers un objet non-vivant (un sport, une culture, un art, etc). | <input type="radio"/> Peur Tout ce qui relève de la crainte, l'appréhension, la terreur. |
| | <input type="radio"/> Colère Cette émotion doit être utilisée pour la colère (exprimée ou froide, rentrée), la rage, la fureur, etc. |

Je préfère ne pas choisir parce que...

| | | |
|--|---|---|
| <input type="radio"/> Constat Texte purement descriptif. | <input type="radio"/> Trop d'émotions Plusieurs émotions, difficile de trancher | <input type="radio"/> Émotion absente L'émotion qui conviendrait n'est pas dans la liste. |
|--|---|---|

Autre

Pourquoi ne préférez-vous pas choisir d'émotion ici ?

FIGURE 7.1 – Nouvelle interface d’annotation

Conclusion

Au cours de ce travail de recherche, nous avons réalisé un état de l’art, proposé un modèle, que nous avons amélioré après une première phase d’expérimentation, lancé un processus d’annotation et réussi à confirmer certains de nos partis pris de départ.

Les enjeux et les contraintes de ce projet étaient multiples et complexes : projet novateur ne pouvant donc pas s’appuyer entièrement sur des travaux antérieurs, recours à des annotateurs bénévoles ayant pour effet de limiter les possibilités d’action, pluridisciplinarité de la problématique.

Nous estimons avoir produit un modèle relativement satisfaisant mais perfectible, dont les applications sont à replacer dans le contexte du marché de l’édition numérique et du public visé. Comme nous l’avons évoqué au chapitre 7, ce modèle ne nous paraît pertinent que dans le cadre du programme de recherche de Short Edition. Il constitue néanmoins une base et son adaptation pour traiter des corpus littéraires plus “traditionnels” est envisageable.

Dans cette perspective, nous estimons que de nouveaux paramètres seraient à prendre en compte, notamment en ce qui concerne la granularité de l’annotation. Dans le cas de Short Edition, cette question s’est posée au moment de la mise en annotation, et dans un premier temps, nous avons décidé de traiter seulement les textes relevant du format *Très Très Court*. Si le projet venait à se transposer à

d'autres œuvres, il serait nécessaire par exemple de définir les fenêtres minimale et maximale à annoter. Dans le cas du seuil minimal, il est aisé de proposer la phrase ou le paragraphe comme unité. Mais dans le cas du seuil maximal, comment l'établir ? Il est bien entendu illusoire d'envisager l'attribution d'une seule émotion à un roman entier (même une combinaison de deux ou trois émotions paraît faible), cependant, le chapitre étant une unité au volume fort variable d'un auteur à un autre (certains n'y ayant même pas recours), il nous semble également difficile à adopter. Dans l'optique d'une généralisation et d'une utilisation plus large du modèle, il conviendrait donc de redéfinir les paramètres du modèle, voir même d'en fixer de nouveaux.

Bibliographie

- BOURION, E. (1995). Le réseau associatif de la peur. *In L'analyse thématique des données textuelles : l'exemple des sentiments.*
- CAMBRIA, E., LIVINGSTONE, A. et HUSSAIN, A. (2011). The hourglass of emotions. *In* ESPOSITO, A., ESPOSITO, A. M., VINCIARELLI, A., HOFFMANN, R. et MÜLLER, V. C., éditeurs : *COST 2102 Training School*, volume 7403 de *Lecture Notes in Computer Science*, pages 144–157. Springer.
- EENSOO, E. et VALETTE, M. (2013). Sémantique textuelle et TAL : un exemple d'application à l'analyse des sentiments. Article à paraître en septembre 2014 dans DOCUMENTS, TEXTES, ŒUVRES. Perspectives sémiotiques (Driss ABLALI, Sémir BADIR, Dominique DUCARD, éd.), Presses Universitaires de Rennes.
- FRAISSE, A. et PAROUBEK, P. (2014). Toward a unifying model for opinion, sentiment and emotion information extraction. *In* CHAIR), N. C. C., CHOUKRI, K., DECLERCK, T., LOFTSSON, H., MAEGAARD, B., MARIANI, J., MORENO, A., ODIJK, J. et PIPERIDIS, S., éditeurs : *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland. European Language Resources Association (ELRA).

- GORCY, G. (1995). *Euphorie et Spleen, deux thèmes antonymiques ? In L'analyse thématique des données textuelles : l'exemple des sentiments.*
- HEIDEN SERGE, MAGUÉ JEAN-PHILIPPE et PINCEMIN BÉNÉDICTE (2010). TXM : Une plateforme logicielle open-source pour la textométrie conception et développement. *In JADT 2010 : 10th International Conference on the Statistical Analysis of Textual Data*, page 12 p., Rome, Italie. ENS-Lyon.
- KAPPAS, A. (2013). Social regulation : messy layers. *Frontiers in Psychology.*
- MATHIEU, Y. Y. (2008). Navigation dans un texte à la recherche des sentiments. *Linguisticae Investigationes*, pages 313–322.
- PAK, A. (2012). *Analyse de sentiments automatique, adaptative et applicative.* Thèse, Université Paris Sud - Paris XI.
- PLUTCHIK, R. (2001). The nature of emotions. *American Scientist*, 89:344–350.
- SADEK, M. D. (1992). A study in the logic of intention. *In* NEBEL, B., RICH, C. et SWARTOUT, W. R., éditeurs : *KR*, pages 462–473. Morgan Kaufmann.

Annexes

.1 Matrice de contraste

La matrice de contraste des émotions se présente sous la forme d'un tableau à double entrée. Les paramètres de chaque émotion sont définis dans la diagonale, et les émotions sont contrastées entre elles aux intersections.

| E : événement A : action O : objet | | | | | | |
|------------------------------------|---|----------------------------------|---|---|--|---|
| | Joie | Sérénité | Affection | Tristesse | Spleen | Peur |
| Joie | Cause : - connue - E/A/O - désirée Temporalité : passée ou présente Intentions : satisfaites Valence : positive | | | | | |
| Sérénité | Cause : - connue - E/A/O - désirée Temporalité : passée Intentions : satisfaites Valence : positive | | | | | |
| Affection | Cause : type + Temporalité | Cause : type + Temporalité | Cause : - connu - objet - désiré Temporalité : présente Valence : positive | Cause : - connue - E/O - désirée Temporalité : passée et impossible dans le futur Intentions : insatisfaites Valence : négative | | |
| Tristesse | Temporalité | Intentions + Temporalité | Cause : type + Temporalité | | | |
| Spleen | Cause : connaissance | Cause : connaissance | Cause : connaissance | Cause : connaissance | Cause : - inconnue Temporalité : absente Valence : négative | |
| Peur | Cause : désir | Cause : désir | Cause : type + désir + Temporalité | Cause : désir | Cause : connaissance | Cause : - connue - E/A/O - non-désirée Temporalité : non-pertinente Valence : négative |

.2 Guide d'annotation

Le guide d'annotation présente les consignes, les émotions leurs définitions, et des exemples commentés tirés du corpus (version initiale et version après modification du modèle).

Guide d'annotations pour les émotions des textes de Short Edition

1 Projet

Ce qui nous intéresse, ce sont les émotions exprimées et vécues *dans le texte*. Il est important de ne pas les confondre avec celles que le lecteur ressent à la lecture du texte. On cherchera donc à caractériser les émotions entre les personnages, celles vécues par le narrateur, etc.

2 Critères de définition

Afin de classer les textes selon les émotions, nous avons établi des critères qui permettent de les définir et de les distinguer entre elles.

La *cause* correspond à ce qui est à l'origine de l'émotion. La cause peut être

- Connue : quand le sujet est capable de discerner la cause de son émotion
ou

- Inconnue : le sujet ne sait pas ce qui cause son émotion.

Il y a plusieurs types de causes :

- Les événements : dans le cas où le sujet subit, ou perçoit des choses qui ont lieu dans son environnement.

- Les actions : le sujet agit.

Enfin, si la cause représente un intérêt pour le sujet, s'il souhaite la voir se réaliser, on dira qu'elle est désirée. Sinon, si le sujet préférerait ne pas être confronté à la cause, on dira qu'elle est non-désirée.

La *temporalité de la cause* correspond au moment où la cause a eu lieu par rapport au moment où l'émotion est ressentie. Ce n'est pas le moment où le sujet ressent l'émotion.

La temporalité peut être :

- Absente ou non-pertinente : dans le cas où la cause est inconnue.

- Passée

- Présente : la cause a lieu simultanément avec le ressenti de l'émotion.

- Possible dans le futur ou seulement dans le futur

3 Les émotions

Les émotions choisies sont au nombre de six. Bien entendu, cela ne recouvre pas l'ensemble des émotions que l'être humain est capable de ressentir et d'exprimer. Il faut donc les considérer comme des sortes de "noms de catégorie". Elles peuvent être exprimées explicitement ou de façon plus subtile.

1. **Joie** : la cause de l'émotion est connue, cela peut être soit un événement, soit une action. Elle est désirée et a eu lieu dans le passé ou se déroule au moment de l'énonciation.

En résumé : Le sujet ressent du bonheur. Il est content, joyeux, plein d'enthousiasme.

Exemples :

- L'ermite <http://short-edition.com/oeuvre/tres-tres-court/l-ermite>

En lisant les quelques lignes de l'article, Pierre sourit. Il est heureux. Même si une partie de ses économies s'est envolée, il ne peut s'empêcher de croire que c'est la meilleure chose à faire... Il ne sait toujours pas ce que Madeleine était devenue mais malgré cela, il est

en paix avec lui-même. Et c'est avec bonheur qu'il regarde une tranche de sa vie, posée en évidence sur son vieux bahut.

Commentaire : la présence de termes comme "heureux", "sourit" permet de choisir la joie.

- Jour de chance <http://short-edition.com/oeuvre/tres-tres-court/jour-de-chance>
Joseph n'a jamais vraiment cru aux signes du destin, au karma ou à toutes ces considérations ésotériques dont lui parlent ses amis écrivains. Il est plutôt terre à terre, un peu trop même selon certains. Aujourd'hui pourtant c'est différent, une odeur de succès flotte dans l'air, Joseph est sûr de lui. Armand Colin et les nombreuses librairies du boulevard ne font qu'accentuer ce sentiment de bonheur qui lui tenaille les tripes, le succès!

Commentaire : ici on peut également relever des termes qui évoquent la joie : "succès", "bonheur".

- Nocturne <http://short-edition.com/oeuvre/tres-tres-court/nocturne-3>
Un magnifique soleil qui s'est levé. Quand on est sortis du bar, il faisait jour. Une petite troupe d'éclopés, de visages si pâles, d'odeur d'alcool et de tabac froid, de vêtements froissés et d'yeux hagards. Les rues sont vides, le bonheur de courir comme pour rattraper le soleil. Le soleil sur la place comme un Botticelli. Et moi qui tourne, tourne, pour voir ces nuages blancs avancer, encombrés d'oiseaux minuscules, noirs sur le bleu et blanc. Et la lumière éblouissante, jaune d'or, brûlante et froide de ce lever de soleil.

Commentaire : dans cet extrait, l'expression de l'émotion n'est pas aussi explicite que dans les deux extraits précédents. Il s'agit plus d'une atmosphère qui est posée grâce à la description.

- L'appel de la nature <http://short-edition.com/oeuvre/tres-tres-court/l-appel-de-la-nature>
Attentive à la redécouverte des émotions et impressions qui exhument mon corps des douleurs de la veille, je voudrais que ce moment d'éternité ne s'arrête jamais, mon exploration de somnambule ivre et assommée par la cacophonie de sensations se prolonge à l'infini... partagée entre l'exaltation, le ravissement, la frénésie, je suis soulagée et grisée, je bouge, mes jambes ont de l'énergie, je marche avec bonheur!

Commentaire : ici, les termes caractérisant la joie sont plus diversifiés : "exaltation", "ravissement", "frénésie", "grisée".

2. **Sérénité** : la cause de l'émotion est connue, cela peut être soit un évènement, soit une action. Elle est désirée et a eu lieu *strictement* dans le passé.

En résumé : Le sujet a réalisé quelque chose qui le satisfait, il est apaisé, détendu, calme, serein, relaxé.

Exemples :

- Le matin de ses treize ans <http://short-edition.com/oeuvre/tres-tres-court/le-matin-de-ses-treize-ans>
Oui car chaque homme, chaque personne est inscrit dans le souvenir du monde même s'il ne l'a pas sauvé! Alors au final, elle avait réussi, son rêve s'était réalisé : elle était éternelle. Enfin, elle avait trouvé la paix qui lui manquait tant, elle était comblée. Elle aurait aimé continuer son histoire mais à quoi bon, le destin avait fait son choix.

Commentaire : les expressions dénotant de la sérénité ici sont relatives à l'accomplissement et à l'apaisement : "elle avait réussi", "son rêve s'était réalisé", "paix", "comblée".

- Transmission <http://short-edition.com/oeuvre/tres-tres-court/transmission>
Amélie racontera demain la suite de l'histoire. En attendant le sommeil qui tarde à venir, l'enseignante pense à tous ces souvenirs ressurgis du passé. Sur le balcon du petit deux-pièces qu'elle occupe au vingtième étage d'une tour dominant la ville, elle observe les étoiles. Sa préférée, la plus proche de toutes, scintille très fort. Bravant l'éclairage urbain, l'étoile lui fait un clin d'œil de complicité.

Commentaire : ici également, la sérénité transparaît dans l'atmosphère décrite : le personnage se remémore des souvenirs, elle observe les étoiles.

3. **Affection** : la cause est connue, ce sont des évènements. Elle est désirée, et a eu lieu avant l'émotion ou simultanément avec celle-ci.

En résumé : Vaut pour toutes les formes d'amour, d'intérêt, de désir (le sentiment amoureux, amour filial, attachement, etc), la passion pour une activité ou un sujet.

Exemples :

- <http://short-edition.com/oeuvre/tres-tres-court/a-ma-mere>

Nous éclations ensemble du même rire, le bonheur nous portait, dans mes yeux tu pouvais lire tout l'amour que j'avais pour toi, toute l'admiration que je te vouais, je t'aimais à la folie Maman.

Commentaire : l'affection est explicite puisque l'on peut lire "l'amour que j'avais pour toi" et "je t'aimais à la folie Maman". L'utilisation du pronom "nous" traduit aussi la présence de deux personnages.

- L'attente <http://short-edition.com/oeuvre/tres-tres-court/l-attente-4>

Comme tous les vendredi matins, je ne pouvais plus dormir. Alors dès 5H30 je me suis levée et j'ai pris un long bain. Pour me détendre. Me détendre de quoi, je ne sais pas. Le vendredi est le meilleur jour de la semaine, je l'attends comme une lycéenne rêvant de son premier rendez-vous. Ça y ressemble un peu puisqu'il est l'amour de ma vie. Comme toujours, j'ai mis une robe. Je n'en ai pas mis pendant 25 ans et la première fois qu'il m'a vue habillée comme ça, il m'a fait des compliments.

Commentaire : en plus d'une mention explicite avec "l'amour de ma vie", les signes de l'affection sont les pronoms ("je" et "il"), ainsi que des expressions du champ lexical de l'affection telles que "premier rendez-vous", "compliments".

- Chassé-croisé <http://short-edition.com/oeuvre/tres-tres-court/chasse-croise>

Comme toutes les fois où il arrivait dans ce hall, il jeta un vague coup d'oeil au grand miroir et comme toutes les fois, les images d'un bonheur passé le frappèrent de plein fouet. Il appela l'ascenseur en faisant mine de s'intéresser à ce que racontait la copine de son frère qui habitait aussi dans l'immeuble. A l'intérieur de cette cage de ferraille, il se sentit mal. Un parfum qu'il ne connaissait que trop bien embaumait l'habitable, ça devait être encore son imagination, il avait souvent l'impression d'être près d'elle sans la voir. La fille descendit au troisième et il se retrouva seul. Comme d'habitude, en arrivant sur son palier, il imaginait qu'elle y serait. Il se l'était imaginé tellement souvent mais ce n'était jamais arrivé, même pas ce soir. Il rentra chez lui en ayant pris une ultime bouffée de cette odeur qui lui était chère, mais il faudrait qu'il arrête, elle ne reviendra jamais.

Commentaire : dans cet extrait l'expression de l'affection n'est pas explicite. Elle se déduit principalement des pronoms "il" (le personnage principal) et "elle".

4. **Tristesse** : la cause est connue, c'est un évènement. Elle est désirée et n'arrivera plus dans le futur ou non -désirée et a eu lieu dans le passé.

En résumé : la tristesse recouvre désespoir, chagrin, peine, affliction.

Exemples :

- La découverte <http://short-edition.com/oeuvre/tres-tres-court/la-decouverte>

Les larmes inondèrent une fois de plus son visage ravagé par la douleur. En avalant un cachet de Valium, Jenny se recoucha, tentant de calmer, encore une fois, son cœur inconsolable de maman orpheline.

Commentaire : la tristesse est exprimée explicitement, avec les "larmes", la "douleur", "cœur inconsolable".

- Carreau cassé <http://short-edition.com/oeuvre/tres-tres-court/carreau-casse>

Que font encore les mères dans ces moments-là ? Des fois elles regardent les jouets entassés pêle-mêle dans leur manque de petite fille, délaissés, eux aussi. Mais cela n'a pas d'importance, parce que de toutes les façons, il n'y a pas de place, de toutes façons, pour les ranger. Ce sont des soirs où il ne se passe rien... Des soirs où les lumières éteintes, on peut douter de soi, soudain s'oublier dans le spectre d'un miroir sans tain. Soudain, dans un de ces soirs-là, il n'y a plus rien, rien.

Commentaire : dans cet extrait, la tristesse n'est pas exprimée explicitement. Elle se déduit de l'ambiance décrite, à travers des termes et des expressions comme "délaissés", "pas de place", "rien", "lumières éteintes", "douter de soi", "plus rien".

- Un cadeau de cent mille euros <http://short-edition.com/oeuvre/tres-tres-court/un-cadeau-de-cent-mille-euros>

Madame m'a lâché la main et elle est morte. Aussi simplement que cela. Je m'y attendais mais j'ai été surprise finalement. A force, je pensais que ça n'arriverait pas, ou alors pas avant longtemps. (...) C'est la fille aînée que j'ai appelée. Elle n'était pas passée depuis huit jours et elle a eu l'air surpris, comme moi. (...) J'ai fait ce qu'elle m'a dit, j'ai appelé

le docteur. « *Ma pauvre Mariette, m'a-t-il dit, on se demande comment elle a duré si longtemps, plus rien ne fonctionnait* ». Pourtant, ai-je pensé, elle souriait, me parlait et elle aimait me tenir la main avant de s'endormir. Puis il a fait le nécessaire pour le reste ou ce sont les enfants... je n'en sais rien. Ça ne me concernait plus. C'est le soir que j'ai pleuré. Je suis allée dans la cuisine et, machinalement, j'ai préparé un potage pour elle. C'est là que le chagrin m'a prise.

Commentaire : dans cet extrait la tristesse est facilement identifiable puisqu'elle est exprimée explicitement, notamment avec l'utilisation de "chagrin", "pleuré".

5. **Spleen** : la cause de l'émotion n'est pas cernée précisément, on déclare donc qu'elle est inconnue. La temporalité de la cause ne peut donc pas être définie.

Il s'agit sans doute de l'émotion la moins aisée à pointer car elle repose rarement sur des caractéristiques explicites.

En résumé : Le sujet est mal sans savoir précisément pourquoi. C'est le blues, la déprime, le moral dans les chaussettes, "l'angoisse d'exister" comme Baudelaire l'a défini.

Exemples :

– Taxi sans retour <http://short-edition.com/oeuvre/tres-tres-court/taxi-sans-retour>
Les haut-parleurs hurlent dans l'habitable, Abattoir Blues, Nick Cave sème ses Mauvaises Graines! L'orage déluge un roulement d'éclairs. Tu pousses l'automobile noire dans la profondeur du fleuve en crue... Tu es parti alors. Personne ne sait où... tu ne sais pas si tu reviendras...

– Antonin <http://short-edition.com/oeuvre/tres-tres-court/antonin>
Il erre au milieu de nulle part. Une bulle vide qui s'élève pour mieux exploser. Des rêves lui courent dans les veines mais lui filent entre les doigts. Ses idées lui échappent, ses certitudes le fuient. Dans sa tête ça fuse, ça bouillonne, ça explose. Seul. Abandonné. Livré à lui-même.

Commentaire : l'ambiance noire, sordide, mélancolique, la solitude qui transparaît permettent d'annoter ces textes comme relevant du spleen.

6. **Peur** : la cause est connue, cela peut être un événement ou une action. Elle est non-désirée. La temporalité est passée car la peur est ressentie vis à vis de quelque chose dont on a eu l'expérience par le passé (c'est le principe de "Chat échaudé craint l'eau froide").

En résumé : Tout ce qui relève de la crainte, l'appréhension, la terreur.

Exemples :

– Terreur nocturne <http://short-edition.com/oeuvre/tres-tres-court/terreur-nocturne>
Et là, un sentiment qu'il commençait à connaître, la peur, le reprenait. Cette peur instinctive qui vous tordait le ventre, avec laquelle vos jambes tremblent sans même savoir pourquoi. Une peur si primale qu'il pleurait, les larmes dévalant ses joues glacées.

Commentaire : la peur est exprimée explicitement. Les autres signes sont les jambes qui tremblent et les larmes.

– Sans issue <http://short-edition.com/oeuvre/tres-tres-court/sans-issue-1>
Il s'arrêta un instant. Un infime instant. Juste le temps de faire taire la souffrance. Mais ce temps ne fut bien sûr pas suffisant. Les dents serrées, il reprit sa course claudicante et arriva enfin à atteindre le petit bosquet qui se trouvait là. Derrière, il entendait les aboiements des chiens lancés à ses trousses.

Commentaire : dans cet extrait, la peur n'est pas explicite. On peut la déduire de l'ambiance décrite, notamment par la présence de "la souffrance", "les dents serrées", "les chiens lancés à ses trousses".

– Une sale histoire d'argent <http://short-edition.com/oeuvre/tres-tres-court/une-sale-histoire-d-argent>

On est une demi-douzaine de clients. Je suis la seule femme. Tout le monde doit s'asseoir par terre. Un récalcitrant se prend un coup de pied dans le pif. Ça pisse le sang. Ça glace le nôtre. Efficace pour calmer les héros. Le pauvre bougre geint sous la douleur et s'en reprend un autre. Il bascule sur le côté. Il sanglote. La peur nous ligote les cordes vocales.

Commentaire : la peur est exprimée explicitement, mais d'autres caractéristiques la pointent : le sang glacé, le coup de pied, la douleur, les sanglots, les cordes vocales ligotées.

4 Consignes complémentaires

Le modèle ne cherche pas l'exhaustivité à tout prix. Il est donc possible de faire le choix de ne pas attribuer d'émotion à un texte. Dans ce cas, il vous sera demandé de justifier ce choix de façon concise. Cela nous permettra d'avoir un retour sur les textes trop délicats à catégoriser, de retenir les raisons qui font cela, et donc d'affiner le modèle pour une éventuelle prochaine version.

Il est également possible de changer l'émotion attribuée après la validation. Mais il n'y a pas de "bonnes" ou de "mauvaises" réponses.

Guide d'annotations pour les émotions des textes de Short Edition

1 Projet

Ce qui nous intéresse, ce sont les émotions exprimées et vécues *dans le texte*. Il est important de ne pas les confondre avec celles que le lecteur ressent à la lecture du texte. On cherchera donc à caractériser les émotions entre les personnages, celles vécues par le narrateur, etc.

2 Critères de définition

Afin de classer les textes selon les émotions, nous avons établi des critères qui permettent de les définir et de les distinguer entre elles.

La *cause* correspond à ce qui est à l'origine de l'émotion. La cause peut être

- Connue : quand le sujet est capable de discerner la cause de son émotion
ou

- Inconnue : le sujet ne sait pas ce qui cause son émotion.

Il y a plusieurs types de causes :

- Les événements : dans le cas où le sujet subit, ou perçoit des choses qui ont lieu dans son environnement.

- Les actions : le sujet agit.

Enfin, si la cause représente un intérêt pour le sujet, s'il souhaite la voir se réaliser, on dira qu'elle est désirée. Sinon, si le sujet préférerait ne pas être confronté à la cause, on dira qu'elle est non-désirée.

La *temporalité de la cause* correspond au moment où la cause a eu lieu par rapport au moment où l'émotion est ressentie. Ce n'est pas le moment où le sujet ressent l'émotion.

La temporalité peut être :

- Absente ou non-pertinente : dans le cas où la cause est inconnue.

- Passée

- Présente : la cause a lieu simultanément avec le ressenti de l'émotion.

- Possible dans le futur ou seulement dans le futur

3 Les émotions

Les émotions choisies sont au nombre de six. Bien entendu, cela ne recouvre pas l'ensemble des émotions que l'être humain est capable de ressentir et d'exprimer. Il faut donc les considérer comme des sortes de "noms de catégorie". Elles peuvent être exprimées explicitement ou de façon plus subtile.

1. **Bonheur** : la cause de l'émotion est connue, cela peut être soit un événement, soit une action. Elle est désirée et a eu lieu dans le passé ou se déroule au moment de l'énonciation.

En résumé : Le sujet ressent du bonheur. Il est content, joyeux, plein d'enthousiasme.

Exemples :

- L'ermite <http://short-edition.com/oeuvre/tres-tres-court/l-ermite>

En lisant les quelques lignes de l'article, Pierre sourit. Il est heureux. Même si une partie de ses économies s'est envolée, il ne peut s'empêcher de croire que c'est la meilleure chose à faire... Il ne sait toujours pas ce que Madeleine était devenue mais malgré cela, il est

en paix avec lui-même. Et c'est avec bonheur qu'il regarde une tranche de sa vie, posée en évidence sur son vieux bahut.

Commentaire : la présence de termes comme “heureux”, “sourit” permet de choisir la joie.

- Jour de chance <http://short-edition.com/oeuvre/tres-tres-court/jour-de-chance>
Joseph n'a jamais vraiment cru aux signes du destin, au karma ou à toutes ces considérations ésotériques dont lui parlent ses amis écrivains. Il est plutôt terre à terre, un peu trop même selon certains. Aujourd'hui pourtant c'est différent, une odeur de succès flotte dans l'air, Joseph est sûr de lui. Armand Colin et les nombreuses librairies du boulevard ne font qu'accentuer ce sentiment de bonheur qui lui tenaille les tripes, le succès!

Commentaire : ici on peut également relever des termes qui évoquent la joie : “succès”, “bonheur”.

- Nocturne <http://short-edition.com/oeuvre/tres-tres-court/nocturne-3>
Un magnifique soleil qui s'est levé. Quand on est sortis du bar, il faisait jour. Une petite troupe d'éclopés, de visages si pâles, d'odeur d'alcool et de tabac froid, de vêtements froissés et d'yeux hagards. Les rues sont vides, le bonheur de courir comme pour rattraper le soleil. Le soleil sur la place comme un Botticelli. Et moi qui tourne, tourne, pour voir ces nuages blancs avancer, encombrés d'oiseaux minuscules, noirs sur le bleu et blanc. Et la lumière éblouissante, jaune d'or, brûlante et froide de ce lever de soleil.

Commentaire : dans cet extrait, l'expression de l'émotion n'est pas aussi explicite que dans les deux extraits précédents. Il s'agit plus d'une atmosphère qui est posée grâce à la description.

- L'appel de la nature <http://short-edition.com/oeuvre/tres-tres-court/l-appel-de-la-nature>
Attentive à la redécouverte des émotions et impressions qui exhument mon corps des douleurs de la veille, je voudrais que ce moment d'éternité ne s'arrête jamais, mon exploration de somnambule ivre et assommée par la cacophonie de sensations se prolonge à l'infini... partagée entre l'exaltation, le ravissement, la frénésie, je suis soulagée et grisée, je bouge, mes jambes ont de l'énergie, je marche avec bonheur!

Commentaire : ici, les termes caractérisant la joie sont plus diversifiés : “exaltation”, “ravissement”, “frénésie”, “grisée”.

2. **Affection** : la cause est connue, ce sont des événements. Elle est désirée, et a eu lieu avant l'émotion ou simultanément avec celle-ci.

En résumé : Vaut pour toutes les formes d'amour, de désir, d'amitié, d'admiration (le sentiment amoureux, amour filial, attachement, etc), envers **un être vivant**.

Attention : lorsque le texte relate une histoire d'amour triste, c'est l'amour qui a la priorité sur la tristesse.

Exemples :

- <http://short-edition.com/oeuvre/tres-tres-court/a-ma-mere>
Nous éclatons ensemble du même rire, le bonheur nous portait, dans mes yeux tu pouvais lire tout l'amour que j'avais pour toi, toute l'admiration que je te vouais, je t'aimais à la folie Maman.

Commentaire : l'affection est explicite puisque l'on peut lire “l'amour que j'avais pour toi” et “je t'aimais à la folie Maman”. L'utilisation du pronom “nous” traduit aussi la présence de deux personnages.

- L'attente <http://short-edition.com/oeuvre/tres-tres-court/l-attente-4>
Comme tous les vendredi matins, je ne pouvais plus dormir. Alors dès 5H30 je me suis levée et j'ai pris un long bain. Pour me détendre. Me détendre de quoi, je ne sais pas. Le vendredi est le meilleur jour de la semaine, je l'attends comme une lycéenne rêvant de son premier rendez-vous. Ça y ressemble un peu puisqu'il est l'amour de ma vie. Comme toujours, j'ai mis une robe. Je n'en ai pas mis pendant 25 ans et la première fois qu'il m'a vue habillée comme ça, il m'a fait des compliments.

Commentaire : en plus d'une mention explicite avec “l'amour de ma vie”, les signes de l'affection sont les pronoms (“je” et “il”), ainsi que des expressions du champ lexical de l'affection telles que “premier rendez-vous”, “compliments”.

- Chassé-croisé <http://short-edition.com/oeuvre/tres-tres-court/chasse-croise>
Comme toutes les fois où il arrivait dans ce hall, il jeta un vague coup d'oeil au grand miroir et comme toutes les fois, les images d'un bonheur passé le frappèrent de plein fouet.

Il appela l'ascenseur en faisant mine de s'intéresser à ce que racontait la copine de son frère qui habitait aussi dans l'immeuble. A l'intérieur de cette cage de ferraille, il se sentit mal. Un parfum qu'il ne connaissait que trop bien embaumait l'habitable, ça devait être encore son imagination, il avait souvent l'impression d'être près d'elle sans la voir. La fille descendit au troisième et il se retrouva seul. Comme d'habitude, en arrivant sur son palier, il imaginait qu'elle y serait. Il se l'était imaginé tellement souvent mais ce n'était jamais arrivé, même pas ce soir. Il rentra chez lui en ayant pris une ultime bouffée de cette odeur qui lui était chère, mais il faudrait qu'il arrête, elle ne reviendra jamais.

Commentaire : dans cet extrait l'expression de l'affection n'est pas explicite. Elle se déduit principalement des pronoms "il" (le personnage principal) et "elle".

3. **Intérêt** : la cause est connue, ce sont des événements. Elle est désirée, et a eu lieu avant l'émotion ou simultanément avec celle-ci.

En résumé : Recoupe la passion, l'intérêt, l'attrait, la passion envers **un objet non-vivant** (un sport, une culture, un art, etc).

Exemples :

- À l'unisson <http://short-edition.com/oeuvre/tres-tres-court/a-l-unisson-1>

Il en rêvait depuis si longtemps ! Apprendre à jouer de la musique, pouvoir, grâce à un instrument traduire des émotions, des sentiments, faire valser ses doigts sur les touches comme autant de lutins dansants qui s'animent sous ses yeux et déversent leurs notes enflammées...

Commentaire : dans cet extrait, le sujet ressent un fort intérêt pour la musique, il rêve d'apprendre à jouer.

- Les champignons : <http://short-edition.com/oeuvre/tres-tres-court/les-champignons>

Mon oncle Albert, le chasseur, avait deux passions : la chasse et la recherche des champignons.[...] Pour récolter un champignon, faut lui faire la cour. Il faut y mettre de la tendresse, du sentiment. Avant de lui demander son nom, faut le regarder... Sans trop d'insistance... pour ne pas te laisser emporter par l'ivresse de la rencontre.

Commentaire : le personnage décrit par le narrateur est passionné par la cueillette des champignons. On peut noter que dans ce texte, le narrateur décrit un personnage, et ce sont donc les émotions de ce personnage qui nous intéressent, puisque ce sont celles qui sont le plus présentes dans le texte.

4. **Tristesse** : la cause est connue, c'est un événement. Elle est désirée et n'arrivera plus dans le futur ou non -désirée et a eu lieu dans le passé.

En résumé : la tristesse recouvre désespoir, chagrin, peine, affliction.

Exemples :

- La découverte <http://short-edition.com/oeuvre/tres-tres-court/la-decouverte>

Les larmes inondèrent une fois de plus son visage ravagé par la douleur. En avalant un cachet de Valium, Jenny se recoucha, tentant de calmer, encore une fois, son cœur inconsolable de maman orpheline.

Commentaire : la tristesse est exprimée explicitement, avec les "larmes", la "douleur", "cœur inconsolable".

- Carreau cassé <http://short-edition.com/oeuvre/tres-tres-court/carreau-casse>

Que font encore les mères dans ces moments-là ? Des fois elles regardent les jouets entassés pêle-mêle dans leur manque de petite fille, délaissés, eux aussi. Mais cela n'a pas d'importance, parce que de toutes les façons, il n'y a pas de place, de toutes façons, pour les ranger. Ce sont des soirs où il ne se passe rien... Des soirs où les lumières éteintes, on peut douter de soi, soudain s'oublier dans le spectre d'un miroir sans tain. Soudain, dans un de ces soirs-là, il n'y a plus rien, rien.

Commentaire : dans cet extrait, la tristesse n'est pas exprimée explicitement. Elle se déduit de l'ambiance décrite, à travers des termes et des expressions comme "délaissés", "pas de place", "rien", "lumières éteintes", "douter de soi", "plus rien".

- Un cadeau de cent mille euros <http://short-edition.com/oeuvre/tres-tres-court/un-cadeau-de-cent-mille-euros>

Madame m'a lâché la main et elle est morte. Aussi simplement que cela. Je m'y attendais mais j'ai été surprise finalement. A force, je pensais que ça n'arriverait pas, ou alors pas

avant longtemps. (...) C'est la fille aînée que j'ai appelée. Elle n'était pas passée depuis huit jours et elle a eu l'air surpris, comme moi. (...) J'ai fait ce qu'elle m'a dit, j'ai appelé le docteur. « Ma pauvre Mariette, m'a-t-il dit, on se demande comment elle a duré si longtemps, plus rien ne fonctionnait ». Pourtant, ai-je pensé, elle souriait, me parlait et elle aimait me tenir la main avant de s'endormir. Puis il a fait le nécessaire pour le reste ou ce sont les enfants... je n'en sais rien. Ça ne me concernait plus. C'est le soir que j'ai pleuré. Je suis allée dans la cuisine et, machinalement, j'ai préparé un potage pour elle. C'est là que le chagrin m'a prise.

Commentaire : dans cet extrait la tristesse est facilement identifiable puisqu'elle est exprimée explicitement, notamment avec l'utilisation de "chagrin", "pleuré".

5. **Spleen** : la cause de l'émotion n'est pas cernée précisément, on déclare donc qu'elle est inconnue. La temporalité de la cause ne peut donc pas être définie.

Il s'agit sans doute de l'émotion la moins aisée à pointer car elle repose rarement sur des caractéristiques explicites.

En résumé : Le sujet est mal sans savoir précisément pourquoi. C'est le blues, la déprime, le moral dans les chaussettes, "l'angoisse d'exister" comme Baudelaire l'a défini.

Exemples :

– Taxi sans retour <http://short-edition.com/oeuvre/tres-tres-court/taxi-sans-retour>
Les haut-parleurs hurlent dans l'habitacle, Abattoir Blues, Nick Cave sème ses Mauvaises Graines! L'orage déluge un roulement d'éclairs. Tu pousses l'automobile noire dans la profondeur du fleuve en crue... Tu es parti alors. Personne ne sait où... tu ne sais pas si tu reviendras...

– Antonin <http://short-edition.com/oeuvre/tres-tres-court/antonin>
Il erre au milieu de nulle part. Une bulle vide qui s'élève pour mieux exploser. Des rêves lui courent dans les veines mais lui filent entre les doigts. Ses idées lui échappent, ses certitudes le fuient. Dans sa tête ça fuse, ça bouillonne, ça explose. Seul. Abandonné. Livré à lui-même.

Commentaire : l'ambiance noire, sordide, mélancolique, la solitude qui transparaît permettent d'annoter ces textes comme relevant du spleen.

6. **Peur** : la cause est connue, cela peut être un événement ou une action. Elle est non-désirée. La temporalité est passée car la peur est ressentie vis à vis de quelque chose dont on a eu l'expérience par le passé (c'est le principe de "Chat échaudé craint l'eau froide").

En résumé : Tout ce qui relève de la crainte, l'appréhension, la terreur.

Exemples :

– Terreur nocturne <http://short-edition.com/oeuvre/tres-tres-court/terreur-nocturne>
Et là, un sentiment qu'il commençait à connaître, la peur, le reprenait. Cette peur instinctive qui vous tordait le ventre, avec laquelle vos jambes tremblent sans même savoir pourquoi. Une peur si primale qu'il pleurait, les larmes dévalant ses joues glacées.

Commentaire : la peur est exprimée explicitement. Les autres signes sont les jambes qui tremblent et les larmes.

– Sans issue <http://short-edition.com/oeuvre/tres-tres-court/sans-issue-1>
Il s'arrêta un instant. Un infime instant. Juste le temps de faire taire la souffrance. Mais ce temps ne fut bien sûr pas suffisant. Les dents serrées, il reprit sa course claudicante et arriva enfin à atteindre le petit bosquet qui se trouvait là. Derrière, il entendait les aboiements des chiens lancés à ses trousses.

Commentaire : dans cet extrait, la peur n'est pas explicite. On peut la déduire de l'ambiance décrite, notamment par la présence de "la souffrance", "les dents serrées", "les chiens lancés à ses trousses".

– Une sale histoire d'argent <http://short-edition.com/oeuvre/tres-tres-court/une-sale-histoire-d-argent>

On est une demi-douzaine de clients. Je suis la seule femme. Tout le monde doit s'asseoir par terre. Un récalcitrant se prend un coup de pied dans le pif. Ça pisse le sang. Ça glace le nôtre. Efficace pour calmer les héros. Le pauvre bougre geint sous la douleur et s'en reprend un autre. Il bascule sur le côté. Il sanglote. La peur nous ligote les cordes vocales.

Commentaire : la peur est exprimée explicitement, mais d'autres caractéristiques la pointent : le sang glacé, le coup de pied, la douleur, les sanglots, les cordes vocales ligotées.

7. **Colère** : la cause de l'émotion est connue, c'est un événement non-désiré qui a eu lieu dans le passé. Le sujet est actif.

En résumé : cette émotion doit être utilisée pour la colère (exprimée ou froide, rentrée), la rage, la fureur, etc.

Exemples :

- Cocotte <http://short-edition.com/oeuvre/tres-tres-court/cocotte>

A mesure que la cocotte-minute chauffait, une colère sourde montait en moi. Des brames stupides m'arrivaient de l'autre pièce, il devait encore être devant un match de foot. La cocotte s'est mise à siffler. Je l'ai empoignée et je me suis dirigée vers le salon. Il était là, échoué, ses yeux exorbités fixés sur l'écran. Il a fini par remarquer ma présence, et son oeil hagard s'est arrêté sur la cocotte. Là, il a ouvert la bouche et il a dit « On bouffe quoi ? ». C'est parti tout seul, la cocotte bouillante a fracassé son crâne d'un coup.

Commentaire : ici le personnage ressent une "colère sourde" qui se manifeste par une réaction violente.

- L'œil au beurre noir <http://short-edition.com/oeuvre/tres-tres-court/l-oeil-au-beurre-noir>
Chez moi, la révolte a pris la forme du silence et de la rumination. Les muscles aux aguets, je mastique ma colère. Je sais que je vais exploser, dans peu de temps. Plus personne ne pourra alors me retenir ni contenir ce que je garde en moi depuis quelques semaines déjà, mimant la servitude et l'acquiescement à l'autorité abusive.

Commentaire : ici aussi le personnage contient sa colère, mais sent qu'il va exploser bientôt.

4 Consignes complémentaires

Le modèle ne cherche pas l'exhaustivité à tout prix. Il est donc possible de faire le choix de ne pas attribuer d'émotion à un texte. Dans ce cas, il vous sera demandé de justifier ce choix de façon concise. Cela nous permettra d'avoir un retour sur les textes trop délicats à catégoriser, de retenir les raisons qui font cela, et donc d'affiner le modèle pour une éventuelle prochaine version.

Il est également possible de changer l'émotion attribuée après la validation. Mais il n'y a pas de "bonnes" ou de "mauvaises" réponses.

.3 Fiche récapitulative

La fiche récapitulative est une version plus condensée et conviviale du guide d'annotation, qui en reprend les informations essentielles (version initiale et version après modification du modèle).

RÉCAPITULATIF

POUR L'ANNOTATION DES ÉMOTIONS

Voici en quelques mots l'essentiel pour être en mesure d'annoter les émotions dans les textes. Vous pouvez conserver cette fiche sous la main et la consulter dès que vous avez un doute !

LES ÉMOTIONS



Joie

Le sujet ressent du bonheur. Il est content, joyeux, plein d'enthousiasme.

Cause connue de type évènement ou action, désirée par le sujet, survenue dans le passé ou simultanée.



Sérénité

Le sujet a réalisé quelque chose qui le satisfait, il est apaisé, détendu, calme, serein, relaxé.

Cause connue de type évènement ou action, désirée, survenue dans le passé et satisfaisant les intentions du sujet.



Peur

Tout ce qui relève de la crainte, l'appréhension, la terreur.

Cause connue non-désirée de type évènement ou action dont le sujet a eu l'expérience par le passé.



Tristesse

Le désespoir, le chagrin, la peine, l'affliction.

Cause connue de type évènement, qui peut être désirée et n'arrivera plus dans le futur ou non-désirée et a eu lieu dans le passé.



Spleen

Le sujet est mal sans savoir précisément pourquoi. C'est le blues, la déprime, le moral dans les chaussettes, « l'angoisse d'exister » comme Baudelaire l'a défini.

Cause inconnue.



Affection

Vaut pour toutes les formes d'amour, d'intérêt, de désir (le sentiment amoureux, amour filial, attachement...), la passion pour une activité ou un sujet.

Cause connue de type évènement, désirée, survenue dans le passé ou simultanée.

COMMENT CHOISIR L'ÉMOTION À ATTRIBUER ?

- On cherche les émotions dans le texte, pas celles ressenties par le lecteur.
- On peut se poser des questions : qui ressent l'émotion ? Qu'est-ce qui déclenche, qu'est-ce qui est à l'origine de l'émotion ?
- Quand plusieurs émotions sont exprimées : on choisit celle ressentie par le narrateur ou le personnage principal, ou on choisit celle qui est majoritairement présente dans le texte (soit la plus exprimée, soit celle exprimée dans la première moitié du texte).
- Quand un texte comporte des émotions qui ne sont pas dans le modèle ou qu'il n'exprime pas d'émotion, on choisit de ne pas annoter et on attribue une raison à ce choix.

Le modèle n'est pas complet, il ne comporte pas toutes les émotions possibles. Il est normal de ne pas toujours parvenir à faire un choix, ou d'en faire un qui sera différent de celui des autres participants. Il n'y a pas de bonne ou de mauvaise réponse.

N'hésitez pas à poser des questions si vous avez un doute.
contact@short-edition.com



RÉCAPITULATIF

POUR L'ANNOTATION DES ÉMOTIONS

Voici en quelques mots l'essentiel pour être en mesure d'annoter les émotions dans les textes. Vous pouvez conserver cette fiche sous la main et la consulter dès que vous avez un doute !

LES ÉMOTIONS



Tristesse

Le désespoir, le chagrin, la peine, l'affliction.

Cause connue de type évènement, qui peut être désirée et n'arrivera plus dans le futur ou non-désirée et a eu lieu dans le passé.



Spleen

Le sujet est mal sans savoir précisément pourquoi. C'est le blues, la déprime, le moral dans les chaussettes, « l'angoisse d'exister » comme Baudelaire l'a défini.

Cause inconnue.



Peur

Tout ce qui relève de la crainte, l'appréhension, la terreur.

Cause connue non-désirée de type évènement ou action dont le sujet a eu l'expérience par le passé.



Colère

Cette émotion doit être utilisée pour la colère (exprimée ou froide, rentrée), la rage, la fureur...

Cause connue de type évènement, non-désirée, survenue dans le passé.



Affection

Vaut pour toutes les formes d'amour, d'intérêt, de désir (le sentiment amoureux, amour filial, attachement...), la passion pour une activité ou un sujet, envers un être vivant.

Cause connue de type évènement, relative à un être vivant, désirée, survenue dans le passé ou simultanée.



Intérêt

Recoupe la passion, l'intérêt, l'attrait, la passion envers un objet non-vivant (un sport, une culture, un art...).

Cause connue de type évènement, relative à un objet inanimé, désirée, survenue dans le passé ou simultanée.



Bonheur

Le sujet ressent du bonheur. Il est content, joyeux, plein d'enthousiasme.

Cause connue de type évènement ou action, désirée par le sujet, survenue dans le passé ou simultanée.

COMMENT CHOISIR L'ÉMOTION À ATTRIBUER ?

■ On cherche les émotions dans le texte, pas celles ressenties par le lecteur.

■ On peut se poser des questions : qui ressent l'émotion ? Qu'est-ce qui déclenche, qu'est-ce qui est à l'origine de l'émotion ?

■ Quand plusieurs émotions sont exprimées : on choisit celle ressentie par le narrateur ou le personnage principal, ou on choisit celle qui est majoritairement présente dans le texte (soit la plus exprimée, soit celle exprimée dans la première moitié du texte).

■ Quand un texte comporte des émotions qui ne sont pas dans le modèle ou qu'il n'exprime pas d'émotion, on choisit de ne pas annoter et on attribue une raison à ce choix.

Le modèle n'est pas complet, il ne comporte pas toutes les émotions possibles. Il est normal de ne pas toujours parvenir à faire un choix, ou d'en faire un qui sera différent de celui des autres participants. Il n'y a pas de bonne ou de mauvaise réponse.

.4 Arbre de décision

L'arbre de décision est constitué de questions fermées permettant aux annotateurs de choisir une émotion avec le moins d'ambiguïté possible (version initiale et version après modification du modèle).

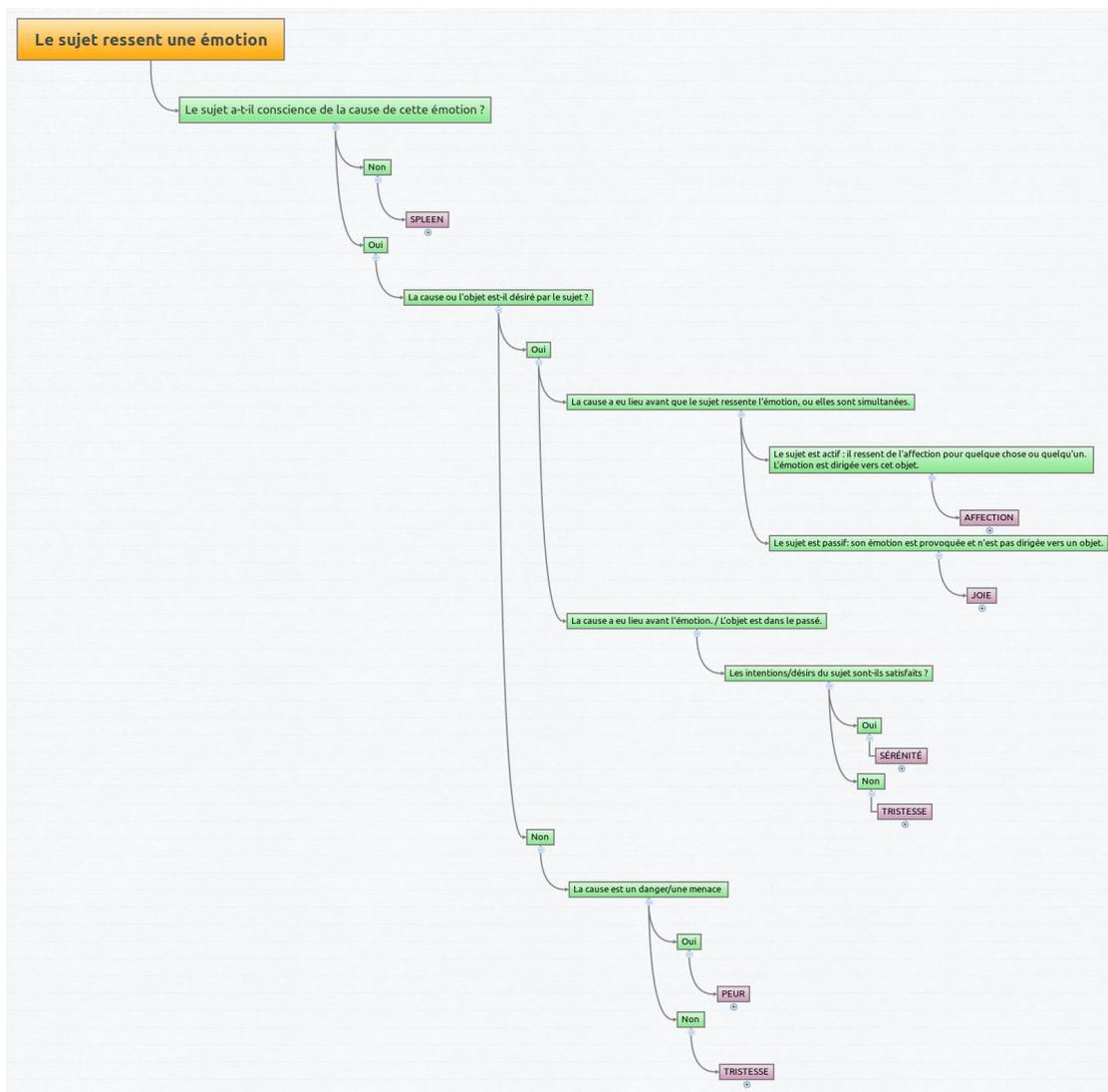


FIGURE 2 – Première version

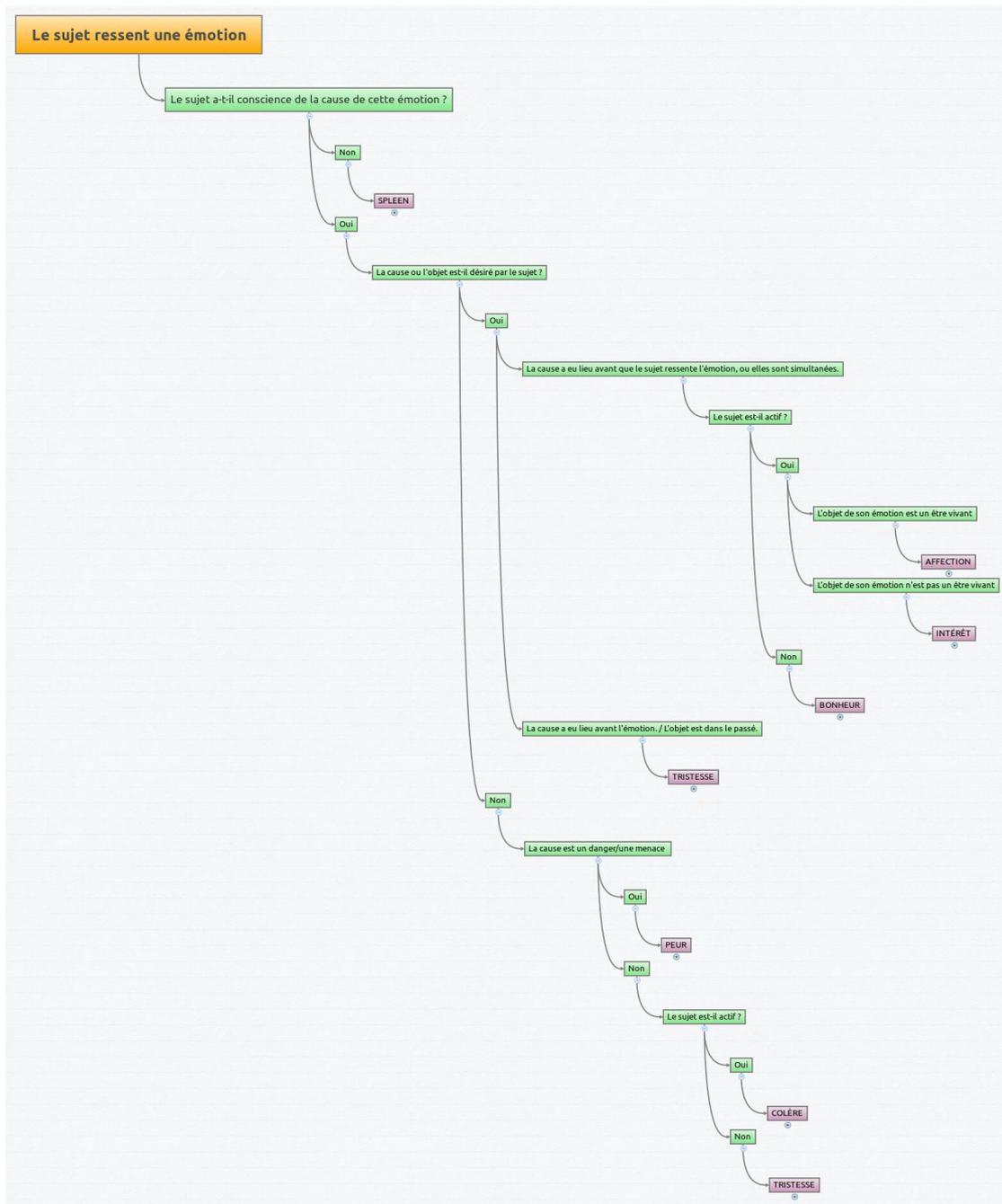


FIGURE 3 – Seconde version

.5 Scripts

.5.1 extract.R

Script d'extraction du contenu textuel du corpus au format XML.

```
#Extraction des textes a l'aide de la structure XML

require(XML)
setwd('/people/gianola/CORPUS_SH/output/all/nvl/')
doc = xmlTreeParse("/people/gianola/CORPUS_SH/output/posts.xml", useInternal=TRUE)
top = xmlRoot(doc)
posts <- xpathApply(top, "//post[contains(category, 'nouvelles')]/content", xmlValue)
# id <- xpathApply(top, "//post[contains(category, 'nouvelles')]/id", xmlValue)
i <- 1
# for (i in 1:length(posts))
for (i in 1:100){
  id <- xpathApply(top, "//post[contains(category, 'nouvelles')]/id", xmlValue)
  id <- id[[i]]
  filename <- paste(id, "txt", sep=".") #concatenation du nom de fichier
  sink(filename, append = TRUE)
#   Extraction du texte filtre par categorie (nouvelles, tres-tres-court, poesie)
  posts <- xpathApply(top, "//post[contains(category, 'nouvelles')]/content", xmlValue)
  texte <- posts[[i]]
#   print(id)
  print(texte) #impression de l'extrait
#   save(texte, file = '/people/gianola/CORPUS_SH/output/texte.RData')
  sink()
}
```

.5.2 normalization.R

Script de nettoyage et normalisation du contenu textuel précédemment extrait du corpus.

```
srcdir <- "/corpus/nvl"

trgtdir <- "/corpus/nvl_normalized"

files <- list.files( srcdir )

setwd( srcdir )
for ( i in 1:length(files) )
{
# ----- set input dir -----
  texte <- readLines( files [ i ], encoding="UTF-8" )
  texte <- tolower(texte)
  texte <- gsub("<.*?>", "", texte)
  texte <- gsub("]]>", "", texte)
  # # ----- handle punctuation marks -----
  texte <- gsub( "^\\ ", "", texte )
  texte <- gsub( '\\U0153', "oe", texte )
  texte <- gsub( '\\U2019', "' ", texte )
  texte <- gsub( "'", "' ", texte )
  texte <- gsub( "-", " - ", texte )
  # added by pap for long hyphens
  texte <- gsub( '\\U2014', " - ", texte )
  # for non breaking spaces
  texte <- gsub( '\\U00A0', " ", texte )
  texte <- gsub( "\\\"", " \\" , texte )
  texte <- gsub( ",", " , ", texte )
  texte <- gsub( "!!", "! !", texte )
  texte <- gsub( "?!", "? !", texte , fixed =TRUE)
  texte <- gsub( "\\.", " .", texte )
  texte <- gsub( " \\ . \\ . \\ .", " ... ", texte , fixed = TRUE)
  # remove any occurrence of multiple white spaces
  texte <- gsub( " +", " ", texte)

# ----- output -----
#setwd( trgtdir )
#cat(texte , file=files [ i ], sep=" ", append = TRUE)
```

```

# for UTF-8 output
# outfile <- file(paste( trgtdir , toString( files [i] ), sep="/"), "wb",
# for latin output (required by script do_Flemm.sh
  outfile <- file(paste( trgtdir , toString( files [i] ), sep="/"), "wb",

#   Encoding( texte ) <- "UTF-8"

#   cat(texte , file=outfile , sep=" ", append=TRUE)

  final_text <- ""
  for( l in 1:length(texte) )
  {
    if( l != 0)
    {
      final_text <- paste( final_text , texte [ l ] )
    }
    else
    {
      final_text <- paste( final_text , " " )
    }
  }

  final_text <- gsub( "\\ +", " ", final_text )
  final_text <- gsub( "^\\ ", "", final_text )

  writeBin( charToRaw( iconv( final_text , from="UTF-8", to="latin1" ) )

close( outfile )
}

```

.5.3 cleaning.sh

Script de nettoyage des données extraites de la base.

```

#!/usr/bin/bash

sed 's/\\t//g;s/\\n//g;s/<p>//g;s/<\\p>//g;s/<br \\/>//g;s/<em>//g;s/<\\em>//g;s/ /g' < ./rsc/reviews_content.csv > ./rsc/clean_rc.csv;

```

.5.4 sorting.pl

Script réalisant la classification des textes selon les émotions attribuées par les annotateurs.

```
#!/usr/bin/perl
use strict;
use warnings;
use Data::Dumper;
use List::Util qw/max/;

# Modele de fichier a fournir : csv au format
# id;title;COUNT(r.id);GROUP_CONCAT(r.emotion_id);COUNT(r.
  no_emotion_comment)
# avec GROUP_CONCAT(r.emotion_id) au format 0,0,0

# my $file = $ARGV[0] or die "Need to get CSV file on the
  command line\n";

open(FILE, '<', './rsc/clean_rc.csv");
open(OUT, '>', './rsc/sorted_reviews.csv");
print OUT "id#title#emotion#content\n";

while (my $line = <FILE>){
  chomp $line;
  # Ignore la ligne d'en-tete
  next if $line =~ /^id/;

  # Recuperation des valeurs apres split
  my @fields = split "#", $line;
  my $tags_total = $fields[2];
  my $no_emotion = $fields[4];
  my $text_id = $fields[0];
  my $title = $fields[1];
  my $content = $fields[5];

  # Recuperation de l'id de chaque emotion (split sur le
    champ GROUP_CONCAT(r.emotion_id)
  my @emotion_id = split ",", $fields[3];

  # Recuperation du champ COUNT(r.no_emotion_comment)
  my %emotions = ( "no_emotion" => $no_emotion);
```

```

# Compteur qui avance dans le tableau des emotions
my $i = 0;

# Initialisation des compteurs pour chaque emotion
my $joie = 0;
my $affection = 0;
my $serenite = 0;
my $tristesse = 0;
my $peur = 0;
my $spleen = 0;

print "$title_\n";
print OUT "$text_id#$title#";

# Si le texte n'a ete annote que deux fois, on va
  verifier si les deux annotations sont concordantes
if ($tags_total == 2){
  print "_!!_Seulement_deux_annotations_!!_";

  # Si seulement un des annotateurs a choisi "Pas d'
    emotion"
  if ($no_emotion == 1){
    print OUT "no_consensus#$content\n";
    print "Pas_de_consensus...\n";
  }

  # Si les deux annotateurs ont choisi "Pas d'emotion
    "
  elsif ($no_emotion == 2){
    print "Consensus_:Pas_d'emotion\n";
    print OUT "no_emotion#$content\n";
  }

  # Si deux emotions ont ete attribuees, on les
    compare pour voir s'il y a consensus
  elsif ($emotion_id[0] eq $emotion_id[1]){
    my $emotion = $emotion_id[0];

    if ($emotion =~ /1/){

```

```

        print "Consensus_:_Joie\n";
        print OUT "joie#$content\n";
    }

    if ($emotion =~ /2/){
        print "Consensus_:_Affection\n";
        print OUT "affection#$content\n";
    }

    if ($emotion =~ /3/){
        print "Consensus_:_Serenite\n";
        print OUT "serenite#$content\n";
    }

    if ($emotion =~ /4/){
        print "Consensus_:_Peur\n";
        print OUT "peur#$content\n";
    }

    if ($emotion =~ /5/){
        print "Consensus_:_Tristesse\n";
        print OUT "tristesse#$content\n";
    }

    if ($emotion =~ /6/){
        print "Consensus_:_Spleen\n";
        print OUT "spleen#$content\n";
    }
}

# S'il n'y a pas de consensus le texte est ignore
else{
    print OUT "no_consensus#$content\n";
    print "Pas_de_consensus...\n";
}
}

# Si plus de deux tags ont ete attribues
else{

```

```

my @list_emotion = ($no_emotion,0,0,0,0,0,0);

# Boucle qui teste chaque id emotion et incremente le
# hash qui compte les emotions
foreach (@emotion_id){

    if ($emotion_id[$i]=~/1/){
        $list_emotion[1]++;
        $joie++;
    }
    elsif ($emotion_id[$i]=~/2/){
        $list_emotion[2]++;
        $affection++;
    }
    elsif ($emotion_id[$i]=~/3/){
        $list_emotion[3]++;
        $serenite++;
    }
    elsif ($emotion_id[$i]=~/4/){
        $list_emotion[4]++;
        $peur++;
    }
    elsif ($emotion_id[$i]=~/5/){
        $list_emotion[5]++;
        $tristesse++;
    }
    elsif ($emotion_id[$i]=~/6/){
        $list_emotion[6]++;
        $spleen++;
    }
    $i++;
}

my $highestvalue = max @list_emotion;

if ($list_emotion[0]==$highestvalue){
    print "Pas_d'emotion\n";
    print OUT "no_emotion#$content\n";
}
elsif ($list_emotion[1]==$highestvalue){

```

```

        print "Joie\n";
        print OUT "joie#$content\n";
    }
    elsif ($list_emotion[2]==$highestvalue){
        print "Affection\n";
        print OUT "affection#$content\n";
    }
    elsif ($list_emotion[3]==$highestvalue){
        print "Serenite\n";
        print OUT "serenite#$content\n";
    }
    elsif ($list_emotion[4]==$highestvalue){
        print "Peur\n";
        print OUT "peur#$content\n";
    }
    elsif ($list_emotion[5]==$highestvalue){
        print "Tristesse\n";
        print OUT "tristesse#$content\n";
    }
    elsif ($list_emotion[6]==$highestvalue){
        print "Spleen\n";
        print OUT "spleen#$content\n";
    }
    }
    print "#####\n";
}

close FILE;
close OUT;

```

.5.5 make_corpus.pl

Script compilant les corpus et générant le fichier de méta-données nécessaire à TXM.

```

#!/usr/bin/perl
use strict;
use warnings;

# Recupere clean_content.csv et reviews.csv
# Dirige le content dans le corpus correspondant

```

```

# Ouverture des fichiers source
open(REVIEWS, '<', "./rsc/sorted_reviews.csv") or die "
    Could not open ./rsc/reviews.csv $!\n";
#open(CONTENT, '<', "./rsc/clean_content.csv") or die "
    Could not open ./rsc/clean_content.csv $!\n";

# Ouverture des fichiers de sortie (un par emotion)
open(NO_EMOTION, '>', "./corpora/no_emotion.txt") or die "
    Could not open ./corpora/no_emotion.txt $!\n";
open(JOIE, '>', "./corpora/joie.txt") or die "Could not
    open ./corpora/joie.txt $!\n";
open(AFFECTION, '>', "./corpora/affection.txt") or die "
    Could not open ./corpora/affection.txt $!\n";
open(SERENITE, '>', "./corpora/serenite.txt") or die "Could
    not open ./corpora/serenite.txt $!\n";
open(TRISTESSE, '>', "./corpora/tristesse.txt") or die "
    Could not open ./corpora/tristesse.txt $!\n";
open(PEUR, '>', "./corpora/peur.txt") or die "Could not
    open ./corpora/peur.txt $!\n";
open(SPLEEN, '>', "./corpora/spleen.txt") or die "Could not
    open ./corpora/spleen.txt $!\n";

# Ouverture du fichier de meta donnees destine a TXM
open(METADATA, '>', "./corpora/metadata.csv") or die "Could
    not open ./corpora/meatadata.csv $!\n";

# Impression du fichier de meta-donnees
print METADATA "\"1\", \"joie\" \"2\", \"affection\" \"
    \"3\", \"serenite\" \"4\", \"peur\" \"5\", \"tristesse\" \"
    n\"6\", \"spleen\" \"n\";

while (my $line = <REVIEWS>){
    chomp $line;
    next if $line =~ /^id/;

    # Recuperation des valeurs id, titre et emotion de
    REVIEWS
    my @reviews = split "#", $line;
    my $review_id = $reviews[0];

```

```

my $title = $reviews[1];
my $emotion = $reviews[2];
my $content = $reviews[3];

# Recuperation des valeurs id et content de CONTENT
#my @content = split "#",$line2;
#my $content_id = $content[0];
#my $content = $content[2];

# Comparaison de l'emotion et redirection dans le
# fichier approprie
if ($emotion eq 'no_emotion'){
    print "$title : $emotion\n";
    print NO_EMOTION $content, "\n";
}

elsif ($emotion eq 'no_consensus'){
    print "$title : Too bad ! Pas de consensus\n";
}

elsif ($emotion eq 'joie'){
    print "$title : $emotion\n";
    print JOIE $content, "\n";
}

elsif ($emotion eq 'affection'){
    print "$title : $emotion\n";
    print AFFECTION $content, "\n";
}

elsif ($emotion eq 'serenite'){
    print "$title : $emotion\n";
    print SERENITE $content, "\n";
}

elsif ($emotion eq 'tristesse'){
    print "$title : $emotion\n";
    print TRISTESSE $content, "\n";
}

```

```

    elsif ($emotion eq 'peur'){
        print "$title : $emotion\n";
        print PEUR $content, "\n";
    }

    elsif ($emotion eq 'spleen'){
        print "$title : $emotion\n";
        print SPLEEN $content, "\n";
    }

}

# Fermeture des differents fichiers ouverts
close REVIEWS;
close CONTENT;
close NO_EMOTION;
close JOIE;
close AFFECTION;
close SERENITE;
close TRISTESSE;
close PEUR;
close SPLEEN;
close METADATA;

```