



**HAL**  
open science

# **Modale, une ressource lexicale de la modalité au prisme des émotions**

Delphine Battistelli, Aline Étienne

► **To cite this version:**

Delphine Battistelli, Aline Étienne. Modale, une ressource lexicale de la modalité au prisme des émotions. 2025. <hal-05002259>

**HAL Id: hal-05002259**

**<https://hal.science/hal-05002259v1>**

Preprint submitted on 23 Mar 2025

**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.



HAL Authorization

# Modale, une ressource lexicale de la modalité au prisme des émotions

## 1. Introduction

Nous proposons de rendre compte dans cet article des résultats obtenus à l'issue de la première phase d'un travail de constitution d'une ressource lexicale de la modalité pour le français. Cette première phase a été menée en déroulant une méthodologie rigoureuse ancrée dans le questionnement suivant : comment, et à quel point, le marquage lexical de la catégorie sémantique de la modalité entre en intersection avec le marquage lexical des émotions dans les textes ? Ce questionnement, très vaste, se situe au carrefour des approches qui s'intéressent à la subjectivité dans le langage. Il est par exemple clairement posé dans des travaux comme ceux d'O. Galatanu (2024) qui s'interroge sur la manière d'aborder la complémentarité des marqueurs appartenant à ces deux champs de la linguistique que sont celui de la modalité et celui de l'émotion. Les résultats dont nous rendons compte ici visitent en quelque sorte ce questionnement général au travers d'une étude quantitative et qualitative menée à partir de marqueurs linguistiques observés dans deux ressources lexicales que nous avons constituées indépendamment l'une de l'autre : l'une pour la modalité, notée RM, construite à partir des exemples fournis dans un ouvrage de référence pour les modalités en français (Gosselin, 2010) ; l'autre pour les émotions, notée RE, construite à partir des occurrences lexicales émotionnelles repérées dans un large corpus de textes selon une approche située dans la lignée des travaux de Micheli (2014). Deux constats ont guidé la mise en place de notre méthodologie : d'une part, l'observation que nous venons de formuler de l'intersection non vide entre RE et RM ; d'autre part, l'intérêt alors de considérer les marqueurs d'émotions comme de potentiels marqueurs de modalité pouvant enrichir RM. Notre démarche permet de vérifier, comme nous le verrons, une intuition largement partagée selon laquelle le domaine de l'émotion correspond en partie aux modalités appréciatives et axiologiques ; mais aussi de « découvrir » d'autres « affinités », peut être moins immédiates, entre les deux champs de l'émotion et de la modalité.

La suite de notre article est organisée comme suit. Dans un premier temps, nous situons brièvement notre approche linguistique de la modalité et des émotions. Puis nous décrivons les deux ressources RM et RE et montrons les résultats de l'observation fine de leur intersection. Nous décrivons ensuite la méthodologie que nous avons mise en place afin de proposer la ressource Modale qui constitue une version enrichie de RM à partir de RE<sup>1</sup>. Enfin, nous revenons sur quelques éléments conceptuels clé dans la définition de la modalité et des émotions examinés à l'aune des résultats obtenus en constituant la ressource Modale. Nous concluons par un bilan de nos travaux à ce jour.

## 2. Conceptualiser la subjectivité dans le langage

La subjectivité telle qu'elle est approchée en linguistique - c'est-à-dire « la subjectivité dans le langage » selon la formule de Benveniste (1966) – peut être définie comme l'ensemble des « procédés linguistiques [...] par lesquels le locuteur imprime sa marque à l'énoncé, s'inscrit dans le message (implicitement ou explicitement) et se situe par rapport à lui [...] » (Kerbrat-Orecchioni, 1980 : 32). Dans cette formulation, on retrouve les principaux éléments de définition à la fois de la catégorie de la modalité et du domaine des émotions. Il n'est donc pas étonnant que l'on soit tentés, comme nous le faisons ici, d'en étudier plus précisément les zones d'intersection.

---

<sup>1</sup> C'est la raison pour laquelle nous utilisons l'acronyme Modale ; il exprime le fait que la ressource a été construite en interrogeant l'intersection entre Modalités et Emotions.

Nous décrivons brièvement dans les sections qui suivent le modèle théorique de la modalité sur lequel nous nous appuyons, et faisons de même en ce qui concerne les émotions.

## 2.1. La catégorie sémantique de la modalité

On s'accorde généralement à définir la modalité comme un domaine sémantique qui englobe un ensemble de valeurs (jussive, désirative, éventuelle, hypothétique, probable, possible, nécessaire, etc.) exprimant l'opinion, l'attitude ou le jugement de l'énonciateur/locuteur par rapport à un certain contenu propositionnel (le *dictum* dans la terminologie de Bally, 1942). Dans les études sur la modalité en linguistique, tous les types de modalités (recensés de façon presque exhaustive par Le Querler (1996)) n'apparaissent pas systématiquement. Comme le signale Meunier (1974 : 8), « le terme [modalité] est saturé d'interprétations qui ressortissent, explicitement ou non, selon les linguistes qui l'utilisent, de la logique, de la sémantique, de la psychologie, de la syntaxe, de la pragmatique ou de la théorie de l'énonciation ». Le plus souvent, on retrouve quatre grands groupes de modalités : les modalités « dynamiques » (qui peuvent être rapprochées des modalités aléthiques telles qu'elles sont définies dans (Le Querler, 1996) et qui renvoient au possible et au nécessaire), les modalités déontiques (qui renvoient à l'obligation et à la permission), les modalités épistémiques (qui marquent le degré de certitude avec lequel l'énonciateur présente son propos ou encore expriment l'engagement de l'énonciateur par rapport à la vérité de l'énoncé) et les modalités appréciatives/axiologiques (l'énonciateur n'est alors plus envisagé comme source de croyance, mais de désir ou de jugement).

Comme le rappelle Gosselin (2019 : 1), les modalités appréciatives, mais aussi axiologiques, relèvent d'une tradition spécifiquement francophone, quoiqu'on puisse considérer qu'elles apparaissent dans la sphère anglosaxonne sous le nom de « bouloimaic modality » ou d'« emotional attitude » chez (Nuyts, 2006 : 12). Dans les travaux francophones, elles sont explicitement désignées chez Brunot (1922 : 541) sous le nom de « modalités sentimentales » et sont reprises sous l'appellation de « modalités appréciatives » et/ou « axiologiques » par ses successeurs avec, parmi eux, Bally (1932) ou Culioli (1968). Pour sa part, Gosselin (2019 : 1) juge important de distinguer les deux concepts : « d'une part les modalités appréciatives qui expriment le caractère désirable ou indésirable d'un procès (état ou événement), d'autre part les modalités axiologiques qui indiquent le caractère louable ou blâmable d'un procès ou d'un jugement ». Dans tous les cas, la prise en compte des modalités appréciatives comme axiologiques implique une conception de la notion de modalité qualifiée de « large » par (Le Querler 1996 : 51), conception que l'on trouve énoncée par (Bally, 1942 : 3) de la façon suivante : « la modalité est la forme linguistique d'un jugement intellectuel, d'un jugement affectif ou d'une volonté qu'un sujet pensant énonce à propos d'une perception ou d'une représentation de son esprit ». Remarquons alors que la définition très générale qui consiste à envisager la modalité comme « l'attitude d'un sujet pensant » semble être communément acceptée dans le cadre de la linguistique énonciative, même si elle lui préfère éventuellement d'autres termes, comme celui de « sujet parlant » (Benveniste, 1966) ou de « sujet énonciateur-locuteur » (Culioli, 1990 : 141). Dans ce cadre, il est à noter que la modalité est clairement étudiée corrélativement à la notion de prise en charge (d'un contenu propositionnel, selon un certain point de vue). Les traces de l'inscription du sujet parlant dans l'énoncé, c'est-à-dire « la subjectivité dans le langage », sont nombreuses (e.g. déictiques, modaux, ...). Et surtout, chaque énonciation est vue comme le produit d'un sujet parlant et ce dernier présente les faits de façon subjective, par rapport à son point de vue, son expérience. C'est là qu'intervient, entre autres, une différence entre des travaux se réclamant de la linguistique énonciative et d'autres travaux, à l'exemple de l'approche anglosaxonne adoptée dans (Wilson & Wiebe, 2003) qui considère les textes comme relevant d'« un mélange de matériaux factuels et subjectifs » (nous traduisons). Dans le cadre de la linguistique énonciative, tout texte est *en soi* le produit d'une subjectivité, celle de l'énonciateur, même s'il apparaît comme « effacé » dans certaines productions.

Que l'on revendique ou non de se situer dans le courant de la linguistique énonciative, on s'entendra pour dire que les marqueurs modaux constituent des traces de la subjectivité du sujet parlant. Eriger cette classe de marqueurs en tant que *catégorie sémantique* constitue en outre le point commun des approches linguistiques qui se penchent sur la diversité conceptuelle dont relève l'analyse des marqueurs de cette classe. Parmi ces approches, nous avons choisi de retenir celle proposée dans le cadre de la théorie modulaire des modalités présentée dans (Gosselin, 2010) où de nombreux lemmes modaux sont donnés à titre d'exemples. Cela nous a permis de constituer

la ressource lexicale RM. Avant de présenter en détail cette ressource, nous décrivons ci-après succinctement les éléments principaux de l'approche théorique de L. Gosselin (2010).

## 2.2. Approche de la modalité retenue

L'approche que nous retenons pour la constitution de la ressource lexicale de la modalité RM est celle de (Gosselin, 2010). Gosselin définit la modalité comme « tout mode de validation / invalidation d'un contenu représenté » (*ibid.* : 309). Six catégories modales (avec leurs continuums de valeurs) sont distinguées (cf. tableau 1).

Catégorie modale	Continuum de valeurs	Exemples de marqueurs (Gosselin, 2010)
aléthique	impossible, contingent, possible, nécessaire	rectangulaire, « il neige », nécessité, chimiquement
épistémique	exclu, contestable, probable, certain	croire, probable, certitude, douteux, vraisemblable
déontique	interdit, facultatif, permis, obligatoire	interdire, obliger, permettre, impérativement
boulique	rejeté, indifférent, désiré	rejeter, souhaitable, désirer, vouloir
appréciative	indésirable, indifférent, désirable	beau, heureux, admirer, souffrance, chouette !
axiologique	blâmable, indifférent, louable	injuste, lâche, voleur, désapprouver, honteux

Tableau 1 : Catégories modales, continuums de valeurs et exemples (Gosselin, 2010)

Un même contenu *c* – ou *dictum* – peut être « affecté » (*loc. cit.*) par différentes modalités, marquant par exemple l'expression du désir de *c* dans (1) ou d'un jugement de valeur renvoyant au caractère désirable de *c* comme dans (2) ou louable quant à *c* comme dans (3) avec *c*= 'marcher dans la campagne'.

- (1) Je souhaite marcher dans la campagne
- (2) Zoé est heureuse de marcher dans la campagne
- (3) Il y a du mérite à marcher dans la campagne

Gosselin distingue en outre les modalités selon qu'elles sont extrinsèques, intrinsèques ou par inférence. Les modalités intrinsèques sont marquées par les lexèmes constitutifs du *dictum* et l'auteur part du postulat selon lequel tout lexème marque au moins une modalité. Ce sont les modalités appréciatives intrinsèques aux lexèmes 'tyran', 'mort' et 'flammes' qu'il faut prendre en compte dans (4) et (5), alors que d'autres sont ajoutées en position initiale dans (6) et en position intégrée dans (7).

- (4) Le tyran est mort dans les flammes (ex. tiré de [Gosselin, 2019 : 8])
- (5) Le tyran a échappé aux flammes (*loc. cit.*)
- (6) Je suis heureux / c'est une chance / heureusement que le tyran est mort dans les flammes (*loc. cit.*)
- (7) Le tyran a eu la chance / le bonheur / de la chance d'échapper aux flammes (*loc. cit.*)

Avec les quelques exemples présentés dans le tableau 1, il apparaît vite que certains marqueurs modaux lexicaux recourent l'expression lexicale des émotions, en particulier pour ce qui concerne les modalités appréciatives (e.g. « heureux », « chance ») et axiologiques (e.g. « mérite »). Ce ne sont cependant pas les seules catégories de modalités à apparaître à l'intersection de l'expression des émotions, surtout si l'on ne se limite pas à l'examen des émotions dans leur mode de dénotation direct.

### 2.3. L'analyse des émotions

Dans une approche linguistique des émotions, l'une des préoccupations est la délimitation conceptuelle du terme émotion vis-à-vis d'autres termes tels qu'affect, état affectif, sentiment ou encore humeur. Il reste que « l'usage [...] de ces termes est assez flou, et leurs significations peuvent se recouvrir » (Plantin 2020 : 2). Quel que soit le terme retenu, on pourra cependant s'accorder sur le fait que ce domaine sémantique rend compte – au moins – de la manière dont, dans un discours, une entité (le locuteur/énonciateur ou bien une autre entité que le locuteur met en discours) s'illustre comme affectée. Certains travaux linguistiques mettent alors tout particulièrement l'accent sur le fait de caractériser « l'expérienceur » de l'émotion, soit l'entité (humain, animal, etc.) qui ressent l'émotion. Cette notion d'expérienceur intervient par exemple dans l'analyse des verbes dits de sentiment de Mathieu (2005) et dans celle des expressions de peur par Dobrovolski & Bresson (1995). Elle intervient également dans la définition des modes d'expression par Micheli (2014), qui utilise la capacité d'une émotion à être plus ou moins facilement auto-attribuée (c'est-à-dire attribuée à l'énonciateur) ou allo-attribuée (c'est-à-dire attribuée à une entité autre que l'énonciateur) pour distinguer entre eux les différents modes d'expression des émotions.

Répertorier tous les moyens linguistiques renvoyant à l'expression d'émotions représente un véritable défi pour les linguistes. C. Kerbrat-Orecchioni souligne à cet égard 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 » (Kerbrat-Orecchioni, 2000 : 50). L'étude de l'expression des émotions nécessite de fait de prendre en compte des observables très nombreux et hétérogènes et ce même s'il devient alors « difficile d'associer [aux émotions] un (ou plusieurs) lieu(x) stable(s) dans le système linguistique » (Micheli *et al.*, 2013 : 9). La plupart des travaux linguistiques tendent ainsi à se focaliser sur des aspects restreints de la dénotation des émotions. Lorsqu'un panel de catégories d'émotions (e.g. joie, peur, colère, embarras, culpabilité) est traité, les travaux se concentrent généralement sur l'étude du lexique émotionnel (voir par exemple le lexique émotionnel constitué à l'aide de critères combinatoires par Diwersy *et al.* (2014)) et même parfois sur l'étude du lexique verbal seulement (voir par exemple l'étude des verbes dits de sentiment proposée par (Mathieu, 2005)). Les analyses fines de marqueurs ne relevant pas du lexique émotionnel strictement se limitent quant à elles le plus souvent à l'examen de quelques structures ou de quelques catégories émotionnelles. Ce phénomène est illustré pour le français, par exemple, par l'étude de syntagmes ayant pour prédicat sémantique « peur » (Dobrovolski & Bresson, 1995), par l'analyse de la structure « être en colère » vs « \*être en peur » (Leeman-Bouix, 1995), ou encore par l'étude de diverses structures (morpho-)syntaxiques exprimant la joie et la rage (Gross, 1995). À notre connaissance, seul Micheli (2014) développe pour le français une approche des émotions incluant une grande hétérogénéité de marqueurs linguistiques (intégrant par exemple des constructions syntaxiques ou la notion d'émotion suggérée) pouvant les exprimer et allant au-delà du strict lexique émotionnel (comme par ex. avec des termes tels que « joie » ou « apeuré »). Les travaux que nous décrivons ci-après s'appuient sur les travaux de Micheli (2014) et constituent notre cadre théorique pour l'analyse des émotions.

### 2.4. Approche des émotions retenue

Notre approche des émotions s'illustre au travers d'un schéma d'annotation manuelle présenté par Etienne *et al.* (2022). Ce schéma est ancré dans les travaux linguistiques de Micheli (2014) tout en restant compatible avec des travaux en psycholinguistique tels que ceux présentés notamment par Blanc (2010) – cf. tableau 2. Dans ce schéma d'annotation sont distingués les quatre modes d'expression suivants auxquels sont associés divers types de marqueurs linguistiques de la notion d'émotion :

- les émotions désignées par un terme du lexique émotionnel (marqueurs linguistiques de niveau lexical) ;
- les émotions comportementales, véhiculées par la description d'un comportement (marqueurs linguistiques de niveau lexical ou propositionnel) ;
- les émotions suggérées qui doivent être inférées de la description d'une situation conventionnellement associée à une émotion (marqueurs linguistiques de niveau lexical ou propositionnel) ;

– et les émotions montrées par des caractéristiques de l'énoncé relevant de plusieurs niveaux d'analyse (marqueurs linguistiques des niveaux lexical, syntaxique et typographique).

Typologie des modes d'expression d'une émotion				
d'après (Blanc, 2010)	Émotion désignée ex. Luc est <u>heureux</u> .	Expression comportementale de l'émotion ex. Luc <u>sourit</u> .		Émotion suggérée ex. Luc a <u>gagné la course</u> .
d'après (Micheli, 2014)	Émotion dite ex. Luc est <u>heureux</u> , il <u>sourit</u> .		Émotion montrée ex. " <u>Oh, chouette !</u> " dit Luc.	Émotion étayée ex. Luc a <u>gagné la course</u> .
d'après (Etienne et al., 2022)	Émotion désignée ex. Luc est <u>heureux</u> .	Émotion comportementale ex. Luc <u>sourit</u> .	Émotion montrée ex. " <u>Oh, chouette !</u> " dit Luc.	Émotion suggérée ex. Luc a <u>gagné la course</u> .

Tableau 2. Typologie des modes d'expression d'une émotion selon (Blanc, 2010 ; Micheli, 2014 ; Etienne et al. 2022)

Le schéma d'Etienne *et al.* (2022) ne se limite ainsi pas à l'étude d'unités appartenant au strict lexique émotionnel. Il permet également d'annoter des segments textuels relevant de la cause ou de la conséquence d'une émotion (à rapprocher de ceux impliqués dans les émotions dites suggérées comme l'a d'ailleurs remarqué Micheli (2014 : 143)), ou encore ceux renvoyant à l'expérienceur de l'émotion (c'est-à-dire l'entité qui ressent l'émotion). Nous ne décrivons cependant ici que trois des notions du schéma, celles qui permettent de saisir ensuite immédiatement les analyses conduites afin de construire la ressource RE :

- **La catégorie émotionnelle.** Concept le plus fréquemment mobilisé dans les travaux portant sur l'étude des émotions (quel qu'en soit le domaine, linguistique (e.g. (Gross, 1995 ; Mathieu, 2005 ; Diwersy *et al.*, 2014)) ou non (e.g. (Piolat & Bannour, 2009 ; Blanc, 2010)), la catégorie émotionnelle permet d'associer une étiquette aux unités émotionnelles annotées. 11 catégories pré-définies ont été retenues dans le schéma : colère, dégoût, joie, peur, surprise, tristesse, admiration, culpabilité, embarras, fierté et jalousie. Il est à noter que, à l'instar de nombre d'approches (e.g. Mathieu, 2005 ; Piolat & Bannour, 2009), ces classes sont entendues comme des « macro- » catégories regroupant dans un lien d'hyponymie des catégories sémantiques plus fines. La « colère » regroupera par exemple la fureur, la rage, l'irritation ou encore l'énervement. De plus, jusqu'à deux catégories émotionnelles peuvent être associées à une même unité émotionnelle (cf. unité textuelle de l'exemple (8) à laquelle sera associée deux étiquettes de catégorie émotionnelle, peur et tristesse) ;
- **Le mode d'expression.** Cette notion renvoie aux moyens linguistiques employés pour exprimer l'émotion (cf. tableau 2). Elle permet une couverture plus large du champ d'observation de l'expression des émotions (c'est-à-dire ne se limitant pas au lexique émotionnel) et en même temps un typage linguistique plus fin des unités émotionnelles repérées. Quatre modes distincts d'expression des émotions, largement inspirés des travaux de Micheli, sont retenus dans le schéma :
  - a) le mode *désigné*. Il correspond aux émotions dénotées directement par l'usage d'un terme du lexique émotionnel (par ex. *admiratif, terreur*) ;
  - b) le mode *comportemental*. Il correspond aux émotions exprimées par le biais de la description d'une manifestation physique de l'émotion. Cette manifestation peut être physiologique (par ex. *rougir, pleurer*) ou consister en des comportements plus complexes (par ex. *frapper quelqu'un* ou *casser quelque chose* sous le coup de la colère) ;
  - c) le mode *suggéré*. Il correspond aux émotions exprimées par la description de situations conventionnellement associées à des ressentis émotionnels (par ex. *revoir un ami après une longue séparation* pourra suggérer une émotion positive telle que la joie) ;
  - d) le mode *montré*. Ce mode le plus hétérogène dans ses réalisations, il correspond à l'ensemble des

caractéristiques linguistiques des énoncés (par ex. interjections, onomatopées, structures syntaxiques particulières comme les énoncés exclamatifs ou averbaux) permettant de traduire l'état émotionnel dans lequel se trouvait l'énonciateur au moment de l'énonciation.

- **Le déclencheur.** Cette notion intervient pour préciser les annotations produites en signalant le ou les termes clés sur lesquels s'ancre le contenu émotionnel véhiculé, autrement dit, le noyau ou cœur de l'unité émotionnelle annotée. Pour une émotion *désignée*, le déclencheur est ainsi systématiquement le terme du lexique émotionnel annoté. Pour les autres modes d'expression, l'annotateur choisit les termes lui paraissant les plus saillants sur le plan de l'expression émotionnelle, sachant qu'une même unité émotionnelle peut avoir plusieurs déclencheurs. Dans l'exemple (8), extrait du guide d'annotation qui accompagne le schéma, l'unité « la sécheresse frappe régulièrement, tuant les animaux et laissant les éleveurs sans ressources » suggère la peur et la tristesse des habitants d'un pays. Cette unité dispose de trois déclencheurs : « sécheresse », « tuant » et « sans ressources ».

(8) La sécheresse frappe régulièrement, tuant les animaux et laissant les éleveurs sans ressources (exemple extrait de [Etienne et Battistelli, 2021 : 23])

### 3. La modalité au prisme des émotions et vice versa

Nous présentons maintenant les ressources lexicales que nous avons construites : d'abord la ressource lexicale en lemmes modaux, notée RM, tirée de l'ouvrage de (Gosselin, 2010, chapitre 7) ; puis, la ressource en lemmes émotionnels, notée RE, tirée du corpus annoté en émotions (Etienne *et al.*, 2022). Nous explorons ensuite les zones d'intersection entre ces deux ressources constituées indépendamment.

#### 3.1. Les marqueurs modaux de référence (RM)

La ressource RM résulte d'un travail systématique de répertoriage des exemples et analyses de Gosselin (2010, chapitre 7), grâce auquel nous avons pu extraire 228 marqueurs de modalité. Ces marqueurs, ci-après nommés « marqueurs de référence », correspondent tous à un lemme de type forme isolée<sup>2</sup> (c'est-à-dire ne contenant ni espace, ni apostrophe) auquel sont associés :

- une ou plusieurs catégories modales parmi celles de l'aléthique, de l'appréciatif, de l'axiologique, du boulique, du déontique et de l'épistémique (soient l'ensemble des catégories modales mobilisées par L. Gosselin) ;
- un niveau de positionnement de l'expression de la modalité qui a trois valeurs : intrinsèque, extrinsèque ou inférence.

Le lemme *agréable* est, par exemple, associé à la catégorie modale de l'appréciatif qu'il dénote de façon intrinsèque, tandis que la potentielle modalité appréciative du lemme *frite* sera à inférer du contexte d'énonciation. Le lemme *certainement* est quant à lui lié à la catégorie modale de l'épistémique qu'il dénote de façon extrinsèque.

Parmi les marqueurs de référence, 18,9 % sont polysémiques au sens où ils articulent l'expression d'au moins deux catégories modales. C'est le cas par exemple du lemme *bon* qui dénote intrinsèquement la catégorie de l'appréciatif et celle de l'axiologique. Le tableau 3 résume les caractéristiques de l'ensemble des marqueurs de référence constitué en indiquant la part que représentent dans cet ensemble chaque catégorie modale, chaque niveau de positionnement de la modalité et le phénomène de polysémie.

---

<sup>2</sup> En réalité, chaque marqueur de modalité analysé ou cité dans (Gosselin, 2010, Chapitre 7) a été répertorié pour mener à la création de deux listes : celle des marqueurs dits de « forme isolée », c'est-à-dire correspondant à un token unique et pouvant donc être aisément repéré ; et celle des marqueurs dits de type « construction », correspondant à des constructions impliquant plusieurs tokens, plus difficiles à repérer dans un corpus car nécessitant des prétraitements (analyse syntaxique, etc.). Par souci de praticité, seule la liste des marqueurs de forme isolée a été employée à ce jour.

		Nb de marqueurs	Prop. (%)
Total		228	
Catégorie modale	aléthique	45	19,7
	appréciatif	75	32,9
	axiologique	44	19,3
	boulque	34	14,9
	déontique	20	8,8
	épistémique	58	25,4
Niveau de positionnement de la modalité	intrinsèque	112	49,1
	extrinsèque	98	43,0
	inférence	18	7,9
Polysémie	oui	43	18,9
	non	185	81,1

Tableau 3 - Caractéristiques des 228 marqueurs de référence issus de (Gosselin, 2010)

Une fois le lexique de marqueurs de référence établi, nous avons effectué automatiquement l'intersection entre ces marqueurs et un ensemble de lemmes participant à l'expression d'émotions, ci-après nommés « lemmes émotionnels ». Ces lemmes sont issus du corpus annoté manuellement en émotions présenté par Etienne *et al.* (2022). Nous offrons, dans la section suivante, une description détaillée de ce corpus et des lemmes sélectionnés pour observer l'intersection entre modalités et émotions.

### 3.2 Les lemmes émotionnels (RE)

L'étude de l'intersection entre les catégories sémantiques de la modalité et des émotions rend nécessaire de disposer d'annotations à l'échelle du marqueur linguistique (par opposition à une annotation à l'échelle de la phrase ou du texte entier), ne se limitant pas, comme nous l'avons déjà dit, au repérage du lexique émotionnel mais qui rendent compte de toute la diversité des marqueurs participant de l'expression des émotions dans les textes. À notre connaissance, seuls le schéma d'annotation des émotions et le corpus annoté selon ce schéma (Etienne *et al.*, 2022) permettent d'avoir une approche de ce type. Notre étude se fonde donc sur ce corpus annoté.

#### Caractéristiques du corpus annoté en émotions

Le corpus décrit par Etienne *et al.* (2022) se compose de 1 594 textes écrits en français (environ 28 000 phrases), destinés aux enfants de 6 à 14 ans, et répartis en trois genres : journalistique (91 % des phrases du corpus), encyclopédique (1 %) et fictionnel (8 %)³. Il a été annoté manuellement en émotions par des experts linguistes à l'échelle du marqueur linguistique, *via* la plateforme d'annotation Glozz⁴ (Widlöcher & Mathet, 2012).

6 790 unités émotionnelles (4 886 dans les textes journalistiques, 1 783 dans ceux de fiction et 121 dans les encyclopédiques) ont été délimitées dans l'ensemble du corpus. Parmi ces unités, les catégories émotionnelles les plus représentées sont : la colère (25,3 % des unités), la peur (21,8 %), la surprise (18,1 %), la joie (18 %) et la tristesse (14 %). Toutes les autres catégories émotionnelles (admiration, fierté, embarras, dégoût, culpabilité et jalousie) représentent chacune moins de 5 % des unités émotionnelles⁵. Il est à noter que les modes d'expression des émotions apparaissent dans le corpus annoté dans des proportions à peu près similaires (33 % des unités relèvent du mode suggéré ; 24,8 % du désigné ; 21,3 % du comportemental ; et 20,9 % du montré).

L'analyse rapportée ici s'appuie sur les 6 790 unités émotionnelles du corpus annoté, et plus précisément sur les 7 204 déclencheurs qui y sont associés. Ces déclencheurs ont été tokenisés puis lemmatisés à l'aide de la librairie opensource Stanza⁶ (Qi *et al.*, 2020). Notre étude porte ainsi sur 10 905 tokens de déclencheurs, répartis entre

³ Ce corpus a été constitué dans le cadre du projet ANR TextToKids (2019-2024) dont la visée était de fournir des éléments d'analyse de la complexité de textes destinés à des enfants jeunes lecteurs (Battistelli *et al.*, 2022 ; Etienne *et al.*, 2024). La complexité d'un texte y a été évaluée en tenant compte de divers critères linguistiques, et notamment celui des catégories émotionnelles en jeu, certaines catégories étant considérées comme basiques (e.g. joie, tristesse), d'autres comme complexes (e.g. embarras, honte) et donc moins accessibles à de jeunes lecteurs.

⁴ La plate-forme Glozz constitue un environnement d'annotation et d'exploration de corpus. Elle est téléchargeable à l'adresse : <http://www.glozz.org/>

⁵ Les catégories les plus représentées sont donc toutes des catégories dites de base (colère, joie, peur, surprise, tristesse) tandis que les moins représentées sont celles dites complexes (admiration, fierté, embarras, dégoût, culpabilité et jalousie). Il est vraisemblable que ces proportions sont liées au corpus utilisé, composé de textes destinés à des enfants jeunes lecteurs, mais ceci reste à vérifier.

⁶ Stanza est un outil multilingue développé par le Stanford NLP Group (<https://corliapi.ortolang.fr/stanza/>). Il permet de réaliser divers traitements tels que la tokenisation, la lemmatisation, l'étiquetage morpho-syntaxique et une analyse syntaxique en relations de dépendances. La tokenisation est le processus de division d'un texte en plus petites unités appelées « tokens »

2 174 lemmes différents - les « lemmes émotionnels » qui constituent ainsi notre ressource RE. Le tableau 4 présente les proportions représentées par les cinq catégories émotionnelles les plus fréquentes du corpus et par les quatre modes d'expression des émotions au sein de ces ensembles de tokens et de lemmes émotionnels.<sup>7</sup>

Notion	Étiquette	Proportion (%) parmi les	
		tokens	lemmes
catégorie émotionnelle	colère	26,2	34,2
	peur	22,0	30,9
	surprise	10,7	17,1
	joie	21,6	32,1
	tristesse	17,7	28,1
mode d'expression	désigné	16,8	17,3
	comportemental	21,0	27,2
	suggéré	50,8	67,3
	montré	11,3	19,0

Tableau 4 – Répartition des catégories émotionnelles et des modes d'expression parmi les déclencheurs d'unités émotionnelles

### 3.3 Analyse du recouvrement entre marqueurs modaux et lemmes émotionnels

Nous avons ensuite effectué automatiquement l'intersection entre RM et RE, soit donc entre les marqueurs modaux de référence et les lemmes émotionnels. Grâce à cette analyse automatique, nous avons pu confirmer dans un premier temps l'existence d'un important recouvrement entre RE et RM. 48,2 % des marqueurs de référence (soit 110 marqueurs) correspondent en effet à un lemme émotionnel. Comme l'illustre le tableau 5, la plupart de ces marqueurs (45,5 %) relèvent de la catégorie modale appréciative. Ce résultat n'est pas surprenant puisque cette catégorie modale est conceptuellement très proche du champ des émotions. On notera que, de manière un peu surprenante, la catégorie épistémique est aussi largement exprimée avec près d'un quart (24,5 %) des marqueurs de référence correspondant à un lemme émotionnel. Le tableau 6 donne, à titre indicatif, les dix marqueurs de référence que l'on retrouve le plus fréquemment parmi les lemmes émotionnels étudiés. On notera dans le tableau 6 la présence du marqueur de modalité aléthique et épistémique « lui ». S'il est indiscutable que ce type de marqueur peut effectivement bien fonctionner comme un marqueur de modalité<sup>8</sup>, il est clair cependant ici que sa présence dans le Top10 des marqueurs de référence est surprenante. En réalité, cela provient d'un effet de bord lié à l'utilisation, pour la lemmatisation (rappelons-nous en effet que le repérage automatique des marqueurs de modalité s'effectue à l'échelle du lemme), de l'outil Stanza. Ce dernier attribue à tous les pronoms (y compris personnels, tels que *je, tu, etc.*) le lemme « lui ». Cela occasionne donc très probablement un sur-repérage du marqueur 'lui'. Au moment où nous avons effectué les analyses présentées ici (et notamment dans le tableau 6), ce marqueur était intégré dans la ressource. Depuis, nous avons décidé d'exclure ce marqueur de la ressource utilisée dans la version du processeur ModalE testable en ligne pour éviter au moins cet effet de bord.

(par ex., la phrase « L'ennui commence à le dévorer. » serait découpée en tokens [« L' », « ennui », « commence », « à », « le », « dévorer », « . »]). Les lemmes sont, comme il est d'usage : la forme pour les lexèmes invariables ; la forme infinitive pour les verbes ; le singulier pour les noms et le masculin singulier pour les adjectifs ; 'le' pour les articles le, la, l', les ; 'un' pour un et des ; 'son' pour les déterminants possessifs (mon, ma, mes, ton, ta, tes ...) ; etc.

<sup>7</sup> La proportion remarquablement importante de lemmes émotionnels (comme de tokens) associée au mode suggéré comparativement aux autres modes s'explique par le fait que les unités textuelles relevant du suggéré sont assez souvent porteuses de plusieurs déclencheurs (voir l'exemple (8)).

<sup>8</sup> (Gosselin, 2010) démontre la pertinence d'un tel marqueur au travers de l'exemple « Moi, personnellement, je trouve que Pierre est grand » dans un usage épistémique (ibid. : 327) ou au travers de l'exemple « C'est moi qui ai posé le dessert sur la table » dans un usage aléthique (ibid. : 319).

	Nb de marqueurs	Prop. (%)	
<b>Total</b>	<b>110</b>		
<b>Catégorie modale</b>	aléthique	18	16,4
	appréciatif	50	45,5
	axiologique	18	16,4
	boulique	16	14,5
	déontique	10	9,1
	épistémique	27	24,5

Tableau 5 - Caractéristiques des marqueurs de référence correspondant à un lemme émotionnel

Marqueur de référence	Catégories modales	Positionnement de la modalité	Nb d'occ.
mal	appréciatif, axiologique	intrinsèque	56
bien	appréciatif, axiologique	extrinsèque	45
lui	aléthique, épistémique	intrinsèque	44
craindre	appréciatif, épistémique	extrinsèque	38
bon	appréciatif, axiologique	intrinsèque	38
guerre	axiologique	inférence	37
heureux	appréciatif	extrinsèque	35
fête	appréciatif	intrinsèque	32
pouvoir	aléthique, déontique, épistémique	extrinsèque	32
jamais	aléthique	extrinsèque	27

Tableau 6 - Top 10 des marqueurs de référence correspondant à un lemme émotionnel

À noter que parmi l'ensemble de marqueurs de référence correspondant à un lemme émotionnel, les marqueurs extrinsèques représentent 49,1 %, les intrinsèques 44,5 % et les inférés 6,4 %. On retrouve donc ici les différents niveaux de positionnement de la modalité dans des proportions similaires à celles de l'ensemble de marqueurs de référence dans sa totalité (cf. tableau 3).

Dans un second temps, le caractère lacunaire du lexique de marqueurs de référence a été mis en avant. L'ensemble de lemmes émotionnels que nous avons étudié est en effet composé de 2 174 lemmes différents (c'est-à-dire sans doublons) dont 94,9 % ne coïncident pas avec un marqueur de référence. Pourtant, parmi ces lemmes, certains participent sans aucun doute de l'expression d'une catégorie modale. Les lemmes émotionnels *admiratif*, *prestigieux*, *émerveiller* ou encore *miracle*, absents du lexique de marqueurs de référence, apparaissent par exemple tous en lien avec le marqueur de référence *admirable* (modalité appréciative intrinsèque).

L'observation de l'intersection entre RM et RE a donc d'une part souligné l'intérêt théorique d'analyser le croisement entre la catégorie sémantique de la modalité et le domaine des émotions et d'autre part mis en lumière la possibilité d'étendre « assez facilement » la ressource de marqueurs modaux de référence constituée à partir de (Gosselin, 2010).

## 4. Vers ModaleE

En vue d'étoffer notre ressource de marqueurs de modalité RM à l'aide de marqueurs tirés de l'analyse des émotions RE, nous avons développé une méthodologie adossée à des critères rigoureux permettant de décider du statut ou non de marqueur modal pour un marqueur venant de RE. Nous décrivons tout d'abord cette méthodologie et présentons ensuite la ressource finale nommée ModaleE ainsi constituée.

### 4.1. Une méthodologie d'enrichissement de la modalité par l'émotion

Notre méthodologie se compose de deux étapes. La première consiste à effectuer l'intersection entre les marqueurs

modaux de référence et les lemmes émotionnels en écartant les lemmes se trouvant dans cette intersection pour la suite du processus. La seconde étape consiste en une analyse manuelle systématique de chaque lemme ne coïncidant pas avec un marqueur de modalité de référence par un évaluateur expert (c'est-à-dire ayant été formé à la notion de modalité telle qu'elle est abordée par L. Gosselin). Pour chacun de ces lemmes, l'évaluateur détermine, à l'aide de quatre critères que nous avons établis, s'il constitue un nouveau marqueur de modalité. Lors de cette analyse, l'évaluateur est amené à associer une ou plusieurs catégories modales au nouveau marqueur modal identifié ainsi qu'un niveau de positionnement de l'expression de la modalité.

Un lemme sera considéré comme un marqueur de modalité s'il respecte au moins l'un des 4 critères suivants :

- (a) le lemme appartient à la même **famille lexicale** (ou famille de mots) qu'un marqueur de référence. L'évaluateur attribue alors au nouveau marqueur la ou les catégories modales du marqueur de référence. Le (non-)respect de ce critère est identifié à l'aide des connaissances linguistiques de l'évaluateur.

Par exemple, le lemme *admiratif* appartient à la même famille lexicale que le marqueur de modalité appréciative de référence *admirable* : *admiratif* est donc un marqueur de modalité appréciative. La même relation unit le lemme *amour* et le marqueur de modalité appréciative de référence *aimer* : *amour* est donc un marqueur de modalité appréciative ;

- (b) le lemme est dans une relation de **synonymie** avec un marqueur de référence. L'évaluateur attribue alors au nouveau marqueur la ou les catégories modales du marqueur de référence. Cette relation peut être vérifiée dans le dictionnaire en ligne du CNRTL<sup>9</sup> ou dans le Dictionnaire Électronique des Synonymes (DES)<sup>10</sup>.

Par exemple, le lemme *prestigieux* a pour synonyme *admirable* dans le DES<sup>11</sup> : *prestigieux* est donc un marqueur de modalité appréciative. La même relation<sup>12</sup> unit le lemme *déshonorant* et le marqueur de modalité axiologique de référence *honteux* : *déshonorant* est donc un marqueur de modalité axiologique ;

- (c) le lemme appartient au même **champ lexical** qu'un marqueur de référence. L'évaluateur attribue alors au nouveau marqueur la ou les catégories modales du marqueur de référence. Les connaissances linguistiques de l'évaluateur sont sollicitées pour juger de cette relation d'appartenance.

Par exemple le lemme *émerveiller* appartient au même champ lexical que le marqueur *admirable* : *émerveiller* est donc un marqueur de modalité appréciative. La même relation unit le lemme *arme* et le marqueur de modalité axiologique de référence *guerre* : *arme* est donc un marqueur de modalité axiologique ;

- (d) dans sa **définition lexicographique**, le lemme est défini en relation avec un marqueur de référence ou bien à l'aide de notions clés intervenant dans la description des catégories modales, telles qu'elles sont définies par L. Gosselin (cf. section 2). Le dictionnaire de référence choisi pour examiner ce critère est le dictionnaire en ligne du CNRTL. L'évaluateur attribue alors au nouveau marqueur soit la ou les catégories modales du marqueur de référence, soit celles dont les concepts clés ont été identifiés dans la définition lexicographique du lemme.

Par exemple :

- le lemme *miracle* est défini comme un « fait extraordinaire qui porte à l'étonnement et à l'admiration. »<sup>13</sup> De par la présence dans cette définition du terme *admiration*, très fortement lié au marqueur de référence *admirable*, *miracle* est donc considéré comme un marqueur de modalité appréciative ;

---

<sup>9</sup> <https://www.cnrtl.fr/portail/>

<sup>10</sup> <https://crisco4.unicaen.fr/des/>

<sup>11</sup> <https://crisco4.unicaen.fr/des/synonymes/prestigieux>, consulté le 30/09/2023.

<sup>12</sup> <https://crisco4.unicaen.fr/des/synonymes/d%C3%A9shonorant>, consulté le 30/09/2023.

<sup>13</sup> <https://www.cnrtl.fr/definition/miracle>, consulté le 30/09/2023.

- le lemme *cauchemar* est défini comme « Ce qui provoque la peur, la panique, l'horreur, l'aversion.»<sup>14</sup> La notion « d'aversion » présente dans cette définition est mobilisée par L. Gosselin pour définir les modalités appréciatives. *Cauchemar* est donc considéré comme un marqueur de modalité appréciative ;
- le lemme *trésor* est défini, de manière littérale comme un «ensemble de choses de valeur [...] accumulées et souvent soigneusement cachés» et de manière figurée comme « ce qui est tenu pour particulièrement beau, noble, précieux (sur un plan affectif, intellectuel, moral).» Ces deux définitions font référence au caractère désirable d'un trésor (« choses de valeur », « beau », ...) liant ainsi ce lemme à la modalité appréciative, telle qu'elle est définie par L. Gosselin.

Contrairement aux catégories modales, le niveau de positionnement de la modalité (entre intrinsèque, extrinsèque et par inférence) associé à chaque nouveau marqueur n'est pas hérité du marqueur de référence ayant permis l'identification du nouveau marqueur. Ce niveau doit être déterminé par l'évaluateur, après une analyse fondée sur les définitions avancées par L. Gosselin (2010). Par exemple, le marqueur de référence *admirable* et le nouveau marqueur *admiratif* sont tous deux des marqueurs de modalité intrinsèque. En revanche, le marqueur de référence *aimer* est un marqueur de modalité extrinsèque tandis que le nouveau marqueur *amour* est un marqueur de modalité intrinsèque.

Bien qu'offrant une couverture large, les critères (a) à (d) décrits ci-dessus ne permettent pas toujours de classer en marqueur de modalité la totalité des lemmes qui, selon l'intuition de l'évaluateur, paraissent pourtant participer de l'expression d'une catégorie modale. Nous avons donc cherché à extraire de l'ouvrage de Gosselin (2010) d'autres critères permettant de compléter l'analyse. Ces critères reposent notamment sur deux paramètres du modèle théorique de L. Gosselin qui, selon les valeurs qu'ils prennent, interviennent pour distinguer les différentes catégories modales entre elles. Ces deux paramètres sont :

- l'instance de validation - notée I. I a trois valeurs possibles : « réalité », pour les modalités aléthiques ; « subjectivité », pour les modalités appréciatives et bouliques ; et « institution », pour les modalités axiologiques et déontiques. L. Gosselin propose un test linguistique pour identifier la valeur de I : si le lemme à analyser est compatible avec la construction « je trouve que », alors I a pour valeur « subjectivité ». Le lemme exprime donc une modalité épistémique, appréciative ou boulique ;
- la direction d'ajustement - notée D. Deux valeurs de D sont retenues ici : « prospective », c'est-à-dire que la modalité exige que le procès sur lequel elle s'applique soit encore à venir - c'est le cas pour les modalités bouliques et déontiques ; et « non-prospective », c'est-à-dire que la modalité accepte que le procès sur lequel elle porte soit aussi bien advenu, en cours ou à venir. L. Gosselin propose à nouveau un test linguistique pour déterminer le caractère prospectif de la modalité : si le lemme n'est pas compatible avec un procès ayant déjà eu lieu (par ex. « il est venu hier »), alors il exprime une modalité prospective.

Ces critères complémentaires sont cependant encore en cours d'élaboration et nous ne les détaillerons pas ici.

## 4.2 Structuration de la ressource Modale en niveaux d'enrichissement

La ressource Modale de marqueurs de modalité que nous avons constituée a été hiérarchisée en « niveaux d'enrichissement » afin de rendre compte des différentes étapes d'élaboration de cette ressource. Le processeur d'analyse de la modalité de la chaîne de traitement automatique des textes (accessible ici : <https://texttokids.ortolang.fr/semantics/>) mobilise cette hiérarchisation en niveaux pour différencier les descripteurs extraits. Ces niveaux sont les suivants :

- niveau 0 : ce niveau correspond à RM, soit aux 228 marqueurs de référence tirés de Gosselin, (2010) ;
- niveau 1 : ce niveau correspond aux 381 marqueurs ajoutés à partir de RE à l'aide des critères (a) à (d), auxquels au moins une catégorie modale et un niveau de positionnement de la modalité ont été associés ;
- niveau 2 : ce niveau regroupe les 168 marqueurs pour lesquels une analyse partielle à l'aide de critères complémentaires a été proposée ;

<sup>14</sup> <https://www.cnrtl.fr/definition/cauchemar>, consulté le 30/09/2023.

- niveau 3 : ce niveau regroupe 302 marqueurs potentiels de modalité (repérés d'après l'intuition de l'évaluateur) mais qui n'ont pas encore été analysés, que ce soit à l'aide des critères (a) à (d) ou de critères complémentaires. Aucune catégorie modale ni aucun niveau de positionnement de la modalité ne leur ont encore été associés.

Comme nous l'avons déjà mentionné, parmi les 2 174 lemmes émotionnels issus du corpus annoté en émotions décrit en section 3.2., seuls 110 (soit 5,1 %) correspondent à un marqueur de modalité de référence. Nous avons appliqué les critères (a) à (d) sur l'ensemble des lemmes restants afin d'y repérer de nouveaux marqueurs de modalité. À ce jour, ce sont 381 nouveaux marqueurs de modalité (marqueurs de niveau 1) qui ont pu être ainsi identifiés (17,5 % des lemmes émotionnels correspondent donc à un marqueur de modalité de niveau 1).

Le tableau 7 présente les caractéristiques de ces nouveaux marqueurs. Les 4 critères définis section 4.1. ont tous permis l'ajout de marqueurs de modalité dans la ressource, avec une utilisation malgré tout plus prononcée du critère (c) (appartenance au même champ lexical qu'un marqueur de référence). Du point de vue des catégories modales, c'est sans surprise celle de l'appréciatif qui est la plus représentée (cf. proximité conceptuelle entre la catégorie appréciative et la notion d'émotion).

	Prop. parmi les nouveaux marqueurs (%)	
<b>Critère</b>	(a) famille lexicale	13,9
	(b) synonymie	23,1
	(c) champ lexical	35,7
	(d) définition	27,3
<b>Catégorie modale</b>	aléthique	5,0
	appréciatif	66,7
	axiologique	31,0
	boulique	7,9
	déontique	3,4
	épistémique	12,9

Tableau 7 - Caractéristiques des marqueurs de modalité de niveau 1

Le tableau 8 présente les 10 marqueurs de niveau 1 les plus fréquents parmi les lemmes émotionnels du corpus annoté.

Marqueur de niveau 1	Cat. modales	Positionnement de la modalité	Critère	Nb d'occ.
peur	appréciatif, épistémique	intrinsèque	champ lexical de 'craindre'	130
accuser	axiologique	intrinsèque	définition	86
dangereux	appréciatif	intrinsèque	définition	80
critiquer	axiologique	extrinsèque	synonyme de 'désapprouver'	57
menacer	appréciatif, axiologique	intrinsèque	définition (en lien avec 'mal')	46
mourir	appréciatif	inférence	champ lexical de 'souffrance'	45
drôle	appréciatif	intrinsèque	champ lexical de 'amuser'	44
tuer	aléthique, axiologique	intrinsèque	champ lexical de 'assassin'	42
remporter	appréciatif	extrinsèque	synonyme de 'réussir'	38
danger	appréciatif	intrinsèque	définition	35

Tableau 8 - Top 10 des marqueurs de niveau 1 correspondant à un lemme émotionnel

Si l'on considère uniquement les niveaux 0 et 1, soit les niveaux considérés comme les plus sûrs dans notre méthodologie en termes de critères définitoires de la modalité, ce sont donc au moins 609 lemmes modaux (= 228+381) qui sont répertoriés au sein de la ressource ModalE.

## 5. Observations et discussion à propos des zones d'interaction conceptuelle entre modalité et émotion

Afin de cerner plus finement les zones d'affinité entre la catégorie de la modalité et le domaine des émotions, nous revenons sur deux notions du schéma d'annotation des émotions, celle renvoyant à la catégorie émotionnelle et celle renvoyant au mode d'expression d'une émotion. Les analyses présentées ci-après s'appuient sur l'ensemble des lemmes émotionnels correspondant également à un marqueur de modalité de niveau 0 ou 1 (soit

un marqueur de référence ou un marqueur ajouté grâce aux critères (a) à (d)). Elles reposent ainsi sur un ensemble de 491 lemmes, à la fois marqueur d'émotion et marqueur de modalité. Près d'un quart (22,6 %) des lemmes émotionnels correspondent donc à un marqueur de modalité de niveau 0 ou 1.

Notion	Étiquette	Prop. (%)
catégorie émotionnelle	colère	35,4
	peur	30,3
	surprise	12,8
	joie	34,8
	tristesse	33,8
mode d'expression	désigné	31,6
	comportemental	21,0
	suggéré	70,5
	montré	20,2

Tableau 9 - Catégories émotionnelles et des modes d'expression parmi les lemmes émotionnels marqueurs de modalité

Le tableau 9 montre la répartition des cinq catégories émotionnelles les plus fréquentes du corpus et des quatre modes d'expression des émotions au sein de l'ensemble des lemmes émotionnels marqueurs de modalité. On constate que ces lemmes expriment la colère, la peur, la joie et la tristesse dans des proportions comparables (environ 30 %). La surprise est en revanche moins représentée (12,8 %) : cela s'explique car, dans le corpus annoté, cette catégorie émotionnelle est majoritairement exprimée par le mode montré, et plus précisément par des marques typographiques telles que les points d'exclamation ou de suspension. Or le lexique de marqueurs de modalité n'inclut pas de telles marques typographiques. La majorité des lemmes émotionnels et marqueurs de modalité relève du mode suggéré, ce qui n'est pas étonnant puisque ce mode représentait déjà la majorité des lemmes émotionnels du corpus (cf. tableau 4). De même, le mode montré ainsi que le mode comportemental apparaissent ici dans des proportions similaires à celles qu'ils représentent au sein de l'ensemble des lemmes émotionnels du corpus. En revanche, le mode désigné apparaît ici sur-représenté par rapport à la proportion qu'il représente parmi les lemmes émotionnels du corpus.

Mode	Catégorie modale	Lemmes émo. et marqueurs de moda.	
		Nb brut	Prop. (%) parmi le mode
désigné	aléthique	2	1,3
	appréciatif	132	<b>85,2</b>
	axiologique	32	20,6
	boulique	8	5,2
	déontique	0	0,0
	épistémique	29	18,7
comp.	aléthique	10	9,7
	appréciatif	43	<b>41,7</b>
	axiologique	36	35,0
	boulique	19	18,4
	déontique	6	5,8
	épistémique	16	15,5
suggéré	aléthique	31	9,0
	appréciatif	204	<b>59,0</b>
	axiologique	99	28,6
	boulique	28	8,1
	déontique	19	5,5
	épistémique	49	14,2
montré	aléthique	12	12,1
	appréciatif	60	<b>60,6</b>
	axiologique	20	20,2
	boulique	7	7,1
	déontique	6	6,1
	épistémique	20	20,2

Tableau 10 - Croisement entre les notions de mode d'expression et de catégorie modale

Le tableau 10 présente, pour chaque mode d'expression, la part de lemmes à la fois marqueurs émotionnels et marqueurs modaux exprimant chacune des catégories modales reprises dans notre travail (nombre brut et

proportion). Pour les modes désigné, suggéré et montré, les catégories modales les plus fréquentes sont l'appréciatif, l'axiologique et l'épistémique. Si aucune de ces catégories ne semble donc particulièrement liée à un mode d'expression, on remarquera tout de même que l'appréciatif est beaucoup plus représenté dans le mode désigné que dans les autres modes. Par ailleurs, si l'appréciatif et l'axiologique restent les deux catégories modales les plus fréquentes du mode comportemental, l'épistémique y est devancé par le boulique. Il semble donc ressortir ici une affinité entre le boulique et le mode comportemental. Cela est certainement dû au fait que le mode comportemental regroupe de nombreux termes (verbes, noms) dénotant des actions intentionnelles (par ex. *faire grève, manifestation*), concept permettant d'inférer la catégorie du boulique (Gosselin, 2010 : 356).

Mode	Marqueur	Nb d'occ.	Cat. modale	Critère
désigné	peur	127	appréciatif, épistémique	champ lexical de 'craindre'
	dangereux	79	appréciatif	définition
	craindre	37	appréciatif, épistémique	<i>L. Gosselin</i>
comp.	accuser	82	axiologique	définition
	critiquer	55	axiologique	synonyme de 'désapprouver'
	reprocher	28	axiologique	synonyme de 'réprocher'
suggéré	mourir	41	appréciatif	champ lexical de 'souffrance'
	tuer	38	aléthique, axiologique	champ lexical de 'assassin'
	remporter	37	appréciatif	synonyme de 'réussir'
montré	bien	22	appréciatif, axiologique	<i>L. Gosselin</i>
	juste	16	axiologique	<i>L. Gosselin</i>
	lui	16	aléthique, épistémique	<i>L. Gosselin</i>

Tableau 11 - Top 3 des lemmes émotionnels marqueurs de modalité, selon le mode d'expression

Le tableau 11 montre, pour chaque mode d'expression, les trois lemmes marqueurs de modalité les plus fréquents. Si l'on ne retrouve pas l'affinité du comportemental pour le boulique, ce tableau illustre pour les autres modes la prédominance des catégories de l'appréciatif, de l'axiologique et de l'épistémique.

Catégorie émotionnelle	Catégorie modale	Lemmes émo. et marqueurs de moda.	
		Nb brut	Prop. (%) parmi la cat. émo.
colère	aléthique	18	10,3
	appréciatif	72	41,4
	axiologique	73	<b>42,0</b>
	boulique	19	10,9
	déontique	15	8,6
	épistémique	24	13,8
peur	aléthique	19	12,8
	appréciatif	79	<b>53,0</b>
	axiologique	41	27,5
	boulique	15	10,1
	déontique	12	8,1
	épistémique	38	25,5
surprise	aléthique	11	17,5
	appréciatif	30	<b>47,6</b>
	axiologique	12	19,0
	boulique	5	7,9
	déontique	2	3,2
	épistémique	20	31,7
joie	aléthique	11	6,4
	appréciatif	127	<b>74,3</b>
	axiologique	17	9,9
	boulique	16	9,4
	déontique	6	3,5
	épistémique	26	15,2
tristesse	aléthique	21	12,7
	appréciatif	98	<b>59,0</b>
	axiologique	36	21,7
	boulique	16	9,6
	déontique	9	5,4
	épistémique	29	17,5

Tableau 12 - Croisement entre les notions de catégorie émotionnelle et de catégorie modale

Le tableau 12 présente, pour chacune des cinq catégories émotionnelles les plus fréquentes du corpus annoté en émotions, la part de lemmes à la fois émotionnels et marqueurs de modalité exprimant chacune des catégories modales reprises dans notre étude (nombre brut et proportion). Les catégories modales de l'appréciatif, de l'axiologique et de l'épistémique sont à nouveau les plus représentées, et ce pour chaque catégorie émotionnelle considérée. Cependant, seule la colère a pour catégorie modale plus fréquente l'axiologique. Il semble donc y avoir une (légère) affinité entre ces deux catégories. Une hypothèse serait que quelque chose qui contrevient aux valeurs morales est particulièrement susceptible de susciter la colère. On note en outre que l'appréciatif est particulièrement représenté au sein des lemmes exprimant la joie (74,3 % contre maximum 59,0 % pour les autres catégories émotionnelles). L'épistémique est pour sa part plus représenté parmi les lemmes exprimant la peur et la surprise (respectivement 25,5 % et 31,7 %, contre maximum 17,5 % pour les autres catégories émotionnelles). Une hypothèse serait que ce qui est incertain est susceptible de susciter la peur et que ce qui est incertain / inattendu est susceptible de susciter la surprise.

Le tableau 13 montre, pour chacune des cinq catégories émotionnelles les plus fréquentes du corpus annoté en émotions, les trois lemmes marqueurs de modalité les plus fréquents. On retrouve dans ce tableau l'association entre les catégories colère et axiologique d'une part, et entre les catégories joie et appréciatif d'autre part. De même, on trouve des marqueurs exprimant de l'épistémique uniquement pour la peur et la surprise.

Cat. émo.	Marqueur	Nb d'occ.	Cat. modale	Critère
colère	accuser	84	axiologique	définition
	critiquer	57	axiologique	synonyme de 'désapprouver'
	reprocher	28	axiologique	synonymie de 'réprouver'
peur	peur	129	appréciatif, épistémique	champ lexical de 'craindre'
	dangereux	80	appréciatif	définition
	menacer	43	appréciatif, axiologique	définition (en lien avec 'mal')
surprise	drôle	26	appréciatif	champ lexical de 'amuser'
	jamais	11	aléthique	<i>L. Gosselin</i>
	incroyable	9	appréciatif, épistémique	famille lexicale de 'croire'
joie	heureux	35	appréciatif	<i>L. Gosselin</i>
	fête	32	appréciatif	<i>L. Gosselin</i>
	content	32	appréciatif	synonyme de 'heureux'
tristesse	mal	33	appréciatif, axiologique	<i>L. Gosselin</i>
	triste	28	appréciatif	champ lexical de 'souffrance'
	mourir	24	appréciatif	champ lexical de 'souffrance'

Tableau 13 - Top 3 des lemmes émotionnels marqueurs de modalité, selon la catégorie émotionnelle

## 6. Bilan

Le positionnement théorique qui sous-tend l'ensemble de notre démarche est double. Dans le champ de la linguistique des émotions (e.g. Chatar Moumni, 2013 ; Micheli, 2014), il s'insère dans l'examen des différents modes d'expression des émotions en se penchant sur la manière dont une catégorie sémantique comme celle de la modalité participe de ces modes d'expression. Cela permet alors un regard différent sur les émotions de celui habituellement opéré, cette fois en mobilisant une catégorie sémantique, alors qu'il est surtout d'ordre lexicosyntaxique généralement (e.g. Gross, 1995 ; Leeman, 1995, Diwersy et al., 2014). Dans le champ de la linguistique énonciative (e.g. Bally, 1932 ; Culioli, 1968), il participe à la description des « modalités sentimentales » (Brunot, 1922 : 541) dont la place, parmi un ensemble de catégories modales, a été conceptualisée dans ce champ.

Les liens entre modalité et émotion sont manifestes dans plusieurs travaux de linguistique sans que les termes « modalité » ou « émotion » ne soient d'ailleurs forcément explicitement mentionnés. Parmi les marqueurs de modalité cités dans Gosselin (2010), certains impliquent clairement une valeur modale et une valeur émotionnelle (par ex. « heureux » pour la modalité appréciative et l'émotion de joie ; « honteux » pour la modalité axiologique et l'émotion d'embarras). Par ailleurs, trois critères intervenant dans Micheli (2014) pour définir les émotions étayées (le degré de probabilité des conséquences d'une situation, le degré de maîtrise de cette situation et la signification normative de cette situation) font écho aux modalités épistémique – voire déontique – et axiologique.

Nous tentons dans notre démarche d'aller au bout de l'éclairage d'une catégorie par une autre. La ressource lexicale de la modalité que nous avons présentée ici en constitue une étape. Elle comporte actuellement 1079 lemmes modaux - dont 609 pour les deux premiers niveaux de description qui sont les plus aboutis - et est testable en ligne à l'adresse [<https://texttokids.ortolang.fr/semantics/>], onglet Modalité. En partant de l'ensemble des lemmes émotionnels du corpus annoté en émotions décrit dans Etienne et al. (2022), plusieurs observations peuvent être faites à propos des liens conceptuels entre modalité et émotion :

- L'expression d'une émotion est très liée à l'expression d'une modalité puisque près d'un quart (22,6 %) des lemmes émotionnels du corpus annoté correspondent à un marqueur de modalité du niveau 0 ou 1 ;
- Tous modes d'expression et toutes catégories émotionnelles confondus, ce sont principalement les catégories de l'appréciatif, de l'axiologique et de l'épistémique qui sont mobilisées dans l'expression des émotions ;
- Quelques affinités plus spécifiques semblent se dégager ; il y aurait ainsi :

- une affinité entre le mode comportemental pour l'expression d'une émotion et la catégorie modale du boulique (actions intentionnelles permettant d'inférer la catégorie boulique) ;
- une affinité entre la catégorie émotionnelle colère et la catégorie modale axiologique (importance des valeurs morales pour l'expression de la colère) ;
- une affinité entre joie et appréciatif ;
- une affinité entre les catégories émotionnelles peur et surprise et la catégorie modale épistémique (caractère plus ou moins certain/incertain, attendu/inattendu lié à la peur et à la surprise).

Ces affinités seront explorées prochainement plus avant à partir d'un autre corpus qui sera d'une part annoté automatiquement en émotions grâce au processeur présenté dans (Etienne, 2023 ; Etienne et al., 2024) et d'autre part soumis au processeur ModalE pour l'extraction des modalités.

## Références bibliographiques

- BALLY Charles, *Linguistique générale et linguistique française*, Paris, Leroux, 1932.
- BATTISTELLI Delphine, ÉTIENNE Aline, RAHMAN Rashedur, TEISSEDRE Charles & LECORVE Gwénoilé, « Une chaîne de traitements pour appréhender la complexité des textes pour enfants d'un point de vue linguistique et psycho-linguistique ». In Actes de la 29<sup>e</sup> Conférence sur le Traitement Automatique des Langues Naturelles. Volume 1 : conférence principale, Avignon, France, 2022, p. 236–246.
- BENVENISTE Emile, « De la subjectivité dans le langage », in *Problèmes de linguistique générale*, 1, Paris, Gallimard, 1966, p. 259-260.
- BLANC Nathalie, « La compréhension des contes entre 5 et 7 ans : quelle représentation des informations émotionnelles ? » *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale* 2010/64 (4), p. 256-265.
- BRUNOT Ferdinand, *La pensée et la langue*. Paris, Masson 1922.
- CHATAR-MOUMNI Nizha, « L'expression verbale des émotions : Présentation ». *Langue française*, 2013/180(4), p. 3-11
- CULIOLI Antoine, « La formalisation en linguistique », *Cahiers pour l'analyse* 9, 1968, p. 106-117.
- CULIOLI Antoine, « Pour une linguistique de l'énonciation, Opérations et représentations, Gap, Ophrys, Collection L'homme dans la langue animée par Janine Bouscaren (HDL), Tome 1 (PLE 1), 1990.
- DIWERSY Sasha, GOOSSENS Vannina, GRUTSCHUS Anke, KERN Beate, KRAIF Olivier, MELNIKOVA Elena & NOVAKOVA Iva, « Traitement des lexies d'émotion dans les corpus et les applications d'EmoBase ». *Corpus*, 2014/13, p. 269-293.
- ÉTIENNE Aline, *Analyse automatique des émotions dans les textes : contributions théoriques et applicatives dans le cadre de l'étude de la complexité des textes pour enfants* (Université de Nanterre - Paris X). Thèse de doctorat. 2023
- ÉTIENNE Aline & BATTISTELLI Delphine, *Annotation manuelle des émotions dans des textes écrits avec la plateforme Glozz*. Rapport de recherche. MoDyCo ; Université Paris Nanterre. <https://hal.archives-ouvertes.fr/hal-03263194>, 2021
- ÉTIENNE Aline, BATTISTELLI Delphine, & LECORVÉ Gwénoilé, « A (Psycho-)Linguistically Motivated Scheme for Annotating and Exploring Emotions in a Genre-Diverse Corpus». in *Proceedings of the 13th Conference on Language Resources and Evaluation (LREC)*. 2022. h
- ÉTIENNE Aline, BATTISTELLI Delphine, & LECORVÉ Gwénoilé, « Emotion Identification for French in Written Texts: Considering Modes of Emotion Expression as a Step Towards Text Complexity Analysis». In *Proceedings of the 14th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis*, p. 168–185, Bangkok, Thailand. Association for Computational Linguistics. 2024
- GALATANU Olga, « L'énonciation : l'espace-temps de modalisation et d'argumentation », communication orale, communication orale, 36<sup>ème</sup> colloque international du Cerlico, 23-24 mai 2024, Nantes. 2024
- GOSELIN Laurent, *Les modalités en français : La validation des représentations*. Brill. 2010.
- GOSELIN Laurent, « Les modalités appréciatives et axiologiques. Sémantique des jugements de valeur ». *Cahiers de Lexicologie : La sémantique en France : Etat des lieux (II)*, dir. J.-Cl. Anscombre, n°111, p. 97-119. 2019
- GROSS Maurice, « Une grammaire locale de l'expression des sentiments ». *Langue française*, 1995/105(1), p. 70-87.
- KERBRAT-ORECCHIONI Catherine, « Quelle place pour les émotions dans la linguistique du XX<sup>e</sup> siècle ». Plantin, C., Doury, M. & Traverso, V.(éds), *Les émotions dans les interactions*. Lyon: Presses Universitaires de Lyon, p. 33-74. 2000

- LEEMAN-BOUIX Danielle, « Pourquoi peut-on dire Max est en colère mais non \*Max est en peur ? Hypothèses sur la construction être en N. » *Langue française*, 1995/105(1), p. 55-69.
- LE QUERLER Nicole, *Typologie des modalités*, Caen, Presses Universitaires de Caen, 1996.
- MATHIEU Yvette Yannick, « Annotation of emotions and feelings in texts », in ACII'05: Proceedings of the First international conference on Affective Computing and Intelligent Interaction, p. 350 – 357. 2005
- MEUNIER André, « Modalités et communication ». *Langue française*, n°21, 1974. Communication et analyse syntaxique. p. 8-25.
- MICHELI Raphaël, *Les émotions dans les discours. Modèle d'analyse, perspectives empiriques*. De Boeck Supérieur. 2014.
- NUYTS Jan, « Modality: Overview and linguistic issues », in W. Frawley, (dir.), *The expression of modality*, Berlin / New York, Mouton de Gruyter, p. 1-26. 2006.
- PIOLAT Annie & BANNOUR Rachid, « EMOTAIX : Un Scénario de Tropes pour l'identification automatisée du lexique émotionnel et affectif », *L'Année Psychologique*, 109, p.657-700. 2009.
- PLANTIN Christian, « Une méthode d'approche de l'émotion dans le discours et les interactions », SHS Web of Conferences. 81. 01001. 10.1051/shsconf/20208101001. 2020.
- QI Peng, ZHANG Yuaho., ZHANG Yuhui, BOLTON Jason, & MANNING Christopher. D, « Stanza : A Python Natural Language Processing Toolkit for Many Human Languages ». in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. 2020.
- WIDLÖCHER Antoine & MATHET Yann, « The glozz platform: A corpus annotation and mining tool». In *Proceedings of the 2012 ACM symposium on Document engineering*, p. 171-180. 2012.
- WILSON Theresa & WIEBE Janyce, « Annotating Opinions in the World Press ». In *Proceedings of the Fourth SIGdial Workshop of Discourse and Dialogue*, p. 13–22. 2003.