



**HAL**  
open science

# Compréhension de textes par les enfants et émotions : point(s) de vue psycholinguistique(s) et leur mise en œuvre en TAL

Aline Étienne, Delphine Battistelli, Gwénoél Lecorvé

## ► To cite this version:

Aline Étienne, Delphine Battistelli, Gwénoél Lecorvé. Compréhension de textes par les enfants et émotions : point(s) de vue psycholinguistique(s) et leur mise en œuvre en TAL. [Rapport de recherche] MoDyCo; IRISA. 2019. hal-02421351

**HAL Id: hal-02421351**

**<https://hal.science/hal-02421351>**

Submitted on 20 Dec 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Rapport de recherche

# Compréhension de textes par les enfants et émotions : point(s) de vue psycholinguistique(s) et leur mise en œuvre en TAL

Aline Étienne<sup>1,2</sup>, Delphine Battistelli<sup>1</sup>, Gwénoél Lecorvé<sup>2</sup>

<sup>1</sup> Université Paris-Nanterre, CNRS, MoDyCo

<sup>2</sup> Univ Rennes, CNRS, IRISA



Travail réalisé avec le soutien du projet ANR TREMoLo (ANR-16-CE23-0019).

# Table des matières

Résumé.....	5
Introduction.....	6
1 . Problématique et méthodologie.....	7
1.1 . Comment définir les émotions ?.....	7
1.2 . La dimension émotionnelle des textes : un objet d'étude multidisciplinaire.....	8
1.3 . Méthodologie.....	9
2 . État de l'art et choix théoriques.....	11
2.1 . Les émotions dans la compréhension de textes.....	11
2.1.1 . La compréhension de textes par les adultes.....	11
2.1.2 . La compréhension de texte par les enfants.....	12
a ) <i>Les émotions, une dimension centrale dans la compréhension de textes par les enfants</i> .....	12
b ) <i>Les étapes développementales dans la compréhension des émotions</i> .....	14
c ) <i>Le développement de la théorie de l'esprit chez les enfants</i> .....	16
d ) Les différents types de verbalisation des émotions.....	16
2.2 . Les émotions dans la langue.....	17
2.2.1 . Les émotions : un « défi » pour les sciences du langage.....	17
2.2.2 . Une typologie des « modes de sémiotisation » des émotions.....	19
2.2.3 . Les émotions et la prise en charge énonciative.....	22
2.3 . Les émotions dans le traitement automatique des langues (TAL).....	24
3 . Analyse linguistique.....	27
3.1 . Le corpus.....	27
3.2 . Analyse d'exemples.....	28
3.2.1 . Exemple d'émotion « désignée ».....	28
3.2.2 . Exemple d'émotion « comportementale ».....	29
3.2.3 . Exemple d'émotion « montrée ».....	29
3.2.4 . Exemple d'émotion « étayée ».....	30
3.2.5 . Exemples mêlant différents modes d'expression des émotions.....	30
3.3 . Quels critères pour le traitement automatique des émotions ?.....	35
4 . Automatisation de l'analyse linguistique.....	37
4.1 . La récolte des données.....	37
4.2 . Construire un corpus annoté en émotions.....	37
4.2.1 . Présentation du logiciel Glozz.....	39
4.2.2 . Élaboration des fichiers pour l'annotation.....	40
4.2.3 . Élaboration du schéma d'annotation.....	43
5 . Résultats.....	50
5.1 . Difficultés d'annotation.....	50
5.1.1 . La distinction entre émotions des personnages et émotions du lecteur.....	50
5.1.2 . Le repérage de l'entité qui éprouve l'émotion.....	51
5.1.3 . L'identification de l'émotion en jeu.....	53
5.1.4 . Le cas des émotions étayées.....	55
5.2 . Résultats quantitatifs de l'annotation.....	56
5.2.1 . Test des requêtes dans Glozz.....	56
5.2.2 . Présentation des résultats quantitatifs.....	58
6 . Conclusion et perspectives.....	67

Bibliographie.....	70
Index des tableaux et des figures.....	73
A. Annexes.....	74
A.1. Exemples traités dans Micheli (2014).....	74
A.2. Cartographie des tâches de la campagne SemEval.....	87
A.3. Structures des pages du <i>P'tit Libé</i> .....	89
A.4. Script parsePtitLibe.py.....	92
A.5. Détail de la démarche d'aspiration des numéros du <i>P'tit Libé</i> .....	95
A.6. Liste des numéros du <i>P'tit Libé</i> aspirés.....	97
A.7. Obtenir le « corpusHashcode » pour les fichiers .aa.....	100
A.8. Explication technique du calcul des indices dans un fichier .ac.....	101
A.9. Lexique élaboré lors du stage sur le projet TextToKids.....	103
A.10. Script txt2GlozzFiles.py.....	105
A.11. Script fouilleGlozz.py.....	111
A.12. Catégories émotionnelles selon les numéros du <i>P'tit Libé</i> .....	117
A.13. Modes d'expression des émotions selon les numéros du <i>P'tit Libé</i> .....	119

## Résumé

Ce mémoire présente les premiers éléments d'une recherche visant à caractériser l'expression linguistique des émotions dans les textes pour enfants à l'aide de critères ne relevant pas strictement du lexique émotionnel et potentiellement implémentables dans une chaîne de traitement automatique des émotions. Ce travail s'insère dans le contexte de l'étude du développement des capacités langagières des enfants, en particulier de leur capacité à comprendre un texte. Pour mener à bien notre recherche, une typologie des modes d'expression linguistique des émotions a été précisée et un corpus annoté en émotions a été construit (97 numéros issus du journal en ligne *Le P'tit Libé* adressé aux 7-12 ans, représentant environ 196 426 tokens). Grâce aux 2043 unités exprimant une émotion délimitées et à une analyse linguistique fine d'exemples extraits du corpus, deux types principaux de critères linguistiques ont été identifiés : d'une part, des critères lexicaux – mais ne relevant pas du lexique émotionnel directement – : interjections, termes renvoyant à des manifestations physiques ou comportementales des émotions, termes décrivant des situations associées de manière conventionnelle à une émotion ; d'autre part, des critères d'ordre syntaxique : énoncés averbaux, exclamation, dislocations. Ces critères pourront entrer dans le développement d'une chaîne d'identification automatique des émotions dans les textes. L'analyse quantitative des résultats de l'annotation suggèrent par ailleurs l'utilisation préférentielle de certains éléments linguistiques pour l'expression de certaines catégories émotionnelles (par ex. la surprise, la joie et la colère seraient bien plus souvent réalisées par l'intermédiaire d'interjections ou de structures syntaxiques comme la dislocation que la peur ou la tristesse).

## Introduction

Avec le développement des nouvelles technologies et en particulier d'Internet, les enfants ont accès à de plus en plus de documents qui ne sont malheureusement pas toujours adaptés à eux. Ce sont notamment les émotions présentes dans un texte et en particulier la façon dont elles sont réalisées linguistiquement qui rendront un document plus ou moins compréhensible pour les enfants. En plus de son intérêt en psycholinguistique, la question de l'expression linguistique des émotions s'avère être aussi un enjeu pour les domaines de la linguistique et du traitement automatique des langues. Nous étudierons donc dans le présent mémoire les différents modes de réalisation linguistique des émotions à partir d'un corpus de textes à destination des enfants, constitué de numéros du journal d'actualité en ligne le *P'tit Libé*. Nous nous intéresserons également à l'implémentation de critères ne relevant pas du lexique émotionnel au sein d'une chaîne de traitement automatique des émotions.

Après avoir explicité notre problématique et notre méthodologie (section 1), nous présenterons les choix théoriques que nous avons effectués à partir d'une sélection de travaux de psycholinguistique, linguistique et traitement automatique des langues (section 2). Nous procéderons ensuite à une analyse linguistique fine de quelques exemples issus du *P'tit Libé*, qui nous permettra de dégager des pistes pour l'étude de la dimension émotionnelle des textes pour enfants (section 3). Ces pistes seront intégrées dans un schéma d'annotation, qui sera appliqué sur l'ensemble de notre corpus via la plateforme Glozz, nous faisant ainsi avancer sur la voie de l'automatisation de l'analyse linguistique (section 4). Quelques résultats observables sur le corpus annoté ainsi constitué seront détaillés (section 5) avant de dresser le bilan de notre démarche (section 6).

# 1 . Problématique et méthodologie

Après avoir défini ce que nous entendrons par « émotion » dans notre travail, nous exposerons notre problématique en nous intéressant à l'enjeu que représente l'étude de la dimension émotionnelle des textes en psycholinguistique, en linguistique et dans le traitement automatique des langues. La méthodologie que nous souhaitons mettre en œuvre pour traiter notre sujet sera ensuite brièvement développée.

## 1.1 . Comment définir les émotions ?

Pour Micheli (2014), la « complexité du domaine affectif » (p. 35) tient notamment dans la difficulté définitoire posée par le terme « émotion » et ses « voisin[s] », les termes « sentiments », « affect », voire « passion », « humeur » ... (pp. 34-35). En effet, ces termes sont aussi bien utilisés comme « mots du langage ordinaire » que comme « concepts savants » dans différents domaines, leur définition pouvant varier selon la discipline considérée (linguistique ou psychologie, par exemple) (p. 36).

Perçu comme un « état affectif spécifique » (ibid) en psychologie, le concept d' « émotion » correspond plutôt à une « classe lexicale spécifique » (ibid) en linguistique. Selon le domaine considéré, ce concept recouvre donc des réalités variées mais il semble malgré tout pouvoir être délimité par des éléments qui apparaissent de manière récurrente dans ses définitions, que ce soit en psychologie, en linguistique ou dans le « langage courant ». Ainsi, une « émotion » est un phénomène ayant de multiples composantes (« physiologique, cogniti[ve], expressi[ve] et comportemental[e] »<sup>1</sup>), caractérisé par sa « brièveté », son « intensité » et sa « saillance » (p. 42). Le terme « sentiment » est généralement vu comme distinct de l' « émotion » et désigne, le plus souvent, un phénomène plus long, moins intense et rarement associé à des « manifestations physiologiques » (p. 41). Le terme d' « affect » sert quant à lui souvent, en linguistique, en psychologie ou dans le « langage courant » comme terme « générique » pour englober tous les différents phénomènes affectifs (p. 42).

La pluralité des définitions possibles pour l' « émotion » est confirmée par Piolat et Bannour (2009). Cependant, chez les psychologues, « il y a tout de même un certain consensus sur le fait qu'une « émotion » renvoie à une réalité psychologique complexe ». L'émotion « implique » en effet :

1. « des processus d'évaluation cognitive du flux d'informations internes ou externes à l'individu »,
2. « des sensations d'excitation et plaisir (ou déplaisir) qui y sont associées »,
3. « des changements physiologiques »,
4. et un « comportement émotif, habituellement dirigé vers un but, [qui] est adaptatif. »

Cette dimension « adaptat[ive] » de l'émotion fait écho à la définition avancée par Lazarus (1984), selon laquelle les émotions naissent de l'évaluation cognitive qu'un individu fait de son environnement, de son rapport à son environnement.

Enfin, Chatar-Moumni (2013) indique que, si en « psychologie moderne », « les émotions sont des réactions déclenchées par des stimuli ou des événements externes ou internes à l'individu », leur

---

1 Micheli s'appuie principalement sur la théorie de l'évaluation cognitive (« *appraisal* ») : Scherer, Klaus (2005) : « What Are Emotions ? And How Can They Be Measured ? », *Social Science Information*, n°44, pp. 695-729

définition linguistique peut être tout autre. Elles sont alors considérées comme « des entités actives dans le sens où elles agissent sur et participent à la co-construction des actes dans la parole », par exemple pour l'argumentation rhétorique (Micheli, 2010<sup>2</sup>). Des ponts peuvent ensuite être faits entre les deux domaines, notamment par des linguistes qui s'inspirent des théories psychologiques sur les émotions pour développer des modèles linguistiques.

⇒ En gardant en mémoire ces définitions variées (psychologiques ou linguistiques), nous utiliserons majoritairement le terme d' « émotion » dans notre travail. Tout d'abord parce que ce terme est celui employé par les travaux de psycholinguistique (cf. section 2.1) et la plupart des travaux de linguistique (cf. section 2.2) sur lesquels nous nous appuyons. Nous pourrions alors reprendre, notamment, la distinction entre les émotions primaires et les émotions complexes utilisée dans (Davidson, 2006) et (Blanc et Quenette, 2017), distinction très utile pour la compréhension de texte par les enfants (cf. section 2.1.2). L'utilisation du terme « émotion » sera par ailleurs essentiel pour nous situer dans le domaine du traitement automatique des langues, où le terme « sentiment », notamment, renvoie à une tout autre notion et catégorie de travaux (cf. section 2.3).

## **1.2 . La dimension émotionnelle des textes : un objet d'étude multidisciplinaire**

Comme le suggère la section précédente, les émotions sont des objets d'étude aussi bien en psychologie (ou psycholinguistique) qu'en linguistique ou encore en traitement automatique des langues (TAL). Dans chacune de ces disciplines, les émotions représentent un réel défi définitoire (cf. infra 1.1). Plus particulièrement, elles se laissent difficilement caractériser lorsque l'attention du chercheur est focalisée sur l'épineux problème de leur expression linguistique., et c'est précisément sur cette question que portera notre travail.

Un des piliers de notre réflexion est constitué par des travaux de psycholinguistique qui se sont intéressés à la compréhension de texte. Ces travaux ont mis en évidence le rôle majeur joué par la dimension émotionnelle d'un texte (i.e. les émotions présentes dans le texte) pour la compréhension du texte, tant chez les adultes que chez les enfants (cf. sections 2.1). Parmi les études menées sur des enfants, certaines (Blanc, 2010 ; Blanc et Creissen, 2017) ont pu montrer que, selon la manière dont elles sont exprimées, les émotions sont plus ou moins accessibles pour les plus jeunes. La manière d'exprimer les émotions influence ainsi la capacité des enfants à comprendre le texte dans lequel elles se trouvent (cf. section 2.1.2). Les émotions d'un texte, et en particulier la façon dont elles sont réalisées linguistiquement constituent donc un enjeu pour la compréhension de texte par les enfants. Il apparaît alors indispensable de pouvoir caractériser finement l'expression linguistique des émotions pour ensuite étudier en détail la dimension émotionnelle d'un texte, afin d'y repérer, par exemple, des sources potentielles de difficulté de compréhension pour les enfants.

La caractérisation fine de la réalisation linguistique des émotions représente malheureusement un problème assez dense. L'expression des émotions implique en effet plusieurs (si ce n'est tous les) niveaux de la langue. Le niveau lexical, plus facile d'accès, est généralement celui étudié de la manière la plus exhaustive, souvent par la constitution de lexiques de termes d'émotions (par ex. celui de Diwersky *et al.* (2014) ; cf. section 2.2). L'étude des autres dimensions linguistiques,

---

2 Micheli R. (2010), *L'émotion argumentée. L'abolition de la peine de mort dans le débat parlementaire français*, Paris : Cerf.



notamment de la syntaxe, est généralement limitée à l'observation de phénomènes très précis, sur quelques exemples choisis (par ex. étude la construction « être en colère vs \*être en peur » de Leeman (1995) ; cf. section 2.2.1). Et lorsque des travaux s'intéressent en profondeur à l'expression des émotions ne reposant pas (uniquement) sur l'utilisation du lexique des émotions (Micheli, 2014 ; cf. section 2.2.2), ils n'évoquent jamais les textes à destination des enfants.

Enfin, l'étude automatique de la dimension émotionnelle des textes est également un enjeu pour le TAL, où la reconnaissance automatique des émotions est rarement traitée. Quand elle l'est, c'est le plus souvent via un lexique de termes d'émotion et seulement occasionnellement dans les textes pour enfants (cf. section 2.3). Trouver des critères linguistiques, autres que les termes du lexique émotionnel, implémentables dans un système de traitement automatique des émotions constituerait donc un apport important en TAL, et ce d'autant plus si ces critères sont testés sur des textes jeunesse.

Ainsi, l'objectif de notre travail sera de caractériser l'expression des émotions dans les textes pour enfants, en dégagant des critères linguistiques ne relevant pas du lexique des émotions, idéalement implémentables dans une chaîne de repérage et d'identification automatique des émotions. Notre corpus d'étude est constitué de textes de presse jeunesse, issus du site d'actualité en ligne *Le P'tit Libé*, adressé aux 7-12 ans.

### 1.3 . **Méthodologie**

Plusieurs étapes seront nécessaires pour mener notre recherche.

Grâce à un travail bibliographique sur la compréhension de textes, en particulier par les enfants (cf. section 2.1.1), et sur la réalisation linguistique des émotions (cf. section 2.2.2), nous retiendrons des pistes théoriques pour aborder l'expression des émotions dans les textes à destination des enfants.

Nous procéderons ensuite à l'analyse linguistique fine d'exemples extraits de textes pour enfants (en l'occurrence de textes de presse jeunesse, issus de notre corpus) pour voir dans quelle mesure les pistes théoriques peuvent s'appliquer à notre corpus (cf. section 3.2). Nous dégagerons ainsi des catégories et des critères pertinents pour l'étude de la dimension émotionnelle de nos textes (cf. section 3.3.).

Ces éléments nous serviront à systématiser l'analyse du corpus en élaborant un schéma d'annotation, qui permettra de construire un corpus annoté (cf. section 4). La constitution d'un tel corpus sera utile pour confirmer la pertinence des catégories et des critères retenus à l'échelle d'un corpus entier, et non plus seulement lors de l'analyse de quelques exemples. Les données quantitatives qui accompagneront notre corpus annoté (nombre d'unités annotées selon leur type etc.) nous permettront de trouver d'éventuels liens entre des caractéristiques linguistiques et des catégories émotionnelles (par ex. telle structure syntaxique sert souvent à exprimer telle émotion) (cf. section 5).

A partir des critères validés sur le corpus et des pistes révélées par l'annotation à grande échelle, des règles pour la reconnaissance automatique des émotions ou des passages textuels émotionnels pourraient être construites (par ex. si on trouve telle structure syntaxique, alors il y a une émotion). Par ailleurs, un corpus annoté de grande taille constitue une ressource précieuse qui peut être explorée par d'autres outils. Les méthodes de la textométrie pourraient par exemple être utilisées

pour contraster entre eux les différents passages textuels (par ex. émotionnel vs non émotionnel, ou passage exprimant la peur vs passage exprimant la joie etc.). Nous dégagerions de cette manière d'autres critères linguistiques (hors lexique émotionnel) pertinents pour la caractérisation des émotions et potentiellement implémentables dans un système de reconnaissance automatique des émotions à base de règles (i.e. un système symbolique). Un grand corpus annoté permettrait également de tester (et d'entraîner) des systèmes d'apprentissage automatique dans le but d'étudier automatiquement la dimension émotionnelle d'un texte.

## 2 . État de l'art et choix théoriques

L'étude de notre problématique est reliée à de nombreux travaux, sur lesquels nous nous sommes appuyés pour opérer des choix théoriques. Dans un premier temps, nous présenterons quelques recherches de psycholinguistique mettant en avant le rôle des émotions dans la compréhension de texte (tant chez les adultes que chez les enfants). Nous nous pencherons ensuite sur la place qui leur est consacrée en linguistique puis dans le domaine du traitement automatique des langues, où nous avons trouvé des sources d'inspiration méthodologique.

### 2.1 . Les émotions dans la compréhension de textes

La compréhension de texte repose sur la capacité à construire une « représentation mentale de la situation décrite par le texte » (et non du texte lui-même), soit d'un « modèle situationnel ». Ce modèle est multidimensionnel : il intègre aussi bien des informations temporelles, spatiales et causales que des informations sur les personnages et leurs motivations (nous traduisons de l'anglais, Zwaan et Radvansky, 1998). L'étude de la compréhension de texte, aussi bien chez les adultes que chez les enfants, a par ailleurs mis en avant que les émotions jouent un rôle central dans l'élaboration de cette « représentation mentale » (ibid.).

#### 2.1.1 . La compréhension de textes par les adultes

Dans une étude de la compréhension de texte par des adultes, Dijkstra *et al.* (1995) distinguent deux types d'émotions présentes dans les textes (nous traduisons de l'anglais) :

1. les « émotions des personnages » : ces émotions, qui « surviennent au sein du monde fictif », correspondent aux émotions ressenties par les personnages dans l'histoire. Elles peuvent être désignées par un « description explicite » (présence d'un label émotionnel) ou nécessiter des « inférences pour reconstruire la réaction émotionnelle d'un personnage » (ex. : « *ses lèvres [...] commençant à pâlir* » nécessiter d'inférer que la pâleur des lèvres du personnage est causée par sa peur)
2. les « émotions du lecteur » : ces émotions correspondent à celles ressenties par le lecteur et sont elles-mêmes de deux types.
  - (1) Les émotions « fictives », déclenchées par ce qui se passe dans l'histoire. Le lecteur « s'identifie à un ou plusieurs personnages » et éprouve de l'« empathie » pour eux. Les émotions du lecteur proviennent alors « des émotions ressenties par les personnages ». Les « émotions des personnages et l'empathie sont donc fortement liées ».
  - (2) Les émotions « artefact », qui « proviennent des aspects structurels, stylistiques ou compositionnels de l'œuvre ». Soit elles portent sur « le monde imaginaire » et sont alors « générées par les structures narratives » (ex. « *suspense, surprise, curiosité* ») ; soit elles portent sur « l'artefact (i.e. le texte littéraire) lui-même » (ex. : « *admiration pour le savoir-faire littéraire de l'auteur* » lors d'une « *description de paysage* » particulièrement réussie).

Lors de cette étude, il a été montré que les « émotions des personnages facilitent la compréhension du texte » car elles déclencheraient des « inférences<sup>3</sup>, qui aident à construire des représentations en mémoire et facilitent l'interprétation du récit ». Elles permettent donc de mieux se souvenir des événements de l'histoire (nous traduisons de l'anglais, Dijkstra *et al.*, 1995).

Cette hypothèse est celle avancée par Dyer (1983, nous traduisons de l'anglais). Selon lui, « les descriptions affectives influencent le traitement [des textes] de deux manières : (a) elles déclenchent des inférences; et (b) elles aident à construire des souvenirs », soit des « représentations en mémoire » de l'histoire. L'importance des émotions est due aux liens qu'elles entretiennent avec les « objectifs des personnages », leurs « attentes » ainsi que les « relations interpersonnelles » et l'« empathie ». Par exemple, l'intensité d'une émotion ressentie par un personnage face à un événement pourra révéler l'importance de cet événement.

⇒ Dans ce mémoire, notre approche linguistique nous poussera à nous intéresser uniquement aux émotions « des personnages » (ou, pour étendre un peu la notion, à celles du locuteur ou de l'énonciateur telles qu'elles transparaissent dans les énoncés) et non à celles du lecteur. L'étude de la compréhension de texte par les adultes a pu mettre en évidence l'importance des émotions des personnages qui, en reliant différents événements entre eux, permettent de maintenir la cohérence de la représentation que le lecteur se fait du texte. Notre travail portant sur la compréhension de texte par les enfants, nous avons jugé essentiel de nous intéresser aux travaux de psycholinguistique portant spécifiquement sur l'impact des émotions dans la compréhension de texte par des enfants.

### 2.1.2 . **La compréhension de texte par les enfants**

Si plusieurs études de psycholinguistique menées sur des enfants ont pu montrer l'importance des émotions dans la compréhension de texte, certaines caractéristiques des émotions ou de leur verbalisation semblent pouvoir constituer des facteurs potentiels de difficulté de compréhension pour les enfants.

#### ***a ) Les émotions, une dimension centrale dans la compréhension de textes par les enfants***

D'après plusieurs recherches traitant de la compréhension de texte par les enfants, les émotions des personnages permettent à la fois une meilleure mémorisation des événements auxquels ils sont associés (Davidson *et al.*, 2001 ; Davidson, 2006) et l'élaboration d'une construction cohérente de la situation décrite par un texte (Blanc 2009 ; Blanc et Creissen, 2017 ; Blanc et Quenette, 2017).

Pour Davidson *et al.* (2001), les émotions favorisent la mémorisation de l'information probablement grâce à leur lien avec les objectifs des personnages (objectif en cours de réalisation, atteint ou raté) et/ou leur lien avec les conséquences, plus ou moins importantes, d'un événement pour les personnages et la suite de l'histoire. Ainsi, un objectif atteint sera souvent associé à une émotion positive (*ex. joie*), tandis qu'un objectif raté déclenchera plutôt une émotion négative chez

---

3 Une inférence peut être définie comme une « opération qui consiste à admettre une proposition en raison de son lien avec une proposition préalable tenue pour vraie ». Il s'agit donc d'une sorte de conclusion que l'on peut tirer à partir de divers éléments que l'on admet être vrais (d'après le dictionnaire en ligne du CNRTL, <https://www.cnrtl.fr/definition/inf%C3%A9rence>).

le personnage (*ex. tristesse ou colère*). Par ailleurs, un évènement associé à une émotion aura généralement plus de conséquences pour le personnage ou l'histoire (nous traduisons de l'anglais ; *ex. « Ron était triste parce que son meilleur ami déménageait. »*) qu'un évènement non émotionnel (nous traduisons de l'anglais ; *ex. « Ron et son frère ont pris un goûter. »*).

Nathalie Blanc (2009) évoque elle aussi l'importance, pour comprendre une histoire, des buts des personnages et de leur lien avec les émotions. La capacité d'un enfant à faire des « inférences émotionnelles » (p. 71) dépendrait de sa capacité à « repérer le but du personnage et le résultat des actions qu'il a engagées pour y parvenir » (*ibid* ; Blanc s'appuie sur le modèle développemental « de l'évolution de la production inférentielle » proposé par van den Broek *et al.*<sup>4</sup>, 2005). De plus, les émotions des personnages pourraient, dans une certaine mesure, être définies par rapport aux buts des personnages. La joie correspondrait à « vouloir quelque chose et l'obtenir » ou « ne pas vouloir quelque chose et ne pas l'obtenir » ; la peur à « ne pas vouloir quelque chose et ne pas être certain de pouvoir l'éviter » ou « vouloir quelque chose mais ne pas être sûr de l'obtenir » ; et la colère et la tristesse à « vouloir quelque chose mais ne pas l'obtenir » ou « vouloir éviter quelque chose mais ne pas y parvenir » (Stein et Liwag, 1997<sup>5</sup> ; cité par Blanc, 2009, p. 72). Ce lien entre « but des personnages » et « émotion des personnages » aurait même une incidence sur la focalisation de « l'attention des enfants sur certains passages de l'histoire ». Ainsi, « pour la colère, ce seraient les antécédents (e.g., les évènements qui ont fait obstacle à l'obtention du but) alors que pour la peur, l'attention se porterait sur les informations à venir, un évènement qui n'est pas encore survenu ou qui est sur le point de se produire. Lorsque le personnage est heureux ou alors triste, ce sont à la fois les antécédents et les conséquences qui sont considérées » (*ibid* ; d'après les travaux de Liwag et Stein, 1995<sup>6</sup>). Les émotions présentes dans un texte « sont donc vraisemblablement des indicateurs de cohérence que les enfants utilisent pour comprendre l'histoire » (*ibid*). Elles peuvent parfois être les « conséquences » d'un évènement (*ex. « la colère [d'un personnage] lorsqu'un camarade lui a cassé son jouet préféré »*), ou parfois au contraire la « cause » d'un autre évènement (*ex. « la colère incite alors le personnage [...] à s'emparer et casser en retour le jouet de son camarade »*) et sont alors un élément indispensable pour recréer l'enchaînement des évènements de l'histoire.

Ces différents éléments font écho à la théorie de Dyer (1983 ; cf. *infra* 2.2.1) concernant l'impact global des affects sur la construction et la mémorisation de la représentation mentale de la situation décrite par un texte.

Les émotions des personnages constituent ainsi une dimension centrale des textes. Entretenant des liens étroits, d'une part avec les buts des personnages et, d'autre part, avec les relations de causalités qui unissent les différents évènements d'une histoire, elles apparaissent indispensables pour saisir l'organisation d'un récit et donc pour élaborer une représentation cohérente de la situation décrite par un texte (ce qui correspond, rappelons-le, à *comprendre* le texte dans un certain nombre d'approches psycholinguistiques). Les enfants étant des êtres en développement, il est

---

4 Van Den Broek, P., et al. (2005). Assessment of comprehension abilities in young children. In *Children's reading comprehension and assessment* (pp. 125-148). Routledge.

5 Stein, N. L., & Liwag, M. D. (1997). Children's understanding, evaluation, and memory for emotional events. *Developmental spans in event comprehension and representation: Bridging fictional and actual events*, 1, 199-235.

6 Liwag, M. D., & Stein, N. L. (1995). Children's Memory for Emotional Events: The Importance of Emotion-Related Retrieval Cues. *Journal of Experimental Child Psychology*, 60(1), 2-31.

cependant probable que certaines caractéristiques des émotions rendent les textes (ou les émotions elles-mêmes) plus ou moins intelligibles pour les enfants.

### **b) Les étapes développementales dans la compréhension des émotions**

Une première source potentielle de difficulté pour la compréhension par les enfants des émotions est liée aux types d'émotions en jeu. Denise Davidson (2006) a ainsi étudié l'impact de plusieurs types d'émotions sur la compréhension du texte (et de l'émotion) par les enfants. Elle distingue trois types d'émotions (en se basant sur les travaux de Ekman, 1992<sup>7</sup> et Lewis, 2000<sup>8</sup> ; nous traduisons de l'anglais) :

- les émotions « de base », ou émotions primaires. En « nombre limité » (colère, dégoût, joie, peur, surprise et tristesse) et relativement universelles (inter-individus et inter-cultures) elles sont maîtrisées « tôt dans la vie » (dès la « première année de vie ») et sont associées à des « expressions faciales stéréotypiques » et « universellement reconnues » ;
- les émotions « de conscience de soi » (« *self-conscious* » ; ou « émotions sociales » (Blanc et Quenette, 2017)). D'apparition un peu plus tardive (vers un an et demi), elles ne seraient pas associées aux mêmes situations et/ou aux mêmes réactions selon les cultures. Elles mettent en jeu des compétences cognitives plus complexes, notamment pour le « traitement de l'information sociale ». Elles nécessitent de focaliser son attention sur autrui, notamment pour envisager l' « évaluation et le jugement des individus par eux-mêmes et par les autres », et d'être capable de « constituer une auto-représentation stable (moi) », pour ensuite « focaliser son attention sur ces représentations ou s'auto-évaluer ». L'individu doit donc être en mesure de « se représenter soi-même et un autre dans un espace commun et de se préoccuper des opinions d'un autre », de se concentrer sur son « soi public » et non son « soi interne ». Des émotions correspondant à cette catégorie sont par exemple la honte (embarras, gêne) et la jalousie ;
- les émotions « de conscience de soi évaluatives » (« *self-conscious evaluative* » ; ou « sociales évaluatives », (Blanc et Quenette, 2017)). Rarement maîtrisées avant 3 ans, elles ont les mêmes caractéristiques que les émotions « de conscience de soi » mais nécessitent, en plus, la capacité à « se comparer ou à comparer son propre comportement à des standards, des règles ou des objectifs culturels internalisés ». Ces éléments culturels font l'objet d'un apprentissage plus ou moins long chez l'enfant. Des émotions correspondant à cette catégorie sont par exemple la fierté et la culpabilité.

Par ailleurs des étapes développementales de compréhension des émotions par les enfants sont dégagées (Davidson, 2006, s'appuie sur les travaux de Griffin<sup>9</sup>, 1995 et Harter<sup>10</sup>, 1999 ; nous traduisons de l'anglais) :

- avant 5-6 ans, compréhension « pré-intentionnelle » des émotions. Les enfants « expliquent l'émotion en se référant exclusivement aux événements comportementaux », la dimension

---

7 Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6, 169–200 .

8 Lewis, M. (2000). Self-conscious emotions: embarrassment, pride, shame and guilt. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed., pp. 623–636), New York, NY: The Guilford Press.

9 Griffin, S. (1995). A cognitive-developmental analysis of pride, shame, and embarrassment in middle childhood. In J. P. Tangney & K.W. Fischer (Eds.), *Self-conscious emotions* (pp. 219–236), New York, NY: The Guilford Press.

10 Harter, S. (1999). *The construction of the self*. New York, NY: The Guilford Press.

comportementale (manifestation physique) des émotions est donc cruciale pour leur compréhension ;

- vers 6-7 ans : les enfants sont en mesure de prendre en compte la dimension psychologique (« intentions de l'acteur », la personne qui ressent l'émotion » et de l'observateur », la « personne qui est témoin de l'émotion ») pour « expliquer l'émotion ». Ils peuvent ainsi « fai[re] référence à un état interne » des personnes impliquées dans l'émotion, et peuvent « être conscients des intentions de l'acteur ou de l'observateur mais pas des deux » ;
- vers 8-9 ans : pour expliquer une émotion, les enfants peuvent faire appel aux intentions de multiples acteurs et observateurs, ainsi qu'à leurs rôles respectifs dans l'émotion. Ils sont également capables d' « incorporer les normes ou les standards sociaux qui peuvent exister dans les émotions individuelles » dans leurs explications.

L'étude menée par Davidson (2006) sur des enfants de 6, 8 et 10 ans confirme ces différentes étapes développementales. Tout d'abord, les enfants, en particulier ceux de moins de 10 ans, expliquent beaucoup les émotions par des informations comportementales, surtout pour les émotions « de conscience de soi ». Ensuite, lorsque les enfants de moins de 10 ans sont confrontés à des émotions « de conscience de soi » (« évaluatives » ou non), ils ne semblent pas en avoir une compréhension fine et ont donc tendance à les mentionner (voire peut-être les retenir) avec des termes liés aux émotions « de base » (*ex.* : « *heureux pour fier, en colère pour gêné/embarassé* »).

L'étude de Blanc et Quenette (2017) sur la capacité d'enfants francophones (8 à 10 ans) à faire des inférences émotionnelles confirme ce dernier résultat. Les enfants retrouvent généralement la « coloration émotionnelle » mais ont plus de mal à saisir l'émotion « spécifique » suggérée par la situation, assimilant ainsi la fierté à la joie et la culpabilité à la colère.

Par ailleurs, une étude (Baron-Cohen, 2010) portant sur le développement de la compréhension des termes d'émotion par des enfants anglophones de 4 à 16 ans, confirme aussi que les émotions « de base » sont accessibles plus tôt que les émotions « complexes » (puisque les termes associés aux émotions « de base » sont beaucoup mieux compris et beaucoup plus tôt que ceux associés aux émotions « complexes ») (nous traduisons de l'anglais, Baron-Cohen, 2010).

⇒ Nous retiendrons ainsi qu'une source potentielle de difficulté de compréhension pour les enfants est la catégorie de l'émotion : une émotion « de base » sera accessible très tôt, tandis qu'une émotion « de conscience de soi »/« sociale », nécessitant des compétences cognitives plus développées et des connaissances socio-culturelles plus importantes, ne sera pas maîtrisée avant 8-10 ans.

De plus, puisque la distinction entre les émotions « de conscience de soi »/« sociales » et les émotions « de conscience de soi »/« sociales » « évaluatives » n'a, a priori, pas d'impact sur la compréhension des enfants (Davidson, 2006 ; Blanc et Quenette, 2017), nous fusionnerons, dans une optique de simplification, ces deux types d'émotions et les appellerons des émotions « complexes ». Seules deux catégories d'émotions seront donc évoquées dans la suite de notre travail : les émotions « de base » (primaires) et les émotions complexes.

### **c )Le développement de la théorie de l'esprit chez les enfants**

La théorie de l'esprit est une notion de neuropsychologie correspondant à une « aptitude cognitive » qui nous permet d'attribuer des « états mentaux » (« états affectifs ou cognitifs ») ou des « représentations mentales » à autrui, à partir d'indices comportementaux (actions, « attitudes », « expressions émotionnelles ») ou de « connaissance[s] supposée[s] de la réalité » qu'a l'autre. Cette aptitude repose donc sur la capacité à faire des inférences. C'est grâce à elle que nous pouvons « prédire », « anticiper » et « interpréter le comportement ou l'action de nos pairs dans une situation donnée ». Elle joue ainsi un rôle clé dans « la régulation des conduites » et le « bon déroulement des interactions sociales » (Duval *et al.*, 2011). La théorie de l'esprit appartient à la « cognition sociale qui mobilise un ensemble de processus mentaux tels que la perception de soi et des autres et l'utilisation des connaissances sur les règles régissant les interactions interpersonnelles pour décoder le monde social » (Beer *et al.*, 2006<sup>11</sup>, cité par Duval *et al.*, 2011).

Deux types de théories de l'esprit peuvent être distingués : la théorie de l'esprit « cognitive », qui consiste à attribuer des croyances, des pensées, des raisonnements à autrui ; et la théorie de l'esprit « affective », qui consiste à attribuer des émotions, des sentiments à autrui. Il y a également deux niveaux de théorie de l'esprit : le 1<sup>er</sup> ordre (ex. : « *Je pense que la femme pense à son travail.* ») et le 2<sup>nd</sup> ordre (ex. : « *Je pense que la femme pense que l'homme pense aux vacances.* ») (l'« ordre zéro » correspondant à un raisonnement du type « X pense que/à ... »).

Les différentes étapes du développement « typiqu[e] » de la théorie de l'esprit chez les enfants seraient : avant 18 mois, perception d'ordre zéro ; dès 4 ans, raisonnement de 1<sup>er</sup> ordre ; et à partir de 6-7 ans, raisonnements de « 2<sup>nd</sup> ordre ». Si « la théorie de l'esprit elle-même ne se développe peut-être pas au-delà du début de l'enfance », « l'interaction entre la théorie de l'esprit et d'autres processus cognitifs telles que les fonctions exécutives continue[rait] de mûrir pendant l'adolescence », comme le suggère l'étude de Dumontheil *et al.* (2010, nous traduisons de l'anglais).

⇒ La théorie de l'esprit apparaît donc indispensable pour comprendre autrui (ses émotions, ses comportements, ses raisonnements etc.) et se mettre à sa place. Nous supposons qu'il s'agit là d'une compétence cognitive essentielle pour comprendre les émotions des personnages dans un texte (se mettre à la place des personnages) voire même pour comprendre certaines émotions « complexes », qui font appel à une image de soi ou au regard de l'autre (ex. *la honte*). Même si cette notion ne sera pas mobilisée explicitement dans le travail que nous présentons dans ce mémoire, elle pourra l'être de manière plus explicite dans nos travaux de recherche à venir.

### **d ) Les différents types de verbalisation des émotions**

Une seconde source potentielle de difficulté pour la compréhension par les enfants des émotions est liée aux différents types possibles de verbalisation d'une émotion. D'après Davidson *et al.* (2001, nous traduisons de l'anglais), un événement associé à une émotion (« information émotionnelle ») est mieux retenu par les enfants (de 7, 9 et 11 ans) quand un label émotionnel explicite est présent, c'est-à-dire quand l'émotion est clairement désignée par un élément du

---

11 Beer, J. S. et Ochsner, K. N. Social cognition: a multi level analysis. *Brain research*, 2006, vol. 1079, no 1, p. 98-105.



lexique. Par ailleurs, dans un texte comportant des « informations émotionnelles » et des « informations non-émotionnelles » (événements non-associés à une émotion), la présence d'un label émotionnel permettrait non seulement un meilleur rappel des informations émotionnelles mais également des informations non-émotionnelles.

L'impact sur la compréhension par les enfants de la façon dont les émotions sont verbalisées dans les textes, suggéré dans (Davidson *et al.*, 2001), a fait l'objet d'autres travaux en psycholinguistique. Nathalie Blanc *et al.* (Blanc, 2010 ; Blanc et Creissen, 2017) utilisent ainsi une typologie des différentes manières de verbaliser l'émotion :

1. **l'émotion « désignée »** : « l'émotion du personnage est explicitement mentionnée [dans le texte] ». Par exemple « *il/elle est triste* » (Blanc et Creissen, 2017).
2. **l'expression comportementale de l'émotion** : « l'émotion est véhiculée au travers du comportement du personnage ». Par exemple, « *il/elle éclate en sanglots* » (Blanc et Creissen, 2017).
3. **l'évènement inducteur de l'émotion** : simplement « suggérée » par un évènement, l'émotion du personnage doit être « inférée » (« inférence émotionnelle ») (Blanc et Creissen, 2017). Par exemple, « *Le puma remit la récompense promise au lapin Mudubina et le nomma Premier Ministre.* » : le « *lapin* » éprouve de la joie (Blanc, 2010).

Confrontés à ces différentes formes de verbalisation, les enfants (6 à 10 ans) comprennent (mémorisation, représentation) mieux les émotions « explicitement [fournies] » dans l'histoire (émotions « désignées » ou décrites par des « expressions comportementales ») que celles simplement « suggérées » par la situation (Blanc et Creissen, 2017). Les enfants de 6 ans et moins auraient plus de facilité avec les émotions décrites par des « expressions comportementales »<sup>12</sup> (Blanc, 2010).

⇒ La façon dont les émotions sont verbalisées dans les textes (« désignées » directement par un élément du lexique, décrites par des « expressions comportementales » ou « suggérées » par des évènements) constituerait un facteur déterminant de la compréhension des émotions, et donc des textes, par les enfants. Il paraît alors judicieux de s'intéresser davantage à l'expression linguistique des émotions.

## 2.2 . Les émotions dans la langue

### 2.2.1 . Les émotions : un « défi » pour les sciences du langage

Un rapide tour d'horizon du traitement des émotions dans la linguistique du XXe siècle (Kerbrat-Orecchioni, 2000) montre que « les émotions posent au linguiste de vrais problèmes et lui lancent un vrai défi, à cause surtout de leur caractère éminemment [...] fuyant et insaisissable ». Ces « problèmes » sont attestés par la présence des « catégories floues, des notions polymorphes et des marqueurs indécis » employés pour l'étude des émotions dans le langage. Les difficultés se situent tant au niveau du signifié des émotions, puisque « le champ affectif de l'affectivité (expressivité ?

---

12 La dimension comportementale des émotions serait aussi cruciale dans la compréhension par les adultes, puisque les émotions inférées lors de la lecture seraient stockées en mémoire sous la forme de comportements (le lecteur s'imaginant tel ou tel comportement d'un personnage (Gygax, 2010).

Émotionnalité ?) est plus que tout autre « indéfinissable », qu’au niveau de leur signifiant, avec la « fantastique diversité des moyens que peut investir le langage émotionnel, puisque tout mot, toute construction peuvent venir en contexte propice se charger d’une connotation affective ». Face à ces difficultés, on a « le sentiment que les émotions sont à la fois dans le langage partout, et nulle part. » Cette complexité de l’étude des émotions en sciences du langage est réaffirmée par Micheli *et al.* (2013), qui indique que les émotions peuvent être associées à des « observables » très « abondants » et « hétérogènes » et qu’il est alors « difficile d’associer [aux émotions] un (ou plusieurs) lieu(x) stable(s) dans le système linguistique ».

Une grande part des travaux de linguistique portant sur les émotions se focalisent sur seulement un niveau de la langue et rares sont ceux s’intéressant aux « constructions » permettant d’exprimer les émotions (Chatar-Moumni, 2013). Diwersky *et al.* (2014) se sont par exemple concentrés sur le lexique et ont constitué un « lexique émotionnel », en faisant notamment appel à des critères combinatoires (combinaison des termes d’émotion avec certains verbes comme « ressentir », « éprouver » etc. ; combinaison avec certains collocatifs aspectuels « élan, transport, bouffée » ...). Certains ont procédé à des analyses sémantiques fines d’émotions ou de structures spécifiques : étude des expressions où la « peur » est le « prédicat sémantique » (Bresson et Dobrovol’skij, 1995) ; analyse approfondie de la structure « être en colère » vs « \*être en peur », qui montre que chaque élément du « contexte » peut être porteur de sens et doit donc être pris en compte, et pas seulement le terme d’émotion lui-même (Leeman, 1995).

Un des travaux dont nous pourrions nous inspirer est celui de Gross (1995), qui présente une étude poussée de diverses structures (morpho-)syntaxiques exprimant les « sentiments » en français. Cette analyse détaillée permet de modéliser les différentes variations de verbalisation sous forme de graphes (automates finis), capables de reconnaître ensuite ces différentes structures (méthode symbolique).

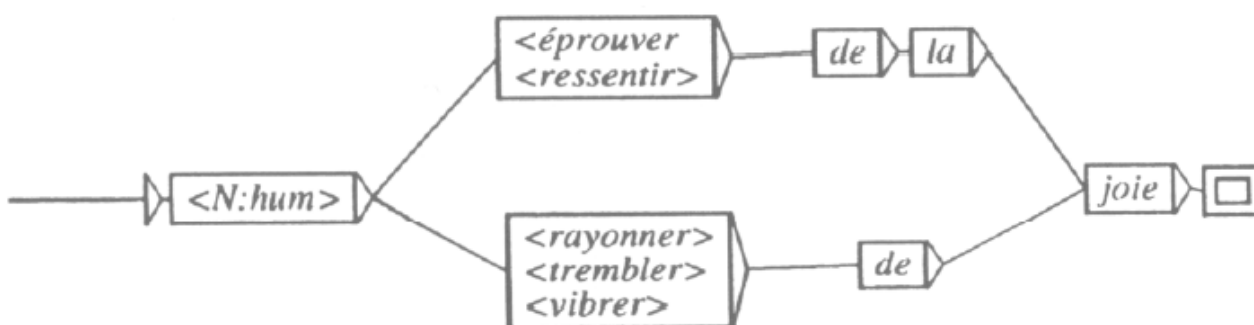


Figure 1 : Exemple d'automate fini (Gross, 1995)

L’automate fini de la figure 1 reconnaîtra automatiquement des segments comme « Julie rayonne de joie » ou encore « Pierre ressent de la joie ». L’étude approfondie de nos propres données permettra, idéalement, d’établir des modélisations similaires pour reconnaître automatiquement les passages émotionnels d’un texte. Cependant, tandis que l’étude de Gross ne porte que sur deux émotions, la « joie » et la « rage », nous nous intéresserons à de nombreuses autres catégories, en nous focalisant sur les éléments linguistiques ne relevant pas strictement du lexique des émotions.

### 2.2.2 . Une typologie des « modes de sémiotisation » des émotions

Par rapport à la multitude de travaux n'abordant qu'un aspect de la verbalisation des émotions (*ex. le lexique*) ou ne traitant que d'une ou quelques émotions spécifiques (*ex. la joie et la rage*), le travail de Micheli (2014) apparaît assez complet puisqu'il propose une typologie pour organiser les différentes façons de « sémiotiser » les émotions (l'acte de « sémiotiser » est défini par l'auteur (*ibid* p. 11) comme « rendre quelque chose manifeste au moyens de signes, verbaux ou coverbaux »). Cette typologie rend compte des moyens potentiellement mobilisés pour exprimer les émotions en trois catégories<sup>13</sup> :

1. **les émotions « dites »** (*ibid*, pp. 35-47). Un énoncé prototypique qui « dit » une émotion est défini comme comportant un « terme d'émotion », une « expression référentielle » (i.e. une « réalis[ation] linguistique ») qui désigne « l'être supposé ressentir l'émotion ». Cet « être » est une « entité humaine ou humanisable ». Souvent, il y a également « une expression qui indique ce sur quoi porte l'émotion » (l'« objet ») ou ce qui la « déclenche » (la « cause »). Enfin, une « relation prédicative » met en relation, « au niveau syntaxique » l'émotion avec l'entité qui la ressent. La variété des énoncés qui « disent » l'émotion sera due à la catégorie grammaticale du terme d'émotion et à la position syntaxique qu'il occupe, de même pour l'expression renvoyant à l'entité affectée par l'émotion.

*ex. Les araignées effrayent Paul. | La peur de Paul pour les araignées est très grande. | Paul est épouvanté par les araignées. etc.*

Une des spécificités des émotions « dites » (par rapport aux émotions « montrées » et aux émotions « étayées ») est qu'elles permettent très facilement l'« auto- et l'allo-attribution » (attribution à soi-même ou à un tiers) de l'émotion, simplement en modifiant l'« expression référentielle » utilisée pour désigner l'entité qui ressent l'émotion.

*ex. Je suis heureuse. | Ma mère est heureuse. | Tout le monde est heureux. etc.*

Une autre spécificité est que l'interprétation d'un énoncé qui « dit » l'émotion ne requiert (quasiment) aucune inférence de la part de l'allocutaire puisque l'émotion et l'entité qu'elle affecte, ainsi que lien qui les unit sont explicitement verbalisés.

2. **les émotions « montrées »** (*ibid*, pp. 62-99). Cette catégorie, parfois désignée par le terme d'« émotion connotée », est beaucoup plus problématique car très « hétérogèn[e] ». Le point commun qui unit tous les énoncés qui « disent » l'émotion réside dans l'« interprétation indicielle » nécessairement effectuée pour comprendre l'émotion qu'ils contiennent. Certaines caractéristiques de l'énoncé, de par leur « rapport stable de cooccurrence » avec le ressenti d'émotion par le locuteur, seront vus comme des « indices » de l'état émotionnel du locuteur par l'allocutaire. Le raisonnement effectué par l'allocutaire sera donc le suivant : « si un énoncé présente [ces] caractéristiques, alors le locuteur qui l'énonce éprouve probablement une émotion » puisqu'il y a « une relation de cooccurrence, voire d'effet à cause, entre l'énonciation d'un énoncé présentant ces caractéristiques et le fait d'éprouver une émotion » (« inférence abductive »). Au vu de l'étendue et de la complexité des marqueurs pouvant être impliqués dans les émotions « montrées », l'auteur se focalise sur :

---

13 Les nombreux exemples analysés par Micheli (2014) lors de l'explicitation des différents « modes de sémiotisation » des émotions ont été recensés dans un exemplier, disponible en annexe (cf. A.1).

- (i) les « marqueurs lexicaux », représentés par des mots ayant un « sens affectif non dénotatif », autrement dit connotatif (ex. « *diktat* ») et les interjections.
- (ii) les « marqueurs syntaxiques », correspondant globalement à des structures en rupture avec le « modèle de la phrase canonique » (soit un sujet et un prédicat, ce dernier réalisé par un verbe conjugué, au sein d'une phrase simple affirmative), « imaginaire » et associé à un locuteur qui ne ressent « aucune émotion particulière ». Plusieurs types d'énoncés sont évoqués : les exclamatifs, les « elliptiques » (soit une « réduction syntaxique »), les « averbaux » (soit une « réduction syntaxique »), les « disloqués à droite » (soit un « réordonnement syntaxique ») et les « clivés », (soit un « réordonnement syntaxique »).

ex. « *Le juge et le bourreau sont sortis. Je suis seul. Seul avec deux gendarmes. Oh ! L'horrible peuple avec ses cris d'hyène. - Qui sait si je ne lui échapperai pas , si je ne serai pas sauvé ? Si ma grâce ?... Il est impossible qu'on ne me fasse pas grâce ! Ah ! Les misérables ! Il me semble qu'on monte l'escalier...* » (Victor Hugo, *Dernier jour d'un condamné*, chap. XLIX, pp. 139-140)

Cet exemple montre l'importance des « faisceau(x) de caractéristiques » pour l'interprétation de l'émotion « sémiotisée ». En l'occurrence, il y a une émotion « dite » (« horrible » renvoyant à la peur) qui permet d'interpréter de manière spécifique l'émotion « montrée » (interjections « oh », « ah » ; exclamation ; énoncés « averbaux »).

- (iii) les « marqueurs transphrastiques et textuels » ou marqueurs discursifs, renvoyant à des « modes spécifiques d'agencement des énoncés au sein du texte » des « types d'organisations transphrastiques » (p. 99), comme par exemple les périodes qui rythment un discours.

Puisque les émotions « montrées » correspondent à des caractéristiques de l'énoncé qui révèlent l'état émotionnel de l'énonciateur lors de la production de cet énoncé, si l'énonciateur (« source du point de vue manifesté par [l'] énoncé ») et le locuteur (« producteur de l'énoncé ») ne font qu'un, « l'émotion montrée est forcément auto-attribuée ». Pour être « allo-attribuée », l'émotion « montrée » doit donc être prise en charge par un autre énonciateur (cf. infra 2.2.3).

3. **les émotions « étayées »** (ibid, pp. 105-122). Un énoncé (ou un discours) « étaye » une émotion lorsqu'il décrit une situation, et plus précisément lorsqu'il propose une « schématisation de la situation », qui est associée à une émotion par les normes et conventions socio-culturelles. L'allocutaire, reconnaissant la « situation schématisée » est alors en mesure d' « inférer » l'émotion qui y est conventionnellement associée. Ce type d'émotion semble lié à la notion d'argumentation (en tout cas, tel que Micheli le présente) puisqu'il s'agit, pour le locuteur, de sélectionner certains aspects d'une situation et de faire en sorte que la « schématisation » ainsi construite « ser[ve] de fondement » à l'émotion qu'il veut exprimer. Une émotion « étayée » est donc, au même titre qu'une émotion « montrée », une émotion « inférée ». Cependant, contrairement aux émotions « montrées », l'allocutaire part des « causes de l'émotion (une situation schématisée, c'est-à-dire présentée sous un certain jour) vers l'émotion elle-même (à savoir l'émotion qu'il y a lieu de ressentir face à cette situation) », il s'agit donc d'une inférence « déductiv[e] ».

Pour analyser les discours qui « étayent », les critères suivants et leur réalisation linguistique peuvent être observés : les « personnes impliquées » et leur « rôle » dans la situation ; la « distance (temps et espace) » entre la situation et le locuteur ; les « conséquences » de la situation et leur « degré de probabilité » ; l'attribution causale et agentive » (i.e. quels sont la « cause » et le « responsable »?) ; le « potentiel de maîtrise » (i.e. la situation peut-elle être contrôlée et par qui?) ; la « ressemblance » avec d'autres situations, elles-mêmes associées à des émotions ; et la « signification normative » (i.e. « la situation schématisée est-elle présentée comme compatible ou incompatible avec des valeurs ou des normes entretenues par tel ou tel groupe de référence ? »). Selon l'émotion « étayée », certains critères seront plus ou moins importants. Par exemple, pour la « honte » et la « culpabilité », la « signification normative » sera particulièrement pertinente.

*ex. « C'est qu'à l'heure qu'il est, tout près de nous, là, sous nos yeux, on massacre, on incendie, on pille, on extermine, on égorge les pères et les mères, on vend les petites filles et les petits garçons... » (Victor Hugo, « Pour la Serbie », Actes et Paroles, Robert Laffont, 1985, p. 949.)*

Si aucun terme d'émotion n'est présent, la situation décrite évoque pourtant l'« indignation », l'« horreur », notamment car la situation est présentée comme proche de locuteur (« à l'heure qu'il est », « tout près de nous ») et positionne « les petites filles » et « les petits garçons », individus que la société est censée protéger à cause de leur vulnérabilité, des victimes.

Une émotion « étayée » peut être « auto-attribu[ée] », soit par la combinaison d'un énoncé qui « dit » l'émotion en la reliant au locuteur et d'un énoncé qui « étaye » cette même émotion (*ex. « Je suis en colère parce que ... », « Je flippe complètement parce que ... »*) ; soit de manière « implicite », si le locuteur est aussi « l'énonciateur ». Elle peut également être « allo-attribu[ée] », soit par la combinaison d'un énoncé qui « dit » l'émotion en la reliant à autrui et d'un énoncé qui « étaye » cette même émotion (*ex. « Il est très en colère parce que ... », « Tu flippes complètement parce que ... »*) ; soit de manière « implicite », si le locuteur présente la « schématisation » de la situation « du point de vue d'autrui », qui est alors l'énonciateur.

⇒ Cette typologie, très linguistique, comporte de grandes similitudes avec celle utilisée par Blanc *et al.* (Blanc, 2010 ; Blanc et Creissen, 2017) pour étudier l'impact des émotions et leur mode d'expression sur la compréhension de texte par les enfants (cf. infra 2.1.2.d). Les émotions « désignées » (Blanc, 2010 ; Blanc et Creissen, 2017) sont très proches des émotions « dites » (Micheli, 2014) et les émotions « suggérées » par un événement (Blanc, 2010 ; Blanc et Creissen, 2017) des émotions « étayées » (Micheli, 2014). La différence majeure entre ces deux approches concerne les « expressions comportementales l'émotion », considérées comme une catégorie à part entière dans (Blanc, 2010 ; Blanc et Creissen, 2017) mais incluses dans les émotions « dites » pour (Micheli, 2014 ; p. 51). Les travaux de Blanc *et al.* n'évoquent pas d'éléments se rapprochant des émotions « montrées » de Micheli. Pour notre étude, portant sur l'expression linguistique des émotions dans les textes pour enfants et ayant en point de départ la question de la compréhension de texte par les enfants, nous avons choisi de fusionner ces deux typologies pour obtenir, en quelque

sorte, notre propre catégorisation des modes d'expression des émotions dans les textes. Nous avons ainsi les catégories suivantes :

1. **les émotions « désignées »**, correspondant à des labels émotionnels qui dénotent l'émotion (ex. « Il a peur », « Elle est heureuse ») ;
2. **les émotions « comportementales »**, qui décrivent une manifestation physique d'une émotion (comportement physique) (ex. « il tremble [de peur] », « elle éclate en sanglots ») ;
3. **les émotions « montrées »**, correspondant à des indices textuels révélant l'état émotionnel de l'énonciateur (comportement discursif) (ex. « Oh ! ») ;
4. **les émotions « étayées »**, qui englobent les situations associées de manière socio-culturelle à une émotion (ex. : recevoir une récompense, associé à la joie ou la fierté).

Le tableau ci-dessous récapitule les trois typologies différentes : celle de Micheli (2014), celle de Blanc *et al.* (Blanc, 2010 ; Blanc et Creissen, 2017) et celle que nous avons retenue en combinant les deux précédentes. Chaque catégorie est illustrée par un exemple avec le même contenu prédicatif : l'entité « Paul » ressent de la joie.

	<b>Catégories</b>			
<b>Micheli</b> (2014)	<b>Émotion « dite »</b> ex. Paul est heureux.		<b>Émotion « montrée »</b> ex. « Ah ! Quel endroit merveilleux », dit Paul.	<b>Émotion « étayée »</b> ex. Après un long voyage, Paul arrive enfin dans la maison de vacances de ses rêves.
<b>Blanc et al.</b> (2010 ; 2017)	<b>Émotion « désignée »</b> ex. Paul est heureux.	<b>« Expression comportementale de l'émotion »</b> ex. Paul sourit.		<b>Émotion « suggérée »</b> ex. Après un long voyage, Paul arrive enfin dans la maison de vacances de ses rêves.
<b>Catégories retenues</b>	<b>Émotion « désignée »</b> ex. Paul est heureux.	<b>Émotion « comportementale »</b> ex. Paul sourit.	<b>Émotion « montrée »</b> ex. « Ah ! Quel endroit merveilleux », dit Paul.	<b>Émotion « étayée »</b> ex. Après un long voyage, Paul arrive enfin dans la maison de vacances de ses rêves.

### 2.2.3 . Les émotions et la prise en charge énonciative<sup>14</sup>

Il nous semble que les liens entre la verbalisation des émotions et la notion de prise en charge énonciative gagneraient à être étudiés de manière plus systématique. Cette notion de linguistique, assez complexe et fluctuante, renvoie, de manière globale, à la façon dont un énonciateur transforme un contenu propositionnel (un contenu prédicatif) en un énoncé. Pour opérer cette transformation, le contenu doit être situé dans l'espace-temps (plus ou moins proche du point où se situe l'énonciateur) ; doté d'une valeur aspectuelle (processus fini, en cours etc.) ; doté d'une

<sup>14</sup> Par manque de temps, cette problématique de la prise en charge énonciative n'a malheureusement pu être qu'effleurée dans notre étude. Elle sera approfondie dans le contexte de mes travaux de thèse à venir.

modalité, soit d'une « attitude », un positionnement entre la source d'une information et l'information elle-même (autrement dit : quelle est la source de l'information ? À quel point cette information est-elle vraie ?). Il peut alors être important de différencier le locuteur, celui qui parle ou écrit, et l'énonciateur, celui qui prend en charge l'énoncé, qui est la source du point de vue exprimé. Dans son discours, un locuteur pourra donc faire appel à d'autres énonciateurs, par exemple dans du discours rapporté (Coltier *et al.*, 2009).

Rabatel (2013) a déjà quelque peu abordé la question avec le phénomène d'« empathie linguistique » qui « renvoie au fait qu'un locuteur prête sa voix à un autre [...] pour envisager un évènement, une situation, à sa place ». Il s'agit donc d'un procédé très particulier (vu dans l'étude de (Rabatel,2013) comme un ressort rhétorique pour l'argumentation) où l'émotion ressentie par un énonciateur second, construit par le locuteur/énonciateur premier, est « mise en spectacle » (ce procédé ne recouvre donc pas les énoncés comme « *Paul a peur* » ou « *Paul dit qu'il a peur* »). La question de la prise en charge apparaît également dans l'œuvre de Micheli (2014) pour l'attribution des émotions « montrées » et « étayées » (cf. infra 2.2.2). Si elle n'est pas nommée, l'importance de la prise en charge peut également être devinée dans Leeman (1995), puisque, d'après cette étude, le « N » dans « être en N » (ex. « *être en colère* ») correspond à une sorte de « compte-rendu de perception » exprimé par le locuteur et renvoie à la façon dont l'entité sujet du prédicat « être en N » est perçue par le locuteur.

Pour illustrer l'intérêt d'étudier la relation entre les émotions et la prise en charge, les exemples ci-dessous montre différents énoncés que l'on peut obtenir à partir d'un même contenu prédicatif (*être\_triste(x)* ;  $x=Zoé$ ) en faisant varier le niveau de prise en charge :

- (1) Charles aurait dit que Zoé est triste.
- (2) Charles dit que Zoé est triste.
- (3) Zoé a l'air triste.
- (4) Zoé est triste.
- (5) Que Zoé est triste !

Dans (1) et (2), le locuteur pose une autre entité, un autre énonciateur (Charles) comme la source de l'information « Zoé est triste » et se distancie ainsi de cette information. La distance est plus forte en (1) avec l'emploi du conditionnel « aurait » qui introduit une certaine incertitude de la part du locuteur. De (1) à (5), on a une certitude croissante du locuteur sur l'information « Zoé est triste » ainsi qu'une impression croissante à la fois de l'intensité de la tristesse de Zoé et de l'implication du locuteur (le locuteur paraît plus affecté par cette tristesse en (5)).

⇒ Bien que de nombreux travaux linguistiques aient porté sur la question des émotions dans le langage, très peu se sont intéressés, de manière systématique (i.e. sans se concentrer sur un seul type de structure ou sur l'expression d'une seule émotion) à l'expression des émotions en dehors du lexique des émotions. Ce que nous appellerons par la suite l'étude « hors-lexique des émotions » concernera donc tous les éléments linguistiques ne relevant pas du lexique des émotions mais permettant malgré tout (souvent en étant combinés à d'autres éléments, parfois appartenant eux-mêmes au lexique des émotions) d'exprimer une émotion. Ainsi, les interjections, les mots connotant une émotion ou encore les termes servant à décrire une situation relèvent tous du niveau lexical de la langue mais n'appartiennent pas au lexique émotionnel et pourront donc retenir notre

attention. Enfin, les travaux de linguistique s'intéressant à l'expression des émotions dans les textes pour enfants semblent très rares. (Micheli, 2014) utilise par exemple principalement des extraits de discours politiques ou de littérature classique. Une partie de notre travail consistera donc à vérifier la pertinence des différents critères linguistiques de l'expression des émotions dans notre corpus de textes pour enfants, en vue notamment de repérer ceux pouvant être implémentés dans une chaîne automatique de repérage des émotions.

### 2.3 . Les émotions dans le traitement automatique des langues (TAL)

Si le traitement des émotions en linguistique est déjà problématique, il semble l'être encore plus dans le domaine du traitement automatique des langues (TAL). L'origine de cette difficulté réside probablement, à l'instar d'autres catégories sémantiques, dans la complexité à identifier des éléments linguistiques précis (i.e. bien définis), renvoyant à la catégorie sémantique de l'émotion (si tant est que l'on puisse la désigner à l'aide de ce terme). Complexité qui semble se refléter, une fois encore (cf. section 1.1), dans la diversité des termes employés pour désigner, de près ou de loin, cette catégorie dans le domaine du TAL. Une cartographie<sup>15</sup> des tâches proposées par la campagne internationale d'évaluation SemEval (mise en place annuellement depuis 2007), consacrée à l'analyse sémantique automatique révèle par exemple que les termes « affect », « sentiment » (« *sentiment analysis* » ou analyse de sentiments), « subjectivité » ou parfois même « opinion » (« *opinion mining* » ou fouille d'opinion) sont très largement employés, sans qu'aucune définition théorique explicite de ces termes ne soit proposée. Lors de la description d'une tâche, il est simplement précisé si le but est de retrouver la polarité de textes ou de portions de texte, une position subjective ou une catégorie émotionnelle. On remarque toutefois que ces travaux font le plus souvent référence à la notion de polarité, selon une distinction positif/négatif/neutre (voire pour/contre/neutre). Le terme « émotion » n'est véritablement mentionné que dans la tâche 3 de 2019 nommée « EmoContext : Contextual Emotion Detection in Text », visant à repérer quatre catégories explicites (joie, colère, tristesse et autre). Elle est définie comme correspondant à de la détection d'émotion, discipline vue comme une branche de l'analyse de sentiments selon (Mohammad, 2015), cité dans la description des tâches de SemEval 2016. Cette cartographie (cf. A2) suggère en tout cas que peu de travaux en analyse dite de sentiments sont consacrés à la détection automatique d'émotions en tant que telles. Nous avons identifié dans la littérature quelques travaux qui font référence directement à l'analyse automatique des émotions dans les textes. Ce sont ceux que nous présentons maintenant plus en détail car ils fournissent des éléments potentiellement intéressants sur lesquels nous appuyer pour notre travail.

Par exemple, Mohammad (2011) a pour objectif de « quantifier et de suivre les émotions » (nous traduisons de l'anglais) présentes dans des livres. Pour réaliser cette tâche, il cherche à retrouver automatiquement, dans des contes de fées et des romans en anglais, des émotions spécifiques (nous traduisons de l'anglais ; « joie », « tristesse », « colère », « peur », « surprise », « dégoût », « anticipation/attentes » et « confiance »). Pour cela, il a recours à un lexique émotionnel (le NRC Emotion Lexicon, créé via le « Mechanical Turk » d'Amazon, contenant environ 14 200 mots-types). Un « analyseur sémantique simple » (nous traduisons de l'anglais) utilisant ce lexique a été

---

15 Ce travail a été réalisé dans le cadre du cours de M2 TAL (Univ. Paris 3, Univ. Paris Nanterre, Inalco) « Corpus annotés et développement de ressources linguistiques » assuré par Delphine Battistelli, au premier semestre de l'année 2018-2019.



élaboré : au sein d'un texte donné, le système retrouve les mots appartenant au lexique émotionnel et calcule ainsi le pourcentage de mots émotionnels représenté par chaque catégorie d'émotion (par exemple, si au total 100 mots du lexique ont été retrouvés dans un texte et que, parmi eux, 15 expriment la « joie », cette catégorie représentera 15%). Ce système serait insuffisant pour analyser la dimension émotionnelle au niveau de la phrase et permettrait seulement de déterminer si, par contraste avec d'autres passages, « une large portion de texte a plus d'expressions émotionnelles » (nous traduisons de l'anglais). Aucun résultat précis (mesure de rappel et de précision, f-mesure) n'a été communiqué quant à la capacité du système à retrouver correctement les différentes catégories émotionnelles. Alm *et al.* (2005) ont pour leur part l'ambition de repérer, au sein de textes pour enfants, les passages émotionnels et de les associer à la ou les bonnes catégories émotionnelles grâce à un classifieur automatique (apprentissage automatique supervisé). Six catégories sont visées : « colère », « dégoût », « joie », « peur », « surprise », « tristesse » (nous traduisons de l'anglais). Pour paramétrer le classifieur linéaire, Alm *et al.* (2005) intègrent, en plus des informations sur les termes du lexique émotionnel contenu par les portions textuelles, des informations sur la présence de discours direct, les marques typographiques, les catégories grammaticales des mots et la présence d'interjection. Malheureusement, faute d'avoir un corpus suffisamment grand annoté en émotions spécifiques (i.e. avec les six catégories citées précédemment), les auteurs se contentent de retrouver les passages « émotionnels » vs « non émotionnels » (nous traduisons de l'anglais) et d'en identifier la polarité (positif, négatif, neutre). La distinction entre les passages « émotionnels » vs « non émotionnels » a donné des résultats peu satisfaisants pour la reconnaissance de passages « émotionnels » avec seulement 0,56 de précision et 0,42 de rappel (contre 0,66 et 0,75 pour les passages « non émotionnels » ; les f-mesures étaient respectivement de 0,47 et 0,70). La classification automatique des passages entre neutre, émotionnel positif et émotionnel négatif a obtenu des résultats encore inférieurs : précision de 0,64 et rappel de 0,75 pour les passages neutres (f-mesure de 0,69) ; précision de 0,13 et rappel de 0,19 pour les passages positifs (f-mesure 0,13) ; et précision de 0,45 et rappel de 0,27 pour les passages négatifs (f-mesure 0,32). Les auteurs soulignent la difficulté de trouver des traits pertinents pour reconnaître automatiquement les émotions dans les textes ainsi que la nécessité d'avoir accès à des corpus plus étendus pour tester et développer leur classifieur automatique. Une autre étude intégrant des éléments ne relevant pas uniquement du lexique émotionnel est celui de Chardon *et al.* (2013). Cette recherche constitue l'un des premiers travaux à montrer « empiriquement » (nous traduisons de l'anglais) l'impact des informations discursives (relations discursives ou rhétoriques qui structurent un texte) sur la subjectivité d'un texte, en étudiant des commentaires en ligne (avis de films, commentaires d'articles de presse) en français. Pour annoter les textes avec les informations discursives, la plate-forme Glozz a été employée : c'est également celle que nous utiliserons pour élaborer notre corpus annoté (cf. section 4). Cependant, cette étude porte sur la polarité du texte et non sur la reconnaissance d'émotions spécifiques.

Du côté de la linguistique de corpus enfin, on peut notamment citer les travaux de textométrie réalisés sur plusieurs types de corpus par Eensoo, Valette *et al.* : messages de blogs de santé (Eensoo et Valette, 2012), commentaires d'articles de presse (Valette et Eensoo, 2014) ou encore tweets (Eensoo *et al.*, 2016). L'analyse textométrique et linguistique effectuée leur a permis de dégager des critères ne relevant pas du lexique des émotions, utilisables pour l'analyse de la « subjectivité » (Eensoo et Valette, 2012 ; Eensoo *et al.*, 2016) ou du « discours évaluatif » (Valette et Eensoo, 2014). L'usage des pronoms (utilisation « massive » du « je » dans les messages

« tristes » ; présence importante de l'interlocuteur « tu » dans les messages « gais » ; Eensoo et Valette, 2012) ou l'usage des « marqueurs de structuration argumentative » (plus présents dans les textes « gais » (Eensoo et Valette, 2012) ; de nature différente selon la posture du locuteur vis-à-vis des Roms, (Eensoo et Valette, 2014)) peuvent ainsi être des indicateurs du « sentiment » (Eensoo et Valette, 2012) ou de l' « opinion » (Eensoo et Valette, 2014) exprimés dans les textes. Les mêmes types de critères ont été dégagés dans (Eensoo *et al.*, 2016) pour la classification de tweets selon leur polarité et selon différentes « classes de subjectivité », notamment l' « opinion » (i.e. « l'expression intellectuelle et réfléchie »), le « sentiment » (i.e. « l'expression intellectuelle-affective ») et l' « émotion » (i.e. « l'expression purement affective »). Ces résultats peuvent d'ailleurs être mis en lien avec des études de psychologie (ex. Chung et Pennebaker, 2007) ayant montré que l'utilisation que nous faisons des « mots-outils » peuvent refléter notre état émotionnel ou encore notre statut social. Boulard *et al.* (2012) ont également utilisé les techniques de textométrie sur un corpus de productions orales spontanées d'enfants (136 enfants, âgés de 3 à 13 ans) pour analyser le développement de la « fonction narrative » chez l'enfant. Le corpus a été analysé par le logiciel Cordial et 136 descripteurs morphosyntaxiques et syntaxiques ont été retenus. Ces descripteurs, ainsi que les informations sur le lexique, ont été utilisés pour typer les regroupements textuels obtenus grâce à différentes méthodes statistiques multidimensionnelles (Analyse Factorielle des Correspondances pour analyser le vocabulaire ; Analyse en Composantes Principales et Classification Ascendante Hiérarchique pour l'étude de la morphosyntaxe). Cette analyse des données textuelles a permis de contraster les différents textes entre eux, mettant ainsi en avant des « styles narratifs » différents (un style plus « élaboré », avec des phrases plus longues, des mots plus abstraits, des coordinations et des subordinations etc. ; un style « moins élaboré », avec des phrases courtes, style plus familier avec des pronoms déictiques « tu », « je » et des interjections etc.). Ces styles, influencés par le contexte de production (discussion en groupe avec les parents, seul avec les parents ou entre enfants) et l'âge des enfants, indiqueraient que le développement de la fonction narrative chez l'enfant se stabiliserait vers 10 ans. Si ce dernier travail est quelque peu éloigné de notre thématique directe (émotion dans les textes pour enfants), il permet néanmoins d'une part d'apporter d'autres éléments concernant les capacités langagières des enfants ; et d'autre part d'illustrer le potentiel de l'analyse linguistique et des méthodes textométriques pour caractériser et contraster différentes catégories de textes.

⇒ Il y a donc, à notre connaissance, très peu de travaux en TAL qui portent sur la reconnaissance de différentes catégories d'émotions (alors qu'il existe pléthore de travaux dans le domaine connexe de la fouille d'opinion et de sentiments plus axé, lui, sur la question de la polarité positive/négative/neutre de textes) ou qui intègrent (ou essaient d'intégrer) des éléments ne relevant pas du lexique émotionnel. Ceux s'intéressant de surcroît aux émotions dans les textes pour enfants, en français, sont encore plus rares (nous n'avons trouvé aucune référence correspondant à ces critères). Notre travail sera donc novateur, en cela qu'il s'intéresse à la caractérisation des émotions par des éléments linguistiques autres que les termes d'émotion, dans les textes pour enfants en français. Nous pourrions malgré tout nous inspirer des travaux approchants déjà menés (cités plus haut), en particulier ceux ayant montré que les éléments linguistiques qui sortent du cadre du lexique émotionnel véhiculent des indications précieuses sur la subjectivité (position du locuteur, polarité du texte) des textes.

### 3 . Analyse linguistique

Après une brève caractérisation de notre corpus, nous analyserons plusieurs exemples (extraits de notre corpus) pour voir dans quelle mesure les critères linguistiques d'expression des émotions évoqués dans la section 2 (cf. section 2.2.2) sont pertinents pour l'étude des textes pour enfants. Nous essayerons de dégager ensuite des (pistes) de critères implémentables dans un système de repérage et d'identification automatique des émotions.

#### 3.1 . Le corpus

Les données nécessaires à notre travail ont été récoltées dans le cadre de mon stage de M2, effectué au laboratoire MoDyCo (UMR 7114, CNRS Université Paris Nanterre) sous la direction de D. Battistelli et G. Lecorvé, et financé par le projet PEPS TextToKids (S2IH INS2I, coordonné par G. Lecorvé). Ce projet a pour objectif l'élaboration d'outils et de ressources pour « faciliter la rédaction de textes à destination d'enfants, dans la perspective de leur expliquer des événements d'actualité [...] et dans le respect de leurs capacités linguistiques »<sup>16</sup>. Puisque des études de psycholinguistique ont montré l'importance de la dimension émotionnelle des textes pour leur compréhension, et ce en particulier chez les enfants (cf. section 2.1.2), mon stage a eu pour objet l'étude des émotions dans des textes de la presse pour enfants et notamment la constitution d'un corpus annoté en émotions (cf. section 4). En l'occurrence, les textes sur lesquels nous avons travaillé sont ceux du *P'tit Libé*, un journal d'actualité en ligne (issu du journal *Libération*), s'adressant aux 7-12 ans. Chaque semaine paraît un numéro qui « explique [...] en profondeur un sujet dont on entend parler dans les médias « des grands » », dans l'optique d' « aider [les enfants] à devenir [de] petit[s] citoyen[s] éclairé[s] »<sup>17</sup>. Le premier numéro, en octobre 2015, traitait par exemple de « la crise des migrants »<sup>18</sup>.

Une précédente stagiaire sur le même projet s'est intéressée aux notions temporelles présentes dans les textes du *P'tit Libé* (Bourgouin, juin 2018). Elle rappelle notamment que les textes qui composent le corpus sont d'un genre particulier : ils appartiennent au « discours journalistique ». Ce discours « spécialisé » possède des caractéristiques qui lui sont propres, par exemple le fait d'être « principalement descriptif et explicatif » (Predescu *et al.* (2011)<sup>19</sup>, cité par Bourgouin, 2018). Leur but est d'informer l'allocutaire, « par le biais de descriptions et d'explications d'événements ou de concepts » (ibid.). L'information doit être exprimée de manière « économique » et « accessible », c'est-à-dire adaptée aux connaissances du public visé (Predescu *et al.* (2011), cité par Bourgouin, 2018). Par ailleurs, les relations entre les événements doivent généralement être suffisamment explicites pour ne pas gêner ni biaiser la compréhension.

Une des premières étapes cruciales de notre travail a été d'analyser en détail quelques extraits de notre corpus du point de vue des émotions supposées y être présentes. Nous souhaitons en effet nous assurer que, malgré l'aspect théoriquement plus « descriptif » et « explicatif » du texte journalistique, des émotions seraient bien réalisées. Nous rappelons que nous nous intéresserons ici exclusivement aux émotions ressenties par les personnages et non celles éveillées chez le lecteur

16 <https://sites.google.com/site/delphinebattistelli/projets-contrats-recherche>

17 <https://ptitlibe.liberation.fr/qui-suis-je/>

18 <https://ptitlibe.liberation.fr/migrants-crise,100840>

19 Predescu, E., Gelan, C. (2011) Discours et langage journalistique, *Diacronia*.

(cf. section 2.1.1). Par « personnages », nous entendons ici les différentes personnes mentionnées ou interviewées dans les articles mais également le locuteur-énonciateur premier : P'tit Libé. En effet, les numéros ne sont pas présentés comme émanant de journalistes (aucun article n'est d'ailleurs signé) mais comme des explications sur l'actualité fournies par P'tit Libé, qui se présente comme « l'enfant de *Libération* » (« Salut, moi c'est P'tit Libé ! »<sup>20</sup>). Ensuite, bien que le site *P'tit Libé* soit adressé a priori aux 7-12 ans, nous nous demandons si nous allions trouver des émotions complexes (*ex. honte, fierté*) et/ou des émotions suggérées par des situations, les deux étant plus difficiles à comprendre pour les moins de 10 ans (cf. section 2.1.2.b). Enfin, nous souhaitons vérifier que les différents critères linguistiques d'expression des émotions (termes d'émotions, structures syntaxiques etc.) sont tous représentés dans nos textes.

### 3.2 . Analyse d'exemples

Pour analyser les différents exemples extraits des articles du *P'tit Libé*, nous utiliserons la typologie des modes d'expression des émotions que nous avons dégagée en combinant les travaux de linguistique de Micheli (2014) et ceux de psycholinguistique de Blanc *et al.* (Blanc, 2010 ; Blanc et Creissen, 2017). Cette typologie comporte les quatre catégories suivantes :

1. les **émotions « désignées »**, correspondant aux termes du lexique des émotions.  
*ex. Paul est heureux. | Les araignées effraient Julie.*
2. les **émotions « comportementales »**, incluant les éléments qui décrivent les manifestations physiques des émotions, les comportements adoptés par les personnes qui ressentent une émotion.  
*ex. Paul sourit. | Julie crie et part en courant en voyant l'araignée.*
3. les **émotions « montrées »**, regroupant toutes les caractéristiques des énoncés indiquant que l'énonciateur éprouve une émotion (terme, interjections, structures syntaxiques).  
*ex. « Quelle chance ! », dit Paul. | « Ah ! Non, quelle horreur ! Une araignée ! », dit Julie.*
4. les **émotions « étayées »**, renvoyant à tous les éléments permettant de décrire une situation associée de manière conventionnelle à une émotion.  
*ex. Paul vient de gagner au loto. | La grosse araignée noire s'approche de Julie.*

Lors de l'analyse de nos exemples, nous chercherons également les deux types d'émotion mentionnés dans les travaux de psycholinguistique (Davidson, 2006 ; Blanc et Quenette, 2017) : les émotions « de base » (colère, dégoût, joie, peur, surprise, tristesse) et les émotions complexes (culpabilité, embarras, fierté, jalousie).

#### 3.2.1 . Exemple d'émotion « désignée »

- (1) « Nino a 8 ans, et il est très inquiet pour la planète. »  
*N°82 - 30 novembre au 6 décembre 2018 – La disparition des animaux sauvages - « Nino, 8 ans, veut sauver les animaux »*

20 <https://ptitlibe.liberation.fr/qui-suis-je/>

L'émotion en jeu dans l'exemple ci-dessus est la peur ressentie par « Nino » (i.e. l'entité qui ressent l'émotion). Elle est désignée directement par un terme du lexique des émotions, l'adjectif « inquiet » (dont l'intensité est renforcée par l'adverbe « très ») et est mise en relation de manière syntaxique avec l'entité qui la ressent. « Nino » est en effet repris par le pronom personnel « il », dont « inquiet » est l'attribut du sujet. L'objet de sa peur (« pour la planète ») est également exprimé.

### 3.2.2 . Exemple d'émotion « comportementale »

Dans le numéro consacré à l'Euro de football, un article est consacré à une petite fille, Neyla, qui a été sélectionnée par concours pour tenir la main des joueurs sur le terrain, au début des matches de l'Euro de football. L'exemple suivant raconte sa réaction lorsqu'elle apprend qu'elle a été sélectionnée.

(2) « [...] Neyla a eu du mal à y croire, elle a pleuré et crié de joie. »  
*N°10 – juin 2016 – L'Euro de foot - « Neyla, 10 ans, participe à l'Euro »*

La surprise initiale que ressent Neyla à l'annonce de la bonne nouvelle (elle a « du mal à y croire ») est ensuite remplacée par de la joie. Ce sont d'abord les manifestations physiques et comportementales de l'émotion qui sont décrites (Neyla « a pleuré et crié »). Elle est ensuite désignée directement lorsque le terme « joie » intervient comme modifieur des verbes exprimant ces manifestations comportementales (« pleuré et crié de joie »).

### 3.2.3 . Exemple d'émotion « montrée »

L'article dont est extrait l'exemple ci-dessous est présenté comme la retranscription d'un dialogue entre un frère (Méissane) et une sœur (Lina) à propos des rapports entre les femmes et les hommes. Ils abordent notamment la question du harcèlement et de ses causes. Nous nous intéresserons uniquement à l'intervention de Lina, celle de Méissane servant uniquement à contextualiser la réaction de Lina.

(3) « Méissane : Une fille habillée de façon sexy, si elle se fait harceler, c'est un peu de sa faute. Si elle n'a pas envie de se faire harceler, elle ne porte pas ça. Les hommes ne peuvent pas changer parce qu'ils sont comme ça naturellement.  
Lina : Naturellement ?? Ce n'est pas naturel, c'est toute une éducation, je te le dis ! »  
*N°28 - 20 au 26 octobre 2017 - L'affaire Weinstein et le harcèlement sexuel - « Lina et Méissane discutent des relations filles-garçons »*

Malgré l'absence de terme du lexique émotionnel, la colère éprouvée par Lina vis-à-vis des propos de son frère est assez évidente grâce aux caractéristiques mêmes des énoncés. Lina commence par produire un énoncé averbal (i.e. sans verbe conjugué), retranscrit avec deux points d'interrogation (« Naturellement ?? »), ce qui nous laisse penser qu'il a été prononcé avec la même véhémence qu'un énoncé exclamatif. Le second énoncé est également exclamatif. D'après Micheli (2014), ces deux types d'énoncés (averbal et exclamatif) seraient produits sous le coup de l'émotion : ces particularités syntaxiques deviennent donc des indices de l'état émotionnel dans lequel se trouvait l'énonciateur au moment de l'énonciation, ils montrent que l'énonciateur ressentait une émotion. En l'occurrence, l'émotion spécifique en jeu (la colère) est interprétée d'après le contexte. Lina n'est ouvertement pas d'accord avec son frère (elle reprend un de ses

termes, « naturellement », pour le contredire « pas naturel ») et n'a donc pas l'air de trouver agréable d'entendre que si « une fille habillée de façon sexy [...] se fait harceler, c'est un peu de sa faute ».

#### 3.2.4 . Exemple d'émotion « étayée »

L'exemple ci-dessous est extrait du même article que l'exemple en 3.2.1. Neyla est donc la petite fille qui a été sélectionnée par concours pour donner la main aux joueurs lors de l'Euro de football.

- (4) « Neyla a de la chance : elle va entrer sur le terrain de pendant l'Euro de football. Elle fait partie des 750 enfants qui ont gagné le droit de tenir la main des joueurs avant le début des matchs. »

*N°10 – juin 2016 – L'Euro de foot - « Neyla, 10 ans, participe à l'Euro »*

Dans ce passage, aucune émotion n'est désignée directement (aucun terme du lexique des émotions) et aucun énonciateur ne semble montrer d'émotion (aucune marque d'exclamation, interjection ou structure syntaxique particulière). La situation décrite suffit à elle seule à suggérer la joie ressentie par Neyla (ainsi que les autres enfants qui participeront à l'Euro). Cette émotion est étayée puisque la situation dans laquelle se trouve Neyla est schématisée comme étant une « chance », un « droit » qu'elle a « gagné[s] » : tous ces éléments positifs et en particulier la victoire peuvent communément être liés à de la joie. De plus, grâce à leurs connaissances personnelles (connaissances du monde) les enfants savent que l'Euro de football est un grand évènement et que les footballeurs sont des stars : pouvoir participer à cet évènement et être proche des joueurs est donc un moment très fort, qui fait probablement rêver de nombreux enfants et est ainsi aisément associable à de la joie. Par ailleurs, puisque ce qui arrive à Neyla est présenté comme une « chance », et non comme quelque chose pour lequel elle aurait beaucoup travaillé, par exemple, nous avons choisi d'annoter cet extrait comme étayant de la joie et non de la fierté (cette dernière implique une « satisfaction d'amour-propre souvent fondée », d'après la définition du CNRTL<sup>21</sup>).

#### 3.2.5 . Exemples mêlant différents modes d'expression des émotions

Les exemples des sections précédentes illustraient les différentes modes d'expression des émotions et comportaient donc un ou deux modes d'expression au maximum. Les passages que nous allons maintenant observer nous montreront comment plusieurs modes d'expressions (voire plusieurs émotions) peuvent être mêlés, voire interagir entre eux.

Un numéro du *P'tit Libé* traitait du harcèlement à l'école. L'extrait suivant est issu d'un article dans lequel sont abordées les conséquences de cette pratique, tant pour le harceleur que pour la victime. L'extrait porte précisément sur les effets potentiellement engendrés par le harcèlement chez la victime.

- (5) « Pour les victimes, les conséquences peuvent aussi être importantes. Certaines se replient sur elles-mêmes, pensent que tout est de leur faute et ont honte de parler de leur problème aux adultes. D'autres encore souffrent de dépression : elles n'arrivent pas à dormir, ne mangent plus et s'isolent. Dans les cas les plus graves, les victimes peuvent se blesser volontairement ou se suicider. »

*N°31 - 9 au 16 novembre 2017 – Le harcèlement à l'école - « Pourquoi est-ce grave ? »*

---

21 <https://www.cnrtl.fr/definition/fi%C3%A8re>

Une seule entité ressentant de l'émotion est présente dans cet exemple, « les victimes » du harcèlement mais la palette des émotions qui leur sont attribuées est assez étendue. Dans un premier temps, la « honte » et la culpabilité qu'elles peuvent ressentir à cause du harcèlement sont évoquées et désignées directement (« faute », « honte »). A cause de ces émotions, les victimes « se replient sur elles-mêmes » (manifestation comportementale). Vient ensuite la tristesse profonde, désignée par le terme « dépression » et ses nombreuses manifestations comportementales (ne pas « dormir », ne pas « mang[er] », « s'isol[er] », « se blesser volontairement » ou même « se suicider »). Toutes ces émotions négatives, en particulier la tristesse, sont également étayées par l'ensemble de la situation qui est dépeinte dans ce passage. Les personnes harcelées sont positionnées comme des « victimes » ayant un « problème » dont les « conséquences [...] importantes » les font « souffr[ir] » et pouvant les mener à des actes terribles « dans les cas les plus graves ». Cet exemple montre ainsi comment plusieurs modes d'expressions différents peuvent être mêlés pour « sémiotiser » (Micheli, 2014) les émotions. En outre, il témoigne de la présence d'émotions complexes (en l'occurrence, la honte et la culpabilité) dans le *P'tit Libé*, alors que ces émotions sont supposément plus dures à comprendre pour les enfants de moins de 10 ans (cf. section 2.1.2.b).

Dans le numéro portant sur le mal-logement, un article relate l'histoire d'une famille qui a dû être relogée en urgence car leur ancien immeuble était sur le point de s'effondrer. La petite fille de cette famille, Inès, a été interviewée.

- (6) « [Inès] se souvient des murs qui n'arrêtaient pas de faire « crac ». Ça lui faisait très peur. « La maison allait tomber sur nous... » murmure-t-elle en se blottissant sur le canapé. »  
*N°91 – 1 au 7 février 2019 – Le mal-logement - « L'immeuble d'Inès et d'Adam menaçait de s'effondrer »*

La peur ressentie par Inès est ici exprimée au moyen des quatre modes différents. Directement désignée (« très peur »), l'émotion est également étayée par l'état de l'habitation, tel qu'il est décrit par Inès : les murs émettaient des sons inquiétants (des « cracs ») et « la maison allait tomber sur [Inès et sa famille] ». Ces éléments suggèrent sans peine la peur de vivre dans un logement dangereux. Les points de suspension utilisés pour retranscrire les paroles de la petite fille dans le passage au discours direct (entre les guillemets) montrent sa peur, et ce d'autant plus qu'ils sont combinés à la description de l'attitude apeurée de la fillette (manifestation comportementale). En effet, elle « murmure » ces propos rapportés « en se blottissant sur [son] canapé », comme si le simple souvenir (« se souvient ») de son ancien logement suffisait à l'effrayer. De même qu'en (5), cet exemple montre comment plusieurs modes peuvent être mis au service de l'expression d'une émotion, et notamment comment ils peuvent interagir entre eux (points de suspension et verbe « murmurer »). Il suggère de plus l'importance de savoir qui est l'énonciateur (en l'occurrence que Inès a énoncé les propos entre guillemets) pour retrouver l'entité qui éprouve une émotion (notion de prise en charge énonciative, cf. section 2.2.3).

L'extrait suivant provient d'un autre article du numéro sur le mal-logement, dans lequel le porte-parole de l'association Droit au logement, Jean-Baptiste Eyraud, est interrogé sur les conséquences du mal-logement sur les personnes qui en sont victimes.

- (7) « « Quand il y a des grandes fissures dans les murs d'une maison, ses habitants ont peur en permanence qu'elle s'écroule. Être tout le temps angoissé, ce n'est pas une vie ! » explique Jean-Baptiste Eyraud, la voix pleine de colère. Dans les familles,

cette situation crée des tensions. « A force d'être les uns sur les autres dans un logement trop petit ou d'être confrontés au froid, on s'énerve et on se dispute sans cesse », insiste-t-il. »

N°91 – 1 au 7 février 2019 – *Le mal-logement* - « *Quelles sont les conséquences du mal-logement ?* »

Dans cet extrait, plusieurs entités ressentant de l'émotion sont distinguées. Il est tout d'abord question des « habitants » des logements insalubres et de la peur qu'ils éprouvent à cause de leur conditions d'hébergement (présence de « grandes fissures dans les murs » d'une maison qui menace de « s'écroul[er] »). Ces dernières, tout comme dans l'exemple (5), permettent d'étayer la peur des habitants, qui est également désignée à deux reprises (« ont peur » et « angoissé »). A noter que l'intensité de la peur semble renforcée par l'usage d'adverbiaux temporels qui accompagnent les désignations directes de l'émotion et qui indiquent le caractère durable et récurrent de l'émotion (« [ils] ont peur *en permanence* » et « être *tout le temps* angoissé »). Puis la peur se transforme en colère. Les mauvaises conditions d'habitation, comme la petitesse des logements (« trop petit ») ou leur manque d'isolation (« froid »), « cré[ent] des tensions » (étayage), engendrent de la colère au sein des « familles » : « on s'énerve » (émotion désignée) et « on se dispute » (émotion désignée voire expression comportementale de l'émotion). Un autre adverbial temporel indiquant le caractère durable et récurrent, « sans cesse » (« on se dispute sans cesse »), permet une fois encore de renforcer l'ampleur de l'émotion. La seconde entité éprouvant une émotion dans ce passage est « Jean-Baptiste Eyraud ». C'est lui qui rapporte les conditions terribles dans lesquelles vivent certaines personnes et qui « explique » les conséquences que peuvent avoir ces conditions de vie (peur et colère des « habitants »). Ces faits éveillent en lui une colère, audible dans sa « voix » (l'émotion est désignée dans « la voix pleine de colère »). L'émotion est aussi montrée dans ses propos par l'énoncé exclamatif (point d'exclamation) et l'utilisation d'une structure disloquée (« Être tout le temps en colère, ce n'est pas une vie ! »).

D'après Micheli (2014), les structures disloquées sont à même de montrer l'émotion ressentie par l'énonciateur pour deux raisons principales. Premièrement, elles permettent de réordonner les éléments d'un énoncé par rapport à l'ordre canonique de l'énoncé. Du point de vue informationnel, cet ordre renvoie à la structure thème-rhème, qui correspond généralement à la structure syntaxique Groupe Nominal – Groupe Verbal et à la structure logique sujet-prédicat. En changeant l'ordre des éléments, les dislocations agissent « directement sur la structure informationnelle de l'énoncé » (p. 94) et focalisent ainsi l'attention sur ce qui est perçu comme le plus important par l'énonciateur : le thème (ce dont on parle), avec une dislocation gauche (aussi appelée topicalisation) ; ou le rhème (ce qu'on en dit, soit le prédicat), avec une dislocation droite. Deuxièmement, les structures disloquées donnent l'impression d'un discours plus saccadé, « ce qui correspondrait bien à la dimension irruptive de l'expérience d'une émotion » (p. 95). Cependant, pour Micheli, ce sont les dislocations à droite qui sont les plus à même de montrer l'émotion puisqu'elles focalisent l'attention sur le « prédicat affectif » (ex. « *il est odieux ce mec* » ; p. 96). Néanmoins, dans notre exemple la dislocation à gauche permet de mettre en avant l'objet de l'indignation de Jean-Baptiste Eyraud : le fait d'« être tout le temps angoissé ». Le fait qu'il s'agisse d'une dislocation à gauche n'empêche pas non plus l'impression d'un discours saccadé, donné sous le coup de l'émotion. Nous analysons donc cette structure comme un indice de la colère de l'énonciateur, et ce d'autant plus que le prédicat du noyau n'est pas affectif (« [ne pas être] une vie »). Dans notre exemple, la colère de Eyraud est finalement étayée par le verbe « insister » (« insiste-t-il ») qui retranscrit l'implication



émotionnelle de l'énonciateur dans la situation qu'il rapporte (les mauvaises conditions de vie et leurs conséquences). Cette seconde entité qui ressent de l'émotion est réalisée par deux expressions référentielles (« Jean-Baptiste Eyraud » et « il ») qui la positionnent à chaque fois comme énonciateur des propos au discours direct. Cela permet notamment de lui attribuer l'émotion montrée. L'importance de la notion de prise en charge énonciative (cf. section 2.2.3) pour analyser l'expression des émotions est donc ici perceptible, comme en (6). En plus d'illustrer les interactions entre plusieurs modes d'expression de l'émotion (exclamation, structure disloquée et « voix pleine de colère »), cet exemple souligne aussi les rapports que les différentes émotions entretiennent parfois entre elles (la peur des habitants de logements insalubre provoque la colère de Eyraud).

Le passage (8) offre un autre exemple d'interaction entre plusieurs modes d'expression des émotions (ici, il sera question d'émotions montrées et étayées). Il est extrait du numéro portant sur l'Assemblée Nationale, dans lequel un article détaille le processus de création d'une loi en France. Pour en expliquer les différentes étapes, l'article prend l'exemple d'un projet de loi fictif, visant à « obliger les enseignants à distribuer des bonbons aux élèves chaque matin ».

(8) « Parmi tous les textes de loi qui sont écrits, il faut faire un peu de tri et choisir ceux qui sont plus importants que d'autres. Ouf, notre texte [de loi] sur les bonbons a été retenu ! [...] Une fois qu'ils ont discuté de tous les morceaux du texte, les députés votent pour ou contre. Chouette, les députés ont validé notre texte sur les bonbons ! Maintenant, ils le transmettent à leurs collègues sénateurs, qui peuvent à leur tour le modifier. [...] Bonne nouvelle pour nous : les députés et les sénateurs ont décidé que la distribution de bonbons, c'était une super idée ! »

*N°22 – juin 2017 – L'Assemblée Nationale - « Comment crée-t-on une loi ? »*

Dans cet exemple, la joie du locuteur-énonciateur est montrée par les interjections « ouf » et « chouette » (cette dernière pouvant presque être considérée comme désignant la joie car elle est définie comme marquant « la satisfaction, le plaisir » dans le CNRTL<sup>22</sup>), les énoncés exclamatifs (points d'exclamation) ainsi que la dislocation gauche (« la distribution de bonbons, c'était une super idée »). Comme en (7), la structure disloquée rend compte de la « dimension irruptive de l'expérience de l'émotion » (Micheli, 2014, p. 95). Puisqu'elle est effectuée à gauche, la dislocation devrait mettre en avant le thème, soit « la distribution de bonbon ». Cependant, l'énoncé disloqué est ici inséré dans une proposition subordonnée conjonctive : nous faisons l'hypothèse que, dans ce cas précis, la dislocation à gauche focalise l'attention sur le rhème (« c'était une super idée ») car elle permet de l'isoler à la fin d'un énoncé long (« Bonne nouvelle pour nous : les députés et les sénateurs ont décidé que la distribution de bonbons, *c'était une super idée !* »). Le rhème devient alors l'élément le plus saillant et apparaît être l'élément le plus important aux yeux de l'énonciateur. Par ailleurs, la joie ressentie par le locuteur-énonciateur est étayée tout au long de l'extrait. La création d'une loi est présentée comme un processus difficile (plusieurs étapes bien détaillées), où seuls les textes de loi « plus importants que les autres » sont retenus. La première source de joie est donc la sélection du texte de loi. L'acceptation finale de la loi est décrite comme une « bonne nouvelle », ce qui la pose clairement comme une autre source de joie et ce d'autant plus qu'après tout le chemin parcouru, le projet est qualifié de « super idée ». L'étayage de l'émotion de joie repose donc à la fois sur le fait d'obtenir ce que l'on souhaite (faire passer un projet de loi) et sur la

---

22 <https://www.cnrtl.fr/definition/chouette>

reconnaissance de son travail (le projet est une « super idée » et fait partie des textes les « plus importants »). Dans cet extrait, c'est donc l'étayage de la joie qui permettent d'interpréter les indices qui la montrent. De plus, si en (6) et en (7) l'émotion était montrée au sein de discours direct, dont l'énonciateur était explicitement annoncé, elle se trouve ici dans le corps même de l'article. Cela nous force à associer la joie au locuteur-énonciateur principal, en l'occurrence *P'tit Libé* (et non à la ou au journaliste ayant rédigé l'article puisque les numéros sont présentés comme étant racontés par le personnage de *P'tit Libé*, cf. infra 3.1).

Ce dernier exemple est extrait du numéro traitant de la situation des Rohingyas en Birmanie et plus précisément de l'article qui explique pourquoi ce thème était très présent dans les médias, début 2018.

- (9) « Ça fait très longtemps que les Rohingyas sont maltraités dans leur pays. Mais c'est la première fois que c'est aussi grave. A la fin du mois d'août, des jeunes Rohingyas qui en avaient assez d'être rejetés ont attaqué des postes de police et ont tué dix policiers. Du coup, l'armée et la police ont voulu punir toute la population musulmane très sévèrement. Il y a eu des coups de feu, des morts, des blessés, des pillages, et beaucoup de villages ont été brûlés. »  
*N°43 – 2 au 8 février 2018 – Les Rohingyas, peuple chassés de Birmanie - « Pourquoi parlez-vous de ce peuple ? »*

Comme en (7), plusieurs émotions et plusieurs personnes les éprouvant sont ici présentes. La première entité évoquée sont « les Rohingyas ». Il est d'abord question de leur tristesse, étayée par la description de leur situation en Birmanie. Ils « sont maltraités » et « rejetés », et ce depuis « très longtemps », phénomènes pour lesquels il est communément admis qu'ils sont susceptibles d'engendrer de la tristesse chez les victimes. Cette situation est la cause de la colère ressentie par « des jeunes Rohingyas », désignée directement par la locution verbale « en avoir assez » (cette locution exprime le fait d'être excédé<sup>23</sup>). C'est ensuite cette colère qui a poussé les « jeunes Rohingyas » à « attaqu[er] des postes de police » et à « tu[er] dix policiers » (manifestations comportementales de la colère). Ces actions ont éveillé à leur tour la colère chez « l'armée et la police », soit la seconde entité à éprouver des émotions dans cet exemple. Leur colère se manifeste par la volonté de « punir toute la population musulmane » (manifestation comportementale) et son ampleur est étayée par le fait de vouloir les punir « très sévèrement ». La phrase finale décrit le chaos qui s'ensuit, en énumérant des actions terribles (« coups de feu », « pillages », « villages [...] brûlés ») pouvant être considérées à la fois comme des manifestations comportementales de la colère de « l'armée et de la police », et comme des éléments étayant la peur de « toute la population musulmane ». Les conséquences de ses actions (« morts », « blessés ») suggèrent quant à elles la peur ressentie par les Rohingyas. Cet exemple illustre ainsi le mélange des modes d'expression des émotions (de même qu'en (5), (6), (7) et (8)) et des émotions elles-mêmes (comme en (7)). Il montre également l'importance des émotions pour rétablir les liens logiques, les enchaînements entre les différents événements d'une histoire (un de ces liens de conséquence est d'ailleurs explicitement exprimé par « du coup »).

---

23 [https://fr.wiktionary.org/wiki/en\\_avoir\\_assez](https://fr.wiktionary.org/wiki/en_avoir_assez)

### 3.3 . Quels critères pour le traitement automatique des émotions ?

Grâce à l'analyse des exemples (1) à (9) (cf. infra 3.2), extraits de notre corpus de textes pour enfants, nous avons pu faire un certain nombre de constats et de vérifications, obtenant ainsi des pistes pour l'automatisation du traitement automatique des émotions.

Les quelques exemples étudiés ont confirmé la présence, dans les textes du *P'tit Libé*, des deux types d'émotions évoqués dans la littérature psycholinguistique (Davidson, 2006 ; Blanc et Quenette, 2017) : les émotions de base (colère, joie etc.) et les émotions complexes (culpabilité, embarras etc.). Rappelons que la présence de ces dernières n'était pas certaine puisque les études de psycholinguistique sur la compréhension de textes par les enfants montrent que les moins de 10 ans ont plus de difficulté à les comprendre (cf. section 2.1.2.b) or le *P'tit Libé* est adressé aux 7-12 ans. Le repérage automatique des émotions dans les textes pour enfants ne saurait alors se contenter de chercher les émotions de base et devrait aussi prendre en compte les émotions complexes. L'analyse des exemples (5), (7) et (9) a par ailleurs révélé que plusieurs émotions surviennent parfois ensemble. La peur et la colère se trouvent notamment assez entremêlées dans (7) et (9). Cela suggère que certaines émotions apparaissent plus fréquemment ensemble : la présence de l'une pourrait ainsi indiquer une plus forte probabilité de la présence de l'autre. Reconnaître une émotion de manière spécifique (i.e. pas seulement la polarité mais le nom de l'émotion en jeu) permettrait alors de faciliter l'identification automatique d'autres émotions présentes.

L'étude des exemples a également montré que les modes d'expression des émotions que nous avons exposés (cf. infra 3.1), à savoir les émotions désignées (lexique émotionnel), comportementales (descriptions de manifestations physiques et/ou comportementales des émotions), montrées (structures syntaxiques, interjections etc. qui montrent que l'énonciateur éprouve une émotion) et étayées (description d'une situation suggérant une émotion), sont tous les quatre mis en œuvre dans les textes du *P'tit Libé* pour « sémiotiser » (Micheli, 2014) les émotions. Dans notre travail, nous nous intéresserons prioritairement aux modes ne relevant pas du lexique émotionnel (i.e. les émotions comportementales, montrées et étayées). Cependant, grâce aux exemples (5) à (9), nous avons constaté que les différents modes interagissent, par exemple lorsque des indices d'émotion montrée deviennent interprétables grâce à la coprésence d'une mention explicite de la même émotion en (7). Pour le traitement automatique des émotions, il semble donc préférable de n'écarter aucun mode de « sémiotisation » (Micheli, 2014) puisqu'il apparaît que c'est, parfois, un faisceau d'éléments linguistiques (terme du lexique des émotions, structures syntaxiques particulières, marques typographiques etc.) qui exprime les émotions dans les textes, et devra donc être repéré et identifié pour la reconnaissance automatique des émotions.

Par ailleurs, les caractéristiques des différents modes d'expressions que nous avons observées dans notre corpus concordent avec celles avancées par Micheli (2014). Ces observations nous permettent de dresser une première liste de critères pertinents pour l'étude de la dimension émotionnelle des textes. Ces critères se situent à plusieurs niveaux de la langue :

- **le niveau lexical.** Ce niveau intervient dans les émotions désignées (lexique émotionnel) mais également dans les comportementales (ex. *crier, pleurer*), montrées (interjections) et étayées (ex. *maltraiter, super*). Plusieurs lexiques émotionnels existent déjà (par ex. celui de Piolat et Bannour (2009), EMOTAIX, adapté lors de mon stage sur le projet TextToKids) et peuvent être facilement implémentable dans un système automatique pour traiter les

émotions (cf. section 4). En revanche, il n'existe, à notre connaissance, aucun lexique des expressions comportementales des émotions ou de termes décrivant des situations conventionnellement associées à des émotions. D'après l'analyse des quelques exemples que nous avons effectuée, nous supposons qu'il est envisageable de constituer une collection de termes décrivant des manifestations physiques des émotions (par exemple avec des verbes comme *crier, pleurer, se blottir* etc.), qui pourrait être ensuite utilisée de manière analogue à un lexique émotionnel. De même, une collection des différentes interjections pourraient être établie. Cette tâche paraît en revanche difficilement réalisable pour les termes étayant les émotions car ils semblent beaucoup plus changeant (ex. *une maison qui s'écroule, un village qui brûle, un texte de loi sélectionné* etc.). Seule une analyse plus complète du corpus permettrait de déterminer si certains termes sont très récurrents pour l'étayage des émotions ;

- **le niveau syntaxique.** Essentiel pour les émotions montrées, il implique certaines structures syntaxiques comme les dislocations, les énoncés averbaux, exclamatifs, clivés et elliptiques. Sur ces cinq types de critères avancés par Micheli pour caractériser les émotions montrées (cf. section 2.2.2), trois ont été illustrés par nos exemples (dislocation en (7) et (8) ; énoncé averbal en (3) ; énoncés exclamatifs en (3), (7) et (8)). Ils pourraient être reconnus automatiquement grâce à des parseurs syntaxiques (i.e. des systèmes qui retrouvent automatiquement les relations syntaxiques reliant les éléments d'une phrase) ou au repérage de certaines marques typographiques (les points d'exclamation pour les énoncés exclamatifs) ;
- **le niveau discursif.** Comme l'a montré l'observation des exemples, (5) à (9) notamment, les marques de l'émotion sont parfois disséminées entre plusieurs énoncés. D'autres fois, c'est l'enchaînement de plusieurs énoncés qui constitue la marque de l'émotion (par ex. en (3), avec la juxtaposition d'un énoncé averbal et d'un énoncé exclamatif). L'exploration de ce niveau paraît difficilement automatisable pour le moment car elle nécessiterait d'être en mesure de reconstituer automatiquement les liens entre les différents énoncés (par ex. chaînes de coréférence).

L'étude préliminaire de notre corpus montre ainsi que l'analyse de la dimension émotionnelle des texte mobilise un faisceau de critères linguistiques variés, impliquant les différentes catégories sémantiques de l'émotion (joie, colère, culpabilité etc.), le lexique employé (lexique émotionnel, descriptions des manifestations comportementales des émotions ou de situations associées à des émotions, interjections), les structures syntaxiques des énoncés (dislocation, exclamation, énoncés averbaux) et parfois la structure discursive du texte-même. Cependant, pour affiner nos intuitions sur l'automatisation du traitement des émotions dans les textes, nous avons choisi d'observer la façon dont tous ces éléments sont réalisés à l'échelle du corpus dans son ensemble. Nous les avons donc utilisés pour élaborer un schéma d'annotation et constituer un corpus annoté (cf. section 4).

## 4 . Automatisation de l'analyse linguistique

Dans la section précédente, l'analyse linguistique détaillée d'exemples extraits de notre corpus a confirmé la présence d'émotions, de base ou complexes, dans les textes du *P'tit Libé*. La diversité des marqueurs linguistiques mis en jeu (lexique des émotions, structures syntaxiques, termes décrivant une manifestation physique d'une émotion etc.) a également pu être vérifiée. Après avoir récupéré l'ensemble des fichiers à traiter (section 4.1), nous systématiserons l'analyse des données en construisant un corpus annoté en émotions (section 4.2).

### 4.1 . La récolte des données

Les données sur lesquelles nous avons mené notre travail sont issues d'un journal en ligne, *Le P'tit Libé* (cf. section 3.1). Puisque nous souhaitons obtenir un corpus assez conséquent (cf. section 1.3), nous avons élaboré un script Python 3 pour aspirer automatiquement le contenu textuel des numéros du *P'tit Libé*, plutôt que de les récupérer manuellement un à un. Ce script (`parsePtitLibe.py` ; cf. annexe A4) utilise la bibliothèque Python `requests` pour charger les pages Internet et stocker leur contenu dans des variables, et le module `BeautifulSoup` (de `bs4`) pour lire ce contenu en suivant sa structuration, c'est-à-dire en parsant le HTML (le détail de la démarche suivie est disponible en A5.).

Pour chaque numéro, toutes les informations ainsi obtenues (pour le numéro : identifiant, date de parution, titre, description ; pour chaque article qui le compose : titre, contenu de l'article sous forme de paragraphes) ont été écrites dans un fichier texte (fichier `.txt`). Le script produit donc en sortie un fichier `.txt` pour chaque numéro, ayant pour nom le titre du numéro (ex. « *Les\_migrants.txt* »). Le format de sortie choisi est facilement utilisable dans une chaîne de traitement automatique (la plateforme Glozz propose par exemple de transformer les fichiers `.txt` en fichiers adaptés à la plateforme, cf. infra 4.2) et permet de conserver la structuration en articles des numéros, et celles en paragraphes des articles, ce qui rend les fichiers plus lisibles pour un utilisateur humain et préserve l'organisation discursive des articles (cf. annexe A3).

Notre corpus comporte au final 97 fichiers (format `.txt`), correspondant à 97 numéros du *P'tit Libé*, d'octobre 2015 à mars 2019 (cf. liste des numéros aspirés en annexe A6). Sa taille totale est d'environ **196 426 tokens**, ou 184 314 tokens<sup>24</sup> si on comptabilise uniquement le corps des articles (donc sans les informations sur les numéros (identifiant, date, titre, description) et sans les titres des articles). Cela représente environ 1,10 Mo.

### 4.2 . Construire un corpus annoté en émotions

Constituer un corpus annoté en émotions est une tâche essentielle pour étudier la façon dont les émotions sont réalisées dans notre corpus. L'annotation de nos données constituera une ressource à partir de laquelle nous pourrons examiner plus en profondeur certaines questions concernant

---

<sup>24</sup> Ces nombres ont été obtenus à partir de la concaténation des 97 fichiers, segmentée sur tous les non-caractères de mots (ponctuation, espaces etc.) grâce à l'expression régulière « `\w+` », utilisée dans Python 3 avec la fonction « `split()` ».

l'expression linguistique des émotions. Nous nous intéresserons plus particulièrement au contexte dans lequel les émotions apparaissent afin de déterminer si des critères linguistiques, telles que des structures syntaxiques ou des marques typographiques par exemple, peuvent être associés aux passages textuels exprimant une émotion (de manière générale ou pour une émotion en particulier). Le corpus annoté nous permettra également d'observer d'éventuels liens unissant des émotions si certaines apparaissent souvent ensemble. Ce travail fondamental n'est cependant pas aisé. Toute tâche d'annotation nécessite de définir **les unités** à annoter ainsi que les étiquettes, autrement dit **les catégories** à retrouver. Ces points sont (relativement) facilement résolus pour des tâches où les unités et les catégories peuvent être délimitées clairement car plus ou moins consensuelles, comme dans le cas de l'annotation morphosyntaxique par exemple. Ce n'est cependant pas le cas pour l'annotation sémantique, dont relève l'annotation en émotions que nous souhaitons effectuer.

En effet, la sémantique ne se restreint pas à un seul niveau de la langue. Là où l'attribution d'une étiquette morphosyntaxique se fera à l'échelle du mot graphique uniquement<sup>25</sup>, l'attribution d'une étiquette sémantique peut se faire à plusieurs niveaux puisque le sens ne se limite pas au niveau lexical. Plus spécifiquement, pour l'annotation des émotions, il sera nécessaire d'annoter les termes du lexique émotionnel (« *triste* », « *effrayer* ») mais également des structures syntaxiques (dislocations, énoncés averbaux, énoncés elliptiques, ...), des marques typographiques (points de suspension, points d'exclamation) voire des ensembles d'énoncés, pour traiter le niveau discursif (groupe d'énoncés averbaux, ...).

Par ailleurs, les champs de la sémantique ne sont pas aisément délimitables (les frontières peuvent être floues) ni, a fortiori, en nombre totalement clos. Le temps, l'espace ou la cause constituent par exemple des champs qui font consensus quant à leur existence en sémantique des langues, mais à l'intérieur de ces grands champs les typologies divergent (nombre et nom des sous-catégories). Le champ couvert par l'analyse des émotions est un autre exemple de champ sémantique. Outre le fait qu'il est régulièrement désigné sous des termes différents (notamment celui de sentiments), on peut remarquer qu'il souffre peut être encore plus que les précédents cités d'un consensus sur les sous catégories qui le composent. Si l'on regarde du côté de l'analyse morphosyntaxique, le nombre de catégories est, lui, fini (pour le français : nom, verbe, adjectif, déterminant, préposition, pronom, adverbe, conjonction de coordination, conjonction de subordination, interjection) et chaque catégorie possède des propriétés qui leur confèrent des frontières assez stables. En sémantique, circonscrire des catégories peut être beaucoup plus complexe. En l'occurrence, pour l'étude de la dimension émotionnelle des textes, se pose la question de savoir quelles sont les différentes émotions que l'on souhaite identifier. Définir les émotions et leurs types semble alors crucial mais ce travail définitoire apparaît particulièrement délicat, tant du côté de la psychologie que de la linguistique (cf. section 1.1). Pour notre travail, nous nous contenterons ainsi de reprendre les catégories mises en avant par certains travaux de psycholinguistique, en particulier (Davidson, 2006) et (Blanc et Quenette, 2017) dans l'étude de la

---

25 A noter que dans le cas des expressions polylexicales, comme « pomme de terre » ou « haut de gamme », la question de l'unité à traiter est tout de même problématique puisque l'expression polylexicale fonctionne comme un tout, auquel on peut être tenté d'attribuer une étiquette, par exemple "nom" pour pomme de terre et "adjectif" pour haut de gamme

compréhension de textes par des enfants. Nous considérerons ainsi six catégories d'émotions « de base » (colère, dégoût, joie, peur, surprise et tristesse) ainsi que quatre émotions complexes (culpabilité, embarras/honte, fierté et jalousie) (cf. section 2.1.2.b).

Nous présenterons dans un premier temps le logiciel Glozz, qui présente de nombreux avantages pour une annotation aussi complexe que la notre, et nous expliquerons ensuite la façon dont nous avons procédé pour obtenir les fichiers nécessaires à l'utilisation de Glozz. Enfin, nous détaillerons le schéma d'annotation choisi.

#### 4.2.1 . Présentation du logiciel Glozz

Pour annoter le corpus, nous utiliserons la plateforme Glozz (Widlöcher et Mathet, 2012<sup>26</sup>) qui présente de nombreux avantages. Elle permet en effet :

- de faire des annotations avec des chevauchements ou des emboîtements ;
- de prendre en compte des unités de tailles variées (caractères, mots, phrases, groupes de phrases, paragraphes etc.) ;
- d'annoter les relations qu'entretiennent entre eux les différents éléments, que ce soit les unités, les relations elle-mêmes ou les schémas (unité-unité, unité-relation, relation-schéma etc.) ;
- de créer des schémas pour rendre compte d'interactions plus complexes entre différents éléments (unités, relations ou schémas eux-mêmes).

Ci-dessous, une capture d'écran de l'annotation de l'exemple (9) de la section 3.2 (où est détaillée l'analyse de l'exemple), montrant quelques unes des possibilités offertes par Glozz (unités de tailles variables, chevauchement/emboîtement pour la dernière phrase, relations entre unités ; code couleur disponible à la fin de la section 4.2.3).

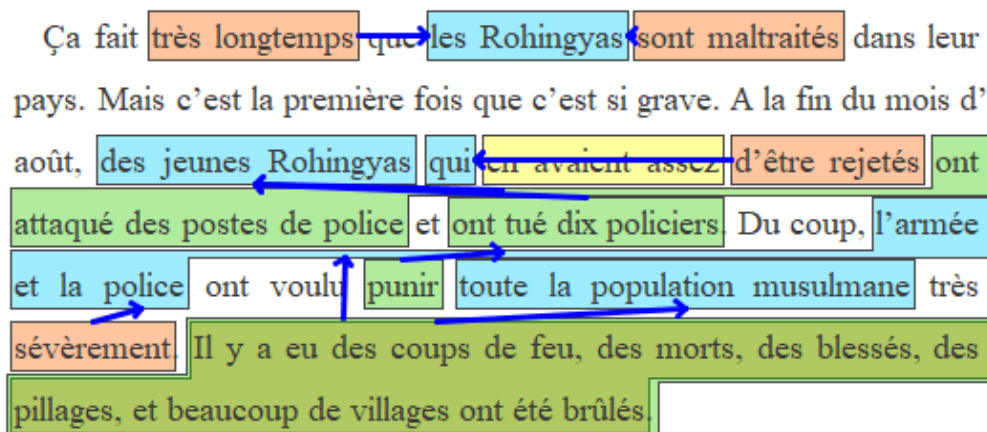


Figure 2 : Exemple d'annotation avec Glozz

Elle se prête donc particulièrement bien à une annotation aussi complexe que celle que nous souhaitons réaliser, où les unités peuvent être de taille variable. La possibilité de relier entre eux des

26 Widlöcher, A. and Mathet, Y. (2012). "The Glozz Platform: A Corpus Annotation and Mining Tool". In Proceedings of the 2012 ACM Symposium on Document Engineering, pp.171-180.

éléments différents nous sera également utile, par exemple pour lier une émotion à l'entité qui la ressent (voire, à plus long terme, à des marqueurs de prise en charge énonciative).

Ce logiciel utilise des formats de fichier bien spécifiques, qui sont détaillés dans le manuel d'utilisation (au demeurant très bien fait). Parmi tous les types de fichiers, seul le .ac n'est pas un document XML :

- **fichier .aam** : décrit le schéma d'annotation à utiliser pour une campagne d'annotation donnée ;
- **fichier .aas** : décrit la feuille de style, nécessaire pour gérer l'affichage des entités définies dans un fichier .aam (il peut y avoir plusieurs feuilles de styles si plusieurs niveaux d'annotation sont différenciés dans le modèle d'annotation, par exemple syntaxique et sémantique) ;
- **fichier .gql** : contient des requêtes dans le langage de requête de Glozz (ces requêtes permettent notamment de rechercher des éléments annotés) ;
- **fichier .aa** : contient les annotations. Il y a un fichier .aa par annotateur ;
- **fichier .ac** : correspond au contenu textuel du fichier à annoter, désormais sur une seule ligne (c'est-à-dire sans retours à la ligne).

Ces deux derniers fichiers peuvent être créés manuellement par le logiciel à partir d'un fichier texte (.txt) mais cette procédure gagne à être automatisée quand il y a de nombreux documents à traiter. Puisque pour notre étude nous avons 97 numéros du *P'tit Libé* à annoter, il semble judicieux d'élaborer un script en mesure de générer automatiquement, à partir d'un fichier .txt, un fichier .ac et un fichier .aa. Les auteurs de Glozz ne proposent pas d'outils pour effectuer cette opération mais les structures des fichiers sont décrites en détail dans le manuel pour permettre justement aux utilisateurs qui le souhaitent de développer leur propre processus d'automatisation. Le schéma d'annotation (.aam) sera construit à la main tandis que la feuille de style (.aas) sera créée via la plateforme, grâce à une interface qui permet d'associer simplement une couleur à un type d'unité (cf. infra 4.2.3). Le système de requêtes proposé par Glozz (requête GQL) sera testé *via* l'interface prévue à cet effet (cf. section 5.2.1).

L'utilisation de Glozz pour constituer notre corpus annoté requiert ainsi d'abord d'avoir un processus automatique pour générer rapidement tous les fichiers (.ac et .aa) nécessaires au fonctionnement du logiciel puis de construire un schéma d'annotation pour savoir quelles informations annoter dans les textes.

#### 4.2.2 . **Élaboration des fichiers pour l'annotation**

Pour générer un fichier .ac à partir d'un fichier .txt, il suffit de supprimer tous les retours à la ligne pour que le contenu textuel du fichier d'origine soit rassemblé sur une seule ligne.

Le fichier .aa dispose quant à lui d'une structure plus complexe, de type XML. Si les balises sont toujours les mêmes, certaines informations sont spécifiques à chaque fichier.



```

<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<annotations> <!--noeud principal-->
  <metadata corpusHashcode="8464-48175182"/> <!--lien avec le fichier .ac-->
  Unité 1 <unit id="aetienne_1561620110278"> <!--id de l'unité-->
    <metadata>
      <author>aetienne</author>
      <creation-date>1561620110278</creation-date>
      <lastModifier>n/a</lastModifier>
      <lastModificationDate>0</lastModificationDate>
    </metadata>
    <characterisation>
      <type>paragraph</type>
      <featureSet/>
    </characterisation>
    <positioning>
      <start>
        <singlePosition index="0"/> <!--début-->
      </start>
      <end>
        <singlePosition index="19"/> <!--fin-->
      </end>
    </positioning>
  </unit>
  Unité 2 <unit id="aetienne_1561620110279">
    <metadata>
      <author>aetienne</author>
      <creation-date>1561620110279</creation-date>
      <lastModifier>n/a</lastModifier>
      <lastModificationDate>0</lastModificationDate>
    </metadata>
    <characterisation>
      <type>paragraph</type>
      <featureSet/>
    </characterisation>
    <positioning>
      <start>
        <singlePosition index="19"/>
      </start>
      <end>
        <singlePosition index="31"/>
      </end>
    </positioning>
  </unit>

```

Figure 3 : Extrait d'un fichier .aa

La première information à générer est le « corpusHashcode », un identifiant qui permet de relier un fichier d'annotation .aa à un fichier .ac. Ce code est composé de deux nombres, séparés par un tiret. Le premier nombre correspond à la taille du fichier .ac, en nombre d'octets. Le second est un "modulo de [son] contenu textuel". Le script Java permettant d'obtenir ce code est fourni dans le guide d'utilisation. J'ai préféré le traduire en Python<sup>27</sup> pour pouvoir l'intégrer directement dans le programme qui constitue les fichiers pour Glozz. L'identifiant produit par la traduction de l'algorithme a été comparé à celui obtenu en créant manuellement le fichier .aa via Glozz. Puisque les résultats étaient identiques, nous avons pu nous assurer que le script Python fonctionnait bien.

Pour faciliter la lecture des documents à annoter manuellement, nous avons décidé de conserver la séparation en paragraphes, qui doivent alors être encodés sous la forme d'unités d'annotation. Ces unités requièrent un identifiant et des bornes de début et de fin. L'identifiant est composé de deux informations : le nom de l'auteur (pour moi, aetienne) et la date de création de l'unité. Cette date de création est exprimée en millisecondes (ms) écoulées depuis le 1er janvier

<sup>27</sup> Les deux versions de l'algorithme sont en annexe A7.

1970 (valeur obtenue aisément en Python grâce à la fonction `time()` du module `time`). Les auteurs ne le précisent pas dans le manuel mais il est important que ce temps de création soit différent pour chaque unité (il suffit qu'il y ait 1 ms d'écart à chaque fois), sinon l'affichage n'est pas rendu correctement dans Glozz. Les bornes de l'unité sont quant à elles liées au fichier `.ac`. La borne de début correspond à l'indice, au sein du fichier `.ac`, du 1er caractère de l'unité, et celle de fin à l'indice du dernier caractère, augmenté de un<sup>28</sup> (une explication technique de la façon dont ces indices ont été trouvés est donnée en annexe A8).

Puisque nous avons une quantité assez importante de fichiers à traiter, nous avons décidé de pré-annoter automatiquement les termes d'émotion de notre corpus, grâce au lexique émotionnel adapté de EMOTAIX (Piolat et Bannour, 2009) constitué pendant mon stage (cf. annexe A9). Pour un fichier `.txt` donné, il s'agit donc d'examiner chaque mot et, si un mot est dans le lexique, de constituer dans le fichier `.aa` une unité d'annotation selon le même principe que pour les paragraphes (identifiant de l'unité et bornes de début et de fin), avec des informations supplémentaires pour encoder le type d'émotion repéré (cf. infra 4.2.3). Le système d'identifiant et de bornes utilisé pour caractériser les unités d'annotation permet la grande liberté de délimitation des unités évoquées précédemment (chevauchement, inclusion) et simplifie, pour nous, l'élaboration des fichiers d'annotation (`.aa`). En effet, grâce à ce système, la succession des unités au sein d'un fichier `.aa` ne doit pas obligatoirement refléter la succession des unités au sein du fichier `.ac` (ou `.txt`) correspondant. Ainsi, nous pouvons créer automatiquement toutes les unités pour les paragraphes puis toutes celles correspondant aux termes d'émotion contenus dans un texte. De manière schématique, le fichier `.aa` généré par notre script sera donc organisé de la manière suivante :

```
<annotations>
  <unit id="aetienne_1562593878266"> 1er paragraphe
    <type>paragraphe</type>
    <positioning>
      <start><singlePosition index="0"/></start>
      <end><singlePosition index="14"/></end>
    </positioning>
  </unit>
  ...
  <unit id="aetienne_1562593878280"> dernier paragraphe
    <type>paragraphe</type>
    <positioning>
      <start><singlePosition index="13394"/></start>
      <end><singlePosition index="16840"/></end>
    </positioning>
  </unit>
  <unit id="aetienne_1562593878281"> 1er terme d'émotion
    <type>Terme d'émotion</type>
    <positioning>
      <start><singlePosition index="168"/></start>
      <end><singlePosition index="175"/></end>
    </positioning>
```

28 La borne finale correspond à l'indice du dernier caractère augmenté de 1, comme cela est le cas quand on veut des « tranches » (*slice*) d'une liste ou d'une chaîne de caractères en programmation. Par exemple, pour une chaîne de caractères de longueur 6, comme « *Python* », les caractères seront numérotés de 0 à 5 inclus. Pour récupérer uniquement les caractères de 2 à 4 inclus, il faudra demander l'intervalle `[2 : 5]`.

```

</unit>
...
<unit id="aetienne_1562593878286"> dernier terme d'émotion
  <type>Terme d'émotion</type>
  <positioning>
    <start><singlePosition index="16592"/></start>
    <end><singlePosition index="16599"/></end>
  </positioning>
</unit>
</annotations>

```

Le script développé (`txt2GlozzFiles_v2.py`, disponible en annexe A10) permet donc, à partir d'un ensemble de fichiers `.txt` et d'un lexique émotionnel donnés, de générer pour chaque document un fichier `.ac` (contenu textuel) et un fichier `.aa` (annotation), utilisables dans le logiciel Glozz. Chaque fichier d'annotation ainsi créé contient, sous la forme d'unités XML, la structure en paragraphes du fichier `.txt` source ainsi qu'une pré-annotation des termes d'émotions présents dans la source, repérés d'après le lexique.

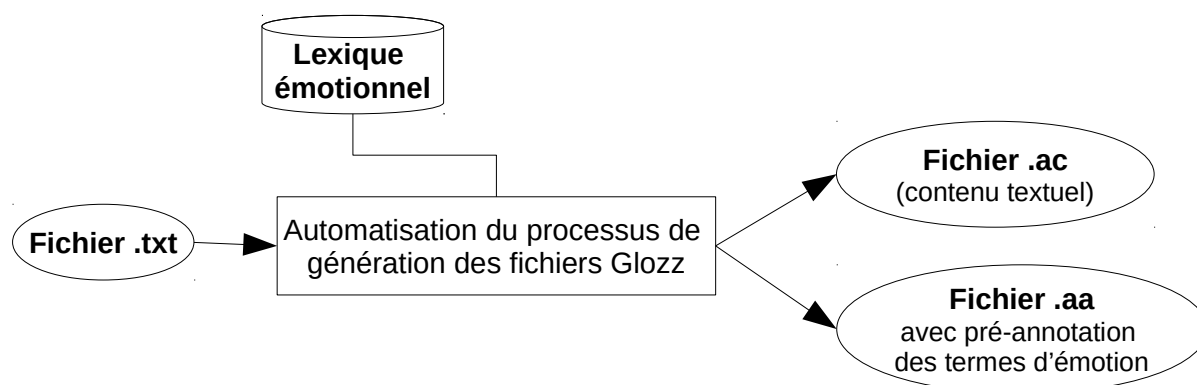


Figure 4 : Schéma de fonctionnement du script `txt2GlozzFiles_v2.py`

Un autre fichier essentiel pour annoter notre corpus via Glozz est le schéma d'annotation.

#### 4.2.3 . Élaboration du schéma d'annotation

Avant de pouvoir annoter les textes, il est nécessaire d'établir un schéma d'annotation, qui prend la forme d'un fichier XML (`.aam`) pour la plate-forme Glozz. Dans ce schéma sont déclarés les différents éléments que l'annotateur pourra délimiter dans le corpus, ainsi que les étiquettes qui leur seront associées. Trois types d'éléments sont définissables dans ce schéma d'annotation Glozz :

1. les **unités** : de taille et de nature variables (mot, phrase etc.), elles correspondent aux portions textuelles délimitées dans le texte par l'annotateur. Elles ont un type (ex. nom) et peuvent avoir des traits (ex. genre, nombre), ayant eux-mêmes des valeurs (ex. masculin, féminin, neutre pour le genre ; singulier et pluriel pour le nombre) ;
2. les **relations** : elles relient deux éléments entre eux, que ce soient des unités, des relations ou des schémas. Elles peuvent également avoir des traits (ex. relation syntaxique) et être orientées ou non (i.e. avoir l'apparence d'une flèche ou non) ;

3. les **schémas** : ce sont des structures plus complexes, mettant en jeu n'importe quel nombre et n'importe quel type d'éléments

Afin d'élaborer le schéma d'annotation, nous devons choisir les informations que nous souhaitons repérer dans le corpus par rapport à nos objectifs d'étude.

Puisque nous nous intéressons à la façon dont les émotions sont exprimées dans les textes, nous annoterons le **mode d'expression** utilisé pour verbaliser une émotion. En combinant les travaux linguistiques (Micheli, 2014) et psycholinguistiques (Blanc, 2010 ; Blanc et Creissen, 2017), nous avons choisi de distinguer quatre modes d'expression (cf. infra 2.2.2) :

1. les émotions « **désignées** », correspondant à des termes du lexique des émotions (label émotionnel) ;
2. les émotions « **comportementales** », traduites par la description d'une manifestation physique d'une émotion (comportement physique) ;
3. les émotions « **montrées** », correspondant à des indices textuels manifestant l'état émotionnel de l'énonciateur (comportement discursif) ;
4. les émotions « **étayées** », qui englobent les situations associées de manière socio-culturelle à une émotion.

Nous souhaiterions savoir si certains types d'émotion sont liés à des caractéristiques linguistiques, des modes d'expression spécifiques (structures syntaxiques, discursives, dénomination...). Par exemple, est-ce que les émotions « de base », dont les manifestations physiques sont un élément définitoire (pour rappel, les émotions « de base », contrairement aux émotions complexes, sont supposées associées à des expressions faciales prototypiques universelles, cf. section 2.1.2.b) sont plus fréquemment exprimées par des expressions comportementales que les émotions complexes ? Seront-elles également plus « montrées », notamment par des interjections (« oh ! », « ah ! »...) ? Pour pouvoir étudier les liens potentiels entre « modes d'expression » et type d'émotion, nous annoterons ainsi le **type d'émotion**, à savoir « de base » ou complexe. Rappelons par ailleurs que cette distinction serait utile pour étudier la compréhension de texte par les enfants (Davidson, 2006 ; Blanc et Quenette, 2017), il sera donc intéressant de voir dans quelle mesure les émotions complexes, théoriquement moins accessibles pour les moins de 10 ans que les émotions « de base », sont représentées dans des textes destinés aux 7-12 ans (les textes de notre corpus, issus du *P'tit Libé*). Nous reprendrons les différentes émotions citées dans les études de psycholinguistique pour nommer plus précisément les émotions repérées. Les **noms d'émotion** suivants seront utilisés :

- émotions « de base » : colère, dégoût, joie, peur, surprise, tristesse ;
- émotions complexes : culpabilité, embarras, fierté, jalousie.

Enfin, nous pensons que le **schéma actanciel** (i.e. qui fait quoi à qui) des segments textuels émotionnels repérés pourra être utile pour l'analyse fine de l'expression des émotions, notamment pour la mettre en lien avec la notion de prise en charge énonciative. En effet, en adaptant le lexique EMOTAIX et en étudiant des exemples tirés de notre corpus, nous avons émis l'hypothèse que **l'entité affectée par l'émotion**, soit la personne ou la chose qui ressent l'émotion, occupe une place centrale dans l'expression et la compréhension des émotions. Cette hypothèse également en lien avec les travaux de Mathieu (2006) sur les verbes d'émotion en français, de Bresson *et al.* (1995) sur les expressions linguistiques de la peur en français (et en allemand) et de Micheli (2014) sur les

émotions « dites » (cf. section 2.2). Nous annoterons ainsi ces entités dans notre corpus, tout en faisant apparaître le lien les unissant directement aux émotions. Plusieurs informations concernant ces entités pourraient être pertinentes pour l'analyse linguistique, comme la **nature du segment textuel** réalisant l'entité affectée. Nous avons également décidé de faire figurer certaines **caractéristiques de l'entité** elle-même, notamment les traits [**±animé**], [**±humain**], [**±collectif**].

- (1) a. Marie est isolée. ⇒ « Marie » : entité [+animé], [+humain]
- b. La maison est isolée. ⇒ « La maison » : entité [-animé] [-humain]
- (2) a. Cette situation fait peur {à on}. ⇒ Entité implicite « {on} » : [+collectif]
- b. On a forcément peur dans ces conditions. ⇒ « On » : entité [+collectif]
- c. Jean a peur. ⇒ « Jean » : entité [-collectif], soit [+individuel]

Ces caractéristiques pourraient donner des pistes pour régler (en partie) le problème de la polysémie, comme en (1) où « isolée » semble prendre un sens émotionnel (tristesse) grâce aux traits [+animé] et [+humain]. Elles pourraient également fournir des pistes pour étudier, par la suite, la prise en charge énonciative (cf. section 2.2.3), comme le suggère l'exemple (2) où, pour un contenu prédicatif similaire (une situation fait peur à une ou des personnes), la formulation change, montrant ainsi plus ou moins l'implication de l'énonciateur par rapport au contenu. L'énonciateur serait plus touché en (2a) et en (2b) qu'en (2c), ce qui pourrait être dû au trait [**±collectif**]<sup>29</sup> de l'entité affectée ou plus particulièrement, l'utilisation du « on ».

Pour encoder de manière plus complète le schéma actanciel des émotions nous aurions pu définir une autre entité correspondant à la cause ou à l'objet des émotions. Par exemple, dans « *X terrorise Y* », Y ressent de la peur, c'est donc l'entité affectée par l'émotion, et X cause cette peur. Cependant, nous avons pensé que la cause (ou l'objet) d'une émotion pouvait ne pas être aussi clairement délimitée et délimitable que l'entité qui la ressent, comme l'atteste l'exemple (3)<sup>30</sup> :

- (3) «La trop grande humidité abîme les poumons des gens», assure Jean-Baptiste Eyraud, le porte-parole de l'association Droit au logement. Certains vivent dans des habitations sans fenêtre, où il y a très peu de lumière et d'aération. Leur corps se fragilise et ils tombent plus souvent malades. Les victimes du mal-logement souffrent aussi souvent de graves problèmes psychologiques. «Quand il y a des grandes fissures dans les murs d'une maison, ses habitants ont peur en permanence qu'elle s'écroule. Être tout le temps angoissé, ce n'est pas une vie !» explique Jean-Baptiste Eyraud, la voix pleine de colère.  
N°91 – 1 au 7 février 2019 – *Le mal-logement - "Quel sont les conséquences du mal-logement ?"*.

La « colère » ressentie ici par « Jean-Baptiste Eyraud » est causée par la situation décrite tout au long du début de l'article et correspond ainsi à une unité textuelle particulièrement longue, difficilement délimitable et qui ne sera pas nécessairement du même ordre dans d'autres passages émotionnels (où la cause pourra être exprimée par un simple syntagme nominal, par exemple). Pour le moment en tout cas, nous avons préféré écarter de notre schéma d'annotation cette information.

Nous proposons à ce stade le schéma d'annotation suivant :

29 Le trait [**±collectif**] est lié à la notion de masse, de générique et non de pluriel. « Les chevaux » dans « Les chevaux sont des mammifères » seront [+collectif] mais [-collectif] dans « Les chevaux de ma grand-mère sont impressionnants. ».

30 On reconnaîtra ici une version étendue de l'exemple (7) de la section 3.2

## Unités :

Les unités sont présentées selon le format :

0. **Type de l'unité** (i.e. le nom qu'elle aura lors de l'annotation)
  - a. Trait 1 : *valeur 1, valeur 2, valeur 3*
  - b. Trait 2 : *valeur 1, valeur 2, valeur 3*
  
1. **Désignée**
  - a. Type d'émotion : *base, complexe*
  - b. Nom de l'émotion : *colère, dégoût, joie, peur, surprise, tristesse, culpabilité, embarras/honte, fierté, jalousie*
  - c. Nature du segment textuel : *adjectif, adverbe, déterminant, interjection, locution verbale, expression polylexicale, nom, nom propre, préposition (syntagme prépositionnel), pronom, verbe*
  
2. **Comportementale**
  - a. Type d'émotion : *base, complexe*
  - b. Nom de l'émotion : *colère, dégoût, joie, peur, surprise, tristesse, culpabilité, embarras/honte, fierté, jalousie*
  - c. Nature du segment textuel : *adjectif, adverbe, déterminant, locution verbale, expression polylexicale, nom, paragraphe, phrase, groupe de phrases, préposition (syntagme prépositionnel), pronom, proposition, verbe*
  
3. **Montrée**
  - a. Type d'émotion : *base, complexe*
  - b. Nom de l'émotion : *colère, dégoût, joie, peur, surprise, tristesse, culpabilité, embarras/honte, fierté, jalousie*
  - c. Nature du segment textuel : *adjectif, adverbe, déterminant, dislocation droite, dislocation gauche, énoncé a verbal, énoncé clivé, énoncé elliptique, exclamation, interjection, locution verbale, expression polylexicale, nom, nom propre, paragraphe, phrase, groupe de phrases, points de suspension, préposition (syntagme prépositionnel), pronom, proposition, verbe*
  
4. **Étayée**
  - a. Type d'émotion : *base, complexe*
  - b. Nom de l'émotion : *colère, dégoût, joie, peur, surprise, tristesse, culpabilité, embarras/honte, fierté, jalousie*
  - c. Nature du segment textuel : *adjectif, adverbe, déterminant, locution verbale, expression polylexicale, nom, nom propre, paragraphe, phrase, groupe de phrases, préposition (syntagme prépositionnel), pronom, proposition, verbe*
  
5. **Entité qui ressent l'émotion**
  - a. Nature du segment textuel : *adjectif, déterminant, expression polylexicale, nom, nom propre, préposition (syntagme prépositionnel), pronom*
  - b. Caractéristique 1 : *animé humain, animé non-humain, inanimé*
  - c. Caractéristique 2 : *collectif, individuel*

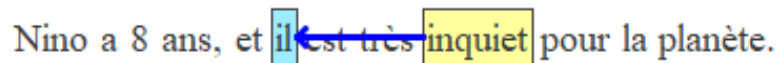
### Relation :

Une seule relation est définie dans notre schéma, celle qui permet de lier une émotion à l'entité qu'elle affecte. Cette relation est orientée de l'émotion vers l'entité.

#### 1. **Affecte**

- a. Relation syntaxique : *sujet, sujet passif, complément direct, complément indirect, complément du nom, attribut du sujet, attribut de l'objet, détermination, épithète, autres*

Pour une phrase comme « Nino a 8 ans, et il est très inquiet pour la planète » (exemple (1) de la section 3.2), on aura, avec l'application de ce schéma d'annotation : « inquiet » comme unité exprimant l'émotion (émotion « désignée », « de base », exprimant la « peur », réalisée par un « adjectif » ) ; et « il » comme entité affectée par l'émotion (« pronom », « animé humain », « individuel »). Il y aura donc une relation qui unira « inquiet » à « il » (relation « attribut du sujet »), comme suit (signification du code couleur en infra 4.2.3) :



Nino a 8 ans, et il est très inquiet pour la planète.

Figure 5 : Exemple d'annotation dans Glozz

Le schéma que nous proposons aborde les émotions en se focalisant en premier lieu sur le mode d'expression, critère jugé à la fois pertinent pour la compréhension de texte par les enfants (Blanc, 2010 ; Blanc et Creissen, 2017) et pour distinguer les émotions d'un point de vue linguistique (Micheli, 2014). Cependant, un angle différent aurait pu être choisi en focalisant l'annotation sur le type d'émotion. Les unités suivantes auraient alors pu être définies :

#### 1. **Émotion de base**

- a. Nom de l'émotion : *colère, dégoût, joie, peur, surprise, tristesse, culpabilité, embarras/honte, fierté, jalousie*
- b. Mode d'expression : *désignée, comportementale, montrée, étayée*
- c. Nature du segment textuel : *adjectif, adverbe etc.*

#### 1. **Émotion complexe**

- a. Nom de l'émotion : *colère, dégoût, joie, peur, surprise, tristesse, culpabilité, embarras/honte, fierté, jalousie*
- b. Mode d'expression : *désignée, comportementale, montrée, étayée*
- c. Nature du segment textuel : *adjectif, adverbe etc.*

#### 2. **Entité qui ressent l'émotion**

- a. Nature du segment textuel : *adjectif, nom etc.*
- b. Caractéristique 1 : *animé humain, animé non-humain, inanimé*
- c. Caractéristique 2 : *collectif, individuel*

Cette seconde proposition de schéma d'annotation aurait permis de mettre en avant une opposition entre des émotions « de base », plus accessibles pour les enfants les plus jeunes, et des émotions complexes, nécessitant des compétences cognitives et socio-culturelles plus poussées

(Davidson, 2006 ; Blanc et Quenette, 2017). Cette seconde version aurait ainsi une approche plus psycholinguistique de la dimension émotionnelle des textes. Étant donné que notre étude s'insère dans les domaines de la linguistique et du TAL, le premier modèle a été jugé plus approprié.

Le schéma d'annotation retenu comporte donc cinq types d'unités : quatre correspondent aux **modes d'expression linguistique** des émotions retenus d'après les travaux de Micheli (2014) et de Blanc *et al.* (Blanc, 2010 ; Blanc et Creissen, 2017) (« désignée », « comportementale », « montrée » et « étayée ») et un renvoie à **l'entité qui ressent l'émotion**. Les quatre unités en rapport avec les modes d'expression présentent toutes les mêmes traits, dont celui de « Nom de l'émotion », qui encode l'**émotion spécifiquement exprimée** par un segment textuel donné. Par exemple, les segments « apeuré », « trembler », « Quelle horreur ! » et « mort » peuvent tous exprimer l'émotion « peur ». Cette structuration - quatre types d'émotions, un trait partagé par tous - peut laisser penser que tous les types d'expression sont aisément associables à des émotions particulières. Cependant, selon le mode de verbalisation considéré, cette association sera plus ou moins évidente et le lien entre l'unité et l'émotion plus ou moins stable. Ainsi, le plus souvent, seules les émotions « désignées » expriment une émotion clairement identifiable : il y a alors un rapport direct entre l'unité linguistique et l'émotion désignée, un rapport de **dénotation**. Les trois autres modes d'expression nécessitent une **interprétation inférentielle**, plus ou moins complexe. Les segments relevant de ces modes ne peuvent donc être liés à des émotions spécifiques que dans un **contexte** donné. Ainsi une même manifestation physique, par exemple un cri, pourra signifier une palette d'émotions très étendue : de la joie (*pousser des cris de joie après une victoire*), de la tristesse (*pousser un cri de détresse après un échec*), de la colère (*crier après quelqu'un qui a cassé quelque chose*), de la surprise (*pousser un cri de surprise en découvrant quelque chose*) ou encore de la peur (*pousser un cri en apercevant un danger*). Le contexte inclut alors **des éléments linguistiques** particuliers (présence d'autres lexèmes comme « détresse », « joie » etc. ou de structures syntaxiques telles que « crier après quelqu'un ») mais également, de manière plus globale, **la situation décrite par le texte** (une victoire, un échec etc.). Si le rapport entre le segment textuel et l'émotion exprimée n'est pas toujours stable, il paraît néanmoins intéressant, pour notre corpus annoté, d'attribuer une émotion précise à ces segments. Cela nous permettra par la suite de remarquer, nous l'espérons, des associations récurrentes, des **schémas prototypiques**, telle que l'association plus fréquente des pleurs à la tristesse, plutôt qu'à la joie.

Un autre trait partagé par les différentes unités est celui de « Nature du segment textuel ». Cette expression, volontairement peu précise, permet d'englober des éléments très variés. Le marquage de cette information est particulièrement intéressant pour les émotions « montrées » puisque cette catégorie très hétérogène inclut aussi bien des mots isolés (*ex. interjections*), que des structures syntaxiques (*ex. dislocations*) ou des groupes de phrases (*ex. enchaînement d'énoncés averbaux*). Bien que ces éléments « montrent » tous potentiellement l'émotion (Micheli, 2014), ils restent de nature différente et nous souhaitons conserver ces particularités pour nos analyses. Une fois l'annotation terminée, nous pourrions fouiller les fichiers .aa (contenant les annotations) pour extraire toutes les émotions « montrées » réalisées par des éléments de même nature, ce qui révélera peut-être des régularités à propos de la verbalisation des émotions (*ex. les structures syntaxiques sont-elles plutôt utilisées pour exprimer certaines émotions, comme la surprise et la colère ?*). Concernant les autres modes d'expression (« désignée », « comportementale » et « étayée »), la nature du segment textuel sera a priori moins pertinente mais nous souhaitons tout de même



encoder cette information pour ne pas passer à côté d'un éventuel lien, par exemple si certaines émotions sont exprimées préférentiellement par des noms ou des verbes.

Par ailleurs, pour rendre plus agréable l'annotation du corpus et sa visualisation, nous avons choisi, via l'interface dédiée dans la plateforme Glozz, un code couleur pour les différents types d'unités : bleu pour l'entité qui ressent l'émotion, jaune pour les émotions désignées, vert pour les comportementales, rouge pour les montrées et orange pour les étayées. Les flèches représentant les relations « Affecte » ont conservé leur couleur par défaut (bleu foncé).

L'objectif de l'annotation du corpus est de croiser les informations émotionnelles (émotion de base ou complexe, nom de l'émotion) avec d'autres types d'informations (mode d'expression, nature du segment textuel) pour montrer que l'expression des émotions ne repose pas uniquement que le niveau lexical. Notre schéma d'annotation intègre de nombreuses informations et nous espérons que sa richesse ainsi que les nombreuses fonctionnalités de la plate-forme Glozz nous permettront de réaliser notre objectif.

## 5 . Résultats

Le schéma d'annotation détaillé dans la section 4.2.3 a été appliqué sur les 97 numéros du *P'tit Libé*, soit sur la totalité du corpus à notre disposition (descriptions de numéro incluses). Nous commencerons par revenir sur quelques difficultés rencontrées lors de l'annotation (section 5.1) puis nous présenterons quelques résultats quantitatifs permettant de dégager des caractéristiques de l'expression des émotions dans notre corpus (section 5.2). En remarque préliminaire, nous souhaitons tout de même préciser que, dans notre corpus, tous les modes d'expression des émotions (désigné, montré, comportemental, étayé), tous les types d'émotion (base et complexe) et toutes les catégories émotionnelles (joie, peur, culpabilité etc.) sont représentés. On remarquera également que tous les numéros annotés comportent au moins une unité émotionnelle.

### 5.1 . Difficultés d'annotation

Lorsqu'on sort des exemples prototypiques (*ex. Paul est heureux. | Les araignées effraient Julie.*), l'annotation de la dimension émotionnelle des textes se révèle souvent complexe, et ce même sur les points fondamentaux que nous avons distingués tout au long de ce mémoire (voir par exemple section 4.2.3) : la séparation entre les émotions des personnages et celles éveillées chez le lecteur ; le repérage de l'entité qui ressent l'émotion et même l'identification précise de l'émotion en jeu.

#### 5.1.1 . La distinction entre émotions des personnages et émotions du lecteur

Dans notre corpus, nous souhaitons annoter les émotions des personnages et non celles éveillées chez le lecteur. Ces personnages sont les personnes interviewées dont les propos sont rapportés, des personnes mentionnées dans les articles etc. voire le locuteur-énonciateur premier, autrement dit le personnage « P'tit Libé », présenté comme celui qui explique aux enfants les différents sujets dont traitent les numéros (cf. section 3.1).

Cette distinction est cependant parfois difficile à faire, notamment à cause du caractère informatif des articles. En effet, le but de ces textes est avant tout d'informer les enfants, de leur expliquer les grands évènements de l'actualité (cf. section 3.1).

- (1) « [En France] on ne mange pas de chat, ni de chien, contrairement à certains pays d'Asie et d'Afrique. En fait, la viande qu'on consomme dépend de la culture du pays dans lequel on habite. Ainsi, par exemple, les Australiens mange du kangourou, les Guinéens du singe et les Péruviens du cochon d'Inde. »  
*N°9 – mai 2016 – De l'animal à la viande - « D'où vient la viande ? »*

Dans l'exemple ci-dessus, on peut supposer que ces quelques lignes suffisent à provoquer du dégoût ou de la tristesse lorsque la consommation de viande de chat ou de chien est évoquée, ou encore de la surprise lorsqu'il est question de manger de la viande de kangourou. Mais nous avons décidé de ne pas annoter ce passage en émotion car, selon nous, l'objectif est purement informatif et aucun « personnage » ne ressent ici d'émotion, puisque aucune marque particulière n'indique un contenu émotionnel (aucun terme du lexique des émotions, aucune structure syntaxique particulière etc.). En revanche, dans l'exemple suivant, issu du même numéro, une émotion a été annotée car

des marques d'émotion accompagnent la référence à la consommation d'un aliment pouvant paraître étrange à des personnes habituées à la gastronomie française.

- (2) « [...] certaines personnes dégustent aussi des insectes ! »  
*N°9 – mai 2016 – De l'animal à la viande - « D'où vient la viande ? »*

Nous avons choisi d'adopter ici le point de vue neutre/explicatif du locuteur-énonciateur premier (P'tit Libé). L'émotion montrée par l'exclamation et suggérée par la situation (manger des insectes) a donc été interprétée comme de la surprise. Ce choix a été conforté par, l'emploi du verbe « déguster » qui empêche, selon nous, de percevoir du dégoût de la part de P'tit Libé. Si nous avions adopté le point de vue du lecteur, nous aurions pu par contre indiquer que ce passage exprime du dégoût.

Tout au long de l'annotation du corpus, nous nous sommes efforcés de toujours prendre le point de vue du personnage (ou de l'énonciateur-locuteur premier) mais il a parfois été impossible de mettre de côté notre position de lecteur et donc les émotions que le texte éveille en nous, comme cela a été le cas dans l'exemple (3).

- (3) « L'affaire Weinstein a encouragé de nombreuses victimes à prendre la parole. En France, une journaliste a lancé sur Twitter, le 13 octobre, un appel à dénoncer les harceleurs et les agresseurs. La consigne était : « Balance ton porc. » Elle a utilisé le mot « porc » pour montrer que les hommes qui font ça agissent comme des animaux qui ne sont pas réputés pour être propres. »  
*N°28 – 20 au 26 octobre 2017 – L'affaire Weinstein et le harcèlement sexuel - « Quelles conséquences a eues la révélation de cette affaire ? »*

Dans cet extrait, l'explication relativement édulcorée, à notre avis, donnée aux enfants quant au choix du terme « porc » ne permet pas de rendre compte du fait que ce mot exprime en fait le dégoût que ressent la journaliste pour les « harceleurs et les agresseurs ». Nous avons donc choisi de d'annoter « porc » comme un terme montrant le dégoût ressenti par la journaliste, même si l'article ne met pas en avant cette interprétation (on notera que notre décision est justifiée par une des définitions de « porc » donnée dans le dictionnaire du CNRTL : « Personne dont la saleté, le comportement physique, l'attitude morale ou intellectuelle suscite un profond dégoût. »<sup>31</sup>).

### 5.1.2 . Le repérage de l'entité qui éprouve l'émotion

De manière plus générale, l'annotation de l'entité qui éprouve l'émotion a posé plus de problèmes que ce que nous avons anticipé. Nous avons en effet pensé que l'entité ressentant l'émotion serait majoritairement exprimée, et associable par un lien syntaxique clair dans le cas des émotions désignées (cf. section 4.2.3).

Pourtant, cette entité n'est pas toujours exprimée, notamment quand elle correspond en fait au locuteur-énonciateur premier.

- (4) « Le studio Ghibli réalise lui-même tous ses films mais ils sont distribués dans les cinémas du monde entier (sauf au Japon) par l'entreprise... Disney ! Décidément, elle est partout. »  
*N°16 – décembre 2016 – L'empire Disney - « Qui sont les concurrents de Disney ? »*

---

31 <https://www.cnrtl.fr/definition/porc>

Dans cet exemple, la surprise est clairement « sémiotisée » (Micheli, 2014). Un effet de suspense est d'abord créé (phrase laissée en suspens avec les points de suspension) puis vient la révélation que Disney distribue les films de son concurrent, le studio Ghibli, à l'étranger. La surprise de cette annonce est marquée par l'exclamation (point d'exclamation) et confirmée par l'utilisation de l'adverbe « décidément » dans la phrase suivante. Tous ces éléments participent ainsi à montrer la (prétendue) surprise qui ne peut être que celle du locuteur-énonciateur premier, P'tit Libé (puisque, dans ce passage, aucune autre entité ne peut prendre en charge cette émotion).

Malgré la présence d'une expression référentielle renvoyant à un candidat pouvant ressentir une émotion, l'incertitude quant à l'entité éprouvant réellement l'émotion exprimée persiste parfois. La question de savoir qui prend en charge un propos paraît alors cruciale. C'est par exemple le cas dans le passage (5), extrait du numéro du *P'tit Libé* consacré aux données personnelles et dans lequel l'exemple fictif d'une petite fille, Inès, est pris pour illustrer les divers moments où on livre ses données personnelles (accepter les cookies, inscription sur les sites, les applications, géo-localisation etc.).

- (5) « Inès se rend compte qu'il y a beaucoup d'informations sur elle en ligne. Elle se demande si ça lui plairait, quand elle aura 17 ans, de revoir des vidéos d'elle à 11 ans en train de danser sur le réseau social Musical.ly. Ou pire, que quelqu'un d'autre tombe dessus... La honte ! »  
*N°52 – 6 au 12 avril 2018 – Les données personnelles - « Où se trouvent nos données personnelles ? »*

Dans cet exemple, il est question de la honte, de l'embarras qu'éprouverait la jeune fille si, plus tard, des personnes venaient à visionner des vidéos qu'elle a mis en ligne à 11 ans. Cette honte est d'abord étayée par la cause de l'émotion (« revoir des vidéos d'elle à 11 ans en train de danser »), qui joue sur le fait qu'en grandissant, on est souvent embarrassé de certaines choses qu'on faisait plus jeune. La honte est ensuite amplifiée (« pire ») par la mention du regard de l'autre (« quelqu'un d'autre »). L'embarras (hypothétique) d'Inès n'est désigné directement qu'à la toute fin de l'extrait par le terme « honte ». Cette émotion est également montrée par l'enchaînement de deux énoncés averbaux (« Ou pire, que quelqu'un tombe dessus » et « La honte »), les points de suspension et l'exclamation qui permettent de recréer, en quelque sorte, le flot des pensées de la jeune fille, dans lequel la sensation de honte monte crescendo. Ces deux derniers énoncés sont pour nous problématiques puisqu'ils peuvent constituer un passage au discours direct libre, dans lequel il est difficile de distinguer la personne qui prend réellement en charge le contenu prédicatif. Si l'on accepte que ces énoncés montrant l'émotion soient une sorte de retranscription des pensées d'Inès (discours direct libre), la jeune fille est l'entité qui éprouve l'émotion. Mais il se pourrait également que ce soit en fait le locuteur-énonciateur premier, P'tit Libé, qui intervienne pour exprimer sa propre opinion sur la situation. D'autant plus que l'objectif de l'article est d'expliquer aux enfants qu'il faut faire attention aux données qu'ils partagent sur Internet. Nous avons finalement attribué l'émotion montrée à Inès (utilisation du lien « Autres » pour la relation « Affecte »), de manière quelque peu arbitraire, optant ainsi pour l'hypothèse du discours direct libre.

Enfin, il arrive que même dans les cas où l'entité et l'émotion sont réalisées par une expression linguistique, sans qu'il y ait de doute sur l'entité qui ressent l'émotion, aucun lien syntaxique ne puisse venir la relier à l'émotion qui l'affecte. C'est par exemple le cas dans l'extrait (6), où il est question de la nouvelle image que Marine Le Pen essaye (en vain) de donner au Front National.

- (6) « Mais dans l'esprit d'une partie de la population, le Front National reste le Front National : qu'il soit dirigé par Jean-Marie ou Marine Le Pen, il fait toujours peur. »  
*N°60 - 1<sup>er</sup> au 7 juin 2018 – Le Front National - « Pourquoi le parti change-t-il de nom ? »*

La peur, désignée ici directement (« il fait toujours peur ») est clairement ressentie par « une partie de la population ». Pourtant, aucun lien syntaxique ne permet de relier les deux segments textuels. Lors de l'annotation, nous avons tous de même uni ces deux unités par la relation « Affecte » définie dans notre schéma d'annotation, la nature de la relation ayant été classée comme « Autres ».

Marine Le Pen fait par exemple attention à ce qu'elle dit, elle essaye de ne pas **choquer**. Mais dans l'esprit d'**une partie de la population**, le Front national reste le Front national : qu'il soit dirigé par Jean-Marie ou Marine Le Pen, il fait toujours **peur**.

Figure 6 : Annotation de l'exemple (6) dans Glozz

### 5.1.3 . L'identification de l'émotion en jeu

Un problème déjà rencontré pendant mon stage sur le projet TextToKids (lors du travail d'adaptation du lexique émotionnel disponible en annexe A9) était de déterminer l'émotion exprimée par certains termes qui impliquent plusieurs individus, chacun pouvant ressentir une émotion différente. Il est alors nécessaire de se demander quelle est la perspective engagée dans le texte pour savoir quelle émotion est exprimée.

Le numéro consacré à l'homophobie (N°88 – 11 au 17 janvier 2019), qui regorgeait d'exemples de ce type, a été pour cela particulièrement difficile à annoter. De nombreuses occurrences des mêmes termes (rejeter, insulter, agresser) parsemaient les articles et mobilisaient plusieurs perspectives pour leur interprétation : celle de la victime, de l'agresseur ou d'un tiers extérieur qui ne fait que constater ce qu'il se passe. La frontière entre les émotions de peur, colère, dégoût, tristesse et embarras voire culpabilité était ainsi souvent floue. Le verbe « agresser », par exemple, renvoie au fait d' « attaquer une ou plusieurs personnes de façon violente et soudaine, et sans être provoqué » (définition CNRTL<sup>32</sup>). D'après cette définition, « agresser » n'est pas un terme du lexique des émotions. Cependant, puisque l'attaquant n'est pas « provoqué » et ne sera ainsi, a priori, pas investi d'une émotion particulière, il est aisé d'imaginer que « agresser » puisse intervenir dans la description d'une situation étayant la peur de la victime, pour qui l'attaque est « violente », « soudaine » et inattendue. Pourtant, ce verbe et ses dérivés (agression, agresseur), présents 14 fois dans l'ensemble du numéro, sont associés à des émotions différentes selon le point de vue adopté. Si l'on prend celui de la victime, « agresser » permet effectivement d'étayer la peur.

- (7) « Les victimes d'**agressions** homophobes sont des femmes et des hommes de tous les âges. Elles ont **été agressées** de jour ou de nuit, à Paris et dans d'autres villes en France. »  
*N°88 – 11 au 17 janvier 2019 – L'homophobie - « Pourquoi en parle-t-on très souvent en ce moment ? »*

32 <https://www.cnrtl.fr/definition/agresser>

Dans l'extrait ci-dessus, les « victimes » sont clairement au centre de l'attention et plusieurs éléments permettent d'étayer la peur qu'elles ressentent : les « agressions » touchent n'importe qui (« des femmes et des hommes de tous les âges »), n'importe quand (« de jour ou de nuit ») et n'importe où (« à Paris et dans d'autres villes en France »). Le rôle même de « victime » qui leur est attribué est conventionnellement associé à la peur. La structure passive (« elles ont été agressées ») permet d'effacer totalement la présence de l'agresseur, focalisant ainsi l'attention sur la victime et renforçant l'impression qu'elles subissent l'action.

Lorsque le point de vue de l'agresseur est privilégié, l'agression peut devenir une manifestation de l'émotion qu'il ressent, de la peur ou du dégoût par exemple. Nous avons opté pour de la peur puisque le P'tit Libé définit l'homophobie comme « la peur des personnes homosexuelles » (article « Qu'est-ce que l'homophobie ? » du même numéro).

(8) « Cette peur [celle des homosexuel.le.s] conduit à du rejet, de la haine, des insultes, de la discrimination, jusqu'à l'**agression** physique et même au meurtre dans le pire des cas. »

*N°88 – 11 au 17 janvier 2019 – L'homophobie - « Qu'est-ce que l'homophobie ? »*

Dans l'exemple (8), les victimes des différents actes évoqués n'apparaissent pas : seuls sont mis en avant les différents comportements qui peuvent être engendrés par l'homophobie, la « peur des homosexuels », et dont l' « agression » fait partie (ainsi que le « rejet », les « insultes », la « discrimination » et le « meurtre »).

Les mots de la famille de « agresser » sont parfois à entendre du point de vue d'une personne tiers, qui porte un regard d'observateur extérieur sur une situation, empêchant l'association du terme à une quelconque émotion, comme le montre l'exemple (9).

(9) « En septembre, [l'association SOS Homophobie] a reçu plus d'un tiers d'appels de victimes en plus par rapport à la même période l'an dernier : « Ça ne veut pas forcément dire qu'il y a plus d'**agressions**. [...] Aujourd'hui, les victimes parlent beaucoup plus et, surtout, elles n'ont plus honte de parler », ajoute Joël Deumier [le président de SOS homophobie] ».

*N°88 – 11 au 17 janvier 2019 – L'homophobie - « Pourquoi en parle-t-on très souvent en ce moment ? »*

Enfin, plusieurs points de vue peuvent simultanément s'appliquer à un même terme. Dans ces cas-là, il n'est pas nécessaire de choisir entre les différentes perspectives puisque la plateforme Glozz permet d'enregistrer plusieurs annotations sur un même segment textuel (créant de cette manière dans le fichier d'annotation deux entités distinctes, ayant les mêmes bornes de début et de fin mais des identifiants différents).

(10) « Depuis plusieurs mois, de nombreuses **agressions** homophobes sont dénoncées sur les réseaux sociaux et dans les journaux. Ça veut dire que des personnes sont **agressées** physiquement (des coups) ou insultées juste parce leurs agresseurs pensent qu'elles sont homosexuelles. »

*N°88 – 11 au 17 janvier 2019 – L'homophobie - « Pourquoi en parle-t-on très souvent en ce moment ? »*

Dans cet extrait, « agressions » intervient pour étayer la colère ressentie par une entité non identifiée dans le texte. Cette colère est également désignée par le verbe « dénoncer ». Le problème de la double perspective concerne le verbe « agresser » (« sont agressées physiquement »).

L'attention est d'abord focalisée sur les « personnes » qui « sont agressées » : l'emploi du passif permet, comme en (7), de mettre l'accent sur la victime, en effaçant l'agresseur. Le verbe « agresser » étaye alors la peur des victimes (tandis que le verbe « insulter » désigne<sup>33</sup> leur embarras/honte). Mais dans la seconde proposition, ce sont les « agresseurs » qui sont mis en avant : on s'intéresse à la raison qui les pousse à commettre ces méfaits. Le fait d'« agresser » une personne parce qu'elle est homosexuelle peut alors être perçu, de la perspective de l'agresseur, comme une manifestation de la peur (comme annoncé précédemment, nous choisissons la peur par rapport à la définition de l'homophobie donnée dans le *P'tit Libé*, cf. exemple (8)).

Depuis plusieurs mois, de nombreuses **agressions homophobes** sont **dénoncées** sur les réseaux sociaux et dans les journaux. Ça veut dire que **des personnes** **sont agressées physiquement (des coups)** ou **insultées** juste parce que **leurs agresseurs** pensent qu'elles sont homosexuelles.

Figure 7 : Annotation de l'exemple (10) dans Glozz

Pour rappel, les entités affectées par une émotion sont en bleu, les émotions désignées sont en jaune, les comportementales en vert et les étayées en orange. Les flèches bleu foncé correspondent à la relation « Affecte » et vont de l'émotion vers l'entité affectée. Les flèches auront ici pour valeur : « Sujet passif » entre « sont agressées physiquement (des coups) », « insultées » et « des personnes » ; « Autres » entre « des personnes sont agressées ... insultées » et « leurs agresseurs ».

#### 5.1.4 . Le cas des émotions étayées

Si la schématisation d'une situation communément associée à une émotion repose parfois sur quelques termes facilement identifiables (ex. « gagner » pour de la joie ; « guerre » pour de la peur), les émotions étayées peuvent s'avérer, à d'autres moments plus dures à délimiter. C'est notamment le cas lorsque la thématique même de l'article ou du numéro est en mesure de suggérer une émotion. Les numéros traitant des attentats, qui expliquent aux enfants ce qu'il s'est passé, ce que ça peut changer pour eux (mesures de l'état d'urgence) et que faire s'ils sont inquiets ou tristes à cause des attaques terroristes (N°3 – novembre 2015 – Les attentats à Paris ; N°11 – juillet 2016 – L'attaque de Nice ; N°84 – 13 décembre 2018 – L'attentat de Strasbourg) ont parfois posé problème pour la délimitation d'émotions étayées : est-ce que décrire les mesures prises consécutivement à un attentat pour protéger la population étaye la peur ressentie par les autorités, la population en général ? D'autres numéros concernés par cette difficulté sont, dans l'ensemble, ceux qui abordent des sujets graves. Par exemple ceux où il est question d'agressions (sexuelles ou physiques), de harcèlement voir de viol (N°16 - décembre 2016 - La pédophilie ; N°28 – 20 au 26 octobre 2017 – L'affaire Weinstein et le harcèlement sexuel ; N°31 – 9 au 16 novembre 2017 – Le harcèlement à l'école ; N°87 - 4 au 10 janvier 2019 - L'homophobie) ou ceux expliquant des conflits armés (N°18 – février 2017 – La guerre en Syrie ; N°43 – 2 au 8 février 2018 – Les Rohingyas, peuple chassé de

33 D'après le CNRTL, insulter correspond à « proférer des paroles, avoir un comportement (interprétable comme) portant atteinte à la dignité de quelqu'un » (<https://www.cnrtl.fr/definition/insulter>).

Birmanie). Des actes, des situations terribles y sont évoquées, décrivant ainsi des scènes dans lesquelles les victimes paraissent en permanence apeurées voire désespérées.

Ce problème de délimitation des émotions étayées confirme les intuitions que nous avons eues, d'une part sur l'implication du niveau discursif dans l'expression des émotions (puisque l'étayage d'une émotion peut se faire tout au long d'un article) ; et d'autre part sur la difficulté de l'automatisation du repérage d'une émotion qui s'étend sur un article tout entier (cf. section 3.3).

En dépit des difficultés soulevées tout au long de cette section, l'annotation du corpus a pu être menée à bien et quelques résultats quantitatifs ont été obtenus.

## 5.2 . Résultats quantitatifs de l'annotation

Une fois le corpus annoté, nous avons souhaité extraire des informations quantitatives. Pour cela, nous pensions tout d'abord utiliser le système de requête proposé par Glozz mais nous avons finalement préféré élaborer un script, à partir duquel ont été obtenus les différents chiffres qui seront présentés.

### 5.2.1 . Test des requêtes dans Glozz

Pour explorer notre corpus annoté, nous avons, dans un premier temps, testé le système de requête de la plateforme Glozz, en utilisant l'interface dédiée. Ce système permet d'exprimer de nombreuses contraintes pour faire des requêtes complexes. On peut par exemple utiliser des expressions régulières pour rechercher un contenu textuel particulier dans les unités ou encore spécifier une certaine distance entre deux unités recherchées. Puisque le cœur de notre travail concerne les émotions exprimées autrement que par le lexique des émotions, nous avons essayé de rechercher toutes les émotions montrées présentes dans un document.

La figure (8) montre l'interface dans laquelle les contraintes sont construites. Plusieurs « boutons » permettent de spécifier différentes caractéristiques que doivent remplir les unités, relations etc. recherchées. En l'occurrence, nous avons simplement spécifié que nous souhaitions retrouver toutes les unités de type « Montree » (notons qu'il s'agit là d'une requête fort simple, par rapport aux possibilités offertes par l'interface).

Une liste de correspondance apparaît ensuite, comme on le voit sur la figure (9). Plusieurs informations concernant les unités retrouvées sont disponibles : les indices de leurs bornes de début et de fin (ce qui permettra de revenir sur le corpus) et leurs identifiants.



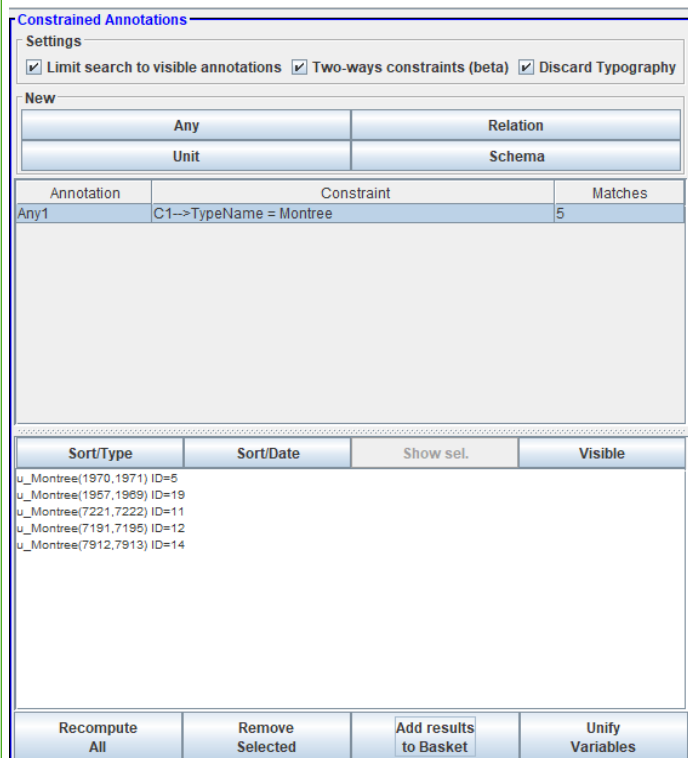
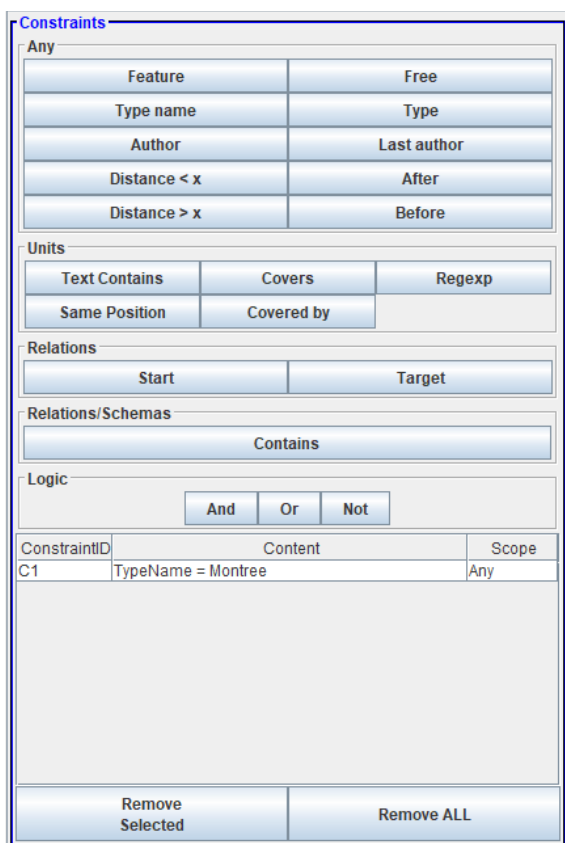


Figure 8 : Interface Glozz pour la création de requêtes Figure 9 : Liste de correspondances pour une requête Glozz

Ces résultats peuvent ensuite être mis dans le « panier » (« Add results to Basket »), ce qui permettra de visualiser, dans le corpus annoté, les différents éléments ayant été retournés par la requête.

La viande n'est pas le seul type d'aliment provenant des animaux.

On mange aussi du poisson (du saumon, du maquereau, de la dorade, par exemple) et des fruits de mer (des huîtres, de la coque, des bulots...). Les poules pondent des œufs ; les vaches, les brebis et les brebis donnent du lait, qui permet de fabriquer du fromage et du beurre notamment. Et ça, c'est pour les aliments les plus sains, mais certaines personnes dégustent aussi des insectes !

Pourquoi mange-t-on de la viande ?

Dans la nature, tous les animaux ont besoin de manger. Mais tout le monde ne mange pas la même chose. Certains se nourrissent que de plantes, on les appelle des végétariens. Les herbivores s'ils ne mangent que de l'herbe et des tiges)

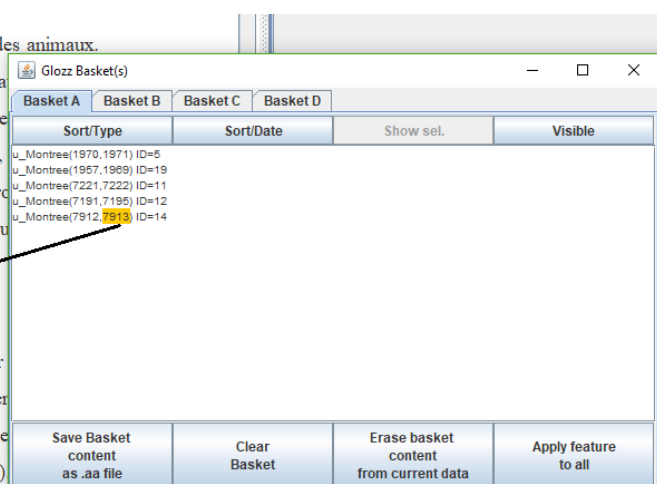


Figure 10 : Visualisation des résultats d'une requête grâce au "panier"

Malheureusement, les résultats d'une requête sont exportables uniquement sous la forme d'un fichier .aa. Ce format d'exportation est pratique pour mettre en évidence seulement certaines annotations, par exemple en créant un nouveau fichier ne contenant que les émotions montrées et les entités qui leur sont liées, ou pour accéder au contexte d'apparition d'une unité, dans l'optique d'une analyse linguistique fine. Un autre inconvénient des requêtes de Glozz est qu'elles ne peuvent pas être effectuées sur plusieurs fichiers en même temps. Or notre corpus est composé de 97 fichiers distincts (pour l'annotation, il y a en fait 97 fichiers .ac et autant de fichiers .aa) : pour rechercher toutes les émotions montrées de notre corpus, il faudrait ouvrir les fichiers un à un et exécuter la requête sur chacun. Les résultats obtenus de cette manière ne serait cependant pas propices à une quantification des émotions montrées puisqu'ils ne pourraient pas être fusionnés (pour mémoire, chaque fichier .aa est lié à un fichier .ac spécifique par un identifiant incluant un modulo de tous les octets qui composent le fichier .ac, cf. section 4.2.2).

Les requêtes de la plateforme Glozz semble donc assez puissantes mais sont plus adaptées à une analyse linguistique fine (accès rapide au contexte des éléments retournés par la requête, possibilité de chercher les annotations correspondants à des contraintes très fortes) ou à la fouille de corpus constitués d'un seul fichier. Pour obtenir nos données quantitatives, nous avons finalement réalisé un script Python, plus adapté à la lecture automatique de nombreux fichiers.

### 5.2.2 . Présentation des résultats quantitatifs

Le script `fouilleGlozz.py` utilise le module `BeautifulSoup` (de `bs4`) pour parcourir l'ensemble des fichiers .aa et récupérer toutes les annotations (le script est disponible en annexe A11). Dans notre schéma d'annotation, nous avons défini une unité pour chaque mode d'expression des émotions, une unité pour les entités affectées par les émotions et une relation pour relier les entités et les émotions. Dans le présent travail, les informations sur les relations et les entités ne seront pas exposées. Elles feront l'objet d'une étude plus poussée lors de ma thèse, où elles pourront être retravaillées, notamment en lien avec la notion de prise en charge énonciative (cf. les difficultés exposées en infra 5.1). Les données quantitatives présentées dans cette section ont été fournies par le script `fouilleGlozz.py`, qui a généré des tableaux .tsv avec le résultat des comptages des unités sur l'ensemble du corpus. Les graphiques ont été obtenus d'après ces tableaux, dans un tableur LibreOffice. Par ailleurs, des tableaux avec le nombre d'unités annotées par numéro sont disponibles en annexe (cf. annexes A13 et A13).

Au total, **2 043 unités d'émotion** ont été délimitées dans notre corpus (de 196 426 tokens). Parmi ces unités, les deux modes d'expression majoritaires sont les émotions désignées (40,3%) et les émotions étayées (40,1%). Les émotions comportementales représentent 11,6 % des unités et les montrées seulement 7,9 % (cf. figure 11).

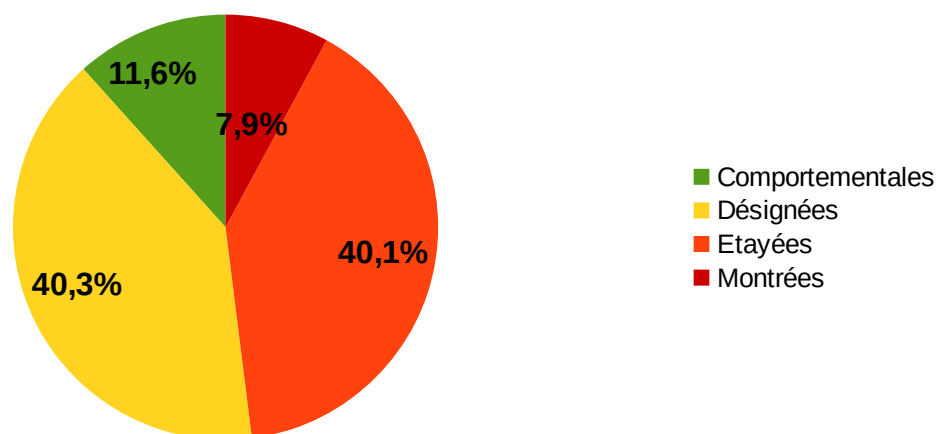


Figure 11: Répartition des émotions entre les différents modes d'expression

Sur l'ensemble des émotions repérées dans le corpus, 89,9 % sont des émotions de base et seulement 10,1 % d'émotions complexes (cf. figure 12). Bien entendu cette différence s'explique en partie par le nombre plus élevé de catégories d'émotions de base (6 émotions de base contre 4 complexes). Pour tenter de limiter ce biais, nous avons calculé le nombre moyen d'unités par catégorie émotionnelle, selon le type d'émotion. Pour les émotions de base, une catégorie aura en moyenne  $1837/6 = 306,2$  unités, contre  $206/4 = 51,5$  unités pour les complexes. Une catégorie d'émotion de base comportera donc en moyenne près de six fois plus d'unités qu'une catégorie d'émotion complexe. Les émotions de base sont ainsi bien plus représentées que les émotions complexes.

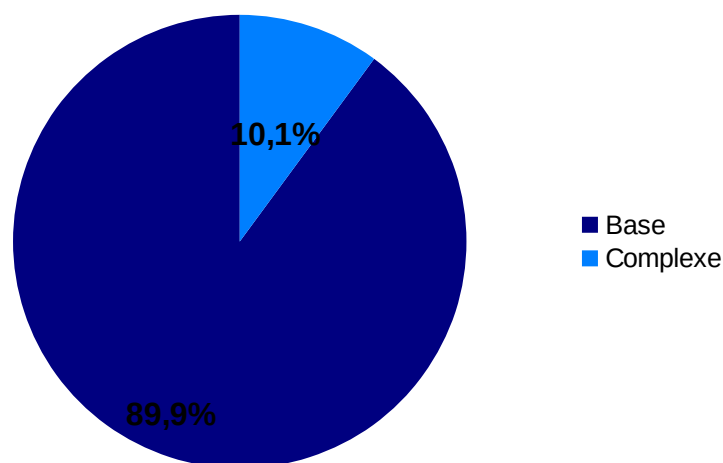


Figure 12: Proportion des différents types d'émotion dans le corpus

Parmi les 10 émotions spécifiques que nous avons souhaité retrouver dans les textes, la peur est la plus représentée (31,4%). Viennent ensuite la colère (21,4%) et la joie (15,8%). La surprise et la tristesse sont présentes dans des proportions similaires (respectivement 10,4 % et 10,2%). Les émotions complexes les plus fréquentes sont l'embarras (4,4%) et la fierté (4,2%). La culpabilité

(1,4%) devance la dernière émotion de base, le dégoût (0,8%). L'émotion la moins présente dans le corpus est la jalousie (0,1%). Ces différentes proportions sont présentées en figure 13.

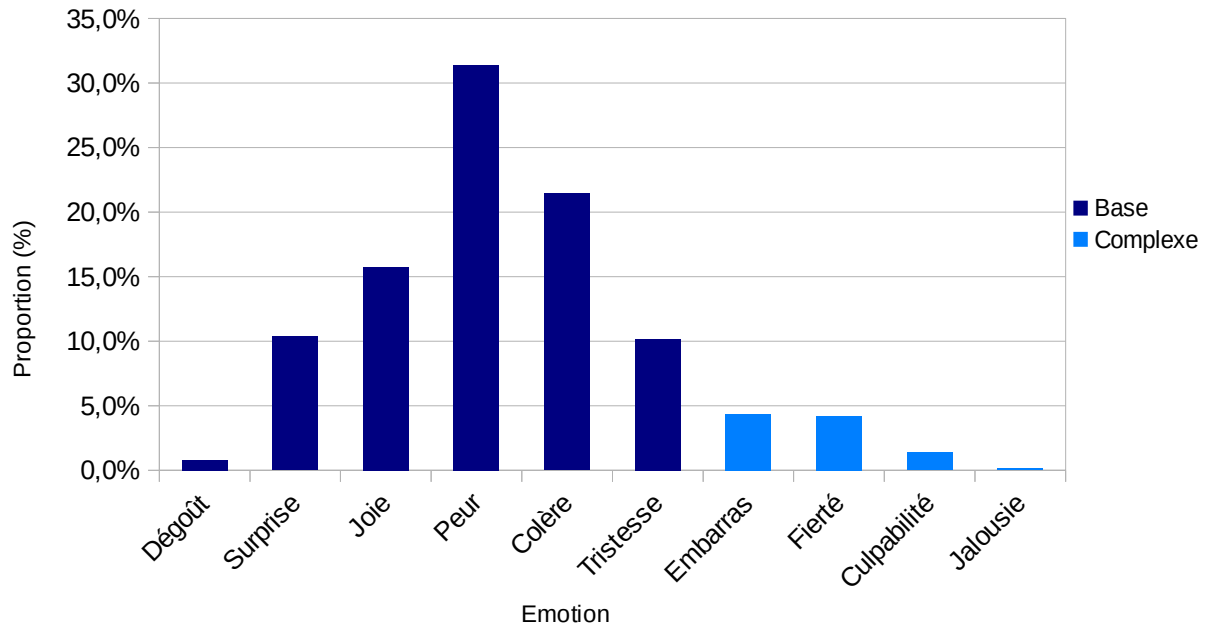


Figure 13: Proportion des différentes catégories émotionnelles

Ces premiers résultats généraux montrent que les émotions de base sont bien plus représentées que les émotions complexes dans notre corpus (près de 90%). Les différentes catégories d'émotion n'apparaissent pas non plus en proportions égales : près d'un tiers des unités annotées renvoient à de la peur, un peu moins d'un quart à de la colère et le dégoût, l'embarras, la fierté, la culpabilité et la jalousie ne dépassent pas les 5 %. La seule émotion positive, la joie représente à peine plus de 15 % des unités (la surprise et la tristesse sont quant à elles autour de 10%). Les unités ne sont pas non plus réparties de manière homogène entre les modes d'expression. Les émotions montrées sont minoritaires (à peine 8%) tandis que les émotions désignées et étayées représentent à elles deux 80 % des unités (environ 40 % chacune). Moins de 12 % des unités sont des émotions comportementales.

La figure 14 montre la proportion des types d'émotion (base ou complexe) pour chaque mode d'expression. Sans surprise, les émotions de base sont toujours majoritaires, et ce de très loin (de 87,6 % à 97,9%). On retrouve donc, pour chaque mode d'expression, la même tendance quand dans le corpus global.

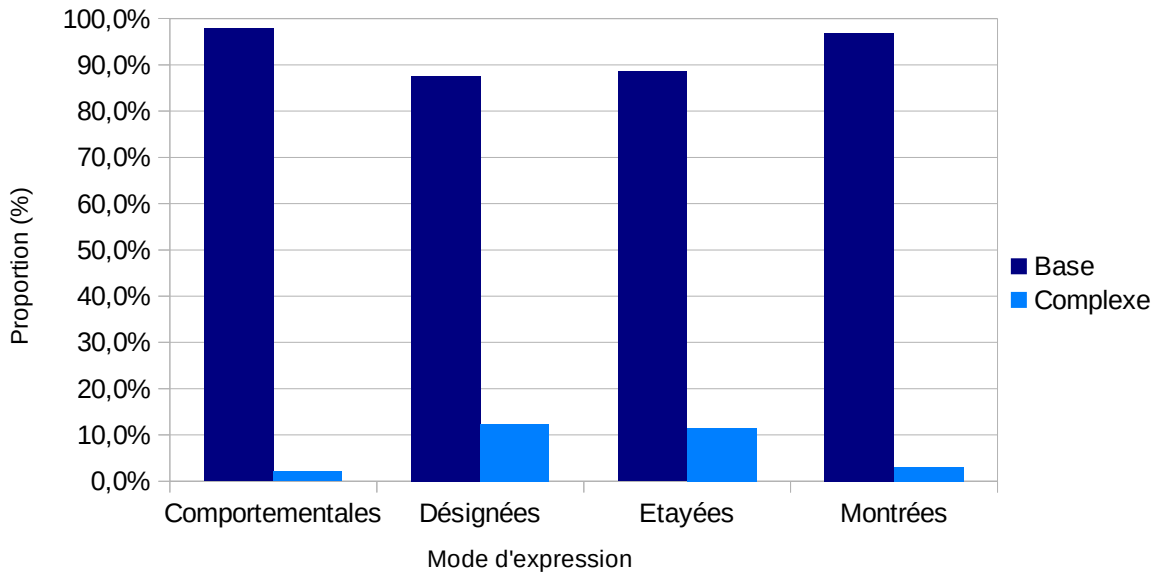


Figure 14: Proportions des types d'émotion selon le mode d'expression

En revanche, lorsqu'on regarde la répartition des émotions selon les modes d'expression (cf. figure 15), la peur n'est pas toujours l'émotion la plus représentée, contrairement à ce que l'on observe sur la totalité du corpus.

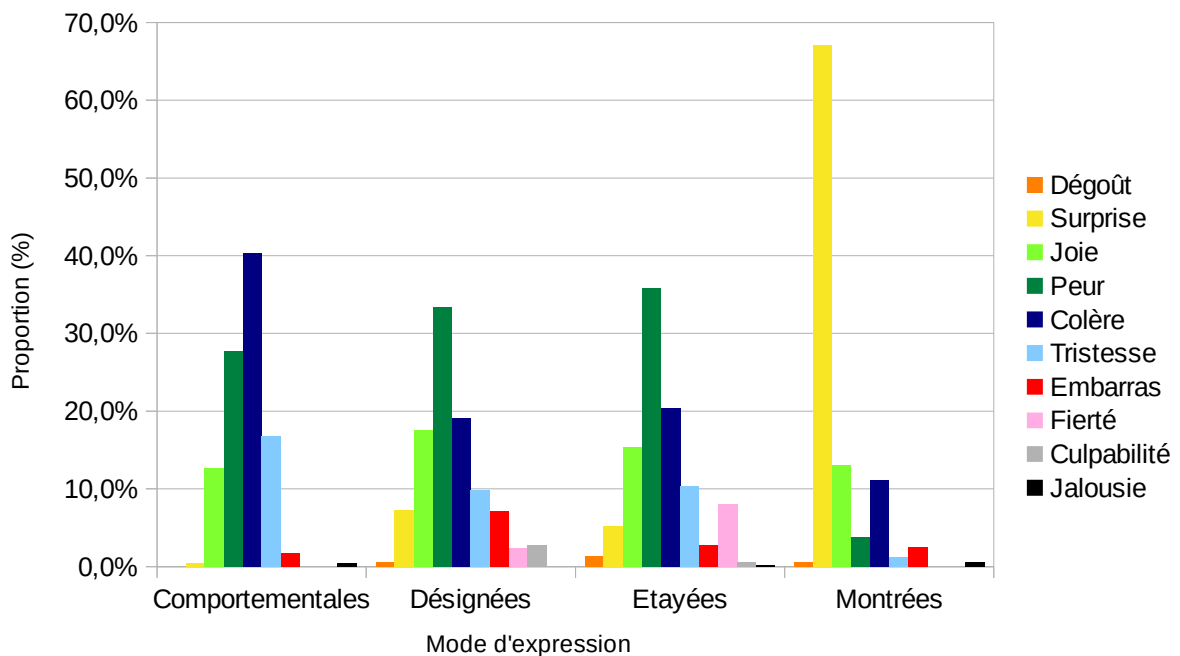


Figure 15: Proportion des différentes catégories d'émotion selon les modes d'expression

Pour les émotions désignées et étayées, la répartition des différentes émotions est similaire à celle constatée sur l'ensemble du corpus. On retrouve toujours en tête la peur (33,4 % des désignées et 35,9 % des étayées), la colère (19,1 % des désignées et 20,4 % des étayées) et la joie (17,6 % des désignées et 15,4 % des étayées). Par contre, la tristesse (9,8 % des désignées, 10,4 % des étayées) dépasse cette fois la surprise. Cette dernière est plus importante parmi les émotions désignées (7,3%) que parmi les étayées (5,2%), où elle est même dépassée par la fierté (8,0%). Dans les émotions étayées comme dans le corpus entier, les émotions les moins représentées sont l'embarras (2,7%), le dégoût (1,3%), la culpabilité (0,6%) et la jalousie (0,1%). Contrairement à ce qui était observé dans l'ensemble du corpus, l'embarras (7,2%) est à peu près autant présent que la surprise (7,3%). Les émotions les moins représentées sont toujours la culpabilité (2,8 %), la fierté (2,4 %), et le dégoût (0,5 %). Malgré quelques variations, la répartition des catégories émotionnelles pour les émotions désignées et les étayées suit donc globalement celle des catégories émotionnelles sur l'ensemble du corpus. Ce n'est par contre pas le cas pour les émotions comportementales et les montrées.

L'émotion la plus exprimée par des manifestations comportementales est la colère (40,3%). La peur ne représente que 27,7 % et la tristesse a cette fois une proportion plus élevée que la joie (16,8 % contre 12,6%). La surprise est très peu présente, seulement 0,4 %, et a la même proportion que la jalousie. L'embarras les dépasse (1,7%). Les émotions montrées « sémiotisent » (Micheli, 2014) quant à elle en grande majorité la surprise (67,1%). Loin derrière se trouve ensuite la joie (13,0%) et la colère (11,2%). La peur ne représente plus que 3,7 % des catégories émotionnelles exprimées, à peine plus que l'embarras (2,5%). La tristesse tombe également, avec seulement 1,2 %. Le dégoût et la jalousie sont toujours parmi les moins présentes (0,6%).

Le tableau 1 donne les valeurs brutes résultant du comptage des différents types d'unités délimitées dans le corpus.

	Comportementales	Désignées	Etayées	Montrées	Total
<b>Dégoût</b>	0	4	11	1	16
<b>Surprise</b>	1	60	43	108	212
<b>Joie</b>	30	145	126	21	322
<b>Peur</b>	66	275	294	6	641
<b>Colère</b>	96	157	167	18	438
<b>Tristesse</b>	40	81	85	2	208
<b>Embarras</b>	4	59	22	4	89
<b>Fierté</b>	0	20	66	0	86
<b>Culpabilité</b>	0	23	5	0	28
<b>Jalousie</b>	1	0	1	1	3
<b>Total</b>	238	824	820	161	2043

Tableau 1: Données brutes

⇒ Ces premiers résultats quantitatifs nous apprennent donc que la majorité des émotions exprimées sont des émotions de base (sur l'ensemble du corpus et au sein de chaque mode d'expression). Les modes d'expression des émotions les plus fréquents sont les émotions désignées et les étayées (env. 40 % chaque), puis viennent les comportementales (env. 12%) et enfin les montrées (env. 8%). Les catégories émotionnelles les plus présentes sur l'ensemble du corpus sont la peur (env. 1/3), la colère (env. 1/4), la joie (env. 15%) et enfin la surprise et la tristesse (env.

10 % chacune). Ce sont également, dans cet ordre, les cinq émotions les plus représentées parmi les émotions désignées et les émotions étayées mais ce n'est plus le cas pour les deux autres modes d'expression. Parmi les expressions comportementales, la colère arrive en tête (env. 40%), suivie de la peur (env. 28%). La tristesse (env. 17%) est plus présente que la joie (env. 13%). La surprise ne représente cette fois plus que 0,5 % (même score que la jalousie et moins que l'embarras, à env. 2%). Enfin, les émotions montrées servent très majoritairement à exprimer de la surprise (env. 67%). La deuxième catégorie émotionnelle la plus exprimée par les émotions montrées est la joie (env. 13%), suivie de la colère (env. 11%). La peur ne représente plus qu'environ 4 % des émotions. Ces différences de répartition des catégories émotionnelles selon le mode d'expression considéré suggère que certains modes sont utilisés de manière privilégiée pour exprimer certaines émotions.

Dans le schéma d'annotation que nous avons utilisé, la nature des segments textuels était enregistrée pour chaque unité (cf. section 4.2.3). Toujours grâce au script, nous avons listé, pour les émotions montrées et les émotions étayées, toutes les natures associées aux unités. Nous avons également compté le nombre d'occurrence de chacune des natures. Les tableaux 2 et 3 montrent les résultats de ces comptages.

	Valeur brute	Pourcentage
<b>Exclamation</b>	109	67,7%
<b>Adverbe</b>	11	6,8%
<b>Dislocation gauche</b>	3	1,9%
<b>Interjection</b>	7	4,3%
<b>Points de suspension</b>	15	9,3%
<b>Énoncé averbal</b>	10	6,2%
<b>Groupe de phrases</b>	2	1,2%
<b>Nom</b>	1	0,6%
<b>Énoncé clivé</b>	3	1,9%
<b>Total</b>	161	100,0%

*Tableau 2: Nature des segments textuels des émotions montrées*

	Valeur brute	Pourcentage
<b>Proposition</b>	278	33,9%
<b>Verbe</b>	115	14,0%
<b>Nom</b>	245	29,9%
<b>Adjectif</b>	76	9,3%
<b>Phrase</b>	46	5,6%
<b>Paragraphe</b>	15	1,8%
<b>Adverbe</b>	11	1,3%
<b>Préposition (SP)</b>	22	2,7%
<b>Groupe de phrases</b>	7	0,9%
<b>Locution verbale</b>	5	0,6%
<b>Total</b>	820	100,0%

*Tableau 3: Nature des segments textuels des émotions étayées*

Nous pouvons ainsi constater que les émotions montrées sont majoritairement réalisées par des structures syntaxiques particulières : l'exclamation (67,7%), les énoncés averbaux (6,2%), les dislocations à gauche (1,9%) et les énoncés clivés (1,9%) représentent au total 77,6 % des segments qui montrent l'émotion (87,0 % si on ajoute les points de suspension). Les marques lexicales (adverbes (6,8%), interjections (4,3%), noms (0,6%)) représentent environ 11,8 % des unités d'émotion montrée et les marqueurs discursif (groupes de phrases) seulement 1,2 %.

Les marqueurs des émotions étayées sont plus diversifiés. Les marqueurs relevant du niveau lexical, les verbes (14,0%), les noms (29,9%), les adjectifs (9,3%), les adverbes (1,3%), les prépositions (2,7%) et les locutions verbales (0,6%) représentent environ 57,8 % des unités. Environ 39,5 % des émotions étayées sont exprimées au niveau de la phrase (proposition, 33,9% ; phrases entières, 5,6%). Enfin, certaines unités sont réalisées au niveau discursif par des groupes de phrases (0,9%) voire des paragraphes entiers (1,8%) mais cela touche une minorité d'unités (env. 2,7%).

⇒ L'observation de ces quelques données quantitatives nous permet donc de confirmer que les structures syntaxiques particulières (exclamation, énoncés averbaux, dislocations, énoncés clivés) sont des critères pertinents pour étudier la façon dont les émotions sont montrées dans les textes. Pour les émotions étayées, l'utilisation massive de termes du lexique laisse espérer qu'il serait possible de dresser une liste partielle de termes étayant les émotions. Après un bref retour sur le corpus, nous avons en effet pu constater que, par exemple, 7 occurrences du terme « guerre » étayant une émotion suggèrent toutes de la peur ; toutes les occurrences de « injuste » et « injustice » étayant une émotion (12 au total) correspondent à de la colère. Les émotions étayées font cependant beaucoup appel à des unités se trouvant au niveau de la phrase ou au niveau discursif (env. 42,2%) ce qui confirme une fois encore (cf. infra 5.1.4) la difficulté de l'étude automatique de ce mode d'expression des émotions (cf. section 3.3).



## 6 . Conclusion et perspectives

L'objectif de notre travail était de trouver, à partir d'un corpus de textes de presse jeunesse (97 numéros du *P'tit Libé*, soit environ 196 426 tokens), des critères pour caractériser l'expression linguistique des émotions. Ces critères devaient respecter deux contraintes : 1) ne pas relever du lexique émotionnel et 2) être (potentiellement) implémentables dans une chaîne de reconnaissance automatique des émotions.

Afin d'étudier la dimension émotionnelle de nos textes, nous avons retenu de la littérature psycholinguistique deux types d'émotions (de base et complexe) et dix catégories émotionnelles (colère, dégoût, joie, peur, surprise, tristesse, culpabilité, embarras, fierté et jalousie) utilisés dans les recherches sur la compréhension de textes par les enfants (cf. section 2.1.2). Nous avons également constitué notre propre typologie des modes d'expression linguistique des émotions en combinant celle proposée en psycholinguistique par Blanc *et al.* (Blanc, 2010 ; Blanc et Creissen, 2017) et celle proposée en linguistique par Micheli (2014) (cf. section 2.2.2). Notre typologie comporte quatre modes d'expression des émotions : les émotions désignées (termes du lexique émotionnel), les émotions comportementales (descriptions de manifestations physiques et comportementales des émotions), les émotions montrées (caractéristiques des énoncés indiquant l'état émotionnel de l'énonciateur) et les émotions étayées (description d'une situation conventionnellement associée à une émotion).

Dans un premier temps, l'analyse linguistique fine d'exemples issus du *P'tit Libé* a confirmé la présence dans le corpus, et donc la pertinence, des types d'émotions, catégories émotionnelles et modes d'expression des émotions que nous avons retenus (cf. section 3.2). En illustrant différents critères hors-lexique émotionnel, ces premiers exemples ont permis de formuler plusieurs hypothèses pour l'analyse automatique des émotions. Certaines structures syntaxiques typiques des émotions montrées (dislocations, énoncés averbaux etc.) seraient reconnaissables par un analyseur syntaxique automatique. Une collection de termes pourrait être constituée pour le repérage des émotions comportementales (*ex. sourire, crier* etc.) (cf. section 3.3).

L'analyse linguistique a ensuite été systématisée grâce à l'élaboration d'un schéma d'annotation intégrant les différentes catégories et les critères mis en œuvre lors de la première étape d'analyse (cf. section 4.2.3). Son application sur notre corpus *via* la plateforme Glozz a conduit au repérage de 2 043 unités exprimant des émotions dans les textes du *P'tit Libé*. Ces unités encodent des informations concernant le type d'émotion (de base ou complexe) et la catégorie émotionnelle (joie, colère, culpabilité etc.) réalisés, ainsi que le mode d'expression de l'émotion (émotions désignées, comportementales, montrées et étayées). La nature des segments textuels émotionnels (i.e. des unités) était également enregistrée (nom, adverbe, énoncé averbal etc.). On remarquera que, malgré l'aspect théoriquement plus « descriptif » et « explicatif » des textes journalistiques évoqué en 3.1, le nombre d'unités émotionnelles annotées dans les textes du *P'tit Libé* laisse supposer que les émotions sont en fait très présentes dans le langage, et ce même dans des genres textuels a priori moins marqués émotionnellement. Le retour sur quelques difficultés rencontrées au cours du processus d'annotation (cf. section 5.1) et l'analyse quantitative des caractéristiques des 2 043 unités émotionnelles délimitées (cf. section 5.3) ont servi à attester la pertinence de critère pour la caractérisation de l'expression linguistique des émotions et à solidifier nos hypothèses pour

l'automatisation de l'analyse des émotions. Au niveau lexical, en dehors des termes du lexique des émotions, certains mots indiqueraient la présence de passages émotionnels : les termes décrivant une manifestation physique ou comportementale d'une émotion (*ex. crier, pleurer, manifester contre* etc.) ; ceux entrant dans la description d'une situation conventionnellement associée à une émotion (*ex. guerre, harceler, gagner* etc.) ; et enfin, les interjections (*ex. ouf, ah* etc.) et certains adverbes (*ex. même, décidément* etc.). Certaines structures syntaxiques sont aussi mobilisées pour exprimer les émotions, comme les dislocations, les énoncés averbaux ou les exclamations. Si ces critères lexicaux et syntaxiques nous semblent assez aisément implémentables dans une chaîne de traitement automatique de la dimension émotionnelle d'un texte, cela n'est pas le cas pour les critères relevant du niveau discursif, notamment lorsque la description d'une situation à même de suggérer une émotion est réalisée à l'échelle de plusieurs phrases, d'un paragraphe entier ou de l'ensemble du document. L'observation quantitative des unités d'émotion de notre corpus suggère par ailleurs l'utilisation privilégiée de certains modes d'expression pour réaliser certaines émotions. Les émotions montrées exprimaient en effet majoritairement de la surprise, de la joie et de la colère, tandis que l'expression comportementale était le plus souvent associée à de la colère, de la peur et de la tristesse. Les émotions désignées et étayées traduisaient quant à elles principalement la peur, la colère et la joie, ce qui correspond aux trois catégories émotionnelles les plus observées sur l'ensemble du corpus. Enfin, l'observation linguistique menée sur notre corpus nous semble souligner l'importance de la notion de prise en charge énonciative pour l'étude de l'expression des émotions.

Le travail que nous avons présenté ici n'est qu'un premier pas dans l'exploration de la problématique que nous avons posée.

Nous souhaiterions perfectionner l'annotation de notre corpus en faisant une campagne d'annotation, à l'issue de laquelle nous pourrions calculer un accord inter-annotateur (i.e. calculer à quel point les différents annotateurs ont annoté les mêmes unités et de la même manière). La qualité du schéma d'annotation pourrait ainsi être vérifiée. Si les annotateurs ne sont jamais d'accord, cela indiquera par exemple que la tâche d'annotation est trop complexe ou mal définie. De plus, avec une campagne d'annotation, les résultats obtenus gagneraient en objectivité. Actuellement, le corpus n'a en effet été annoté que par une seule personne, les résultats de notre travail sont donc potentiellement biaisés par la subjectivité de cette personne. Notons que la plateforme Glozz se prêterait particulièrement bien à une campagne d'annotation puisque, pour un même contenu textuel (fichier .ac), il peut y avoir autant de fichiers d'annotation que l'on veut. Et grâce au système d'identification et de délimitation des unités, il serait facile de voir si deux annotations d'un même contenu textuel concordent ou non. Par ailleurs, l'utilisation d'un analyseur syntaxique automatique pourrait augmenter la partie pré-annotée du corpus en repérant notamment l'entité qui ressent l'émotion (dans les cas où elle est liée syntaxiquement au terme d'émotion). Nous pourrions aussi nous en servir pour repérer certaines constructions syntaxiques montrant l'émotion (dislocations, énoncés averbaux etc.) et désambiguïser des termes du lexique des émotions (*ex. « être fier de quelqu'un » vs « se fier à quelqu'un »*).

Notre travail pourrait être poursuivi en trouvant d'autres critères de l'expression linguistique des émotions (hors lexique émotionnel, bien entendu). Pour cela, une analyse linguistique visant à approfondir la question du lien entre les émotions et la prise en charge énonciative semble

pertinente. L'importance de cette notion de prise en charge a en effet pu être illustrée pour retrouver l'entité qui ressent l'émotion et parfois même pour identifier l'émotion précise en jeu (cf. section 5.1). Maintenant que nous disposons d'un corpus annoté en émotions, nous pourrions également utiliser les outils de la textométrie pour contraster entre elles différentes portions du corpus (par ex. les passages renvoyant à de la colère vs ceux renvoyant à de la joie, etc.), en nous inspirant des travaux de Eensoo et Valette (2012). D'autres critères pourraient ressortir de ces contrastes (par ex. utilisation de conjonctions ou de temps verbaux particuliers dans les passages renvoyant à de la tristesse, etc.).

Enfin, l'étude que nous avons menée nous permet d'imaginer les premiers jalons d'une chaîne de traitement automatique des émotions. Un texte passant par cette chaîne devrait d'abord subir quelques pré-traitements classiques : découpage en mots ou tokens, en phrases et en paragraphes ; étiquetage-morphosyntaxique (i.e. nature des mots) ; et analyse automatique des relations syntaxiques (notamment pour avoir les informations sur les structures disloquées, les énoncés averbaux etc.). Les termes de différents lexiques pourraient ensuite être repérés : lexique émotionnel ; termes liés aux expressions comportementales de l'émotion ; termes étayant fréquemment des émotions ; et interjections (pour les émotions montrées). Des règles utilisant les éléments cités précédemment (mots de lexiques, structures syntaxiques) pourraient ensuite être utilisées pour identifier automatiquement les passages émotionnels d'un texte.

## Bibliographie

- Baron-Cohen, S., Golan, O., Wheelwright, S., & Granader, Y.** (2010). Emotion word comprehension from 4 to 16 years old: a developmental survey. *Frontiers in evolutionary neuroscience*, 2, 109.
- Blanc, N.** (2010). *Lecture et habiletés de compréhension chez l'enfant*. Dunod.
- Blanc, N.** (2010). La compréhension des contes entre 5 et 7 ans: Quelle représentation des informations émotionnelles?. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 64(4), 256.
- Blanc, N., & Quenette, G.** (2017). La production d'inférences émotionnelles entre 8 et 10 ans: quelle méthodologie pour quels résultats?. *Enfance*, (4), 503-511.
- Boulard, A., Poudat, C. & Gauthier, J.-M.** (2012). Des mots pour se dire : développement de la fonction narrative chez l'enfant. Conf. JADT 2012
- Bourgouin, C.** (juin 2018). « Les notions temporelles dans la mise en récit d'événements dans le discours journalistique enfantin. »
- Bresson, D., & Dobrovol'skij, D.** (1995). Petite syntaxe de la « peur ». Application au français et à l'allemand. *Langue française*, 107-119.
- Chardon, B., Benamara, F., Mathieu, Y., Popescu, V., & Asher, N.** (2013, March). Measuring the effect of discourse structure on sentiment analysis. In *International Conference on Intelligent Text Processing and Computational Linguistics* (pp. 25-37). Springer, Berlin, Heidelberg.
- Chatar-Moumni, N.** (2013). L'expression verbale des émotions: présentation. *Langue française*, (4), 3-11.
- Chung, C., & Pennebaker, J. W.** (2007). The psychological functions of function words. *Social communication*, 1, 343-359.
- Coltier, D., Dendale, P., & De Brabanter, P.** (2009). La notion de prise en charge: mise en perspective. *Langue française*, (2), 3-27.
- Creissen, S., & Blanc, N.** (2017). Quelle représentation des différentes facettes de la dimension émotionnelle d'une histoire entre l'âge de 6 et 10 ans? Apports d'une étude multimédia.
- Davidson, D.** (2006). The role of basic, self-conscious and self-conscious evaluative emotions in children's memory and understanding of emotion. *Motivation and Emotion*, 30(3), 232-242.
- Davidson, D., Luo, Z., & Burden, M. J.** (2001). Children's recall of emotional behaviours, emotional labels, and nonemotional behaviours: Does emotion enhance memory?. *Cognition & Emotion*, 15(1), 1-26.
- Dijkstra, K., Zwaan, R. A., Graesser, A. C., & Magliano, J. P.** (1995). Character and reader emotions in literary texts. *Poetics*, 23(1-2), 139-157.
- Diwersy, S., Goossens, V., Grutschus, A., Kern, B., Kraif, O., Melnikova, E., & Novakova, I.** (2014). Traitement des lexies d'émotion dans les corpus et les applications d'EmoBase. *Corpus*, (13), 269-293.

- Dumontheil, I., Apperly, I. A., & Blakemore, S. J.** (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental science*, 13(2), 331-338.
- Duval, C., Piolino, P., Bejanin, A., Laisney, M., Eustache, F., & Desgranges, B.** (2011). La théorie de l'esprit: aspects conceptuels, évaluation et effets de l'âge. *Revue de neuropsychologie*, 3(1), 41-51.
- Dyer, M. G.** (1983). The role of affect in narratives. *Cognitive Science*, 7(3), 211-242.
- Eensoo, E., & Valette, M.** (2012, June). Sur l'application de méthodes textométriques à la construction de critères de classification en analyse des sentiments. In *TALN 2012* (Vol. 2, pp. 367-374). GETALP-LIG.
- Eensoo, E., & Valette, M.** (2014). Approche textuelle pour le traitement automatique du discours évaluatif. *Langue française*, 184(4), 107-122.
- Eensoo, E., & Valette, M.** (2015, June). Une méthodologie de sémantique de corpus appliquée à des tâches de fouille d'opinion et d'analyse des sentiments: étude sur l'impact de marqueurs dialogiques et dialectiques dans l'expression de la subjectivité.
- Eensoo, E., Nouvel, D., Martin, A., & Valette, M.** (2016, June). Combiner analyses textométriques, apprentissage supervisé et représentation vectorielle pour l'analyse de la subjectivité. In *11e Défi Fouille de Texte (DEFT'2015)*, Caen (France).
- Gross, M.** (1995). Une grammaire locale de l'expression des sentiments. *Langue française*, 70-87.
- Gygax, P.** (2010). L'inférence émotionnelle durant la lecture et sa composante comportementale. *L'Année psychologique*, 110(2), 253-273.
- Kerbrat-Orecchioni, C.** (2000). Quelle place pour les émotions dans la linguistique du XXe siècle? Remarques et aperçus. *Collection éthologie et psychologie des communications*, 33-74.
- Lazarus, R. S.** (1984). On the primacy of cognition.
- Leeman, D.** (1995). Pourquoi peut-on dire « Max est en colère » mais non « \*Max est en peur » ? Hypothèses sur la construction " être en N". *Langue française*, 55-69.
- Mathieu, Y. Y.** (2006). A computational semantic lexicon of french verbs of emotion. In *Computing attitude and affect in text: Theory and applications* (pp. 109-124). Springer, Dordrecht.
- Micheli, R., Hekmat, I., & Rabatel, A.** (2013). Les émotions: Des modes de sémiotisation aux fonctions argumentatives. *Semen. Revue de sémio-linguistique des textes et discours*, (35).
- Micheli, R.** (2014). *Les émotions dans les discours: modèle d'analyse et perspectives empiriques*. De Boeck.
- Mohammad, S.** (2011, June). From once upon a time to happily ever after: Tracking emotions in novels and fairy tales. In *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities* (pp. 105-114). Association for Computational Linguistics.
- Mohammad, S.** (2015). Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In *Emotion measurement* (pp. 201-237). Woodhead Publishing.

- Piolat, A., & Bannour, R.** (2009). EMOTAIX: un scénario de Tropes pour l'identification automatisée du lexique émotionnel et affectif. *L'Année psychologique*, 109(4), 655-698.
- Rabatel, A.** (2013). Ecrire les émotions en mode emphatique. *Semen. Revue de sémio-linguistique des textes et discours*, (35).
- Zwaan, R. A., & Radvansky, G. A.** (1998). Situation models in language comprehension and memory. *Psychological bulletin*, 123(2), 162.

# Index des tableaux et des figures

## Index des tableaux

Tableau 1: Données brutes.....	62
Tableau 2: Nature des segments textuels des émotions montrées.....	63
Tableau 3: Nature des segments textuels des émotions étayées.....	63

## Index des figures

Figure 1 : Exemple d'automate fini (Gross, 1995).....	18
Figure 2 : Exemple d'annotation avec Glozz.....	39
Figure 3 : Extrait d'un fichier .aa.....	41
Figure 4 : Schéma de fonctionnement du script txt2GlozzFiles_v2.py.....	43
Figure 5 : Exemple d'annotation dans Glozz.....	47
Figure 6 : Annotation de l'exemple (6) dans Glozz.....	53
Figure 7 : Annotation de l'exemple (10) dans Glozz.....	55
Figure 8 : Interface Glozz pour la création de requêtes.....	57
Figure 9 : Liste de correspondances pour une requête Glozz.....	57
Figure 10 : Visualisation des résultats d'une requête grâce au "panier".....	57
Figure 11: Répartition des émotions entre les différents modes d'expression.....	59
Figure 12: Proportion des différents types d'émotion dans le corpus.....	59
Figure 13: Proportion des différentes catégories émotionnelles.....	60
Figure 14: Proportions des types d'émotion selon le mode d'expression.....	61
Figure 15: Proportion des différentes catégories d'émotion selon les modes d'expression.....	61

# A. Annexes

## A.1. Exemples traités dans Micheli (2014)

Page	Exemple	Référence	Type d'émotion	Émotion spécifique	Marque de l'émotion (1)	Marque de l'émotion (2)	Entité affectée (1)	Entité affectée (2)
34	Une rage m'a pris contre ce peuple.	<i>Dernier jour d'un condamné</i> , Victor Hugo, édition de Roger Borderie, Gallimard, 2000, coll. « Folio ».	Dite	Rage (colère)	Nom	rage	narrateur	me
91	Dérisoire, l'argument lié au prétendu effet dissuasif de la peine capitale. [...] Illusoire, l'argument d'exemplarité. [...] Effroyable, l'argument de la vengeance, car la mort de la victime innocente ne sera jamais rachetée par l'immolation du coupable.	Christian Goux, <i>Journal officiel</i> , compte rendu intégral de la séance du 17 septembre 1981, p. 1169	Montrée	Effroi (peur)	Énoncé <u>averbal</u>	Effroyable, l'argument de la vengeance	locuteur	aucune marque
91	Dérisoire, l'argument lié au prétendu effet dissuasif de la peine capitale. [...] Illusoire, l'argument d'exemplarité. [...] Effroyable, l'argument de la vengeance, car la mort de la victime innocente ne sera jamais rachetée par l'immolation du coupable.	Christian Goux, <i>Journal officiel</i> , compte rendu intégral de la séance du 17 septembre 1981, p. 1169	Dite	Effroi (peur)	Adjectif	Effroyable	locuteur	aucune marque
92	Elle m'a oublié, visage, parole, accent ; et puis, qui me reconnaîtrait avec cette barbe, ces habits et cette pâleur ? [...] Quoi ! Déjà plus père !	<i>Dernier jour d'un condamné</i> , chap. XLIII, p. 116	Montrée	Non spécifiée	Énoncé <u>averbal</u>	Déjà plus père ?	narrateur	aucune marque
93	Ah ! Malheureux rêveur, brise donc d'abord le mur épais de trois pieds qui t'emprisonne ! La mort ! La mort ! La mort !	<i>Dernier jour d'un condamné</i> , chap. XVII, p. 80	Montrée	peur	Énoncé <u>averbal</u>	La mort !	narrateur	aucune marque
95	Quel gougeat, il est odieux ce mec !!!!!	Réaction d'internaute concernant une question délicate posée par Nelson Monfort à une nageuse française (question sur la mort de sa mère, à la sortie du bassin, après une course olympique).	Montrée	haine	Dislocation à droite	il est odieux, ce mec !!!!!	locuteur	aucune marque
95	Quel gougeat, il est odieux ce mec !!!!!	Réaction d'internaute concernant une question délicate posée par Nelson Monfort à une nageuse française (question sur la mort de sa mère, à la sortie du bassin, après une course olympique).	Dite	haine	Adjectif	odieux	locuteur	aucune marque

100	<p>[1] if you reject the notion that this nation's promise is reserved for the few</p> <p>[2] your voice must be heard in this election [<u>applaudissements</u>]</p> <p>[3] if you reject the notion that our government is forever beholden to the highest bidder</p> <p>[4] you need to stand up in this election [<u>applaudissements</u>]</p> <p>[5] if you believe</p> <p>[5a] that new plants and factories can dot our landscape (.)</p> <p>[5b] that new energy can power our future (.)</p> <p>[5c] that new schools can provide ladders of opportunity to this nation of dreamers</p> <p>[6] if you believe in a country where</p> <p>[6a] everyone gets a fair shot</p> <p>[6b] and everyone does a fair share</p> <p>[6c] and everyone plays by the same rules (.)</p> <p>[7] then I need you to vote this November [<u>applaudissements</u>]</p> <p>[8] America I never said this journey would be easy</p> <p>[9] and I won't promise that now (.)</p> <p>[10] yes our path is harder</p> <p>[11] but it leads to a better place [<u>applaudissements</u>]</p> <p>[12] yes our road is longer</p> <p>[13] but we travel it together [<u>applaudissements</u>]</p> <p>[14] we don't turn back [<u>applaudissements</u>]</p> <p>[15] we leave no one behind [<u>applaudissements</u>]</p> <p>[16] we pull each other up [<u>applaudissements</u>]</p> <p>[17] we draw strength from our victories</p> <p>[18] and we learn from our mistakes</p> <p>[19] but we keep our eyes fixed on that distant horizon knowing</p> <p>[19a] that providence is with us</p> <p>[19b] and that we are surely blessed to be citizens of the greatest nation on earth</p>	« péroraison du discours que le Président des États-Unis Barack Obama livre lors de la convention du parti démocrate le 6 septembre 2012 »	Montrée	plaisir, satisfaction mais émotion suscitée	Rythmicité	L'organisation du discours, très structurée, qui crée des moments d'attente, d'anticipation chez l'auditoire.	auditoire	aucune marque
-----	---	--	---------	---	------------	---	-----------	---------------



Page	Exemple	Référence	Type d'émotion	Émotion spécifique	Marque de l'émotion (1)	Marque de l'émotion (2)	Entité affectée (1)	Entité affectée (2)
49	Vous avez pitié du coupable qui va subir sa peine ; ne trouverez-vous pas un peu de compassion aussi en faveur de celui qui a subi la peine de mort sans l'avoir méritée ?	George Berry, Assemblée Nationale, débat sur l'abolition de la peine de mort, 4 novembre 1908	Dite	pitié	Nom	pitié	auditoire	Vous
49	Messieurs les membres de la majorité, ce texte me donne l'impression que vous avez peur.	Michel Rocard, Assemblée Nationale, débat autour de la loi « anti-casseurs », 29 avril 1970	Dite	peur	Nom	peur	auditoire	vous
49	J'ai honte pour l'Europe, j'ai honte pour nos organisations internationales.	Marianne Ruby, 17 mars 2011, à propos de la situation en Libye	Dite	honte	Nom	honte	locuteur	Je
50	J'ai de la haine !!	Association internationale des victimes de l'inceste	Dite	haine	Nom	haine	locuteur	Je
50	Je ressens de la haine envers mon ex... comment lutter contre ce sentiment désagréable ?	Question Yahoo d'internaute	Dite	haine	Nom	haine	locuteur	Je
50	J'éprouve de la haine et du dégoût envers mon mari qui me quitte pour un homme.	Forum Les copains d'avant, publié le 15 avril 2010	Dite	haine	Nom	haine	locuteur	Je
50	J'éprouve de la haine et du dégoût envers mon mari qui me quitte pour un homme.	Forum Les copains d'avant, publié le 15 avril 2010	Dite	dégoût	Nom	dégoût	locuteur	Je
52	Je me serais attendu à tout éprouver sauf le sentiment que j'ai ressenti il y a un instant et que je vous livre avec toute ma force d'homme : vous m'avez fait honte !	Discours de Robert Badinter (16 juillet 1992) lors de la commémoration de la rafle du Vél'd'Hiv (Badinter s'adresse à la foule qui vient de siffler le président Mitterrand)	Dite	honte	Nom	honte	locuteur	me
52	Libérez enfin ma terre/des curés, des journaloux, des militaires/de tous les preneurs de tête/qui provoquent sous ma fenêtre/ma colère.	« L'aquarium », chanson de Renaud (1991)	Dite	colère	Nom	colère	locuteur	ma
52	L'annonce de cette triste nouvelle a suscité de la colère chez les travailleurs, qui n'avaient pas été prévenus par l'entreprise.	A propos de la fermeture d'une usine, site d'information cyberpresse.ca (28 février 2011)	Dite	tristesse	Adjectif	triste	locuteur	aucune marque

Page	Exemple	Référence	Type d'émotion	Émotion spécifique	Marque de l'émotion (1)	Marque de l'émotion (2)	Entité affectée (1)	Entité affectée (2)
52	L'annonce de cette triste nouvelle a suscité de la colère chez les travailleurs, qui n'avaient pas été prévenus par l'entreprise.	A propos de la fermeture d'une usine, site d'information cyberpresse.ca (28 février 2011)	Dite	colère	Nom	colère	les travailleurs	les travailleurs
53	Ses excuses ont augmenté/renforcé/attisé ... ma colère.	Construit	Dite	colère	Nom	colère	locuteur	ma
53	Ses excuses ont calmé/apaisé/diminué ... ma colère.	Construit	Dite	colère	Nom	colère	locuteur	ma
53	L'appréhension m'envahit.	Construit	Dite	peur	Nom	appréhension	locuteur	me
54	La peur m'envahit / La terreur m'envahit.	Construit	Dite	peur	Nom	peur	locuteur	me
54	La peur m'envahit / La terreur m'envahit.	Construit	Dite	peur	Nom	terreur	locuteur	me
54	La joie m'inonde. / Cette nouvelle m'inonde de joie.	Construit	Dite	joie	Nom	joie	locuteur	me
54	Le chagrin m'accable. / Cette nouvelle m'accable de chagrin.	Construit	Dite	tristesse	Nom	chagrin	locuteur	me
56	A 5 heures du matin, il effraya/épouvanta/terrifia/terrorisa... sa femme.	Construit	Dite	peur	Verbe	effraya/épouvanta/terrifia/terrorisa	sa femme	sa femme
58	Je suis furieux, si vous voulez savoir !	Propos de Karl Lagerfeld, au sujet des injures antisémites proférées par son collègue John Galliano, Tribune de Genève, 2 mars 2011.	Dite	colère	Adjectif	furieux	locuteur	Je
58	Depuis 3 nuits Milo se réveille en hurlant en fin de soirée, il paraît terrorisé.	Forum de discussion en ligne (www.familles.com)	Dite	peur	Adjectif (Participe passé)	terrorisé	Milo	il
63	Hélas ! qu'est-ce que la mort fait avec notre âme ? Quelle nature lui laisse-t-elle ? qu'a-t-elle à lui prendre ou à lui donner ? Où la met-elle ? (...) Ah ! Un prêtre ! Un prêtre qui sache cela ! Je veux un prêtre, et un crucifix à baiser ! Mon Dieu, toujours le même !	Dernier jour d'un condamné, chap. XLI, p. 122	Montrée	affolement (peur)	Multitude d'énoncés interrogatifs	la mort fait avec notre âme ? Quelle nature lui laisse-t-elle ? qu'a-t-elle à lui prendre ou à lui donner ? Où la met-	locuteur	aucune marque
63	Hélas ! qu'est-ce que la mort fait avec notre âme ? Quelle nature lui laisse-t-elle ? qu'a-t-elle à lui prendre ou à lui donner ? Où la met-elle ? (...) Ah ! Un prêtre ! Un prêtre qui sache cela ! Je veux un prêtre, et un crucifix à baiser ! Mon Dieu, toujours le même !	Dernier jour d'un condamné, chap. XLI, p. 122	Montrée	affolement (peur)	Énoncés exclamatifs et interjections	Hélas ! (...) Ah !	locuteur	aucune marque

Page	Exemple	Référence	Type d'émotion	Émotion spécifique	Marque de l'émotion (1)	Marque de l'émotion (2)	Entité affectée (1)	Entité affectée (2)
63	Hélas ! qu'est-ce que la mort fait avec notre âme ? Quelle nature lui laisse-t-elle ? qu'a-t-elle à lui prendre ou à lui donner ? Où la met-elle ? (...) Ah ! Un prêtre ! Un prêtre qui sache cela ! Je veux un prêtre, et un crucifix à baiser ! Mon Dieu, toujours le même !	<i>Dernier jour d'un condamné</i> , chap. XLI, p. 122	Montrée	affolement (peur)	Énoncés exclamatifs et interjections	Hélas ! (...) Ah !	locuteur	aucune marque
63	Hélas ! qu'est-ce que la mort fait avec notre âme ? Quelle nature lui laisse-t-elle ? qu'a-t-elle à lui prendre ou à lui donner ? Où la met-elle ? (...) Ah ! Un prêtre ! Un prêtre qui sache cela ! Je veux un prêtre, et un crucifix à baiser ! Mon Dieu, toujours le même !	<i>Dernier jour d'un condamné</i> , chap. XLI, p. 122	Montrée	affolement (peur)	Énoncés averbaux exclamatifs	Un prêtre ! Un prêtre qui sache cela !	locuteur	aucune marque
63	Hélas ! qu'est-ce que la mort fait avec notre âme ? Quelle nature lui laisse-t-elle ? qu'a-t-elle à lui prendre ou à lui donner ? Où la met-elle ? (...) Ah ! Un prêtre ! Un prêtre qui sache cela ! Je veux un prêtre, et un crucifix à baiser ! Mon Dieu, toujours le même !	<i>Dernier jour d'un condamné</i> , chap. XLI, p. 122	Montrée	déception (intense)	Interjection, énoncé averbal et exclamatif	Mon Dieu, toujours le même !	locuteur	aucune marque
68	Je me levai ; mes dents claquaient, mes mains tremblaient et ne savaient où trouver des vêtements, mes jambes étaient faibles.	<i>Dernier jour d'un condamné</i> , chap. II, p. 42	Dite	peur	Description verbale d'indices physiologiques et comportementaux	mes dents claquaient, mes mains tremblaient (...) mes jambes étaient faibles	locuteur	me
73	C'est qu'à l'heure qu'il est, tout près de nous, là, sous nos yeux, on massacre, on incendie, on pille, on extermine, on égorge les pères et les mères, on vend les petites filles et les petits garçons...	« Pour la <u>Serbie</u> », <i>Actes et Paroles</i> , Victor Hugo, éd. Robert Laffont, 1985 (1876) p. 949	Étayée	indignation (colère)	Unités lexicales qui participent à la représentation d'une situation conventionnellement associée à l'indignation	massacre, incendie, pille, extermine, égorge les pères et les mères, on vend les petites filles et les petits garçons	locuteur	aucune marque
73	J'ai honte pour l'Europe, j'ai honte pour nos organisations internationales on parle tous les jours de justice de solidarité on a été infoutus d'être au rendez-vous d'un peuple qui est martyrisé et qui demain le sera encore plus par l'incapacité des puissants de ce monde à s' <u>mettre</u> d'accord	Martine Aubry, 17 mars 2011, à propos de la situation en <u>Lybie</u>	Montrée	honte	Rupture du registre courant par l'utilisation d'un mot familier	infoutu	locuteur	aucune marque

Page	Exemple	Référence	Type d'émotion	Émotion spécifique	Marque de l'émotion (1)	Marque de l'émotion (2)	Entité affectée (1)	Entité affectée (2)
73	J'ai honte pour l'Europe, j'ai honte pour nos organisations internationales on parle tous les jours de justice de solidarité on a été infoutus d'être au rendez-vous d'un peuple qui est martyrisé et qui demain le sera encore plus par l'incapacité des puissants de ce monde à s' <u>mettre</u> d'accord	Martine Aubry, 17 mars 2011, à propos de la situation en <u>Lybie</u>	Étayée	honte	Description d'une situation qui légitime la honte	qui est martyrisé et qui demain le sera encore plus par l'incapacité des puissants de ce monde à s' <u>mettre</u> d'accord	locuteur	aucune marque
73	J'ai honte pour l'Europe, j'ai honte pour nos organisations internationales on parle tous les jours de justice de solidarité on a été infoutus d'être au rendez-vous d'un peuple qui est martyrisé et qui demain le sera encore plus par l'incapacité des puissants de ce monde à s' <u>mettre</u> d'accord	Martine Aubry, 17 mars 2011, à propos de la situation en <u>Lybie</u>	Dite	honte	Nom	(avoir) honte	locuteur	Je
73	aujourd'hui les deux partis du centre le parti radical et le PDC avec 30 % des voix dans l'électorat s'arrogent 90 % des décisions politiques c'est un diktat pur et simple [...] on se retrouve dans une situation où l'on est confronté à un diktat de la droite où on se retrouve marginalisé	Extrait de la prise de parole de Christian Levrat (président du parti socialiste) lors d'un débat radiophonique opposant les présidents des principaux partis politiques suisses le 27 septembre 2010	Montrée	colère	Utilisation d'un nom qui, socialement, peut être associé à des réactions émotionnelles négatives	diktat	Locuteur + ses camarades ?	on
79	Ah, n'importe, c'est horrible !	<i>Dernier jour d'un condamné</i> , chap. III, p. 48	Dite	peur	Adjectif	horrible	locuteur	aucune marque
79	Ah, n'importe, c'est horrible !	<i>Dernier jour d'un condamné</i> , chap. III, p. 48	Montrée	peur	Interjection (et énoncé exclamatif)	Ah !	locuteur	aucune marque
79 – 80	Voilà ce qu'ils vont faire de ton père, ces hommes dont aucun ne me hait, qui tous me plaignent et tous pourraient me sauver. Ils vont me tuer. Comprends-tu cela Marie ? Me tuer de sang-froid, en cérémonie, pour le bien de la chose ! Ah grand Dieu !	<i>Dernier jour d'un condamné</i> , chap. XXVI, p. 101	Montrée	Effroi (peur)	Interjections (et énoncés exclamatifs)	Ah ! Grand Dieu !	locuteur	aucune marque
79 – 80	Voilà ce qu'ils vont faire de ton père, ces hommes dont aucun ne me hait, qui tous me plaignent et tous pourraient me sauver. Ils vont me tuer. Comprends-tu cela Marie ? Me tuer de sang-froid, en cérémonie, pour le bien de la chose ! Ah grand Dieu !	<i>Dernier jour d'un condamné</i> , chap. XXVI, p. 101	Étayée	Effroi (peur)	Description de la mort prochaine	Me tuer de sang-froid, en cérémonie	locuteur	me
80	Le juge et le bourreau sont sortis. Je suis seul. Seul avec deux gendarmes. Oh ! l'horrible peuple avec ses cris d'hyènes. - Qui sait si je ne lui échapperai pas ? Si je ne serai pas sauvé ? Si ma grâce ? ... Il est impossible qu'on ne me fasse pas grâce ! Ah ! Les misérables ! Il me semble qu'on monte l'escalier...	<i>Dernier jour d'un condamné</i> , chap. XLIX, pp. 139-140	Dite	Horreur (peur)	Adjectif	horrible	locuteur (mais désigne le peuple)	aucune marque

Page	Exemple	Référence	Type d'émotion	Émotion spécifique	Marque de l'émotion (1)	Marque de l'émotion (2)	Entité affectée (1)	Entité affectée (2)
80	Le juge et le bourreau sont sortis. Je suis seul. Seul avec deux gendarmes. Oh ! l'horrible peuple avec ses cris d'hyènes. - Qui sait si je ne lui échapperai pas ? Si je ne serai pas sauvé ? Si ma grâce ? ... Il est impossible qu'on ne me fasse pas grâce ! Ah ! Les misérables ! Il me semble qu'on monte l'escalier...	<i>Dernier jour d'un condamné</i> , chap. XLIX, pp. 139-140	Dite	Mépris	Nom	les misérables	locuteur (mais désigne les bourreaux)	aucune marque
80	Le juge et le bourreau sont sortis. Je suis seul. Seul avec deux gendarmes. Oh ! l'horrible peuple avec ses cris d'hyènes. - Qui sait si je ne lui échapperai pas ? Si je ne serai pas sauvé ? Si ma grâce ? ... Il est impossible qu'on ne me fasse pas grâce ! Ah ! Les misérables ! Il me semble qu'on monte l'escalier...	<i>Dernier jour d'un condamné</i> , chap. XLIX, pp. 139-140	Montrée	Peur	Interjections (et énoncés exclamatifs)	Ah ! Oh !	locuteur	aucune marque
80	Le juge et le bourreau sont sortis. Je suis seul. Seul avec deux gendarmes. Oh ! l'horrible peuple avec ses cris d'hyènes. - Qui sait si je ne lui échapperai pas ? Si je ne serai pas sauvé ? Si ma grâce ? ... Il est impossible qu'on ne me fasse pas grâce ! Ah ! Les misérables ! Il me semble qu'on monte l'escalier...	<i>Dernier jour d'un condamné</i> , chap. XLIX, pp. 139-140	Montrée	Peur	Énoncé <u>averbal</u>	l'horrible peuple avec ses cris de hyènes.	locuteur	aucune marque
87	Comment ? Vous ici ?	Exemple repris de <u>Bally</u> (1951)	Montrée	Surprise	Énoncé elliptique	Vous ici ? (au lieu de « Comment se fait-il que vous soyez	locuteur	aucune marque
89	je ne crois pas je n'admettrai jamais que la guerre soit inévitable et fatale jusqu'à la dernière limite [[applaudissements]] jusqu'à la dernière limite de mon pouvoir et jusqu'au dernier souffle de ma vie s'il le faut je ferai tout pour la détourner de ce pays vous m'entendez bien/ TOUT pour écarter le risque prochain présent de guerre je reFUSE de considérer comme possible la guerre aujourd'hui parce qu'elle serait nécessaire ou fatale demain la guerre est possible quand on l'admet comme possible FATALE quand on la proclame (.) fatale et moi jusqu'au bout je me refuse à désespérer de la paix et de l'action de la nation française pour la pacification	Extrait d'un discours de Léon Blum (Président du Conseil), le 6 septembre 1936, devant la fédération socialiste de la Seine	Montrée	Peur	Énoncé elliptique	fatale quand on la proclame fatale	locuteur	aucune marque
89	je ne crois pas je n'admettrai jamais que la guerre soit inévitable et fatale jusqu'à la dernière limite [[applaudissements]] jusqu'à la dernière limite de mon pouvoir et jusqu'au dernier souffle de ma vie s'il le faut je ferai tout pour la détourner de ce pays vous m'entendez bien/ TOUT pour écarter le risque prochain présent de guerre je reFUSE de considérer comme possible la guerre aujourd'hui parce qu'elle serait nécessaire ou fatale demain la guerre est possible quand on l'admet comme possible FATALE quand on la proclame (.) fatale et moi jusqu'au bout je me refuse à désespérer de la paix et de l'action de la nation française pour la pacification	Extrait d'un discours de Léon Blum (Président du Conseil), le 6 septembre 1936, devant la fédération socialiste de la Seine	Étayée	Peur	Utilisation de terme liés à la mort	guerre, fatale	locuteur	aucune marque

Page	Exemple	Référence	Type d'émotion	Émotion spécifique	Marque de l'émotion (1)	Marque de l'émotion (2)	Entité affectée (1)	Entité affectée (2)
89	je ne crois pas je n'admettrai jamais que la guerre soit inévitable et fatale jusqu'à la dernière limite [[applaudissements]] jusqu'à la dernière limite de mon pouvoir et jusqu'au dernier souffle de ma vie s'il le faut je ferai tout pour la détourner de ce pays vous m'entendez bien/ TOUT pour écarter le risque prochain présent de guerre je reFUSE de considérer comme possible la guerre aujourd'hui parce qu'elle serait nécessaire ou fatale demain la guerre est possible quand on l'admet comme possible FATALE quand on la proclame (.) fatale et moi jusqu'au bout je me refuse à désespérer de la paix et de l'action de la nation française pour la pacification	Extrait d'un discours de Léon Blum (Président du Conseil), le 6 septembre 1936, devant la fédération socialiste de la Seine	Étayée	Peur	Utilisation de terme liés à la mort	guerre, fatale	locuteur	aucune marque
95	..... Et Hollande d'ajouter : « Il devrait présenter ses excuses, pas sa candidature »..... Il est ODIEUX ce monsieur, si bien nommé Fraise des bois, par son ex-ancien-nouveau-pote !!!!!	Commentaire d'un article de presse ( <i>Le Figaro</i> ) concernant un discours où François Hollande avait « attaqué le bilan de son concurrent sortant Nicolas Sarkozy »	Montrée	haine	Dislocation à droite	Il est ODIEUX ce monsieur	locuteur	aucune marque
95	..... Et Hollande d'ajouter : « Il devrait présenter ses excuses, pas sa candidature »..... Il est ODIEUX ce monsieur, si bien nommé Fraise des bois, par son ex-ancien-nouveau-pote !!!!!	Commentaire d'un article de presse ( <i>Le Figaro</i> ) concernant un discours où François Hollande avait « attaqué le bilan de son concurrent sortant Nicolas Sarkozy »	Dite	haine	Adjectif	ODIEUX	locuteur	aucune marque
97	la relance des services publics puisque vous savez que par la faute de Nicolas Sarkozy (.) le plus grand plan social du pays a été le fait de lui Nicolas Sarkozy en supprimant les postes de fonctionnaires ce sont 150000 POSTES qui ont été supprimés les rétablit c'est donc au moins 150000 postes immédiatement disponibles pour la population et donc autant de circulation	électoral prononcé lors d'un meeting à Lille par Jean-Luc Mélenchon, candidat du Front de Gauche durant la campagne présidentielle de 2012 (l'extrait commence à 54'32 de l'enregistrement audio-	Montrée	indignation (colère)	Énoncé clivé	ce sont 150 000 POSTES qui ont été supprimés	locuteur	aucune marque
98	j'ai lu que des petits prétentieux sont là qui jaspent autour de nous ah monsieur Mélenchon sont-ils obligés de concéder en vous voyant si nombreux ils ont peur de vous (.) bien sûr ce que vous dites est tellement sympathique ça nous fait rêver (.) mais vous savez bien que ce n'est pas réaliste (.) quand est-ce qu'on va leur demander à eux si c'est réaliste ce qu'ils font parce que moi je retourne la question [...] je me tourne vers le président de la République et je lui dis vous avez donné 3 milliards à la restauration de dégrèvement de TVA et votre fichue politique de contraction de la dépense publique de suppression des postes de fonctionnaires dont nous manquons partout dans la police à l'école partout partout elle elle coûte 500 millions par an elle rapporte 500 millions par an (.) quand est-ce qu'on fait le bilan de la catastrophe de votre politique c'est VOUS qui n'êtes pas réaliste	Extrait d'un « discours électoral prononcé lors d'un meeting à Lille par Jean-Luc Mélenchon, candidat du Front de Gauche durant la campagne présidentielle de 2012 (l'extrait commence à 46'30 de l'enregistrement audio-visuel du discours).	Montrée	Non spécifiée	Énoncé clivé	c'est VOUS qui n'êtes pas réaliste	locuteur	aucune marque

Page	Exemple	Référence	Type d'émotion	Émotion spécifique	Marque de l'émotion (1)	Marque de l'émotion (2)	Entité affectée (1)	Entité affectée (2)
123	Personne ne peut accepter d'être dupe des affirmations gratuites et des vœux pieux, selon lesquels on contrôlera ici plus qu'ailleurs les conséquences innombrables et désastreuses de la légalisation de l'avortement. En France, comme partout, le commerce de la mort deviendra la spécialité lucrative d'avorteurs et d'avortoirs patentés où l'on se préoccupera, comme on le fait déjà en Angleterre, de la vente des fœtus avortés à des usines de traitement de « grasses animales » et où l'on se livrera, comme on le fait déjà dans les pays qui ont légalisé l'avortement, à des expériences dites scientifiques sur les fœtus encore vivants. Cela ne s'appellera plus du désordre, madame le ministre, Cela ne s'appellera même plus de l'injustice. C'est de la barbarie organisée et couverte par la loi, comme elle le fut, hélas ! Il y a trente ans, par le nazisme en Allemagne. Nous affirmons, il y a un an, que reconnaître le droit à l'avortement, c'était reconnaître inéluctablement le droit à l'eugénisme, à l'euthanasie. Nos propos ont été confirmés par les événements, et nous voyons se déclencher et se développer une campagne en faveur de l'euthanasie, comme nous avons vu se déclencher et se développer une campagne en faveur de l'avortement, avec le concours de deux ou trois lauréats du prix Nobel, aux talents mobilisés pour toutes les batailles de l'idéologie de la mort. Le chantage aux chiffres, les provocations, la propagande en faveur de l'avortement libre ne peuvent plus masquer les conséquences catastrophiques, tant physiologiques et psychiques pour la femme et ses futures grossesses que morales, économiques et démographiques pour la société, qu'entraînerait la légalisation de l'avortement.	Extrait de l'intervention du député Jacques Médecin dans un débat sur la légalisation de l'IVG, <i>Journal Officiel</i> , 2e séance du mardi 26 novembre 1974, p. 7036.	Étayée	Peur	Évocation des conséquences de la situation schématisée et de leur « degré de probabilité »	« conséquences innombrables et désastreuses », « conséquences catastrophiques » + description détaillée de ces conséquences ; les conséquences sont présentées comme très probables (usage du futur « deviendra », « préoccupe », « se livrera »)	locuteur	aucune marque
123	Personne ne peut accepter d'être dupe des affirmations gratuites et des vœux pieux, selon lesquels on contrôlera ici plus qu'ailleurs les conséquences innombrables et désastreuses de la légalisation de l'avortement. En France, comme partout, le commerce de la mort deviendra la spécialité lucrative d'avorteurs et d'avortoirs patentés où l'on se préoccupera, comme on le fait déjà en Angleterre, de la vente des fœtus avortés à des usines de traitement de « grasses animales » et où l'on se livrera, comme on le fait déjà dans les pays qui ont légalisé l'avortement, à des expériences dites scientifiques sur les fœtus encore vivants. Cela ne s'appellera plus du désordre, madame le ministre, Cela ne s'appellera même plus de l'injustice. C'est de la barbarie organisée et couverte par la loi, comme elle le fut, hélas ! Il y a trente ans, par le nazisme en Allemagne. Nous affirmons, il y a un an, que reconnaître le droit à l'avortement, c'était reconnaître inéluctablement le droit à l'eugénisme, à l'euthanasie. Nos propos ont été confirmés par les événements, et nous voyons se déclencher et se développer une campagne en faveur de l'euthanasie, comme nous avons vu se déclencher et se développer une campagne en faveur de l'avortement, avec le concours de deux ou trois lauréats du prix Nobel, aux talents mobilisés pour toutes les batailles de l'idéologie de la mort. Le chantage aux chiffres, les provocations, la propagande en faveur de l'avortement libre ne peuvent plus masquer les conséquences catastrophiques, tant physiologiques et psychiques pour la femme et ses futures grossesses que morales, économiques et démographiques pour la société, qu'entraînerait la légalisation de l'avortement.	Extrait de l'intervention du député Jacques Médecin dans un débat sur la légalisation de l'IVG, <i>Journal Officiel</i> , 2e séance du mardi 26 novembre 1974, p. 7036.	Étayée	Peur	La situation est présentée comme incontrôlable	« Personne ne peut accepter d'être dupe des affirmations gratuites et des vœux pieux, selon lesquels on contrôlera ici plus qu'ailleurs les conséquences innombrables et désastreuses de la légalisation de l'avortement. », « inéluctable »	locuteur	aucune marque
123	Personne ne peut accepter d'être dupe des affirmations gratuites et des vœux pieux, selon lesquels on contrôlera ici plus qu'ailleurs les conséquences innombrables et désastreuses de la légalisation de l'avortement. En France, comme partout, le commerce de la mort deviendra la spécialité lucrative d'avorteurs et d'avortoirs patentés où l'on se préoccupera, comme on le fait déjà en Angleterre, de la vente des fœtus avortés à des usines de traitement de « grasses animales » et où l'on se livrera, comme on le fait déjà dans les pays qui ont légalisé l'avortement, à des expériences dites scientifiques sur les fœtus encore vivants. Cela ne s'appellera plus du désordre, madame le ministre, Cela ne s'appellera même plus de l'injustice. C'est de la barbarie organisée et couverte par la loi, comme elle le fut, hélas ! Il y a trente ans, par le nazisme en Allemagne. Nous affirmons, il y a un an, que reconnaître le droit à l'avortement, c'était reconnaître inéluctablement le droit à l'eugénisme, à l'euthanasie. Nos propos ont été confirmés par les événements, et nous voyons se déclencher et se développer une campagne en faveur de l'euthanasie, comme nous avons vu se déclencher et se développer une campagne en faveur de l'avortement, avec le concours de deux ou trois lauréats du prix Nobel, aux talents mobilisés pour toutes les batailles de l'idéologie de la mort. Le chantage aux chiffres, les provocations, la propagande en faveur de l'avortement libre ne peuvent plus masquer les conséquences catastrophiques, tant physiologiques et psychiques pour la femme et ses futures grossesses que morales, économiques et démographiques pour la société, qu'entraînerait la légalisation de l'avortement.	Extrait de l'intervention du député Jacques Médecin dans un débat sur la légalisation de l'IVG, <i>Journal Officiel</i> , 2e séance du mardi 26 novembre 1974, p. 7036.	Étayée	Peur	Analogie avec des « situations historiques, dont la résonance affective est particulièrement forte »	« nazisme en Allemagne »	locuteur	aucune marque
123	Personne ne peut accepter d'être dupe des affirmations gratuites et des vœux pieux, selon lesquels on contrôlera ici plus qu'ailleurs les conséquences innombrables et désastreuses de la légalisation de l'avortement. En France, comme partout, le commerce de la mort deviendra la spécialité lucrative d'avorteurs et d'avortoirs patentés où l'on se préoccupera, comme on le fait déjà en Angleterre, de la vente des fœtus avortés à des usines de traitement de « grasses animales » et où l'on se livrera, comme on le fait déjà dans les pays qui ont légalisé l'avortement, à des expériences dites scientifiques sur les fœtus encore vivants. Cela ne s'appellera plus du désordre, madame le ministre, Cela ne s'appellera même plus de l'injustice. C'est de la barbarie organisée et couverte par la loi, comme elle le fut, hélas ! Il y a trente ans, par le nazisme en Allemagne. Nous affirmons, il y a un an, que reconnaître le droit à l'avortement, c'était reconnaître inéluctablement le droit à l'eugénisme, à l'euthanasie. Nos propos ont été confirmés par les événements, et nous voyons se déclencher et se développer une campagne en faveur de l'euthanasie, comme nous avons vu se déclencher et se développer une campagne en faveur de l'avortement, avec le concours de deux ou trois lauréats du prix Nobel, aux talents mobilisés pour toutes les batailles de l'idéologie de la mort. Le chantage aux chiffres, les provocations, la propagande en faveur de l'avortement libre ne peuvent plus masquer les conséquences catastrophiques, tant physiologiques et psychiques pour la femme et ses futures grossesses que morales, économiques et démographiques pour la société, qu'entraînerait la légalisation de l'avortement.	Extrait de l'intervention du député Jacques Médecin dans un débat sur la légalisation de l'IVG, <i>Journal Officiel</i> , 2e séance du mardi 26 novembre 1974, p. 7036.	Étayée	Peur	Présentation (« schématisation ») de la situation comme « profondément incompatible avec des valeurs présentées comme constitutives du groupe »	« idéologie de la mort »	locuteur	aucune marque

Page	Exemple	Référence	Type d'émotion	Émotion spécifique	Marque de l'émotion (1)	Marque de l'émotion (2)	Entité affectée (1)	Entité affectée (2)
126	<p>Quand le risque sanitaire de pollution de l'eau est avéré même dans un cas de nappe phréatique en état de sécheresse depuis des années...</p> <p>Quand Paris est en péril, que dit l'Élysée ?</p> <p>« Je vous le dis, comme Président de la république, il n'y a aucun risque, parce que c'est vrai »</p> <p>Arrêtons le bavardage et les palabres sans fin : j'en appelle à l'indignation citoyenne sans parti pris politique.</p> <p>Les risques sanitaires sont bien là et d'abord pour les enfants et les personnes âgées. Toute inaction ou laisser-faire de l'État ferait de cet État-là un complice actif dans les catastrophes humaines à venir faute d'avoir protégé le seul bien dont nous ne savons pas renouveler les stocks : l'eau potable.</p>	<p>Entrée de blog, mise en ligne le 6 mars 2011, par Muriel Bodin (avocate) sur le site <a href="http://www.conseil-juridique.net">www.conseil-juridique.net</a></p>	Montrée	indignation (colère)	Nom	indignation	Locuteur + autre ?	Je + citoyenne ?
126	<p>Quand le risque sanitaire de pollution de l'eau est avéré même dans un cas de nappe phréatique en état de sécheresse depuis des années...</p> <p>Quand Paris est en péril, que dit l'Élysée ?</p> <p>« Je vous le dis, comme Président de la république, il n'y a aucun risque, parce que c'est vrai »</p> <p>Arrêtons le bavardage et les palabres sans fin : j'en appelle à l'indignation citoyenne sans parti pris politique.</p> <p>Les risques sanitaires sont bien là et d'abord pour les enfants et les personnes âgées. Toute inaction ou laisser-faire de l'État ferait de cet État-là un complice actif dans les catastrophes humaines à venir faute d'avoir protégé le seul bien dont nous ne savons pas renouveler les stocks : l'eau potable.</p>	<p>Entrée de blog, mise en ligne le 6 mars 2011, par Muriel Bodin (avocate) sur le site <a href="http://www.conseil-juridique.net">www.conseil-juridique.net</a></p>	Étayée	indignation (colère)	Distance entre la locutrice et la situation évoquée	« Paris » (c'est le lieu où exerce la locutrice, la situation est donc très proche)	locuteur	aucune marque
126	<p>Quand le risque sanitaire de pollution de l'eau est avéré même dans un cas de nappe phréatique en état de sécheresse depuis des années...</p> <p>Quand Paris est en péril, que dit l'Élysée ?</p> <p>« Je vous le dis, comme Président de la république, il n'y a aucun risque, parce que c'est vrai »</p> <p>Arrêtons le bavardage et les palabres sans fin : j'en appelle à l'indignation citoyenne sans parti pris politique.</p> <p>Les risques sanitaires sont bien là et d'abord pour les enfants et les personnes âgées. Toute inaction ou laisser-faire de l'État ferait de cet État-là un complice actif dans les catastrophes humaines à venir faute d'avoir protégé le seul bien dont nous ne savons pas renouveler les stocks : l'eau potable.</p>	<p>Entrée de blog, mise en ligne le 6 mars 2011, par Muriel Bodin (avocate) sur le site <a href="http://www.conseil-juridique.net">www.conseil-juridique.net</a></p>	Étayée	indignation (colère)	Personnes impliquées et leur rôle	« les enfants et les personnes âgées », présentés comme des patients (victimes) ; « Président de la république » et « l'Élysée », présentés comme des agents	locuteur	aucune marque
126	<p>Quand le risque sanitaire de pollution de l'eau est avéré même dans un cas de nappe phréatique en état de sécheresse depuis des années...</p> <p>Quand Paris est en péril, que dit l'Élysée ?</p> <p>« Je vous le dis, comme Président de la république, il n'y a aucun risque, parce que c'est vrai »</p> <p>Arrêtons le bavardage et les palabres sans fin : j'en appelle à l'indignation citoyenne sans parti pris politique.</p> <p>Les risques sanitaires sont bien là et d'abord pour les enfants et les personnes âgées. Toute inaction ou laisser-faire de l'État ferait de cet État-là un complice actif dans les catastrophes humaines à venir faute d'avoir protégé le seul bien dont nous ne savons pas renouveler les stocks : l'eau potable.</p>	<p>Entrée de blog, mise en ligne le 6 mars 2011, par Muriel Bodin (avocate) sur le site <a href="http://www.conseil-juridique.net">www.conseil-juridique.net</a></p>	Étayée	indignation (colère)	Présentation des conséquences et de leur degré de probabilité	« risques sanitaires », « pollution de l'eau » (des conséquences graves, négatives) ; « avéré » (une probabilité très élevée)	locuteur	aucune marque
126	<p>Quand le risque sanitaire de pollution de l'eau est avéré même dans un cas de nappe phréatique en état de sécheresse depuis des années...</p> <p>Quand Paris est en péril, que dit l'Élysée ?</p> <p>« Je vous le dis, comme Président de la république, il n'y a aucun risque, parce que c'est vrai »</p> <p>Arrêtons le bavardage et les palabres sans fin : j'en appelle à l'indignation citoyenne sans parti pris politique.</p> <p>Les risques sanitaires sont bien là et d'abord pour les enfants et les personnes âgées. Toute inaction ou laisser-faire de l'État ferait de cet État-là un complice actif dans les catastrophes humaines à venir faute d'avoir protégé le seul bien dont nous ne savons pas renouveler les stocks : l'eau potable.</p>	<p>Entrée de blog, mise en ligne le 6 mars 2011, par Muriel Bodin (avocate) sur le site <a href="http://www.conseil-juridique.net">www.conseil-juridique.net</a></p>	Étayée	indignation (colère)	Attribution <u>agentive</u>	aucune action de la part des pouvoirs publics (« inaction », « laisser-faire »)	locuteur	aucune marque
126	<p>Quand le risque sanitaire de pollution de l'eau est avéré même dans un cas de nappe phréatique en état de sécheresse depuis des années...</p> <p>Quand Paris est en péril, que dit l'Élysée ?</p> <p>« Je vous le dis, comme Président de la république, il n'y a aucun risque, parce que c'est vrai »</p> <p>Arrêtons le bavardage et les palabres sans fin : j'en appelle à l'indignation citoyenne sans parti pris politique.</p> <p>Les risques sanitaires sont bien là et d'abord pour les enfants et les personnes âgées. Toute inaction ou laisser-faire de l'État ferait de cet État-là un complice actif dans les catastrophes humaines à venir faute d'avoir protégé le seul bien dont nous ne savons pas renouveler les stocks : l'eau potable.</p>	<p>Entrée de blog, mise en ligne le 6 mars 2011, par Muriel Bodin (avocate) sur le site <a href="http://www.conseil-juridique.net">www.conseil-juridique.net</a></p>	Étayée	indignation (colère)	Présentation du potentiel de maîtrise	« le seul bien dont nous ne savons pas renouveler les stocks » ; sans action, la situation serait totalement incontrôlable	locuteur	aucune marque

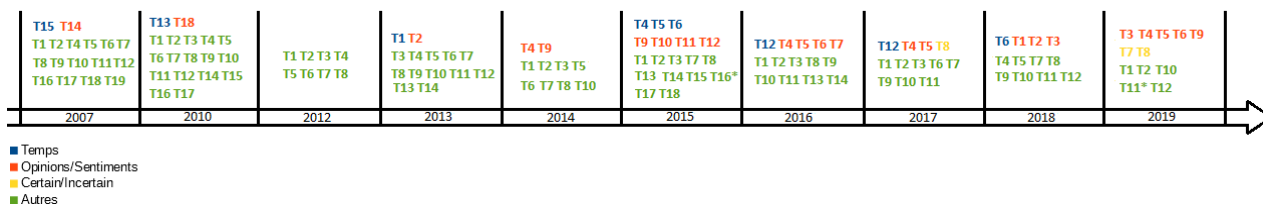
Page	Exemple	Référence	Type d'émotion	Émotion spécifique	Marque de l'émotion (1)	Marque de l'émotion (2)	Entité affectée (1)	Entité affectée (2)
126	<p>Quand le risque sanitaire de pollution de l'eau est avéré même dans un cas de nappe phréatique en état de sécheresse depuis des années...</p> <p>Quand Paris est en péril, que dit l'Élysée ?</p> <p>« Je vous le dis, comme Président de la république, il n'y a aucun risque, parce que c'est vrai »</p> <p>Arrêtons le bavardage et les palabres sans fin : j'en appelle à l'indignation citoyenne sans parti pris politique.</p> <p>Les risques sanitaires sont bien là et d'abord pour les enfants et les personnes âgées. Toute inaction ou laisser-faire de l'État ferait de cet État-là un complice actif dans les catastrophes humaines à venir faute d'avoir protégé le seul bien dont nous ne savons pas renouveler les stocks : l'eau potable.</p>	<p>Entrée de blog, mise en ligne le 6 mars 2011, par Muriel Bodin (avocate) sur le site <a href="http://www.conseil-juridique.net">www.conseil-juridique.net</a></p>	Étayée	indignation (colère)	Présentation de la situation, en rupture avec des valeurs collectives	les pouvoirs publics, censés protéger la population, ne le font pas	locuteur	aucune marque
129	<p>[Il est demandé à Denise, une apprentie vendeuse, d'essayer un manteau et ainsi de servir de mannequin.]</p> <p>Alors, tous les yeux revinrent sur Denise. Elle s'abandonnait aux mains de Marguerite, qui la faisait tourner lentement.</p> <p>- Hein ? qu'en pensez-vous ? demanda Mme Marty à Mme Desforges.</p> <p>[...]</p> <p>- Il n'est pas mal, et de coupe originale... Seulement, il me semble peu gracieux de la taille.</p> <p>- Oh ! intervint Mme Aurélie, il faudrait le voir sur madame elle-même... Vous comprenez, il ne fait aucun effet sur mademoiselle, qui n'est guère étoffée... [...]</p> <p>On sourit. Denise était devenue très pâle. Une honte la prenait, d'être ainsi changée en machine qu'on examinait et dont on plaisantait librement. Mme Desforges [...] ajouta méchamment :</p> <p>- Sans doute, il irait mieux si la robe de mademoiselle était moins large. Et elle jetait à Mouret le regard moqueur d'une Parisienne, que l'attiffement ridicule d'une provinciale égayait. [...]</p> <p>- Puis il faudrait être peignée, murmura-t-il.</p> <p>Ce fut le comble. Le directeur daignait rire, toutes ces demoiselles éclatèrent. Marguerite risqua un léger gloussement [...]; Clara avait lâché une vente, pour se faire du bon sang à son aise ; même des vendeuses de la lingerie étaient venues, attirées par la rumeur. [...] Denise avait encore pâli, au milieu de tout ce monde qui se moquait. Elle se sentait violente, mise à nu, sans défense. Quelle était donc sa faute, pour qu'on s'attaquât de la sorte à sa taille trop mince, à son chignon trop lourd ?</p>	<p>Émile Zola, <i>Au Bonheur des Dames</i>, Paris, Gallimard, coll. « Folio », 1980, pp. 151-153.</p>	Dite	honte	Nom	honte	Denise	la

129	<p>[Il est demandé à Denise, une apprentie vendeuse, d'essayer un manteau et ainsi de servir de mannequin.]</p> <p>Alors, tous les yeux revinrent sur Denise. Elle s'abandonnait aux mains de Marguerite, qui la faisait tourner lentement.</p> <p>- Hein ? qu'en pensez-vous ? demanda Mme Marty à Mme Desforges.</p> <p>[...]</p> <p>- Il n'est pas mal, et de coupe originale... Seulement, il me semble peu gracieux de la taille.</p> <p>- Oh ! intervint Mme Aurélie, il faudrait le voir sur madame elle-même... Vous comprenez, il ne fait aucun effet sur mademoiselle, qui n'est guère étoffée... [...]</p> <p>On sourit. Denise était devenue très pâle. Une honte la prenait, d'être ainsi changée en machine qu'on examinait et dont on plaisantait librement. Mme Desforges [...] ajouta méchamment :</p> <p>- Sans doute, il irait mieux si la robe de mademoiselle était moins large. Et elle jetait à Mouret le regard moqueur d'une Parisienne, que l'attiffement ridicule d'une provinciale égayait. [...]</p> <p>- Puis il faudrait être peignée, murmura-t-il.</p> <p>Ce fut le comble. Le directeur daignait rire, toutes ces demoiselles éclatèrent. Marguerite risqua un léger gloussement [...]; Clara avait lâché une vente, pour se faire du bon sang à son aise ; même des vendeuses de la lingerie étaient venues, attirées par la rumeur. [...] Denise avait encore pâli, au milieu de tout ce monde qui se moquait. Elle se sentait violente, mise à nu, sans défense. Quelle était donc sa faute, pour qu'on s'attaquât de la sorte à sa taille trop mince, à son chignon trop lourd ?</p>	<p>Émile Zola, <i>Au Bonheur des Dames</i>, Paris, Gallimard, coll. « Folio », 1980, pp. 151-153.</p>	Étayée	honte	Présentation des personnages et de leur rôle	Denise, le lieu et l'objet de la honte (« la honte la prenait ») ; Des agents de la honte : Mme Aurélie critique son apparence physique (« Vous comprenez, [le manteau] ne fait aucun effet sur mademoiselle, qui n'est guère étoffée ») ; Mme Desforges, ses vêtements (« Sans doute, il irait mieux si la robe de mademoiselle était moins large. ») ; le directeur, sa coiffure (« Puis, il faudrait être peignée ») ; Des spectateurs (Clara, « des vendeuses de la lingerie »)	Denise	aucune marque
-----	--	---	--------	-------	--	---	--------	---------------

129	<p>[Il est demandé à Denise, une apprentie vendeuse, d'essayer un manteau et ainsi de servir de mannequin.]</p> <p>Alors, tous les yeux revinrent sur Denise. Elle s'abandonnait aux mains de Marguerite, qui la faisait tourner lentement.</p> <p>- Hein ? qu'en pensez-vous ? demanda Mme Marty à Mme Desforges.</p> <p>[...]</p> <p>- Il n'est pas mal, et de coupe originale... Seulement, il me semble peu gracieux de la taille.</p> <p>- Oh ! intervint Mme Aurélie, il faudrait le voir sur madame elle-même... Vous comprenez, il ne fait aucun effet sur mademoiselle, qui n'est guère étoffée... [...]</p> <p>On sourit. Denise était devenue très pâle. Une honte la prenait, d'être ainsi changée en machine qu'on examinait et dont on plaisantait librement. Mme Desforges [...] ajouta méchamment :</p> <p>- Sans doute, il irait mieux si la robe de mademoiselle était moins large. Et elle jetait à Mouret le regard moqueur d'une Parisienne, que l'attiffement ridicule d'une provinciale égayait. [...]</p> <p>- Puis il faudrait être peignée, murmura-t-il.</p> <p>Ce fut le comble. Le directeur daignait rire, toutes ces demoiselles éclatèrent. Marguerite risqua un léger gloussement [...]; Clara avait lâché une vente, pour se faire du bon sang à son aise ; même des vendeuses de la lingerie étaient venues, attirées par la rumeur. [...] Denise avait encore pâli, au milieu de tout ce monde qui se moquait. Elle se sentait violente, mise à nu, sans défense. Quelle était donc sa faute, pour qu'on s'attaquât de la sorte à sa taille trop mince, à son chignon trop lourd ?</p>	<p>Émile Zola, <i>Au Bonheur des Dames</i>, Paris, Gallimard, coll. « Folio », 1980, pp. 151-153.</p>	Étayée	honte	Référence à des normes	le regard d'autrui (« Tous les yeux revinrent sur Denise », « changée en une machine qu'on examinait », « mise à nu ») fait peser sur Denise le poids de sa non-conformité à des normes sociales, attendues par les autres personnages présents (être « peignée », moins « mince »)	Denise	aucune marque
-----	--	---	--------	-------	------------------------	---	--------	---------------

## A.2. Cartographie des tâches de la campagne SemEval

La frise ci-dessous, réalisée pour le cours « Corpus annotés et développement de ressources linguistiques », assuré par Delphine Battistelli au premier semestre de l'année 2018-2019, présente l'ensemble des tâches proposées par la campagne d'évaluation SemEval, de 2007 à 2019. Les tâches concernant l'analyse de sentiments/d'opinion apparaissent en orange. Les intitulés des différentes tâches ainsi que les liens permettant d'accéder à leurs descriptions sont listés ci-dessous.



2007

- tâche 14 : « Affective Text », <https://aclweb.org/anthology/papers/S07-1013/s07-1013>

2010

- t. 18 : « Disambiguating Sentiment Ambiguous Adjectives », <http://semeval2.fbk.eu/semeval2.php?location=tasks#T3>

2013

- t. 2 : « Sentiment Analysis in Twitter », <https://www.cs.york.ac.uk/semeval-2013/task2.html>

2014

- t. 4 : « Aspect Based Sentiment Analysis », <http://alt.qcri.org/semeval2014/task4/>
- t. 9 : « Sentiment Analysis in Twitter », <http://alt.qcri.org/semeval2014/task9/>

2015

- t. 9 : « CLIPEval Implicit Polarity of Events », <http://alt.qcri.org/semeval2015/task9/>
- t. 10 : « Sentiment Analysis in Twitter », <http://alt.qcri.org/semeval2015/task10/>
- t. 11 : « Sentiment Analysis of Figurative Language in Twitter », <http://alt.qcri.org/semeval2015/task11/>
- t. 12 : « Aspect Based Sentiment Analysis », <http://alt.qcri.org/semeval2015/task12/>

2016

- t. 4 : « Sentiment Analysis in Twitter », <http://alt.qcri.org/semeval2016/task4/>
- t. 5 : « Aspect Based Sentiment Analysis », <http://alt.qcri.org/semeval2016/task5/>
- t. 6 : « Detecting Stance in Tweets », <http://alt.qcri.org/semeval2016/task6/>
- t. 7 : « Determining Sentiment Intensity of English and Arabic Phrases », <http://alt.qcri.org/semeval2016/task7/>

2017

- t. 4 : « Sentiment Analysis in Twitter », <http://alt.qcri.org/semEval2017/task4/>
- t. 5 : « Fine-Grained Sentiment Analysis on Financial Microblogs and News », <http://alt.qcri.org/semEval2017/task5/>

2018

- t. 1 : « Affect in Tweets », <https://competitions.codalab.org/competitions/17751>
- t. 2 : « Multilingual Emoji Prediction », <https://competitions.codalab.org/competitions/17344>
- t. 3 : « Irony Detection in English Tweets », <https://competitions.codalab.org/competitions/17468>

2019

- t. 3 : « EmoContext : Contextual Emotion Detection in Text », <https://www.humanizing-ai.com/emocontext.html>
- t. 4 : « Hyperpartisan News Detection », <https://pan.webis.de/semEval19/semEval19-web/>
- t. 5 : « HatEval: Multilingual Detection of Hate Speech Against Immigrants and Women in Twitter », <https://competitions.codalab.org/competitions/19935>
- t. 6 : « OffensEval: Identifying and Categorizing Offensive Language in Social Media », <https://competitions.codalab.org/competitions/20011>
- t. 9 : « Suggestion Mining from Online Reviews and Forums », <https://competitions.codalab.org/competitions/19955>



### A.3. Structures des pages du *P'tit Libé*

Les structures sont présentées de manière schématique de manière à n'inclure que les parties pertinentes pour la récupération des données.

#### A) La page des archives : <https://ptitlibe.liberation.fr/archives/>

```
<!DOCTYPE html>
<html lang="fr">
  [...]
  <body class="">
    [...]
    <main id="main-content" class="main-content">
      <div id="archives" class="flex-vertical">
        [...]
        <div class="archive-card">
          [...]
          <div class="folder-desc">
            <a href="/les-premiers-pas-sur-la-lune,101084">
              <h2>les premiers pas de l'homme sur la Lune</h2>
              <h3>N°115 - 19 au 25 juillet 2019</h3>
            </a>
            [...]
          </div>
        </div>
        ETC.
        <div class="archive-card">
          [...]
          <div class="folder-desc">
            <a href="/migrants-crise,100840">
              <h2>Les migrants</h2>
              <h3>N°1 - octobre 2015</h3>
            </a>
            [...]
          </div>
        </div>
      </div>
    </main>
    [...]
  </body>
</html>
```

Annotations in the diagram:

- Lien vers le numéro (points to `<a href="/les-premiers-pas-sur-la-lune,101084">`)
- Titre du numéro (points to `<h2>les premiers pas de l'homme sur la Lune</h2>`)
- Informations du numéro (points to `<h3>N°115 - 19 au 25 juillet 2019</h3>`)
- Numéro le plus récent (bracketed next to the first archive card)
- Numéro le plus ancien (bracketed next to the second archive card)

**B) Une page de numéro :** <https://ptitlibe.liberation.fr/migrants-crise,100840>

```

<!DOCTYPE html>
<html lang="fr">
  [...]
  <body class="">
    [...]
    <main id="main-content" class="main-content">
      <div id="folder-content" class="flex-vertical">
        <div id="folder-header" class="flex-vertical">
          [...]
          <h3>N°1 - octobre 2015</h3>
          <h1>Les migrants</h1>
          <div class="folder-desc">
            <p>Pour ce premier numéro, «le P'tit Libé» a décidé de raconter ce qu'on appelle la «crise des migrants». [...]</p>
          </div>
        </div>
      </div>
      <div id="folder-contents" class="flex-vertical">
        [...]
        <div id="articles">
          <a href="/p-tit-libe/2015/10/07/l-histoire-de-suyrana-10-ans_1592629" class="card">
            [...]
            <h2>L'histoire de Suyrana, 10 ans</h2>
          </a>
          <a href="/p-tit-libe/2015/10/07/qu-est-ce-qu-un-migrant_1592637" class="card">
            [...]
            <h2>Qu'est-ce qu'un migrant ?</h2>
          </a>
          ETC.
        </div>
      </div>
    </main>
  </body>
</html>

```

Informations du numéro

Titre du numéro

Description du numéro

Lien vers un article

Premier article

Deuxième article

**C) Une page d'article :** [https://ptitlibe.liberation.fr/p-tit-libe/2015/10/07/l-histoire-de-suyrana-10-ans\\_1592629](https://ptitlibe.liberation.fr/p-tit-libe/2015/10/07/l-histoire-de-suyrana-10-ans_1592629)

```

<!DOCTYPE html>
<html lang="fr">
  [...]
  <body class="">
    [...]
    <main id="main-content" class="main-content article-page">
      [...]
      <div id="article" class="flex-vertical">
        [...]
        <h1>L'histoire de Suyrana, 10 ans</h1>
        <div id="article-content">
          [...]
          <p>Son papa Mohammad était directeur du service touristique des douanes. [...]</p>
          <p>Il y a trois ans, son papa a donc décidé de protéger sa famille et de quitter le pays tous ensemble [...]</p>
          gtc.
        </div>
      </div>
    </main>
  </body>
</html>

```

Titre de l'article

Paragraphe de l'article

## D) Extrait d'un fichier de sortie .txt :

```
1 N°1 - octobre 2015
2 Les migrants
3 Pour ce premier numéro, «le P'tit Libé» a décidé de raconter ce qu'on appelle la «crise des migra:
4
5
6 L'histoire de Suyrana, 10 ans
7
8 Suyrana a 10 ans. Elle vivait avec ses deux petites sœurs et ses parents dans une grande maison d
9 Son papa Mohammad était directeur du service touristique des Bahamas. Mais un jour, la guerre a é
10 Il y a trois ans, son papa a donc décidé de protéger sa famille et de quitter le pays tous ensemb
11 Le papa de Suyrana connaissait un Français depuis très longtemps, qui a aidé la famille à venir s
12 Elle habite maintenant avec sa famille dans un petit appartement à Carhaix. Son papa cherche du t
13 Suyrana est aujourd'hui en CE2, elle s'est fait des amis et a appris le français. Elle le parle m
14 LE PARCOURS DE SUYRANA
15
16
17 Qu'est-ce qu'un migrant ?
18
19 Un migrant, c'est quelqu'un qui quitte son pays. Il peut aller dans le pays d'à côté ou très loin
20 Certaines personnes décident de partir parce qu'elles ont envie de découvrir une autre culture, o
21 D'autres migrants ne choisissent pas vraiment de quitter leur pays. Ils auraient préféré rester c
```

## A.4. Script parsePtitLibe.py

```
import os
import requests #charger la page et stocker son contenu dans une variable
from bs4 import BeautifulSoup #parser les pages html
import re
import argparse

###----- DESCRIPTION -----###
# A partir d'urls du P'tit Libé, récupère le contenu textuel des numéros.
# Génère un fichier .txt pour chaque numéro.
# Sans argument, url = page des archives, sortie = répertoire courant
# Avec "-u + nom_de_fichier.txt" : un fichier avec une liste d'urls de numéros à récupérer (une url/ligne)
# Avec "-o + nom de répertoire" : le répertoire de sortie pour créer les fichiers

###----- LANCEMENT -----###
#Le script n'a aucun argument obligatoire, on peut donc lancer : python3 parsePtitLibe.py
#Sinon : python3 parsePtitLibe.py -u nom_de_fichiers_avec_les_urls.txt -o nom_de_répertoire_de_sortie

###----- FONCTIONS -----###

def clean_all(str):
    """
    Entrée : une chaîne de caractères
    Sortie : la chaîne, avec "\xa0" remplacé par un espace "normal"
    """
    return re.sub("\xa0", " ", str)

def getContent(url):
    """
    Entrée : une url
    Sortie : le contenu html de l'url (parsé par BeautifulSoup)
    """

    #requête
    requete = requests.get(url)
    #stockage du contenu de la page
    page = requete.content
    #parsing de la page par BeautifulSoup
    soup = BeautifulSoup(page, features="lxml")

    return soup

def get_urls_from_file(file):
    """
    Entrée : le chemin d'un fichier avec des urls (une url/ligne)
    Sortie : la liste des urls
    """

    liste_numeros = list()
    with open(file, 'r', encoding="utf-8") as in_stream: #ouverture du fichier
        for line in in_stream: #pour chaque ligne
            liste_numeros.append(line.strip())

    return liste_numeros
```

```

def getURLS(url_archives="https://ptitlibe.liberation.fr/archives/"):
    """
    Fontionne avec l'url des archives du P'tit Libé.
    Entrée : Rien (l'url des archives est rentrée par défaut)
    Sortie : La liste des urls des numéros du P'tit Libé présents sur la page des archives.
    """

    archives = getContent(url_archives)

    #on ne veut pas :
    # 1. https://ptitlibe.liberation.fr/journal-enfants-twiki-festival,101000
    # 2. https://ptitlibe.liberation.fr/concours-faces-objets,100994
    # 3. https://ptitlibe.liberation.fr/paris-attacks,100870 (les versions anglaises)
    #ne prend pas non plus en compte les numéros spéciaux
    # ex. : https://ptitlibe.liberation.fr/calendrier-avent-2018,101011
    liste_numeros = [
        "https://ptitlibe.liberation.fr"+num.find("a").get("href") #on récupère le lien et on reconstitue l'url
        for num in archives.find_all("div", {"class": "folder-desc"})
        if (num.find("h3") and
            "N°" in num.find("h3").string and
            "N°100" not in num.find("h3").string and
            "English" not in num.find("h2").string) #pour filtrer les urls qu'on ne veut pas
    ]

    return liste_numeros
## POUR UN NUMERO ##

#On veut récupérer les caractéristiques du numéro (n°, dates, titre, description)
#ainsi que les url renvoyant aux articles qui le composent.
bruit = re.compile(r"<[>]+>") #pour enlever les balises <br/>, <em>, <img>, <a>, <p> etc. si besoin
def getInfosAndURLs(url):
    """
    Pour une url de numéro du P'tit Libé donnée, renvoie :
    - les caractéristiques du numéro
    - le titre
    - la description
    - la liste des urls des articles qui le composent
    """

    #dans les urls qu'on ne veut pas
    filtre = ["quiz", "mon-coin-lecture", "le-coin-lecture", "imprime-ce-numero", "memory"]

    numero = getContent(url) #on récupère la page html

    #on récupère les caractéristiques du numéro
    h3 = numero.find("h3").string #donne la lère occurrence d'une balise h3
    h1 = numero.find("h1").string #titre du numéro
    desc = numero.find("div", {"class": "folder-desc"}).find("p") #description du numéro
    #N.B.: pour la description, il se peut que des balises (ex. : <br/>) se situent au sein de la balise <p>
    #desc.string renverra alors None.
    #Pour éviter les problèmes ultérieurs, dans le retour de résultat de cette fonction,
    # on transforme desc en string et on enlève les balises <p> (et les autres balises) grâce à la regEx bruit

    #on récupère les liens vers les articles
    articles = [
        "https://ptitlibe.liberation.fr"+lien.get("href") #on reconstitue les urls
        for lien in numero.find("div", {"id": "articles"}).find_all("a") #on récupère les balises <a>
        if not re.search("|".join(filtre), lien.get("href")) #on filtre pour enlever les urls qu'on ne veut pas
    ]

    return clean_all(h3), clean_all(h1), clean_all(re.sub(bruit, "", str(desc))), articles

def nettoieTitre(titre):
    """
    nettoie un titre d'article pour que ça soit un nom de fichier 'propre'
    càd pas de caractères accentués, pas d'espaces etc.
    """

    fic = re.sub("«|»|,|:", "", titre) #on commence par enlever les guillemets et les virgules
    #comme l'ordre de remplacement n'importe pas,
    #on utilise maketrans, pour remplacer toute une série de caractères
    translator= fic.maketrans("éèêââùùôç' ' ", "eeeeaauc___")

    return fic.translate(translator)

```

```

## POUR LES ARTICLES ##
#On veut accéder à tous les articles qui composent le numéro (les urls qu'on a récupérées précédemment)
#Pour chaque article, on veut son titre et son contenu textuel.
def getTitleAndText(lien):
    """
    Entrée : un lien d'article
    Sortie : le titre de l'article + le contenu textuel
    """

    article = getContent(lien) #récupération et parsing de la page

    #le contenu textuel est dans les balises <p>, elles-mêmes dans <div id="article-content">
    #/>\ certaines balises <p> sont un peu spéciales et ne nous intéressent pas (class="")
    #certains éléments de la liste sont vides : on les enlève dans le retour de résultat
    contenu = [
        re.sub(bruit, '', str(e)) #on enlève les balises
        for e in article.find("div", {"id": "article-content"}).find_all("p", {"class":""})
    ]

    #pour éviter les problèmes liés à la présence de balises <br> par ex.
    contenu = [clean_all(e.strip()) for e in contenu if e != '']

    return clean_all(article.find("h1").string), [e for e in contenu if e]

###----- EXECUTION -----###
if __name__ == '__main__':

    ### ARGUMENTS ###
    argparser = argparse.ArgumentParser()
    argparser.add_argument("-u", "--url", required=False, help="Fichier avec les urls du P'tit Libé à traiter")
    argparser.add_argument("-o", "--output", required=False, help="Chemin du répertoire de sortie")
    args = vars(argparser.parse_args())
    url = args["url"]
    path_out = args["output"]

    ### SORTIE ###
    if not path_out : # si aucun répertoire de sortie n'a été spécifié
        path_out = ""

    try: # on essaye de créer un répertoire de sortie
        os.mkdir(path_out)
    except OSError : #si le répertoire de sortie spécifié existe déjà
        pass

    ### URLS A TRAITER ###
    if url : # si un fichier avec des urls a été fourni en entrée
        urls_numeros = get_urls_from_file(url) #on récupère la liste d'urls
    else :
        urls_numeros = getURLS() #on prend les urls des numéros de la page des archives (url par défaut)

    ### TRAITEMENT DES URLS ###
    for url in urls_numeros : #pour chaque numéro de la liste

        h3, h1, desc, articles = getInfosAndUrls(url) #on récupère les infos sur le numéro
        print("Getting content of :",h1) #pour savoir où on en est
        #print("\n".join(articles)) #pour savoir quels articles seront récupérés

        fic=nettoieTitre(h1) #on prépare le nom du fichier

        #ouverture du fichier qui accueille le contenu du numéro
        with open(os.path.join(path_out, fic)+".txt", 'w', encoding="utf-8") as out:
            try :
                out.write(h3+"\n"+h1+"\n"+desc+"\n\n") #écriture des infos concernant le numéro
            except TypeError : #parfois problème de lecture du fichier html (on a donc None au lieu de string)
                print("****Problème avec :",h3,h1,desc)
            for lien in articles : #pour chaque article du numéro
                titre, contenu = getTitleAndText(lien) #on récupère le titre et le contenu
                out.write(titre.strip()+"\n\n"+"\n".join(contenu)+"\n\n\n") #on écrit le tout dans le fichier

```

## A.5. Détail de la démarche d'aspiration des numéros du *P'tit Libé*

Sur le site du *P'tit Libé*, trois types de pages nous ont été utiles :

1. la page des archives, qui contient les liens vers tous les numéros du journal depuis octobre 2015 ;
2. les pages de numéros (une page par numéro), qui contiennent des informations sur le numéro (titre, date, description) et les liens vers tous les articles qui le composent ;
3. les pages des articles (un page par article), qui contiennent le titre de l'article et son contenu textuel (sous forme de paragraphes).

Toutes ces pages sont structurées selon le format HTML, soit selon un système de balises (signalées par des chevrons « < » et « > ») qui organisent le contenu informationnel. Par exemple, si une page contient un lien vers une autre page, ce lien sera signalé par une balise « <a> » ayant un attribut « href » qui donne la référence du lien hypertexte :

```
« <a href="/p-tit-libe/2015/10/07/1-histoire-de-suyrana-10-ans_1592629"> »
```

Après nous être assurés que la structure des pages était régulière, nous avons pu automatiser le processus de récupération des numéros du *P'tit Libé*. Notre script `parsePtitLibe.py` part de la page des archives et cherche toutes les balises « <div> » ayant pour classe « folder-desc » (soient les balises de la forme « <div class="folder-desc"> ») : ce sont les balises qui signalent un numéro du journal. Des conditions devaient être remplies par les éléments imbriqués sous cette balise pour qu'elle soit retenue :

- l'élément « <h3> » devait obligatoirement contenir « N° » (ex. « <h3>N°1 - octobre 2015</h3> »), ce qui nous a permis d'éliminer les numéros spéciaux<sup>34</sup> ;
- l'élément « <h2> » ne devait pas contenir le terme « English » (ex. « <h2>Paris attacks - English version</h2> »), ce qui nous a permis d'éliminer les numéros en anglais<sup>35</sup>.

Grâce à ce filtrage, seuls les numéros pertinents ont été conservés. Nous avons alors récupéré la valeur de l'attribut « href » de la balise « <a> » pour obtenir le lien vers les numéros (ex. dans « <a href="/migrants-crise,100840"> », « /migrants-crise,100840 » est le lien vers le numéro du *P'tit Libé* consacré aux migrants). Nous avons ensuite reconstitué les URLs (ajout de « <https://ptitlibe.liberation.fr> » avant le lien) pour pouvoir accéder aux numéros (ex. « <https://ptitlibe.liberation.fr/migrants-crise,100840> » est l'URL reconstitué pour le numéro sur les migrants) et avons ainsi obtenu la liste des URLs de tous les numéros du *P'tit Libé*.

Sur la page de chaque numéro, nous avons récupéré trois types d'informations :

1. les informations sur le numéro, contenues dans la balise « <h3> » de la page (ex. « <h3>N°1 - octobre 2015</h3> ») ;
2. le titre du numéro, contenu dans la balise « <h1> » de la page (ex. « <h1>Les migrants</h1> ») ;

---

34 Par exemple, un numéro qui présente les ateliers organisés par le *P'tit Libé* : <https://ptitlibe.liberation.fr/journal-enfants-txiki-festival.101000>

35 Par exemple : <https://ptitlibe.liberation.fr/paris-attacks.100870>

3. la description du numéro, contenue dans la balise « <p> », imbriquée dans la balise « <div> » ayant pour classe « folder-desc » (« <div class="folder-desc"> ») (ex. « <div class="folder-desc"><p>Pour ce premier numéro, «le P'tit Libé» a décidé de raconter ce qu'on appelle la «crise des migrants». Parce qu'on voit beaucoup d'images de gens qui fuient leur pays sur des bateaux, qu'on entend parler de «réfugiés» et de pays qui se disputent sur la façon de les accueillir, et que tout ça n'est pas forcément facile à comprendre.</p></div> »);
4. la liste des liens vers les articles qui composent le numéro. Chaque lien est représenté par une balise « <a> » et correspond à la valeur de son attribut « href ». Toutes les balises « <a> » sont imbriquées dans la balise « <div> » ayant pour identifiant « articles » (ex. dans « <div id="articles"><a href="/p-tit-libe/2015/10/07/l-histoire-de-suyrana-10-ans\_1592629">[...]</a></div> », « [/p-tit-libe/2015/10/07/l-histoire-de-suyrana-10-ans\\_1592629](/p-tit-libe/2015/10/07/l-histoire-de-suyrana-10-ans_1592629) » est le lien de l'article « L'histoire de Suyrana, 10 ans »). Comme précédemment, il suffit ensuite de reconstituer l'URL pour accéder aux articles (ex. « [https://ptitlibe.liberation.fr/p-tit-libe/2015/10/07/l-histoire-de-suyrana-10-ans\\_1592629](https://ptitlibe.liberation.fr/p-tit-libe/2015/10/07/l-histoire-de-suyrana-10-ans_1592629) »).

Ensuite, nous sommes allés sur la page de chaque article, et ce pour chaque numéro, afin de récupérer le titre de l'article (dans la balise « <h1> », par ex. « <h1> L'histoire de Suyrana, 10 ans</h1> ») et le contenu de l'article. Ce contenu se trouvait au sein de la balise « <div> » ayant pour identifiant « article-content » (<div id="article-content">) et était structuré en paragraphes. Chaque paragraphe était représenté par une balise « <p> » (<p>), dont nous avons donc conservé le contenu textuel (ex. « <div id="article-content"><p>Son papa Mohammad était directeur du service touristique des douanes. Mais un jour, la guerre a éclaté et la vie est devenue dangereuse. Suyrana ne pouvait plus aller à l'école. Il y avait des bombardements plusieurs fois par semaine.</p>[...]</div> »).



## A.6. Liste des numéros du *P'tit Libé* aspirés

Les numéros sont présentés par ordre chronologique.

- N°1 - octobre 2015 - De l'animal à la viande
- N°2 - novembre 2015 - Donald Trump
- N°3 - novembre 2015 - Jean-Luc Mélenchon
- N°4 - décembre 2015 - La BD avec Mortelle Adèle
- N°5 - janvier 2016 - La colère des gilets jaunes
- N°6 - février 2016 - La COP21
- N°7 - mars 2016 - La Corée du Nord
- N°8 - avril 2016 - La Coupe du monde de foot
- N°9 - mai 2016 - La disparition des abeilles
- N°10 - juin 2016 - La disparition des animaux sauvages
- N°11 - juillet 2016 - La Fashion Week
- N°12 - août 2016 - La gauche et la droite en politique
- N°13 - septembre 2016 - La guerre en Syrie
- N°14 - octobre 2016 - La loi travail
- N°15 - novembre 2016 - La méfiance envers les journalistes
- N°16 - décembre 2016 - La pédophilie
- N°17 - janvier 2017 - La protection du patrimoine
- N°18 - février 2017 - La rentrée télé
- N°19 - mars 2017 - La sécurité routière
- N°20 - avril 2017 - La SNCF
- N°21 - mai 2017 - Les armes à feu aux États-Unis
- N°22 - juin 2017 - Les attentats à Paris
- N°23 - juillet 2017 - Les césars et les oscars
- N°24 - septembre 2017 - Les coulisses des dessins animés
- N°25 - 29 septembre au 5 octobre 2017 - Les coulisses de l'école
- N°26 - 6 au 12 octobre 2017 - Les dangers du glyphosate
- N°27 - 13 au 19 octobre 2017 - Les dessous du cinéma
- N°28 - 20 au 26 octobre 2017 - Les dinosaures
- N°29 - 27 octobre au 2 novembre 2017 - Les données personnelles
- N°30 - 3 au 9 novembre 2017 - Les extraterrestres
- N°31 - 9 au 16 novembre 2017 - Les fake news
- N°32 - 17 au 23 novembre 2017 - Les grands prix littéraires
- N°33 - 24 au 30 novembre 2017 - Les grèves
- N°34 - 1er au 7 décembre 2017 - Les impôts
- N°35 - 8 au 14 décembre 2017 - Les jeux de société
- N°36 - 15 au 21 décembre 2017 - Les Jeux olympiques
- N°37 - 22 au 28 décembre 2017 - Les Jeux olympiques d'hiver

N°38 - 29 décembre 2017 au 4 janvier 2018 - Les jeux vidéo  
N°39 - 5 au 11 janvier 2018 - Les JO à Paris  
N°40 - 12 au 18 janvier 2018 - Les manifestations en Iran  
N°41 - 19 au 25 janvier 2018 - Les métiers de Noël  
N°42 - 26 janvier au 1er février 2018 - Les migrants  
N°43 - 2 au 8 février 2018 - Les référendums  
N°44 - 9 au 15 février 2018 - Les régionales  
N°45 - 16 au 22 février 2018 - Les repas de fêtes  
N°46 - 23 février au 1er mars 2018 - Les restaurants étoilés  
N°47 - 2 au 8 mars 2018 - Les Rohingyas, peuple chassé de Birmanie  
N°48 - 9 au 15 mars 2018 - Les rythmes scolaires  
N°49 - 16 au 22 mars 2018 - Les secrets du 29 février  
N°50 - 23 au 29 mars 2018 - Les théories du complot  
N°51 - 30 mars au 5 avril 2018 - Les tubes de l'été  
N°52 - 6 au 12 avril 2018 - Les vaccins  
N°53 - 13 au 19 avril 2018 - Les végétariens  
N°54 - 20 au 26 avril 2018 - Les youtubeurs  
N°55 - 27 avril au 3 mai 2018 - Le Brexit  
N°56 - 4 au 10 mai 2018 - Le changement d'heure  
N°57 - 11 mai au 17 mai 2018 - Le conflit Israël-Palestine  
N°58 - 11 au 17 mai 2018 - Le départ du ministre de l'écologie  
N°59 - 25 au 31 mai 2018 - Le féminisme  
N°60 - 1er au 7 juin 2018 - Le festival d'Avignon  
N°61 - 8 au 14 juin 2018 - Le Front national  
N°62 - 14 au 21 juin 2018 - Le harcèlement à l'école  
N°63 - 22 au 28 juin 2018 - Le mal-logement  
N°64 - 29 juin au 5 juillet 2018 - Le mariage royal  
N°65 - 6 au 12 juillet 2018 - Le pétrole, utile mais polluant  
N°66 - 13 au 19 juillet 2018 - Le plan loup  
N°67 - 20 au 26 juillet 2018 - Le premier anniversaire du président à l'Elysée  
N°68 - 24 au 30 août 2018 - Le prix Nobel de la paix  
N°69 - 31 août au 6 septembre 2018 - Le «P'tit Libé» des philosophes  
N°70 - 7 au 13 septembre 2018 - Le P'tit Libé des philosophes  
N°71 - 14 au 20 septembre 2018 - Le Salon de l'agriculture et les éleveurs  
N°72 - 21 au 27 septembre 2018 - Le succès du rap  
N°73 - 28 septembre au 4 octobre 2018 - Le Tour de France  
N°74 - 5 au 11 octobre 2018 - L'accueil des migrants en France  
N°75 - 12 au 18 octobre 2018 - L'affaire Benalla  
N°76 - 19 au 25 octobre 2018 - L'affaire Weinstein et le harcèlement sexuel  
N°77 - 26 octobre au 1er novembre 2018 - L'antisémitisme

N°78 - 2 au 8 novembre 2018 - L'art contemporain  
N°79 - 9 au 15 novembre 2018 - L'Assemblée nationale  
N°80 - 15 au 22 novembre 2018 - L'attaque de Nice  
N°81 - 23 au 29 novembre 2018 - L'attentat de Strasbourg  
N°82 - 30 novembre au 6 décembre 2018 - L'élection américaine  
N°83 - 7 au 13 décembre 2018 - L'élection du président au Brésil  
N°84 - 13 décembre 2018 - L'élection présidentielle  
N°85 - du 21 au 27 décembre 2018 - L'empire Disney  
N°86 - 28 décembre 2018 au 3 janvier 2019 - L'Euro de foot  
N°87 - 4 au 10 janvier 2019 - L'homophobie  
N°88 - 11 au 17 janvier 2019 - L'indépendance de la Catalogne  
N°89 - 18 au 24 janvier 2019 - Mai 68  
N°90 - 25 au 31 janvier 2019 - météo : un hiver perturbé  
N°91 - 1 au 7 février 2019 - Ouragans : une année record  
N°92 - 8 au 14 février 2019 - Parcoursup et l'avenir des jeunes  
N°93 - 15 au 21 février 2019 - Roland-Garros  
N°94 - 22 au 28 février 2019 - Sécurité : la vie après les attentats  
N°95 - 1er au 7 mars 2019 - Trump et les élections de mi-mandat  
N°96 - 8 au 14 mars 2019 - Vladimir Poutine  
N°97 - 15 au 21 mars 2019 - Zika et les épidémies

## A.7. Obtenir le « corpusHashcode » pour les fichiers .aa

### A) Script en Java

```
String hash = "";
FileInputStream s = null;
try
{
    s = new FileInputStream(f);
    int length = s.available();
    hash = "" + length+"-";
    long code=1;
    for (int i = 0; i < length; i++)
    {
        int n = s.read();
        if (n!=0)
        {
            code*=n;
            code=code%99999999;
        }
        //System.out.println("code="+code);
    }
    hash+=code;
}
}
```

### B) Traduction en Python3

```
f_ac = open(path_f_ac, 'rb').read() #ouverture et lecture en octets du fichier .ac
length = len(f_ac) #stockage de la longueur du fichier
hashcode = str(length)+"-" #constitution de la lère partie de l'identifiant

code = 1 #pour accueillir la 2nde partie de l'identifiant
for i in range(0,length): #pour i de 0 à la longueur du fichier
    n = int(f_ac[i]) #stockage de l'octet de rang i dans le fichier
    if n != 0 : # si n est différent de zéro
        code *= n #multiplication de la valeur de « code » par n
        code = code%99999999 #stockage du modulo de « code » dans « code »
hashcode += str(code) #concaténation de la 2nde partie de l'identifiant
```

## A.8. Explication technique du calcul des indices dans un fichier .ac

Dans le logiciel Glozz, l'annotation est contenue dans un fichier .aa et le contenu textuel des texte annotés dans un fichier .ac. Le lien entre les deux est opéré par la délimitation des unités : elles ont une borne de début et une borne de fin, représentées respectivement par les indices, au sein du fichier .ac, du caractère de début et du caractère de fin des unités. Un peu de réflexion a été nécessaire pour savoir comment trouver ces indices.

Les indices permettant de positionner les paragraphes correctement sont ceux des caractères dans le fichier .ac. Or, ce fichier correspond au contenu textuel du fichier .txt, débarrassé de tous les retours à la ligne : la structuration en paragraphes est donc perdue dans le fichier .ac. Cette information doit alors être récupérée d'après le fichier .txt : il a ainsi été nécessaire de trouver un moyen pour faire le lien entre l'emplacement des paragraphe dans le fichier .txt et dans le fichier .ac. Dans notre script, le contenu textuel des fichiers .txt est stocké sous la forme d'une liste de paragraphes : si on colle bout à bout (si on concatène) tous les éléments de cette liste, on obtient un fichier .ac. Ainsi :

- Soit L une liste composée des paragraphes  $\mathcal{P}$  d'un fichier .txt.
- Soit  $\mathcal{P}_n$  le n-ième paragraphe de L, composé de caractères (de lettres).
- Soit  $C_i$  un caractère de  $\mathcal{P}_n$ , d'indice i au sein de  $\mathcal{P}_n$ .
- Soit A un fichier .ac, assimilable à une chaîne de caractères résultant de la concaténation de tous les éléments ordonnés de L, soit de tous les éléments  $\mathcal{P}$ .

$$L = [ \underbrace{\langle\langle c_0, c_1, \dots, c_i \rangle\rangle}_{\mathcal{P}_1}, \underbrace{\langle\langle c_0, c_1, \dots, c_i \rangle\rangle}_{\mathcal{P}_2}, \dots, \underbrace{\langle\langle c_0, c_1, \dots, c_i \rangle\rangle}_{\mathcal{P}_n} ]$$

$$A = [ \langle\langle c_{0\mathcal{P}_1}, c_{1\mathcal{P}_1}, \dots, c_{i\mathcal{P}_1}, c_{0\mathcal{P}_2}, c_{1\mathcal{P}_2}, \dots, c_{i\mathcal{P}_2}, \dots, c_{0\mathcal{P}_n}, c_{1\mathcal{P}_n}, \dots, c_{i\mathcal{P}_n} \rangle\rangle ]$$

⇒ Dans A, l'indice j du caractère  $C_i$ , d'indice i dans  $\mathcal{P}_n$ , correspondra à i augmenté de la somme des longueurs des paragraphes précédant  $\mathcal{P}_n$  :

$$j = i + \sum_{k=0}^{n-1} \text{len}(P_k)$$

D'après cette définition, dans un paragraphe P :

- pour le 1<sup>er</sup> caractère, d'indice 0 dans P,  $j = \sum_{k=0}^{n-1} \text{len}(P_k)$  , où n est la position de P dans L ;
- pour le dernier caractère, d'indice z dans P,  $j = z + \sum_{k=0}^{n-1} \text{len}(P_k)$  Or,  $n = \text{len}(P) - 1$ .

Donc  $j = \text{len}(P) - 1 + \sum_{k=0}^{n-1} \text{len}(P_k)$  , avec n la position de P dans L.

La borne initiale d'un paragraphe P correspondra ainsi à la somme des longueurs des paragraphes précédant P. La borne finale correspondra à :

$$\text{borne finale} = \text{len}(P) - 1 + \sum_{k=0}^{n-1} \text{len}(P_k) + 1$$

Autrement, dit, et la borne finale correspondra à la somme des longueurs des paragraphes précédents, augmentée de la longueur du paragraphe.

## A.9. Lexique élaboré lors du stage sur le projet TextToKids

Ce lexique a été adapté de EMOTAIX, lexique proposé par Piolat et Bannour (2009). EMOTAIX comporte 56 catégories et environ 2300 termes. Lors de l'adaptation du lexique, uniquement 24 catégories émotionnelles et environ 1420 termes ont été conservés. Seules les catégories utilisées pour pré-annoter notre corpus sont présentées ici.

colere : acariâtre, acariâtreté, bisque, bisquer, bouder, bouderie, boudeur, bougon, bougonner, colère, coléreuse, coléreusement, coléreux, colérique, courroucer, courroux, dispute, disputer, endêver, engueulade, engueuler, enrager, enragé, exaspérant, exaspération, exaspérer, excédant, excéder, furax, fureur, furibard, furibond, furie, furieusement, furieux, fâcher, fâcherie, fâcheusement, fâcheux, fâché, grincher, grincheux, grognon, grommelage, grommelant, grommeler, grommellement, hargne, hargneusement, hargneux, importun, importuner, indignation, indignement, indigner, indigné, intolérable, intolérablement, intolérance, intolérant, invectif, invective, invectiver, invectif, irascibilité, irascible, ire, maronnant, maronner, maugréant, maugréer, maugréur, mauvaise humeur, mécontent, mécontentement, mécontenter, outrance, outrancier, outrancièrement, outrer, outré, pester, querelle, quereller, querelleur, querelleux, rage, rageant, rager, rageur, rogne, ronchon, ronchonnant, ronchonne, ronchonnement, ronchonner, ronchonneur, ronchonneux, scandale, scandaleux, scandaliser, à la limite du supportable, énervant, énervement, énerver, énervé, révoltant, révolte, révolter, révolté

culpabilite : blâmable, blâmé, condamnable, condamné, contrit, contrition, coupable, coupablement, culpabiliser, culpabilité, dans son tort, en tort, faute, fautif, fautivement, regret, regrettable, regretter, remords, repentance, repentant, repentir, repentir, reprochable, reproche, reprocher, s'en vouloir

degout : abject, abjectement, abjection, aversion, blasement, blaser, blasé, débecter, dégoût, dégoûtant, dégoûtation, dégoûter, dégoûté, dégueulasse, dégueulasserie, dégueux, exécration, exécration, exécrer, faire horreur, ignobilité, ignoble, ignoblement, immonde, immondice, odieusement, odieux, rebut, rebutable, rebutant, rebuter, répugnance, répugnant, répugner, vil, vilénie

embarras : affront, aucun respect, complexe d'infériorité, complexé, déshonneur, déshonorant, déshonorer, honte, honteusement, honteux, humiliant, humiliation, humilier, humilié, ignominie, ignominieusement, ignominieux, infamant, infamie, irrespect, irrespectueusement, irrespectueux, mal à l'aise, mal à son aise, mortifiant, mortification, mortifier, offensant, offense, offenser, offensé, outrage, outrageant, outrager, outrageusement, outrageux, outragé, penaud, pudeur, pudique, pudiquement, sentiment d'infériorité, vexant, vexation, vexer, vexé

fierte : amour de soi, amour-propre, digne, dignement, dignité, fier, fierté, fièrement, honneur, honorable, honorant, honorer, honorifique, honorifiquement

jaalousie : envier, envieusement, envieux, jalousement, jalouser, jalousie, jaloux

joie : allègre, allègrement, allégresse, amusant, amusement, amuser, badin, badiner, bien-être, bienheureux, bonheur, bonne humeur, boute en train, boute-en-train, béat, béatitude, content, contentement, contenter, divertir, divertissant, divertissement, drôle, drôlement, délectation, délecter, délice, délicieusement, délicieux, enjouement, enjoué, enthousiasmant, enthousiasme, enthousiasmer, enthousiaste, entrain, entraînant, euphorie, euphorique, euphorisant, exaltant, exaltation, exalter, exalté, extase, extasier, extatique, exultation, exulter, folichon, folichonnerie, folâtre, folâtrément, folâtrer, folâtrerie, fou rire, félicité, gai, gaiement, gaieté, gaité, guilleret, heureusement, heureux, hilarant, hilare, hilarité, joie, jouir, jouissance, jouissant, jouissif, jovial, jovialement, jovialité, joyeusement, joyeuseté, joyeux, jubilant, jubilation, jubilatoire, jubiler, liesse, marrade, marrant, marrer, nirvana, optimisme, optimiste, optimistement, plaisir, pleurer de

rire, plénitude, pouffer, ravi, ravir, ravissement, rieur, rigolade, rigoler, rigolo, rire, rire aux larmes, rire aux éclats, régal, régalant, régaler, réjouir, réjouissance, réjouissant, satisfaction, satisfaire, satisfait, sémillant, ébaubi, ébaubir, ébaudissement, également, égayant, égayer, merveille, merveilleusement, merveilleux

peur : abominable, abominer, affolant, affolement, affoler, affolé, affres, affreusement, affreux, alarmant, alarme, alarmer, alarmisme, alarmiste, alerte, alerter, angoissant, angoisse, angoisser, angoissé, anxieusement, anxieux, anxiété, apeurement, apeurer, apeuré, appréhender, appréhensif, appréhension, atroce, atrocement, atrocité, cauchemar, cauchemarder, cauchemardesque, couard, couardise, craindre, crainte, craintif, danger, dangereusement, dangereux, dangersité, effarouchable, effarouchant, effarouchement, effaroucher, effrayant, effrayer, effroi, effroyable, effroyablement, en abomination, frayeur, froussard, frousse, horreur, horrible, horriblement, horrifiant, horrifier, inquiet, inquiétant, inquiéter, inquiétude, insécurité, intimidant, intimider, lâche, lâcheté, menace, menacer, menacé, menaçant, monstre, monstrueusement, monstrueux, monstruosité, méfiance, méfiant, méfier, paniquant, paniquante, panique, paniquer, peur, peureusement, peureux, phobie, phobique, phobiquement, pleutre, pleutrement, pleutrerie, poltron, poltronnerie, préoccupant, préoccupation, préoccuper, préoccupé, pusillanime, pusillanimité, péril, périlleux, péteuse, péteux, pétochard, pétoche, qui-vive, redoutable, redoutablement, redouter, se déballonner, sinistre, sinistrement, souci, soucier, soucieusement, soucieux, soupçonneux, stress, stressant, stresser, stressé, suspens, suspense, terreur, terrible, terriblement, terrifiant, terrifier, terrorisant, terrorisation, terroriser, timide, timidement, timidité, timoré, trac, trouillard, trouille, venette, épeurer, épouvantable, épouvantail, épouvante, épouvamment, épouvanter

surprise : ahuri, ahurir, ahurissant, ahurissement, déconcertant, déconcerter, décontenancement, décontenancer, décontenancé, décontenancée, effarade, effarant, effarement, effarer, effaré, interloquer, interloqué, ne pas s'y attendre, pantois, sidérant, sidération, sidérer, surprenant, surprendre, surpris, surprise, tourneboulage, tourneboulement, tournebouler, troublant, trouble, troubler, ébahi, ébahir, ébahissement, époustouflant, époustoufler, étonnamment, étonnant, étonnement, étonner, étonné

tristesse : misère, misérable, misérablement, abattement, abattre, abattu, accablant, accablement, accabler, affliction, affligeant, affliger, attristant, attrister, calamiteux, calamité, chagrin, chagrinant, chagriner, consterner, contrister, couci couça, crève-coeur, deuil, dramatique, dramatiquement, dramatisant, dramatiser, drame, décourageant, découragement, décourager, délaissement, délaisser, délaissé, démoralisant, démoralisation, démoraliser, déplorable, déplorablement, déploration, déplorer, déprimant, déprime, déprimer, déprimé, désabusement, désabusé, désespoir, désespérance, désespérant, désespérer, désespéré, désolant, détresse, endeuiller, ennui, ennuyant, ennuyer, ennuyeusement, ennuyeux, ennuyé, esseulement, esseulé, gémir, gémissant, gémissement, inconsolable, inconsolation, inconsolé, infortune, infortuné, isolant, isolement, isoler, isolé, lamentation, lamenter, lugubre, malheur, malheureusement, malheureux, maussade, maussaderie, morfondement, morfondre, morfondu, morne, morose, morosité, mélancolie, mélancolique, mélancoliquement, navrant, navrement, navrer, neurasthénie, neurasthénique, nostalgie, nostalgique, nostalgiquement, patraque, peine, peiner, pessimisme, pessimiste, pessimistement, prostration, prostré, rabat-joie, sans espoir, spleen, taciturne, taciturnement, triste, tristement, tristesse, tristoune



## A.10. Script txt2GlozzFiles.py

```
#!/usr/bin/python3
# -*- coding: utf-8 -*-

import os
import re
import sys
import time, datetime
from collections import defaultdict

###-----INFOS-----###
#Pour un ensemble de fichiers .txt donné, génère les fichiers .ac et .aa nécessaires pour le logiciel Glozz.
#Utilise un lexique de la forme "terme\tcatégorie\POS" ou "terme\tcatégorie", selon si on veut les POS ou non.
#Grâce au lexique, les fichiers .aa contiennent une pré-annotation des termes du lexique émotionnel.
#Les arguments à donner sont, dans l'ordre :
# chemin du répertoire avec les fichiers .txt
# chemin du répertoire avec les fichiers .ac
# chemin du répertoire avec les fichiers .aa
# identifiant de l'annotateur
# "oui" ou "non" selon si le lexique contient les informations de POS ou non
# chemin du lexique

#Lancement : python3 txt2GlozzFiles_v2.py ./txt ./ac ./aa aetienne oui ./lexique_formes_cat_pos.tsv
# OU
# python3 txt2GlozzFiles_v2.py ./txt ./ac ./aa aetienne non ./lexique_formes_cat.tsv

###-----FONCTIONS-----###
def get_files(rep):
    """
    Récupère la liste des fichiers contenus dans un répertoire donné.
    Entrée : nom du répertoire
    Sortie : liste des fichiers
    """
    files = [
        os.path.join(rep, fname) #reconstitution du chemin
        for fname in os.listdir(rep) #pour chaque fichier dans le répertoire donné
        if fname.endswith(".txt") #filtre : uniquement les .txt
    ]
    return files
separateur = re.compile("\n+") #pour retrouver les paragraphes
def get_file_content(file):
    """
    Récupère le contenu textuel d'un fichier donné.
    Entrée : nom du fichier moins l'extension
    Sortie : nom du fichier + liste des paragraphes
    """
    fname = re.compile("(\\w+)\\.txt")
    m = re.search(fname, f)
    nom = m.group(1) #récupération du nom du fichier, sans son extension
    content = open(file, 'r').read().strip() #lecture et récupération du contenu textuel
    para = separateur.split(content) #découpage en paragraphes

    return nom, [e for e in para if e] #filtre pour enlever les éléments vides
```

```

def get_lexicon(f_path, filtre="", choix=""):
    """
    Récupère le contenu d'un lexique donné.
    Entrée :
    - chemin du lexique
    - éventuellement, liste des catégories qu'on veut prendre en compte
    - avec information de POS ('pos') ou non ('')
    Sortie :
    - dictionnaire {"forme 1" : "catégorie", "forme 2" : "catégorie"}
    - ET, éventuellement, dictionnaire {"forme 1": "POS"}
    """

    dico_terms_cat = dict()
    dico_terms_pos = dict()

    with open(f_path, 'r', encoding="utf-8") as in_stream : #ouverture du lexique
        in_stream.readline() #on évacue la lère ligne
        for line in in_stream : #pour chaque ligne du fichier
            elements = line.strip().split("\t") #on récupère le terme et la catégorie

            terme = elements[0]
            cat = elements[1]
            if choix == "pos":
                pos = elements[2]

            if filtre : #s'il faut filtrer les catégories
                if cat in filtre :
                    dico_terms_cat[terme] = cat #ajout du couple terme/catégorie
                    if choix == "pos" :
                        dico_terms_pos[terme] = pos

    if choix == "pos": #si on veut l'information de POS
        return dico_terms_cat, dico_terms_pos
    else :
        return dico_terms_cat

def get_type(emotion):
    """
    Pour une émotion spécifique donnée, renvoie son type (base, complexe ou inconnu).
    Entrée : "émotion"
    Sortie : "son type"
    """

    base = ["joie", "colere", "surprise", "degout", "tristesse", "peur"] #liste des émotions de base

    if emotion in base :
        return "Base"
    elif emotion == "comportement":
        return "Inconnu"
    else :
        return "Complexe"

#FICHIERS AC
def txt2ac(rep_ac, nom, paragraphes):
    """
    Génère un fichier .ac
    Entrée :
    - rep de sortie
    - nom du fichier txt (par rapport au dico_fic_para)
    - et liste des paragraphes qui le composent (par rapport au dico_fic_para)
    """

    try: #on essaye de créer le répertoire de sortie
        os.mkdir(rep_ac)
    except OSError: #si ce n'est pas possible, c'est qu'il existe déjà
        pass

    with open(os.path.join(rep_ac, nom+".ac"), 'w', encoding = "utf-8") as out_ac:
        out_ac.write("".join(paragraphes))

```

```

#FICHIERS AA
def build_corpusHashCode(nom,rep_ac):
    """
    Génère un identifiant pour lier un fichier .aa à un fichier .ac.
    Entrée :
    - le nom du fichier (même nom pour le fichier .aa et .ac)
    - le répertoire où se trouve le fichier .ac à associer au fichier .aa
    Sortie : l'identifiant qui permettra d'associer le fichier .aa au fichier .ac
    """

    path_f_ac = os.path.join(rep_ac,f+".ac") #reconstitution du fichier .ac à associer au fichier .aa
    f_ac = open(path_f_ac,'rb').read() #ouverture et lecture du fichier en flux d'octets
    length = len(f_ac) #taille du fichier

    hashcode = str(length)+"-" #première partie de l'identifiant

    code = 1 #va accueillir la deuxième partie de l'identifiant
    for i in range(0,length):
        n = int(f_ac[i]) #lecture octet par octet
        if n != 0 : #si l'octet ne vaut pas 0
            code *= n #on multiplie le code par la valeur de l'octet
            code = code%99999999 #et on fait le modulo

    hashcode += str(code) #ajout de la 2e partie de l'identifiant

    return hashcode

def get_index_para(paragraphes):
    """
    Trouve la position des paragraphes (indices de début et de fin dans le fichier .ac).
    Entrée : listes des paragraphes qui composent un fichier .txt
    Sortie : dico {"paragraphe" : (indice de début, indice de fin)}
    """

    #Pour conserver la séparation en paragraphes dans Glozz,
    # il faut encoder la position des paragraphes dans le fichier .aa
    dico_para_indices = dict()
    longueur_cumulative = 0 #somme itérative des longueurs des paragraphes, pour calculer les indices des caractères
    for para in paragraphes : #pour chaque paragraphe
        longueur_para = len(para) #on stocke sa longueur
        debut = longueur_cumulative
        fin = longueur_cumulative + longueur_para

        dico_para_indices[para] = (debut,fin) #on stocke les infos

        longueur_cumulative += longueur_para #incrémentement de la longueur cumulative

    return dico_para_indices

def time_ms():
    """
    Donne la date sous la forme du temps, en millisecondes (nb entier), écoulé depuis le 1/1/1970
    """

    # time.time() donne le temps écoulé depuis le 1/1/1970 en secondes
    # on multiplie par 1000 pour avoir cette durée en millisecondes
    # on transforme le résultat en entier
    return int(time.time() * 1000)

def get_times(paragraphes):
    """
    Donne un temps de création pour chaque paragraphe (nécessaire pour les unités du fichier .aa)
    Entrée : la liste des paragraphes d'un fichier
    Sortie : le dico {"paragraphe" : "temps de création en ms"}
    """

    dico_para_time = dict()
    temps = time_ms() #temps initial, celui pour le 1er paragraphe
    for para in paragraphes:
        dico_para_time[para] = temps
        temps += 1

    return dico_para_time

```

```

def build_unit_para(id,debut_fin,temps_creation):
    """
    Construit l'unité XML correspondant à un paragraphe pour le fichier .aa.
    Entrée :
    - l'identifiant d'annotation (ex. : astienne)
    - le tuple d'indices (debut de l'unité, fin de l'unité)
    - le temps de creation de l'unité
    """
    t = str(temps_creation)

    #Construction de l'unité
    unit = '<unit id="'+id+"_"+t+'">\n\t<metadata>\n\t\t<author>'+id+'</author>\n\t\t<creation-date>'+t+
    '</creation-date>\n\t\t<lastModifier>n/a</lastModifier>\n\t\t<lastModificationDate>0</lastModificationDate>'+
    '\n\t\t<metadata>\n\t\t<characterisation>\n\t\t\t<type>paragraph</type>\n\t\t\t<featureSet/>\n\t\t</characterisation>'+
    '\n\t\t<positioning>\n\t\t\t<start>\n\t\t\t\t<singlePosition index="'+
    str(debut_fin[0])+'"/>\n\t\t\t</start>\n\t\t\t<end>\n\t\t\t\t<singlePosition index="'+
    str(debut_fin[1])+'"/>\n\t\t\t</end>\n\t\t</positioning>\n\t\t</unit>'

    return unit

regEx_mot = re.compile(r"(\w+)") #pour trouver les mots
def find_terms(dico_para_indices, dico_terms_cat, dico_formes_pos=""):
    """
    Pour un paragraphe donné, trouve les termes qui appartiennent au lexique
    et stocke l'indice de début et de fin de ces termes.
    Ainsi que la POS, si on le souhaite
    Entrée :
    - dictionnaire {"paragraphe" : (indice de début, indice de fin)} (indices au sein du fichier .ac)
    - dictionnaire {"terme 1" : "catégorie", "terme 2" : "catégorie"}
    - éventuellement, dictionnaire {"forme" : "POS"}
    Sortie :
    - dictionnaire {(indice de début du terme, indice de fin du terme) : "categorie"}
    - ET, éventuellement, dictionnaire {(indice de début du terme, indice de fin du terme) : "POS"}
    """
    dico_indicesTermes_cat = dict()
    dico_indicesTermes_pos = dict()

    for paragraphe in dico_para_indices : #pour chaque paragraphe
        for m in re.finditer(regEx_mot,paragraphe) : #on cherche tout ce qui peut correspondre à un mot
            mot = m.group(1) #le mot
            debut_mot = m.start(1) #indice de début dans le paragraphe
            fin_mot = m.end(1) #indice de fin dans le paragraphe

            if mot in dico_terms_cat : #si le mot est dans le lexique
                #Pour avoir les indices par rapport au fichier .ac
                indice_debut = debut_mot + dico_para_indices[paragraphe][0] #ajout de l'indice de début du paragraphe
                indice_fin = fin_mot + dico_para_indices[paragraphe][0]
                #pour le couple d'indices, on stocke la catégorie
                dico_indicesTermes_cat[(indice_debut,indice_fin)] = dico_terms_cat[mot]

                if dico_formes_pos : #si on veut l'info de POS
                    dico_indicesTermes_pos[(indice_debut,indice_fin)] = dico_formes_pos[mot]

    if dico_formes_pos : #si veut l'info de POS
        return dico_indicesTermes_cat, dico_indicesTermes_pos
    else :
        return dico_indicesTermes_cat

def get_terms_time(dico_para_time,dico_indicesTermes_cat):
    """
    Donne un temps de création pour chaque unité de terme d'émotion (nécessaire pour le fichier .aa).
    Entrée :
    - dictionnaire {"paragraphe" : temps de création}
    - dictionnaire {(indice de début de terme, indice de fin de terme):"catégorie"}
    Sortie : dictionnaire {(indice de début de terme, indice de fin de terme): temps de création}
    """
    temps = max(dico_para_time.values()) #on part du temps du dernier paragraphe
    dico_indicesTermes_temps = dict()
    for couple in dico_indicesTermes_cat:
        temps += 1
        dico_indicesTermes_temps[couple] = temps

    return dico_indicesTermes_temps

```

```

def build_unit_emotion(id,debut_fin,temps_creation,cat,dico_cat_type,pos=""):
    """
    Construit l'unité de terme d'émotion pour le fichier .aa.
    Entrée :
    - l'identifiant d'annotation
    - le tuple d'indices (debut de l'unité, fin de l'unité)
    - le temps de creation de l'unité (en chaine de caractères)
    - la catégorie émotionnelle du terme
    - le dictionnaire de correspondance { "joie" : "base", "fierte" : "complexe"}
    - éventuellement, la POS du terme
    """
    if cat == "comportement":
        mode = "Comportementale" #mode d'expression
        nom = "inconnu"

    else :
        mode = "Designee" #mode d'expression
        nom = cat #nom de l'émotion dénotée

    feat_type = dico_cat_type[cat] #émotion de base, complexe ou inconnue ?
    t = str(temps_creation)

    feat_pos = ""
    if pos :
        feat_pos = '\t\t\t<feature name="nature">'+pos+'</feature>\n'

    #Construction de l'unité
    unit = '<unit id="'+id+"_"+t+">\n\t\t<metadata>\n\t\t<author>'+id+'</author>\n\t\t<creation-date>'+t+'</creation-date>\n\t\t<lastModifieur>n/a</lastModifieur>\n\t\t<lastModificationDate>0</lastModificationDate>'+
    '\n\t\t<metadata>\n\t\t<characterisation>\n\t\t<type>'+mode+'</type>\n\t\t<featureSet>\n\t\t\t<feature name="type">'+
    feat_type+'</feature>\n\t\t\t<feature name="nom">'+cat+'</feature>\n\t\t'+feat_pos+'\t\t</featureSet>\n\t\t'+
    '</characterisation>\n\t\t<positioning>\n\t\t\t<start>\n\t\t\t\t<singlePosition index="'+str(debut_fin[0])+
    '">\n\t\t\t\t<start>\n\t\t\t\t<end>\n\t\t\t\t\t<singlePosition index="'+str(debut_fin[1])+
    '">\n\t\t\t\t</end>\n\t\t\t</positioning>\n\t\t</unit>'

    return unit

def txt2aa(rep_aa,rep_ac,nom,id,dico_para_indices,dico_para_time,dico_indicesTermes_cat,
dico_indicesTermes_temps,dico_cat_type,dico_indicesTermes_pos=""):
    """
    Génère le fichier .aa.
    Entrée :
    - répertoire où créer les fichiers aa
    - répertoire des fichiers .ac correspondants
    - le nom du fichier
    - identifiant d'annotation (ex. : "aetienne")
    - dico {"paragraphe" : (indice début, indice fin)}
    - dico {"paragraphe" : 'temps'}
    - dictionnaire {(indice début, indice fin) : "nom de l'émotion"}
    - dictionnaire {(indice début, indice fin) : temps} pour les termes d'émotion repérés
    - dictionnaire de correspondance {"nom d'émotion": "base, complexe ou inconnu"}
    - éventuellement, dictionnaire {(indice de début, indice de fin): "POS"}
    """
    debut_aa = '<?xml version="1.0" encoding="UTF-8" standalone="no"?>\n<annotations>\n<metadata corpusHashCode="'
    corpusHashCode=build_corpusHashCode(nom,rep_ac)
    debut_aa_suite = ''>\n'
    fin_aa = '\n</annotations>'

    units_para = [
        build_unit_para(id,dico_para_indices[para],dico_para_time[para])
        for para in dico_para_indices
    ]

    units_emotion = [
        build_unit_emotion(id,couple,dico_indicesTermes_temps[couple],
            dico_indicesTermes_cat[couple],dico_cat_type,dico_indicesTermes_pos[couple])
        if dico_indicesTermes_pos
        else build_unit_emotion(id,couple,dico_indicesTermes_temps[couple],dico_indicesTermes_cat[couple],dico_cat_type)
        for couple in dico_indicesTermes_cat
    ]

    try: #on essaye de créer le répertoire de sortie
        os.mkdir(rep_aa)
    except OSError: #si ce n'est pas possible, c'est qu'il existe déjà
        pass

    with open(os.path.join(rep_aa,nom+".aa"),'w',encoding = "utf-8") as out_aa:
        out_aa.write(debut_aa+corpusHashCode+debut_aa_suite)
        out_aa.write("\n".join(units_para)+"\n")
        out_aa.write("\n".join(units_emotion))
        out_aa.write(fin_aa)

```

```

###-----EXECUTION-----###
if __name__ == '__main__':

    #####Données de départ#####
    rep_txt = sys.argv[1]

    #Répertoires de sortie
    rep_ac = sys.argv[2]
    rep_aa = sys.argv[3]

    #Identifiant pour les fichiers aa
    # id = "aetienne"
    id = sys.argv[4]

    #Chemin du lexique
    avec_POS = sys.argv[5]
    f_path = sys.argv[6]

    #Liste des émotions du lexique qu'on veut annoter
    filtre = [
        'joie','surprise','peur','tristesse','colere','degout',
        'fierte','culpabilite','embarras','jalousie',
        'comportement',
    ]
    #####

    #Récupération des fichiers .txt
    fichiers = get_files(rep_txt)

    #Récupération du contenu des fichiers
    dico_fic_para = dict()
    for f in fichiers :
        nom, para = get_file_content(f)
        dico_fic_para[nom] = para
    #Ecriture des fichiers ac
    for f in dico_fic_para :
        txt2ac(rep_ac,f, dico_fic_para[f])

    #Récupération du lexique émotionnel
    if avec_POS == "oui" :
        dico_formes_cat, dico_formes_pos = get_lexicon(f_path,filtrer,"pos")
    else :
        dico_formes_cat = get_lexicon(f_path,filtrer)

    #Correspondance entre nom d'émotion et type (base ou complexe)
    dico_cat_type = {
        emotion:get_type(emotion)
        for emotion in filtrer
    }

    #Ecriture des fichiers .aa
    for f in dico_fic_para : #pour chaque fichier
        #Pour faciliter la lecture dans Glozz, on conserve la structure en paragraphes
        dico_para_indices = get_index_para(dico_fic_para[f]) #emplacement des paragraphes
        dico_para_time = get_times(dico_fic_para[f]) #temps de création des unités de paragraphe

        #Pour faciliter l'annotation, on pré-annote les termes d'émotions
        if avec_POS == "oui" :
            dico_indicesTermes_cat, dico_indicesTermes_pos = find_terms(dico_para_indices, dico_formes_cat, dico_formes_pos)
            dico_indicesTermes_temps = get_terms_time(dico_para_time,dico_indicesTermes_cat)
            txt2aa(rep_aa,rep_ac,f,id,dico_para_indices,dico_para_time,dico_indicesTermes_cat,
                dico_indicesTermes_temps,dico_cat_type,dico_indicesTermes_pos)
        else :
            dico_indicesTermes_cat = find_terms(dico_para_indices, dico_formes_cat) #emplacement des termes
            dico_indicesTermes_temps = get_terms_time(dico_para_time,dico_indicesTermes_cat)
            txt2aa(rep_aa,rep_ac,f,id,dico_para_indices,dico_para_time,dico_indicesTermes_cat,
                dico_indicesTermes_temps,dico_cat_type)

```

## A.11. Script fouilleGlozz.py

```
#!/usr/bin/python3
# -*- coding: utf-8 -*-
import os
import sys
from bs4 import BeautifulSoup as BS #pour parser le XML
from collections import Counter
from collections import defaultdict

###-----INFOS-----###
# Ce script permet de fouiller en même temps plusieurs fichiers d'annotation Glozz (.aa).
# Pour un répertoire de fichier .aa et un répertoire de fichiers .ac correspondant
# le script génère des tableaux (fichiers .tsv) avec les résultats quantitatifs de l'annotation

# Lancement : python3 fouilleGlozz.py répertoire_aa répertoire_ac répertoire_sortie
# ex. : python3 fouilleGlozz.py ../GlozzFiles/aa ../GlozzFiles/ac ../GlozzFiles/sorties

###-----OBJETS-----###
# Des structures spéciales (objets) ont été créées pour stocker les informations d'annotation.
class Emotion:
    """
    Une unité annotée comme émotion dans Glozz.
    Attributs :
    - unit_id = identifiant de l'unité
    - author = annotateur
    - creation_date = temps de création de l'unité
    - unit_type = type de l'unité (Designee, Montree, Comportementale ou Etayee)
    - unit_nature = nature de l'unité (Nom, Adj etc.)
    - start = indice, dans le fichier .ac, du début de l'unité
    - end = indice, dans le fichier .ac, de la fin de l'unité
    - unit_text = texte correspondant à l'unité
    - emotion_type = type d'émotion (Base ou Complexe)
    - emotion_name = nom de l'émotion (colère, joie etc.)
    """
    def __init__(self, u, ac, id_aa):
        self.unit_id = u["id"]+id_aa
        self.author = u.find('author').text
        self.creation_date = u.find('creation-date').text
        self.unit_type = u.find('type').text
        self.unit_nature = u.find('feature', {"name":"nature"}).text
        self.start = int(u.find('start').find("singleposition")["index"])
        self.end = int(u.find('end').find("singleposition")["index"])
        self.unit_text = ac[self.start:self.end]
        self.emotion_type = u.find('feature', {"name":"type"}).text
        self.emotion_name = u.find('feature', {"name":"nom"}).text
```

```

class Entite :
    """
    Une unité annotée comme entité ressentant l'émotion dans Glozz.
    Attributs :
    - unit_id = identifiant de l'unité
    - author = annotateur
    - creation_date = temps de création de l'unité
    - unit_type = type de l'unité (toujours 'Entite')
    - unit_nature = nature de l'unité (Nom, Adj etc.)
    - start = indice, dans le fichier .ac, du début de l'unité
    - end = indice, dans le fichier .ac, de la fin de l'unité
    - unit_text = texte correspondant à l'unité
    - entity_carac1 = caractère [+/- animé] [+/-humain] de l'entité
    - entity_carac2 = caractère [+/- collectif] de l'entité
    """
    def __init__(self, u, ac, id_aa):
        self.unit_id = u["id"]+id_aa
        self.author = u.find('author').text
        self.creation_date = u.find('creation-date').text
        self.unit_type = u.find('type').text
        self.unit_nature = u.find('feature', {"name":"nature"}).text
        self.start = int(u.find('start').find("singleposition")["index"])
        self.end = int(u.find('end').find("singleposition")["index"])
        self.unit_text = ac[self.start:self.end]
        self.entity_carac1 = u.find('feature', {"name":"caracteristique_1"}).text
        self.entity_carac2 = u.find('feature', {"name":"caracteristique_2"}).text

class Relation:
    """
    Une relation entre une émotion et l'entité qui la ressent, dans Glozz.
    Attributs :
    - relation_id = l'identifiant de la relation
    - author = l'annotateur
    - creation_date = la date de création de la relation
    - relation_type = le type de relation (toujours 'Affecte')
    - relation_nature = la nature de la relation syntaxique entre l'émotion et l'entité
    - term1_id = identifiant du 1er terme
    - term2_id = identifiant du 2e terme
    """
    def __init__(self, r, ac, id_aa):
        self.relation_id = r["id"]+id_aa
        self.author = r.find('author').text
        self.creation_date = r.find('creation-date').text
        self.relation_type = r.find('type').text
        self.relation_nature = r.find('feature', {"name":"syntaxe"}).text
        terms = r.find_all('term')
        self.term1_id = terms[0]["id"]
        self.term2_id = terms[1]["id"]

```



```

###-----FONCTIONS-----###
def get_annotatons(aa,ac) : #pour récupérer les informations d'annotation
"""
Pour un fichier .aa et un fichier .ac donnés,
renvoie :
- la liste des unités d'émotions annotées (objets Emotion())
- la liste des unités d'entités annotées (objets Entite())
- la liste des relations annotées (objets Relations())
"""
emotions, entites, relations = list(), list(), list()
soup = BS(aa,'lxml') #on parse le fichier xml (fichier aa)

# dans des fichiers différents, des unités peuvent avoir le même identifiant
# pour éviter les confusions, on va utiliser l'identifiant du fichier .aa
id_aa = soup.find('metadata')['corpushashcode'] #identifiant du fichier .aa
for u in soup.find_all("unit") : #pour chaque unité d'annotation
    typ = u.find('type').text #on filtrera sur le type d'unité
    if typ == "paragraph":
        next
    elif typ == "Entite" : #les entités qui ressentent les émotions
        #un objet Entite(), pour conserver toutes les infos
        unite_entite = Entite(u,ac, id_aa)
        entites.append(unite_entite)
    else : #les émotions
        #un objet Emotion(), pour conserver toutes les infos
        unite_emotion = Emotion(u,ac,id_aa)
        emotions.append(unite_emotion)
for r in soup.find_all("relation") : #pour chaque relation
    #un objet Relation(), pour conserver toutes les infos
    relation = Relation(r, ac, id_aa)
    relations.append(relation)
return emotions, entites, relations
def count_intersection(listel, liste2) : #pour écrire les résultats
"""
Pour deux listes données,
renvoie le nombre (string) d'éléments en commun
"""
return str(len(set(listel)&set(liste2)))
###-----EXECUTION-----###
if __name__ == '__main__' :

    # Répertoires
    rep_aa = sys.argv[1]
    rep_ac = sys.argv[2]
    rep_out = sys.argv[3]

    # Fichiers à traiter
    #filtre des fichiers .autosav
    aa_files = [e for e in os.listdir(rep_aa) if e.endswith(".aa")]
    ac_files = os.listdir(rep_ac)

```

```

### Rappel sur la structure de nos données : ###
# fichier .aa = fichier XML avec les annotations (balises)
# fichier .ac = contenu textuel, sur une seule ligne
# ces fichiers sont appariés et ont le même nom
# (ex. 'De_l_animal_a_la_v viande.aa' et 'De_l_animal_a_la_v viande.ac')
# il est donc important de faire quelques vérifications
try :
    assert len(aa_files) == len(ac_files)
except AssertionError :
    print("Problème : pas le même nombre de fichiers .ac et de fichiers .aa ?")

for aa, ac in zip(aa_files, ac_files) :
    try :
        assert aa[:-3] == ac[:-3]
    except AssertionError :
        print("Attention ! Les fichiers aa et ac ne sont pas dans le même ordre !")
### Récupération des informations d'annotation ###
dico_aa_emotions, dico_aa_entites, dico_aa_relations = dict(), dict(), dict()

for aa_file, ac_file in zip(aa_files, ac_files) : #pour chaque couple de fichiers
    #lecture et stockage du fichier d'annotation
    aa = open(os.path.join(rep_aa, aa_file), 'r').read()
    #lecture et stockage du fichier avec le texte
    ac = open(os.path.join(rep_ac, ac_file), 'r').read()

    #récupération des annotations
    emotions, entites, relations = get_annotatons(aa, ac)

    #stockage dans les dictionnaires appropriés
    dico_aa_emotions[aa_file] = emotions
    dico_aa_entites[aa_file] = entites
    dico_aa_relations[aa_file] = relations

### Structures pour afficher les résultats ###
#listes de tous les objets Emotion()
emotions_tout = [em for file in dico_aa_emotions for em in dico_aa_emotions[file]]

# Dictionnaire modes
dico_modes = defaultdict(lambda :list())
for emotion in emotions_tout :
    dico_modes[emotion.unit_type].append(emotion.unit_id)

# Dictionnaire émotions de base
dico_bases = defaultdict(lambda :list())
for emotion in emotions_tout :
    if emotion.emotion_type == "Base" :
        dico_bases[emotion.emotion_name].append(emotion.unit_id)
print(dico_bases.keys())

# Dictionnaire émotions complexes
dico_complexes = defaultdict(lambda :list())
for emotion in emotions_tout :
    if emotion.emotion_type == "Complexe" :
        dico_complexes[emotion.emotion_name].append(emotion.unit_id)
print(dico_complexes.keys())

```

```

### Ecriture des résultats ###
try: #on essaie de créer le répertoire de sortie
    os.mkdir(rep_out)
except OSError: #si ce n'est pas possible, c'est qu'il existe déjà
    pass

## Résultats par fichier ##
# Modes d'expressions
with open(os.path.join(rep_out,"resultats_modes.tsv"),"w",encoding="utf-8") as out :
    out.write("Fichier\tComportementale\tDesignee\tEtayee\tMontree\tTotal\n")
    for file in dico_aa_emotions :
        out.write(file+"\t")
        cnt_modes = Counter([emotion.unit_type for emotion in dico_aa_emotions[file]])
        out.write(
            str(cnt_modes.get("Comportementale","0"))+"\t"+str(cnt_modes.get("Designee","0"))+
            "\t"+str(cnt_modes.get("Etayee","0"))+"\t"+str(cnt_modes.get("Montree","0"))+
            "\t"+str(sum([cnt_modes[key] for key in cnt_modes]))+"\n"
        )
    out.write(
        "Total\t"+
        "\t".join([str(len(dico_modes[mode])) for mode in sorted(dico_modes)])+"\t"+
        str(len(emotions_tout))
    )

# Types et noms d'émotion
with open(os.path.join(rep_out,"resultats_emotions.tsv"),"w",encoding="utf-8") as out :
    out.write(
        "Fichier\tBase\tColere\tDegout\tJoie\tPeur\tSurprise\tTristesse\t"+
        "Complexe\tCulpabilite\tEmbarras\tFierite\tJalousie\tTotal\n"
    )
    for file in dico_aa_emotions :
        out.write(file+"\t")
        cnt_types = Counter([emotion.emotion_type for emotion in dico_aa_emotions[file]])
        cnt_noms = Counter([emotion.emotion_name for emotion in dico_aa_emotions[file]])
        out.write(
            str(cnt_types.get("Base","0"))+"\t"+
            str(cnt_noms.get("colere","0"))+"\t"+str(cnt_noms.get("degout","0"))+"\t"+
            str(cnt_noms.get("joie","0"))+"\t"+str(cnt_noms.get("peur","0"))+"\t"+
            str(cnt_noms.get("surprise","0"))+"\t"+str(cnt_noms.get("tristesse","0"))+"\t"+
            str(cnt_types.get("Complexe","0"))+"\t"+
            str(cnt_noms.get("culpabilite","0"))+"\t"+str(cnt_noms.get("embarras","0"))+"\t"+
            str(cnt_noms.get("fierite","0"))+"\t"+str(cnt_noms.get("jalousie","0"))+"\t"+
            str(sum([cnt_types[key] for key in cnt_types]))+"\n"
        )
    out.write(
        "Total\t"+
        str(sum([ len(dico_bases[base]) for base in dico_bases]))+"\t"+
        "\t".join([str(len(dico_bases[base])) for base in dico_bases])+"\t"+
        str(sum([len(dico_complexes[comp]) for comp in dico_complexes]))+"\t"+
        "\t".join([str(len(dico_complexes[comp])) for comp in dico_complexes])+"\t"+
        str(len(emotions_tout))
    )

```

```

## Résultats globaux ##
bases = [e for emotion in dico_bases for e in dico_bases[emotion]]
complexes = [e for emotion in dico_complexes for e in dico_complexes[emotion]]

with open(os.path.join(rep_out,"resultats_modes_emotions.tsv"),"w",encoding="utf-8") as out :
    out.write("\tComportementale\tDesignee\tEtayee\tMontree\tTotal\n")

    out.write(
        "Base\t"+
        "\t".join([count_intersection(bases,dico_modes[mode]) for mode in sorted(dico_modes)])+"\t"+
        str(len(bases))+"\n"
    )
    for base in dico_bases :
        out.write(base[0].upper()+base[1:]+"\t")
        for mode in sorted(dico_modes) :
            out.write(count_intersection(dico_bases[base],dico_modes[mode])+"\t")
        out.write(str(len(dico_bases[base]))+"\n")

    out.write(
        "Complexe\t"+
        "\t".join([count_intersection(complexes,dico_modes[mode]) for mode in sorted(dico_modes)])+"\t"+
        str(len(complexes))+"\n"
    )
    for comp in dico_complexes :
        out.write(comp[0].upper()+comp[1:]+"\t")
        for mode in sorted(dico_modes) :
            out.write(count_intersection(dico_complexes[comp],dico_modes[mode])+"\t")
        out.write(str(len(dico_complexes[comp]))+"\n")

## Résultats pour montrées et étayées ##
montrees_nature = [em.unit_nature for em in emotions_tout if em.unit_type == "Montree"]
cnt_montrees_nature = Counter(montrees_nature)

etayees_nature = [em.unit_nature for em in emotions_tout if em.unit_type == "Etayee"]
cnt_etayees_nature = Counter(etayees_nature)

with open(os.path.join(rep_out,"montrees_etayees_natures.tsv"),"w",encoding = "utf-8") as out :
    out.write("Montrees\n")
    for n in cnt_montrees_nature :
        out.write(n+"\t"+str(cnt_montrees_nature[n])+"\n")
    out.write("\n\nEtayees\n")
    for n in cnt_etayees_nature :
        out.write(n+"\t"+str(cnt_etayees_nature[n])+"\n")

```

## A.12. Catégories émotionnelles selon les numéros du *P'tit Libé*

Fichier	Colere	Degout	Joie	Peur	Surp.	Trist.	Culp.	Emb.	Fierte	Jal.	Total
De_l_animal_a_la_viande.aa	0	4	3	0	9	0	0	0	0	0	16
Donald_Trump.aa	2	0	4	3	2	0	0	1	0	0	12
Garros.aa	1	0	5	4	3	0	0	0	2	0	15
la_BD_avec_Mortelle_Adele.aa	0	0	5	1	3	0	0	0	0	0	9
la_colere_des_gilets_jaunes.aa	22	0	6	13	0	0	0	0	0	0	41
La_COP21.aa	0	0	1	13	4	0	0	0	0	0	18
la_Coree_du_Nord.aa	7	0	0	26	1	2	0	1	2	0	39
la_Coupe_du_monde_de_foot.aa	0	0	6	0	7	8	0	0	9	0	30
la_disparition_des_abeilles.aa	2	0	1	5	0	0	1	0	0	0	9
la_disparition_des_animaux_sauvages.aa	10	0	4	17	5	2	0	0	0	0	38
la_Fashion_Week.aa	0	0	2	3	2	2	0	0	0	0	9
La_gauche_et_la_droite_en_politique.aa	0	0	0	0	2	0	0	0	0	0	2
La_guerre_en_Syrie.aa	0	0	0	19	0	0	0	0	0	0	19
La_loi_travail.aa	21	0	0	9	0	0	0	0	0	0	30
la_mefiance_envers_les_journalistes.aa	6	0	0	17	2	0	4	2	0	0	31
la_pedophilie.aa	1	0	4	18	1	9	3	7	0	0	43
la_protection_du_patrimoine.aa	2	0	5	2	7	1	0	0	1	0	18
la_rentree_tele.aa	0	0	3	3	0	0	0	0	0	0	6
la_securete_routiere.aa	2	0	7	0	0	2	0	0	0	0	11
la_SNCF.aa	3	0	0	1	2	0	0	0	0	0	6
Les_attentats_a_Paris.aa	0	0	0	12	3	11	0	0	0	0	26
les_cesars_et_les_oscars.aa	2	0	10	1	5	0	0	3	13	0	34
les_coulistes_des_dessins_animes.aa	0	0	4	0	5	3	1	0	0	0	13
Les_coulistes_de_l_ecole.aa	0	0	1	0	0	0	0	0	0	0	1
les_dangers_du_glyphosate.aa	4	0	0	12	0	0	2	0	0	0	18
les_dessous_du_cinema.aa	0	0	1	4	7	1	0	0	0	0	13
les_dinosaures.aa	0	0	3	7	6	0	0	0	0	0	16
les_donnees_personnelles.aa	7	0	1	2	2	0	1	8	0	0	21
les_extraterrestres.aa	2	0	7	7	7	0	0	0	0	0	23
les_fake_news.aa	5	0	0	0	0	0	0	1	1	0	7
les_grands_prix_litteraires.aa	0	0	6	0	1	4	0	0	0	0	11
les_greves.aa	27	1	0	2	1	0	0	0	0	0	31
les_impots.aa	5	0	0	0	0	0	0	0	0	0	5
les_jeux_de_societe.aa	0	0	6	1	2	2	0	2	0	0	13
Les_Jeux_olympiques.aa	0	0	2	0	4	0	0	0	10	0	16
les_Jeux_olympiques_d_hiver.aa	0	0	7	4	2	2	0	0	5	0	20
les_jeux_video.aa	0	0	3	0	0	0	0	0	0	0	3
Les_JO_a_Paris.aa	2	0	0	1	1	0	0	0	0	0	4
les_manifestations_en_Iran.aa	23	0	1	4	0	0	1	0	0	0	29
les_metiers_de_Noel.aa	1	0	21	2	5	5	0	2	4	0	40
Les_migrants.aa	1	0	1	47	0	3	0	0	0	0	52
les_referendums.aa	13	0	1	3	0	0	1	0	0	0	18
Les_regionales.aa	1	0	1	3	0	0	0	0	1	0	6
les_repas_de_fetes.aa	1	0	11	0	1	0	0	0	0	0	13
les_restaurants_etoiles.aa	0	0	23	2	2	2	0	0	4	0	33
les_Rohingyas_peuple_chasse_de_Birmanie.aa	9	0	2	18	0	9	0	0	0	0	38
les_rythmes_scolaires.aa	0	0	1	0	0	1	0	0	0	0	2
Les_secrets_du_29_fevrier.aa	1	0	5	0	3	1	0	0	1	0	11
les_theories_du_complot.aa	11	0	0	2	1	0	0	0	0	1	15
Les_tubes_de_l_ete.aa	0	0	18	0	7	2	0	1	2	0	30
les_vaccins.aa	2	0	1	6	0	0	0	0	0	0	9
les_vegetariens.aa	0	3	1	6	2	5	1	0	1	0	19

les_youtubeurs.aa	0	0	9	0	2	0	0	0	0	0	11
Le_Brexit.aa	3	0	0	0	0	0	0	0	0	0	3
le_changement_d_heure.aa	0	0	0	1	6	1	0	0	0	0	8
le_depart_du_ministre_de_l_ecologie.aa	9	0	0	2	4	5	0	0	0	0	20
le_feminisme.aa	19	0	0	4	4	2	0	2	0	0	31
le_festival_d_Avignon.aa	3	0	4	4	5	0	0	0	0	0	16
le_Front_national.aa	1	0	0	13	8	0	0	0	0	0	22
le_harcelement_a_l_ecole.aa	3	0	1	43	1	22	1	18	0	1	90
le_mariage_royal.aa	4	0	7	0	0	3	0	0	1	0	15
le_petrole_utile_mais_polluant.aa	11	0	0	7	4	3	0	0	0	0	25
le_plan_loup.aa	10	0	0	19	1	0	0	0	0	0	30
le_premier_anniversaire_du_president.aa	9	0	4	1	1	0	0	0	3	0	18
le_prix_Nobel_de_la_paix.aa	5	0	0	5	1	0	0	0	0	0	11
le_P_tit_Libe_des_philosophes.aa	23	0	15	4	1	23	1	1	0	0	68
le_P_tit_Libe_des_philosophes1.aa	0	0	9	11	6	3	0	0	0	0	29
le_Salon_de_l_agriculture_et_les_eleveurs.aa	4	0	0	0	0	0	0	0	0	0	4
le_succes_du_rap.aa	7	1	11	4	0	3	0	0	3	0	29
le_Tour_de_France.aa	3	0	1	0	8	1	0	0	7	0	20
logement.aa	8	0	2	22	1	4	0	4	0	0	41
Luc_Melenchon.aa	11	0	0	0	1	1	1	0	0	0	14
L_accueil_des_migrants_en_France.aa	0	0	0	8	0	2	0	0	0	0	10
L_affaire_Benalla.aa	1	0	0	4	2	0	3	1	0	0	11
L_affaire_Weinstein_et_le_harcelement_sexuel.aa	15	5	0	14	0	1	3	3	0	0	41
L_antisemitisme.aa	12	0	0	11	0	3	0	0	0	0	26
L_art_contemporain.aa	0	0	0	0	3	0	0	0	3	0	6
L_Assemblee_nationale.aa	0	0	8	0	0	0	0	0	0	1	9
L_attaque_de_Nice.aa	1	0	0	16	0	8	0	0	0	0	25
L_attentat_de_Strasbourg.aa	0	0	0	11	1	5	0	0	0	0	17
L_election_americaine.aa	3	0	0	0	8	0	0	0	0	0	11
L_election_du_president_au_Bresil.aa	13	0	7	19	3	0	0	0	0	0	42
L_election_presidentielle.aa	0	0	0	0	5	0	0	0	0	0	5
L_empire_Disney.aa	0	0	1	1	7	0	0	0	1	0	10
L_Euro_de_foot.aa	2	0	13	1	3	0	0	0	11	0	30
L_homophobie.aa	3	2	1	26	0	18	3	29	0	0	82
L_independance_de_la_Catalogne.aa	1	0	0	0	1	4	0	1	1	0	8
Mai_68.aa	19	0	7	11	0	0	0	0	0	0	37
mandat.aa	3	0	1	12	2	0	1	2	0	0	21
meteo_un_hiver_perturbe.aa	1	0	10	0	2	0	0	0	0	0	13
Ouragans_une_annee_record.aa	0	0	6	12	0	2	0	0	0	0	20
Palestine.aa	14	0	0	6	1	1	0	0	0	0	22
Parcoursup_et_l_avenir_des_jeunes.aa	6	0	5	4	0	9	0	0	0	0	24
Securite_la_vie_apres_les_attentats.aa	8	0	1	17	1	0	0	0	0	0	27
Unis.aa	3	0	2	12	2	3	0	0	0	0	22
Vladimir_Poutine.aa	4	0	2	4	0	4	0	0	0	0	14
Zika_et_les_epidemies.aa	0	0	1	12	0	0	0	0	0	0	13
Total	16	212	322	641	438	208	89	86	28	3	2043

### A.13. Modes d'expression des émotions selon les numéros du *P'tit Libé*

Fichier	Comportementale	Designee	Etayee	Montree	Total
De l animal a la viande.aa	0	4	7	5	16
Donald Trump.aa	0	11	1	0	12
Garros.aa	0	5	7	3	15
la BD avec Mortelle Adele.aa	0	4	0	5	9
la colere des gilets jaunes.aa	5	21	14	1	41
La COP21.aa	1	2	11	4	18
la Coree du Nord.aa	6	16	17	0	39
la Coupe du monde de foot .aa	6	6	13	5	30
la disparition des abeilles.aa	0	6	3	0	9
la disparition des animaux sauvages.aa	2	15	14	7	38
la Fashion Week.aa	2	4	1	2	9
La gauche et la droite en politique.aa	0	1	1	0	2
La guerre en Syrie.aa	4	1	14	0	19
La loi travail .aa	8	14	8	0	30
la mefiance envers les journalistes.aa	3	19	8	1	31
la pedophilie.aa	1	28	14	0	43
la protection du patrimoine.aa	2	8	4	4	18
la rentree tele.aa	1	4	0	1	6
la securite routiere.aa	1	4	6	0	11
la SNCF.aa	1	4	0	1	6
Les attentats a Paris .aa	6	10	10	0	26
les cesars et les oscars.aa	2	12	19	1	34
les coulisses des dessins animes.aa	0	4	4	5	13
Les coulisses de l ecole.aa	0	1	0	0	1
les dangers du glyphosate.aa	1	7	10	0	18
les dessous du cinema.aa	2	3	3	5	13
les dinosaures.aa	2	4	8	2	16
les donnees personnelles.aa	0	11	6	4	21
les extraterrestres.aa	0	15	4	4	23
les fake news.aa	0	4	3	0	7
les grands prix litteraires.aa	0	6	4	1	11
les greves.aa	13	13	4	1	31
les impots.aa	2	2	1	0	5
les jeux de societe.aa	0	7	4	2	13
Les Jeux olympiques .aa	0	1	13	2	16
les Jeux olympiques d hiver.aa	1	8	11	0	20
les jeux video.aa	0	3	0	0	3
Les JO a Paris.aa	0	1	2	1	4
les manifestations en Iran.aa	7	12	10	0	29
les metiers de Noel.aa	2	23	11	4	40
Les migrants.aa	13	15	24	0	52
les referendums.aa	4	7	6	1	18
Les regionales .aa	0	4	2	0	6
les repas de fetes.aa	1	6	5	1	13
les restaurants etoiles.aa	0	6	20	7	33
les Rohingyas peuple chasse de Birmanie.aa	11	8	19	0	38

les rythmes scolaires.aa	0	2	0	0	2
Les secrets du 29 fevrier .aa	0	9	0	2	11
les theories du complot.aa	3	5	7	0	15
Les tubes de l ete.aa	2	8	15	5	30
les vaccins.aa	0	8	1	0	9
les vegetariens.aa	0	11	8	0	19
les youtubeurs.aa	0	6	2	3	11
Le Brexit.aa	2	0	1	0	3
le changement d heure.aa	0	4	0	4	8
le depart du ministre de l ecologie.aa	0	8	10	2	20
le feminisme.aa	4	15	12	0	31
le festival d Avignon.aa	2	5	6	3	16
le Front national.aa	1	16	5	0	22
le harcelement a l ecole.aa	9	27	53	1	90
le mariage royal.aa	1	6	8	0	15
le petrole utile mais polluant.aa	3	13	6	3	25
le plan loup.aa	3	12	14	1	30
le premier anniversaire du president a l Elysee.aa	2	8	8	0	18
le prix Nobel de la paix.aa	1	6	4	0	11
le P tit Libe des philosophes.aa	16	29	20	3	68
le P tit Libe des philosophes1.aa	0	13	14	2	29
le Salon de l agriculture et les eleveurs.aa	1	0	3	0	4
le succes du rap.aa	4	11	14	0	29
le Tour de France.aa	0	5	12	3	20
logement.aa	3	12	22	4	41
Luc Melenchon.aa	0	7	7	0	14
l accueil des migrants en France.aa	2	5	3	0	10
l affaire Benalla.aa	0	7	4	0	11
l affaire Weinstein et le harcelement sexuel.aa	4	12	17	8	41
l antisemitisme.aa	1	12	12	1	26
l art contemporain.aa	0	1	4	1	6
L Assemblee nationale.aa	0	0	3	6	9
L attaque de Nice .aa	5	10	10	0	25
l attentat de Strasbourg.aa	3	7	7	0	17
L election americaine.aa	0	1	4	6	11
l election du president au Bresil.aa	7	13	21	1	42
L election presidentielle .aa	0	1	0	4	5
L empire Disney.aa	0	3	0	7	10
L Euro de foot .aa	4	8	16	2	30
l homophobie.aa	14	41	27	0	82
l independance de la Catalogne.aa	1	4	2	1	8
Mai 68.aa	14	12	11	0	37
mandat.aa	2	12	6	1	21
meteo un hiver perturbe.aa	1	2	7	3	13
Ouragans une annee record.aa	2	9	9	0	20
Palestine.aa	2	6	13	1	22
Parcoursup et l avenir des jeunes.aa	3	8	12	1	24
Securite la vie apres les attentats.aa	2	13	12	0	27
Unis.aa	3	9	8	2	22
Vladimir Poutine.aa	1	5	8	0	14
Zika et les epidemies.aa	0	7	6	0	13
<b>Total</b>	<b>238</b>	<b>824</b>	<b>820</b>	<b>161</b>	<b>2043</b>