



HAL
open science

Apports de la linguistique et du TAL à l'analyse des émotions dans les textes pour enfants

Aline Étienne, Delphine Battistelli, Gwéno   Lecorv  

► **To cite this version:**

Aline   tienne, Delphine Battistelli, Gw  no   Lecorv  . Apports de la linguistique et du TAL    l'analyse des   motions dans les textes pour enfants. Colloque Langage et   MOTions 3  me Edition, Nov 2020, Bordeaux, France. hal-03028735

HAL Id: hal-03028735

<https://hal.science/hal-03028735v1>

Submitted on 27 Nov 2020

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 destin  e au d  p  t et    la diffusion de documents scientifiques de niveau recherche, publi  s ou non,   manant des   tablissements d'enseignement et de recherche fran  ais ou   trangers, des laboratoires publics ou priv  s.

Introduction

Les émotions des personnages jouent un rôle central dans la compréhension de textes par les enfants [2]. Nous présentons ici le corpus annoté que nous avons constitué en vue de caractériser l'expression des émotions dans les textes à tous les niveaux de la langue : lexical, syntaxique, discursif. Cette étude s'inscrit dans le cadre de travaux ayant pour objectif le développement d'outils à même de repérer automatiquement des portions de texte difficiles à comprendre pour les enfants jeunes lecteurs (TextToKids ANR AAPG 2019).

Notre étude

Corpus

97 numéros (octobre 2015 à mars 2019) du journal d'actualité en ligne Le P'tit Libé, adressé aux 7-12 ans, et qui vise à expliquer « l'actu des grands » aux enfants.

Méthodologie

Trois grandes étapes :

- (A) **Synthèse des pistes et résultats** dans la littérature psycholinguistique et linguistique pour l'étude des émotions dans les textes pour enfants.
- (B) **Analyse linguistique fine** d'extraits de notre corpus, afin de valider (A).
- (C) **Proposition d'un schéma d'annotation** des émotions. Objectif : confirmer, à l'échelle d'un corpus entier, la pertinence des critères et catégories, issus de (A) et validés en (B).

Schéma d'annotation

Dans notre schéma d'annotation, nous intégrons :

- **le type d'émotion** (« de base » vs. « complexe ») [2] [4] ;
- **10 catégories émotionnelles**. 6 émotions de base (COLÈRE, DÉGOÛT, JOIE, PEUR, SURPRISE et TRISTESSE) et 4 émotions complexes (CULPABILITÉ, EMBARRAS, FIERTÉ, JALOUSIE) [2] [4] ;
- **l'entité** qui ressent l'émotion [5], pouvant être [\pm humain] et [\pm collectif] (ex. « je » vs. « on ») ;
- **la relation** (syntaxique) entre l'unité émotionnelle et le segment dénotant l'entité ressentant l'émotion (ex. dans « Il effraye Jo », « Jo » est l'objet d'« effraye »). [5]
- **la nature du segment textuel** exprimant l'émotion et de celui exprimant l'entité qui ressent l'émotion (verbe, nom, adjectif, proposition etc.) [5].

4 modes d'expression des émotions (selon [1] [3] [5]) :

- ▶ « expression **comportementale** de l'émotion » : émotion exprimée par la description d'une manifestation physiologique (ex. pleurer, frissonner) ou comportementale (ex. taper du poing) de l'émotion ; Ci-après en vert.
- ▶ émotion « **désignée** » : émotion désignée par le lexique émotionnel (ex. triste, effrayer) ; Ci-après en jaune.
- ▶ émotion « **étayée** » : émotion inférée. L'inférence émotionnelle est étayée par la description d'une situation conventionnellement associée à une émotion (ex. des retrouvailles entre deux amis peuvent étayer la joie) ; Ci-après en orange.
- ▶ émotion « **montrée** » : les caractéristiques de l'énoncé-même montrent que le locuteur était sous le coup d'une émotion lors de l'énonciation (ex. interjections, exclamation). Ci-après en rouge.

Exemples annotés selon notre schéma d'annotation

(a) *En 2014, l'équipe du Brésil a été éliminée en demi-finale [...]. Cette défaite a rendu les Brésiliens tristes [...]. Ils pleuraient et déchiraient les tickets du match à la sortie du stade.*

⇒ 'TRISTESSE' de l'entité « les Brésiliens » exprimée selon trois modes : comportemental, désigné et étayé.

(b) *Grâce aux gaz à effet de serre, la température moyenne sur notre planète est de 15 °C. Sans eux, elle serait de... -18 °C, c'est très froid !*

⇒ 'SURPRISE' de l'entité 'locuteur' exprimée selon un mode : montré.

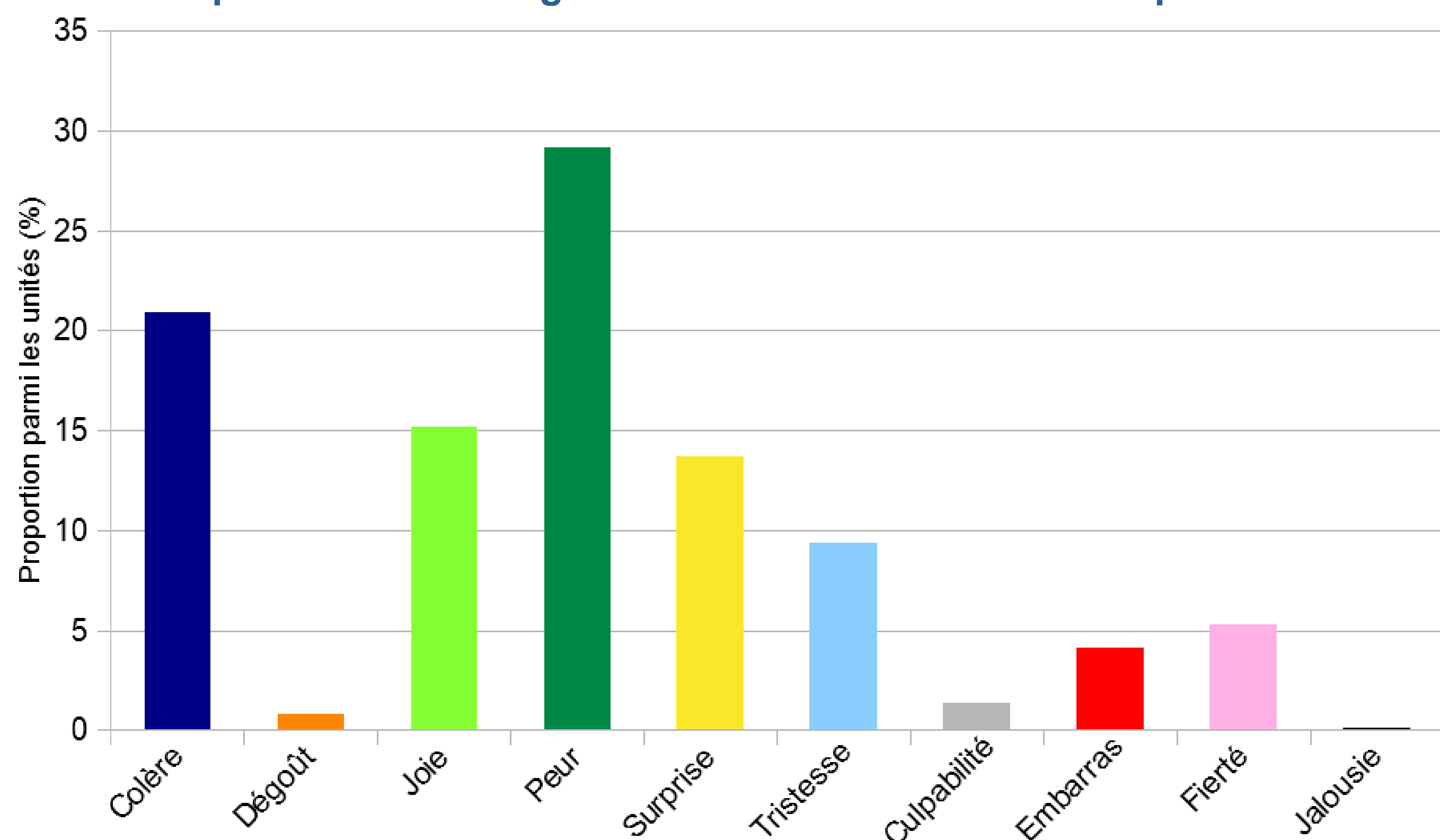
(c) *Au mois de décembre, elle a dû quitter sa maison à cause de la guerre.*

⇒ 'PEUR' de l'entité « elle » exprimée selon deux modes : comportemental et étayé.

Résultats

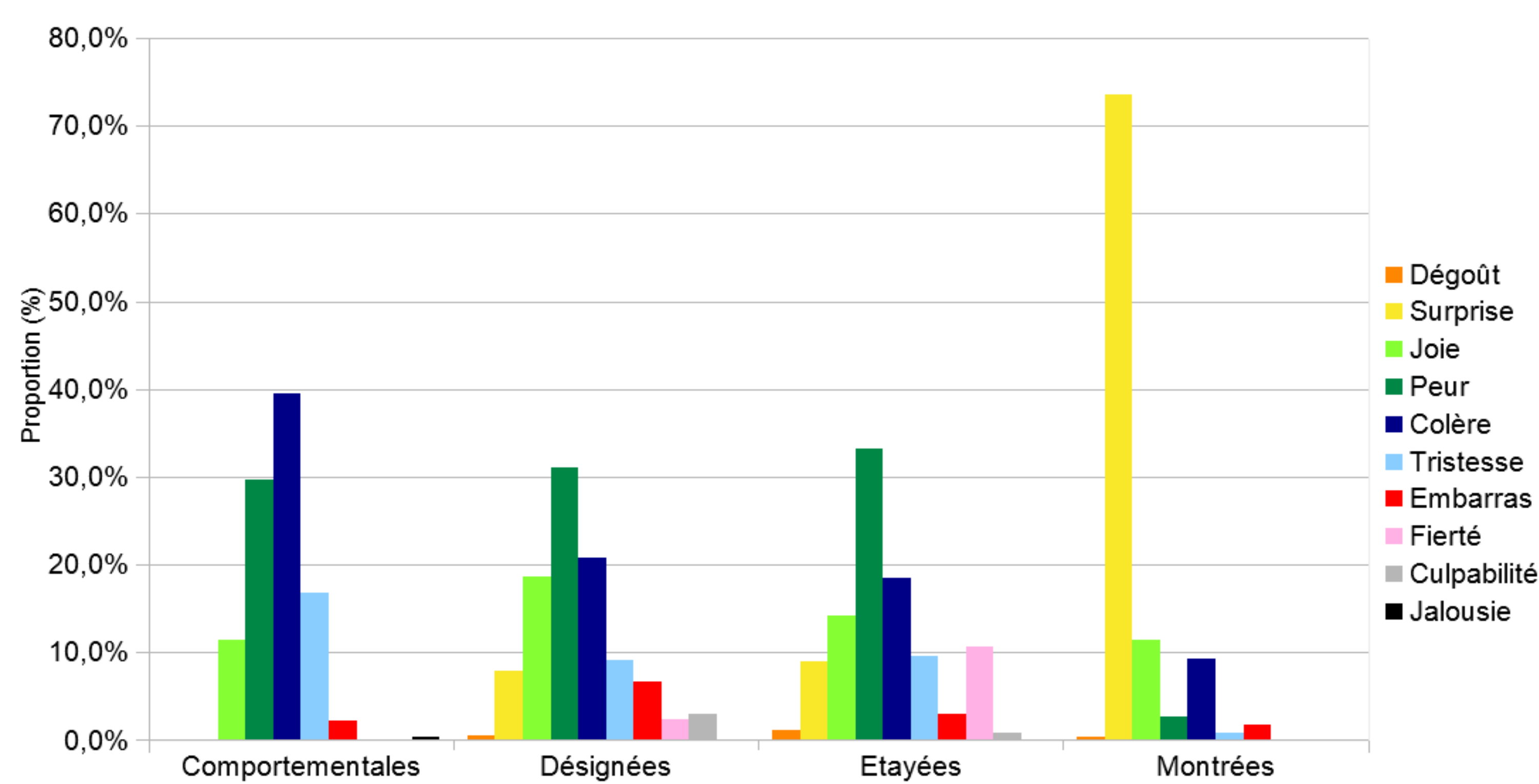
- 2368 unités textuelles émotionnelles au total : 41,5% d'émotions étayées, 37,1% de désignées, 11,8% de comportementales et 9,6% de montrées.
- Principales natures de segments textuels émotionnels : nom (24%), verbe (21%), proposition (19%), adjectif (16%), point d'exclamation (0,6%).

Proportions des catégories émotionnelles dans le corpus



- Sur l'ensemble du corpus : majorité d'émotions négatives (peur et colère), cohérente avec les thèmes du P'tit Libé (thèmes de l'actualité).
- **Représentation des émotions variable selon le mode d'expression** : forte prédominance de la 'SURPRISE' et peu de 'PEUR' dans les émotions montrées ; prépondérance de la 'COLÈRE' et quasi-absence de la 'SURPRISE' dans les émotions comportementales.

Proportions des catégories émotionnelles selon le mode d'expression



- Ces observations quantitatives et l'analyse qualitative des annotations mettent en évidence :
 - ▶ **le recours important à l'émotion étayée**, catégorie complexe et hétéroclite ;
 - ▶ **l'association privilégiée entre des modes d'expression et certaines émotions** ;
 - ▶ **la pertinence des marqueurs linguistiques ne relevant pas du lexique émotionnel** pour la caractérisation des émotions ;
 - ▶ **des critères experts pour le repérage automatique des émotions** au sein d'une chaîne de traitement automatique des langues (TAL).

Discussion

- Apport de la linguistique : meilleure connaissance des divers modes d'expression de différents types d'émotions.
- Apport du TAL : systématisation de l'analyse linguistique via un schéma d'annotation.

Trois perspectives immédiates à ce stade :

- ▶ **En linguistique** : analyse approfondie des émotions étayées et enrichissement du schéma d'annotation pour tenir compte du contexte modal de l'expression d'une émotion (ex. « Il a fui. » vs « Il **a dû** fuir » | « Elle a peur. » vs « **Je crois** qu'elle a peur. »).
- ▶ **En TAL** : grâce au corpus annoté, développement de programmes informatiques pour repérer automatiquement divers modes d'expression d'émotions variées dans les textes.
- ▶ **En psycholinguistique** : mobilisation du corpus annoté dans des protocoles expérimentaux pour étudier l'impact du mode d'expression, y compris le contexte modal, sur la compréhension que les enfants ont d'un texte comportant des émotions.

Références

- [1] 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.
- [2] Blanc, N., & Quenette, G. (2017). La production d'inférences émotionnelles entre 8 et 10 ans: quelle méthodologie pour quels résultats?. *Enfance*, (4), 503-511.
- [3] 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. *Psychologie française* 62(3), 263-277.
- [4] 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.
- [5] Micheli, R. (2014). *Les émotions dans les discours: modèle d'analyse et perspectives empiriques*. De Boeck