

Annotation et modélisation des émotions dans un corpus textuel : une approche évaluative

Jonas Noblet

(1) LIDILEM, Université Grenoble Alpes, Bâtiment Stendhal, CS40700, 38058 Grenoble cedex 9, France

(2) Société Ixiade, 11 Rue Aimé Béréy, 38000 Grenoble, France

jonas.noblet@univ-grenoble-alpes.fr

RÉSUMÉ

L'émotion est un phénomène capital dans le fonctionnement de l'être humain en société. Elle reste pourtant un sujet encore largement ouvert, notamment dans ses manifestations textuelles. La présente communication examine un corpus industriel manuellement annoté selon une approche évaluative de l'émotion. Cette conception théorique aujourd'hui peu exploitée propose une perspective différente, en complément des approches traditionnelles. Partant du constat que les annotations que nous avons collectées présentent un fort désaccord, nous avons émis l'hypothèse que celles-ci suivent néanmoins des tendances statistiques stables. Par le biais de modèles de langue entraînés sur ces annotations, nous montrons qu'il est possible de modéliser le processus d'étiquetage, et que la variabilité est guidée par des caractéristiques linguistiques sous-jacentes. Réciproquement, nos résultats indiquent que les modèles de langue semblent en mesure de distinguer les situations émotionnelles sur la base des critères évaluatifs.

ABSTRACT

Annotation and modeling of emotions in a textual corpus : an evaluative approach

Emotion is a crucial phenomenon in the functioning of human beings in society. However, it remains a widely open subject, particularly in its textual manifestations. This paper examines an industrial corpus manually annotated following an evaluative approach to emotion. This theoretical framework, which is currently underutilized, offers a different perspective that complements traditional approaches. Noting that the annotations we collected exhibit significant disagreement, we hypothesized that they nonetheless follow stable statistical trends. Using language models trained on these annotations, we demonstrate that it is possible to model the labeling process and that variability is driven by underlying linguistic features. Conversely, our results indicate that language models seem capable of distinguishing emotional situations based on evaluative criteria.

MOTS-CLÉS : Annotation, Émotion, Accord inter-juges, Modèles de langue.

KEYWORDS: Annotation, Emotion, Inter-rater agreement, Language models.

1 Introduction

Comme le montre [Damasio \(1995\)](#), l'émotion joue un rôle central dans la cognition humaine : elle est un composant essentiel au fonctionnement de l'être humain en société. L'émotion interagit avec l'attention, la mémoire, le raisonnement et la prise de décision (voir [Lemaire \(2021\)](#), [Houwer & Hermans \(2010\)](#), [Fox *et al.* \(2018\)](#) pour des revues). En même temps que se sont développées les

sciences cognitives à partir de la deuxième moitié du XX^e siècle, l'étude de l'émotion a elle aussi crû en importance, recueillant plus de financements et générant plus de publications (Dukes *et al.*, 2021).

La présente communication est le fruit d'une collaboration entre le laboratoire LIDILEM et la société Ixiade, dans le cadre de la thèse CIFRE menée par Noblet (2026). Les recherches dont nous présentons les résultats s'appuient sur un corpus issu des projets¹ industriels menés dans le cadre de l'activité économique de l'entreprise. Ces projets ont pour visée d'établir l'acceptabilité² de concepts innovants en sondant l'opinion à leur sujet. Malgré la nature spécifique de ce corpus, nous proposons une réflexion en mesure de contribuer plus largement à la compréhension des émotions et de leur traitement dans le texte.

La suite de l'article se divise en trois parties. La première introduit des éléments de contexte pour notre étude qui orientent son propos. La seconde présente les données – corpus et annotation – sur lesquelles se base la présente contribution. Enfin, la dernière détaille les dernières analyses que nous avons conduites, notamment à l'aide d'un modèle de langue fine-tuné.

2 Éléments de contexte

L'émotion, un champ disciplinaire encore débattu

Depuis son développement jusqu'à aujourd'hui, une théorie majeure domine le champ d'étude des émotions : la théorie des émotions de base (*basic emotion theory*) développée par Paul Ekman (Ekman, 1984, 1992 ; Ekman & Cordaro, 2011). Celle-ci, conçue dans le prolongement des travaux de Darwin (1890) et du mouvement behavioriste, postule l'existence d'un ensemble d'émotions élémentaires, biologiquement évoluées. Selon cette conception, les émotions prennent la forme de catégories clairement délimitées et universelles, communes à tous les êtres humains. Une version qui s'est beaucoup diffusée propose six émotions de base : la colère, le dégoût, la peur, la joie, la tristesse et la surprise (Ekman *et al.*, 1987).

Peu après son apparition, cette théorie fait l'objet de plusieurs critiques. Ce sont à la fois les méthodes expérimentales et les prémices – l'émotion comme phénomène universel divisé en catégories distinctes – qui sont remises en question (Russell, 1994 ; Barrett, 2018 ; Siegel *et al.*, 2018). De là naît l'approche constructiviste : celle-ci affirme que l'émotion est le produit d'un apprentissage en lien avec un contexte socioculturel qui prend appui sur deux dimensions fondamentales, la valence et l'excitation (*arousal*), qui structurent les états internes. La valence spécifie si une expérience subjective est positive ou négative et l'activation définit le degré de stimulation physiologique. Ce paradigme est aussi qualifié de *dimensionnel*, en opposition aux catégories discrètes d'Ekman.

En parallèle de ces deux perspectives sur l'émotion, une troisième a été développée : l'approche évaluative. Celle-ci énonce que ce ne sont pas les événements *per se*, mais l'interprétation que chacun en fait, qui est à la source des émotions. Selon les théories évaluatives, un stimulus – un tremblement de terre, une silhouette dans le noir, une notification de rendez-vous manqué – est évalué et produit une réponse adaptée, conscientisée comme la peur ou sous une autre forme. Dans ce cadre, les émotions

1. Nous réservons le terme *projet* pour désigner spécifiquement les études de test de concepts – bien ou service – innovants conduites par Ixiade ; chaque projet est associé à un concept innovant, porté par un client donné (Noblet, 2025).

2. La notion d'acceptabilité désigne l'usage potentiel d'une innovation telle que l'individu ciblé se la représente, avant les premières manipulations (Terrade *et al.*, 2009).

sont la représentation subjective d'un état synchronisé du corps et de la pensée, en relation forte avec une situation donnée. Cette vision de l'émotion n'est pas incompatible avec les deux précédentes. C'est moins tant par sa perspective théorique que par l'accent mis sur la perception et l'interprétation du contexte qu'elle se distingue des deux précédentes.

Tour d'horizon des corpus annotés

Les émotions sont un processus qui se manifeste sous de nombreuses modalités corporelles, cognitives et expressives. D'abord étudiées à partir des expressions du visage, elles ont par la suite été examinées dans l'expression langagière. Nous passons en revue quelques-uns des corpus annotés qui ont contribué à guider la réflexion théorique à ce sujet, en les mettant en relation avec les théories psychologiques que nous venons de décrire brièvement. Le langage lui-même peut prendre différentes formes ; nous prenons le parti, en lien avec notre sujet, de ne considérer que les bases textuelles en écartant les corpus oraux et multimodaux.

En ouverture de cette section, nous avons présenté le modèle des émotions élémentaires d'[Ekman et al. \(1987\)](#), l'un des plus influents dans le domaine de l'étude des émotions. De fait, plusieurs corpus s'en inspirent au début des années 2000 pour définir leur schéma d'annotation, comme le relèvent [Bostan & Klinger \(2018\)](#) dans leur revue. À cet égard, nous pouvons citer les travaux d'[Alm et al. \(2005\)](#) et [Strapparava & Mihalcea \(2007\)](#) qui reprennent directement les six catégories initialement avancées par Ekman. Ce modèle de l'émotion reste par la suite une inspiration majeure, employé directement ([Li et al., 2017](#)) ou avec quelques ajustements ([Mohammad, 2012](#) ; [Liu et al., 2019](#))³. S'agissant de l'approche dimensionnelle, elle a également été appliquée aux données textuelles, quoique dans des proportions plus réduites ([Preoțiuc-Pietro et al., 2016](#) ; [Buechel & Hahn, 2017](#)). Nous présentons dans le Tableau 1 les corpus les plus importants identifiés par [Bostan & Klinger \(2018\)](#) (taille > 10k exemples) auxquels nous en avons ajouté d'autres, francophones ou postérieurs à la revue.

Ces dernières années, les travaux qui visent à identifier l'émotion tendent à employer des paradigmes d'annotation plus élaborés qu'auparavant. Annoté par contribution collective, le corpus CrowdFlower ([Van Pelt & Sorokin, 2012](#)) s'appuie sur 16 étiquettes non standards qui reprennent et élargissent les catégories d'Ekman. Le corpus GoEmotions, l'un des plus massifs à ce jour, adopte quant à lui un jeu de 28 étiquettes pour caractériser l'émotion dans des commentaires Reddit. Il fait suite aux travaux de [Cowen & Keltner \(2021 ; 2020 ; Cowen et al., 2021\)](#) qui, par le biais de nouveaux outils statistiques et notamment les réseaux de neurones, réimaginent la théorie des émotions de base de Paul Ekman ([Keltner et al., 2019](#)). Privilégiant une approche différente, [Troiano et al. \(2022\)](#) fondent le schéma d'annotation du corpus x-enVENT sur la théorie évaluative. En plus de 15 étiquettes pour représenter les états affectifs, ils en emploient 22 pour caractériser le processus d'évaluation du contexte émotionnel. Ces 22 dimensions ont chacune été évaluées sur une échelle de Likert de 1 à 5. La contrepartie d'une telle procédure est la taille du corpus, puisque seul un millier de phrases environ a pu être étiqueté.

Dans l'espace francophone, le volume de ressources textuelles annotées est sensiblement plus faible : nous trouvons principalement des bases lexicales comme les corpus EMA ([Gobin et al., 2017](#)) et FANCat ([Syssau et al., 2021](#)). Un projet qui se distingue par son envergure est celui porté par [Étienne \(2023\)](#), sous la direction de D. Battistelli et G. Lecorvé. Dans le cadre de cette thèse a été compilé un ensemble d'articles de presse pour enfants qui ont ensuite été annotés librement selon 10 catégories

3. En particulier, certains corpus font appel à une variante du modèle d'Ekman développée par [Plutchik \(1980, 2001\)](#).

Corpus	Granularité	# Annotation	Taille	Source
<i>Corpus anglophones</i>				
Tales	phrases	6 C + V	15 302	Alm et al. (2005)
CrowdFlower	tweets	14 C	40 000	Van Pelt & Sorokin (2012)
TEC	tweets	8 C	21 051	Mohammad (2012)
DailyDialog	conversation	6 C	13 118	Li et al. (2017)
EmoBank	phrases	V + A + D	10 548	Buechel & Hahn (2017)
DENS	extraits litt.	8 C + N	9 710	Liu et al. (2019)
GoEmotions	posts Reddit	27 C + N	54 263	Demszky et al. (2020)
XED	sous-titres	8 C + N + V	23 940	Öhman et al. (2020)
x-enVENT	phrases	15 C + contexte	1 001	Troiano et al. (2022)
<i>Corpus francophones</i>				
EMA	mots	6 C + V + A	1 286	Gobin et al. (2017)
FANCat	mots	10 C + V + A	1 031	Syssau et al. (2021)
TextToKids-Émotions	textes	10 C + contexte	1 594	Étienne (2024)

TABLE 1 – Sélection de ressources pour l’analyse des émotions. Les annotations font référence aux schémas suivants : [*n* C] nombre de catégories discrètes ; [N] catégorie neutre ; [V] valence ; [A] excitation ; [D] dominance ; [contexte] annotation détaillée du contexte d’énonciation.

de l’émotion, mais également selon le type d’entités en jeu et selon une typologie développée par [Blanc & Quenette \(2017\)](#) et [Micheli \(2014\)](#).

Au regard de ce bref tour d’horizon, il nous semble clair que le champ d’étude de l’expression de l’émotion dans le texte est encore très ouvert. Tant d’un point de vue théorique – nous avons commencé par montrer que le concept même d’émotion était sujet à débat, sans même mentionner les travaux qui traitent de l’aspect purement linguistique de la question – que du point de vue des ressources disponibles, limitées pour la langue française. C’est dans ce contexte que nous présentons ci-après un nouveau corpus annoté, basé sur une ressource textuelle et une procédure d’étiquetage originales.

3 Données expérimentales

Un corpus d’opinions face aux innovations

Comme présenté en introduction, les données linguistiques que nous étudions ont été produites dans un environnement industriel. Celles-ci sont issues de la plate-forme communautaire Yoomaneo⁴, développée en 2020 par Ixiade ([Masson & Parmentier, 2023](#)). Il s’agit d’une application web qui vise à mesurer l’acceptabilité des concepts innovants en les soumettant au jugement de sa base d’utilisateurs. Malgré une visée motivée par des considérations économiques, la variété des innovations testées justifie un intérêt plus large d’ordre scientifique.

En pratique, tout projet débute par un questionnaire d’éligibilité, qui permet de s’assurer que le profil des participants correspond à celui ciblé. Si l’innovation examinée est un nouveau tableau électrique, ce sont des électriciens qui sont recherchés. Les critères de sélection peuvent avoir trait au métier, à

4. <https://webapp.yoomaneo.com/>

l'âge, à la catégorie socioprofessionnelle, etc. Suite à cela, l'innovation testée est exposée via une courte vidéo et une série de questions ouvertes est posée. Celles-ci suivent un protocole standard, réadapté en fonction des concepts et des interrogations spécifiques des clients porteurs des projets. Les questions explorent la perception spontanée, la compréhension du concept, de son usage et de ses fonctions, l'adéquation aux valeurs et aux structures sociales. Toutes demandent aux participants de rédiger un message – un *post* – pour répondre. Lorsque le projet est terminé les participants sont dédommagés par une carte cadeau dont le montant varie selon l'implication et la nature du projet.

Le corpus ainsi formé regroupe 2446 posts, correspondant à 22 concepts testés, produits par 238 membres de la plate-forme distincts. Pour l'annotation selon le protocole décrit ci-après, les posts sont divisés en verbatims⁵, formant un ensemble total de 4980 verbatims. Le corpus fait actuellement l'objet d'une procédure d'anonymisation en vue de sa publication.

La composante évaluative recontextualisée

Pour analyser l'émotion dans les productions textuelles, nous nous appuyons sur le modèle théorique EMINOSA. Celui-ci résulte des travaux de recherches réalisés par le laboratoire LIP/PC2S⁶ en partenariat avec Ixiade, menés successivement par [Sbai \(2013\)](#), [Dupré \(2016\)](#), et [Loeser \(2019\)](#). Dans ses fondements, il est inspiré du modèle multi-componentiel de l'émotion développé par [Scherer et al. \(2001\)](#)⁷. Ce dernier divise l'émotion en cinq composantes qui interagissent entre elles : évaluative, conative, expressive, physiologique et subjective. Outre la composante évaluative et la composante subjective qui caractérise les états internes, les trois autres prennent en charge le mouvement du corps⁸. Ce modèle donne un rôle prépondérant à la composante évaluative, aussi qualifiée de *cognitive* en raison de ses liens avec d'autres mécanismes mentaux comme l'attention, la mémoire ou le raisonnement.

Dans le cadre de nos travaux, nous nous sommes nous-même intéressé plus particulièrement à la composante évaluative. Tout comme le modèle de [Scherer et al.](#), EMINOSA subdivise cette composante – que nous nommons E_{cog} – en plusieurs dimensions :

- la Familiarité – le concept innovant testé évoque des choses connues et, par extension, son fonctionnement est bien compris ;
- l'Aggréabilité – indépendamment de l'Utilité, le concept est plaisant / attractif⁹ ;
- l'Utilité – le concept remplit une fonction ou améliore l'existant ;
- la Légitimité – Le concept est en accord avec les valeurs personnelles / professionnelles.

Là où le modèle multi-componentiel et les autres modèles évaluatifs envisagent la réaction émotionnelle pour un stimulus indifférencié, EMINOSA a été développée spécifiquement pour caractériser les réactions face aux concepts innovants. C'est ainsi, par exemple, que la dimension « adéquation aux buts en cours » (*goal relevance*), telle que proposée par [Scherer et al.](#), prend la forme de l'Utilité. Par ailleurs, le modèle EMINOSA a été testé pour lui-même lors de plusieurs expériences indépendantes ([Loeser, 2019](#)).

5. Segment de texte, généralement de la taille d'une phrase.

6. Laboratoire Inter-universitaire de Psychologie. Personnalité, Cognition, Changement Social.

7. D'autres travaux ont également servi de référence pour EMINOSA, dont ceux de [Frijda \(1986\)](#) et [Roseman \(2013\)](#).

8. La composante conative intègre la projection au mouvement – l'action telle qu'elle est conçue en amont ; la composante motrice les mouvements involontaires (rythme cardiaque, sudation) et la composante expressive les mouvements volontaires (expression faciale, posture).

9. Cette définition inclut les propriétés ergonomiques et la facilité / difficulté d'utilisation.

Annotation de la composante évaluative dans le texte

Nous décrivons ci-après le processus d’annotation des opinions selon les dimensions de la composante E_{cog} . Celui-ci a fait l’objet d’une communication antérieure, dans laquelle il est développé plus en détails en même temps que des analyses préliminaires (Noblet, 2025). Comme pour d’autres corpus, la démarche d’annotation textuelle est motivée par la problématique de mieux expliquer et référencer les affects – dans notre cas envers des concepts innovants.

Pour identifier la manifestation des dimensions E_{cog} dans notre corpus, nous avons recouru à des annotateurs humains. Pour les quatre dimensions, les annotateurs ont attribué une valeur à chaque verbatim :

- +1 si la dimension est présente positivement (pour l’Utilité par exemple, l’évaluateur exprime que le concept est utile) ;
- -1 si la dimension est présente négativement (le concept est inutile) ;
- 0 si la dimension est absente ou présente de manière ambiguë.

Pour l’Exemple 1, un annotateur aura tendance à associer la valeur +1 à l’Agréabilité (intuitif, facile d’utilisation), -1 à l’Utilité (ne résout pas le problème du poids), et 0 pour la Familiarité et la Légitimité.

- (1) *Cela semble très intuitif, facile d’utilisation mais cela ne résout pas le problème du poids.*

Quatre études pilotes et une campagne d’annotation à large échelle ont été menées. Pour cette dernière, 6 annotateurs ont été recrutés qui ont tous traité l’ensemble des verbatims de 21 des 22 projets¹⁰. La plupart des corpus annotés s’appuient sur un nombre réduit d’annotateurs ; en mobiliser six permet de mieux appréhender la variabilité des pratiques d’annotation. En incluant les études pilotes, 4873 verbatims ont été annotés donnant lieu à 32 732 annotations.

4 Analyses et résultats

Statistiques générales

Préalablement à des analyses plus fines, nous présentons quelques traitements statistiques qui donnent un aperçu général du corpus annoté. En première observation, nous constatons que les annotations E_{cog} sont fortement déséquilibrées. Comme le montre la Figure 1, la part de verbatims annotés positivement et négativement varie grandement en fonction des dimensions, projets et annotateurs considérés. De plus, les annotations négatives sont très peu fréquentes : seuls 3% des verbatims environ sont annotés négativement, toutes dimensions confondues, comme l’indique la Figure 1b. Remarquons également que la dimension de Familiarité dans son ensemble est faiblement mise en évidence. Inversement, l’Utilité domine les annotations, se manifestant deux à trois fois plus que les autres dimensions. Cette dernière tendance semble appuyer le modèle évaluatif récemment développé par Moors (2017) : cette chercheuse soutient que l’utilité attendue des actions disponibles occupe une place centrale dans la réaction émotionnelle.

10. Le projet écarté est l’un des premiers ayant été déployé sur la plate-forme, selon des modalités notablement différentes de ceux qui ont suivi.

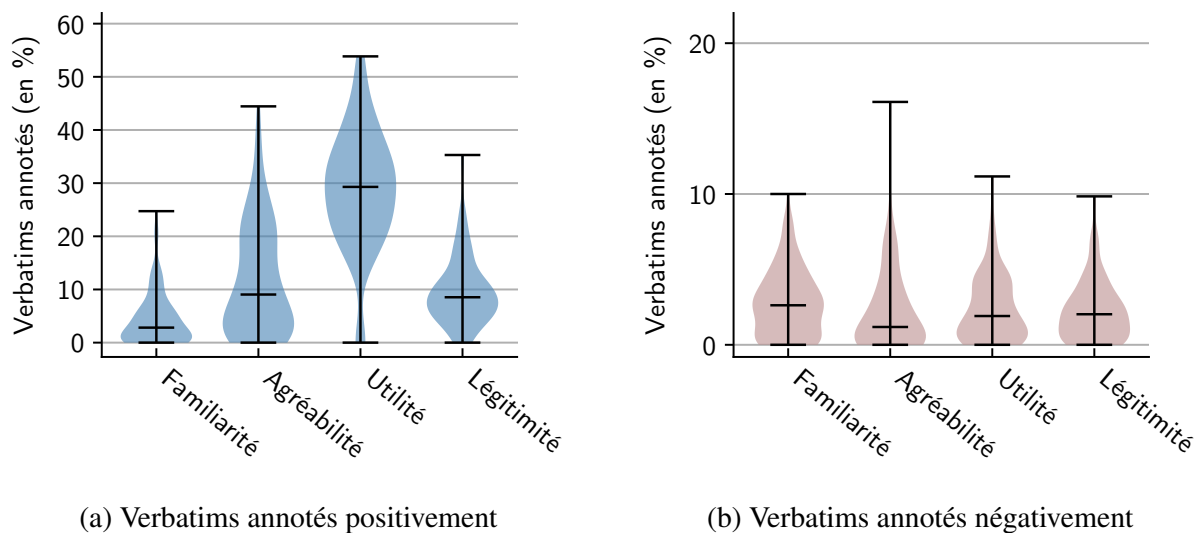


FIGURE 1 – Proportion des verbatims annotés, par annotateur et par projet.

Une autre spécificité de notre corpus est la grande part d'exemples qui ne sont pas annotés – qu'aucun annotateur n'a étiquetés positivement ou négativement, pour une catégorie donnée – s'élevant à 45% des verbatims pour l'Utilité et jusqu'à 76% pour la Familiarité. Les autres corpus que nous avons référencés appliquent généralement une étape de sélection des exemples pertinents pour l'analyse des émotions (Demszky *et al.*, 2020 ; Mohammad, 2012 ; Étienne, 2024). Nous avons préféré maintenir notre corpus en l'état : identifier quels verbatims doivent ou non être annotés est une partie intégrante de la problématique de la reconnaissance des émotions dans le texte. Nous privilégions ainsi le critère écologique – des données pleinement représentatives de leur contexte d'énonciation – à celui du contraste – des dimensions bien démarquées.

En outre, nous avons relevé un fort désaccord entre les annotations. Le degré d'accord pour chaque dimension mesuré à l'aide de l'alpha de Krippendorff (Krippendorff, 2011) est représenté dans la première ligne du Tableau 2. Notons toutefois que ce désaccord n'est pas total. Si nous ignorons les annotations nulles et que nous jugeons seulement l'accord entre annotations positives et négatives (deuxième ligne du Tableau 2), l'accord est très bon ($\alpha \geq 0,8$). À l'inverse, si nous ignorons la polarité et que nous tenons uniquement compte de la présence ou de l'absence des dimensions pour un verbatim donné (troisième ligne du Tableau 2), nous observons un fort désaccord, reflétant le score global.

Modalité	Familiarité	Agréabilité	Utilité	Légitimité
Global	0,29	0,31	0,53	0,38
Polarité	0,80	0,93	0,91	0,92
Pertinence	0,30	0,28	0,51	0,36

TABLE 2 – Accord inter-annotateurs pour les dimensions d'E_{cog}, pour l'ensemble des annotations, suivant différentes modalités de mesure.

Après avoir examiné les données en surface, lors d'une communication précédente (Noblet, 2023), nous avons émis l'hypothèse qu'il existe un modèle probabiliste sous-jacent qui gouverne les pratiques d'annotation ; les annotations E_{cog}, bien qu'aléatoires, suivraient des tendances statistiques explicables.

Des traits linguistiques, comme un lexique explicitement associé aux dimensions ¹¹ ou un fort degré de confiance apparent semble aller de pair avec une annotation unanime. Schématiquement, plus une dimension est marquée avec intensité dans un verbatim donné, plus il sera annoté. Le Tableau 3 reproduit des exemples que nous avons donné à l'appui de cette hypothèse, annoté à différentes proportions pour la dimension de Légitimité.

Légitimité	Exemples
1,00 (6/6)	<i>Cette solution s'appuie sur une approche "objet connecté", clairement dans l'air du temps.</i>
0,83 (5/6)	<i>De plus, si ce produit peut se décliner pour corriger nos postures au quotidien même au travail, cela répond au souci de préservation de sa santé sur le long-terme.</i>
0,67 (4/6)	<i>On est en plein dans les objets connectés du futur, reliés complètement à nos smartphones (ce n'est pas un mal)</i>
0,50 (3/6)	<i>IoT, bien être, sport à la maison</i>
0,33 (2/6)	<i>Préventif : permet de soigner mais aussi d'anticiper les douleurs</i>
0,17 (1/6)	<i>Repond a un besoin, propose une solution complete, suscite l'interet.</i>

TABLE 3 – Exemples correspondant à différents niveaux de Légitimité moyenne positive

Modélisation des annotations par réseaux Transformers

Pour éprouver l'hypothèse selon laquelle les dimensions E_{cog} sont annotées en suivant un gradient stable, nous avons affiné, sur la base des annotations moyennées, plusieurs modèles de type Transformer pré-entraînés (Vaswani *et al.*, 2017). Si ces modèles sont en mesure de retrouver les proportions annotées sur des corpus test, cela attesterait de la présence d'indices linguistiques qui motiveraient la variation dans les jugements.

À cette fin, nous nous sommes basés sur les modèles Flaubert (flaubert-base-cased, 138 M de paramètres) (Le *et al.*, 2020) et GTE (gte-multilingual-base, 305 M de paramètres) (Zhang *et al.*, 2024), développé par la société Alibaba. Ce dernier a été choisi en raison de ses bonnes performances au test d'évaluation MTEB (Ciancone *et al.*, 2024). À partir de ces modèles, nous avons conçu trois architectures :

- (1) modèle Flaubert gelé connecté à un classifieur logistique ;
- (2) modèle GTE gelé connecté à un classifieur logistique ;
- (3) modèle GTE gelé à l'exception de la dernière couche, connecté à un classifieur logistique.

Les modèles ont été entraînés séparément sur chaque dimension en suivant un protocole de validation croisée à l'échelle des projets – 4 projets ont servi successivement de tests tandis que les autres ont contribué à l'entraînement. Nous représentons les performances d'annotation automatique sous la forme de matrices de confusion – la configuration (3) est illustrée en Figures 2, les configurations (1) et (2) sont données en annexe (Figures 3, 4).

Remarquons de prime abord que le modèle Flaubert seul n'est aucunement en mesure de distinguer les différentes valeurs d'annotation, quelle que soit la dimension considérée ¹². À l'inverse, le modèle

11. En se référant à la typologie de Étienne *et al.* (2020), l'appellation dimensions *désignées* pourrait être employée .

12. Comme le modèle GTE, il est toutefois possible de l'affiner pour trouver des performances correctes.

GTE, même sans entraînement supplémentaire, est en mesure de produire des représentations qui permettent déjà de discriminer grossièrement des gradients d’annotation. Nous pouvons en effet constater que les annotations produites suivent la diagonale des matrices de confusion. À l’exception de la Familiarité, le modèle a néanmoins tendance à beaucoup délaissé les étiquettes négatives (cadres supérieurs des matrices) au profit de la classe majoritaire, l’annotation nulle.

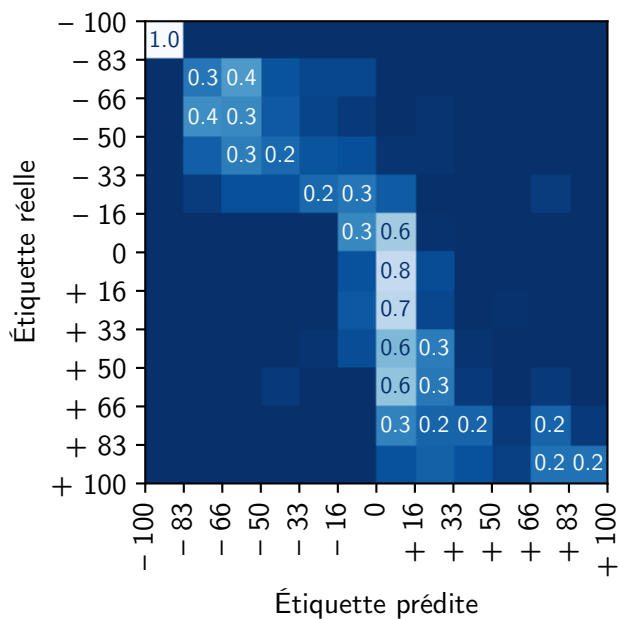
Les performances les plus élevées sont obtenues en donnant un « coup de pouce » au modèle GTE, en entraînant sa dernière couche. Dans cette configuration, nous constatons que les étiquettes produites par le modèle suivent plutôt fidèlement la diagonale, ce qui indique que le modèle parvient à retrouver les proportions d’annotation (voir Figure 2). En particulier, pour des verbatims annotés positivement en forte proportion (extrémité inférieure droite), l’annotation nulle est correctement abandonnée par le modèle au profit d’une annotation marquée positivement (pour l’utilité, les prédictions du modèle sont supérieures à 0,33 en majorité, voir Figure 2c).

Le Tableau 4 présente une estimation des performances du modèle GTE affiné (configuration 3), en délimitant les valeurs par des seuils arbitraires. Bien que cette estimation ne permette pas de rendre compte avec justesse des résultats obtenus car nous présumons une échelle continue, elle offre un aperçu synthétique plus directement interprétable. Comme nous l’avons précédemment remarqué, à l’exception de la Familiarité, toutes les autres dimensions sont faiblement annotées par le modèle automatique pour la polarité négative ; la précision est bonne mais le rappel très faible (inférieur à 0,25). La dimension de Familiarité se distingue des trois autres : elle est mieux annotée pour la polarité négative que pour la polarité positive. Cette tendance s’explique par des verbatims annotés en plus grand nombre pour la polarité négative. Nous retrouvons ici la spécificité du corpus, constitué d’opinions au sujet de concepts innovants souvent peu familiers. En dernier lieu, notons que la Légitimité est la dimension la plus problématique – malgré une bonne précision, le rappel est faible pour les étiquettes positives (49,7) et très faible pour les étiquettes négatives (10,5). Ce rappel peu élevé illustre là encore la propension du modèle à adopter l’étiquette nulle majoritaire dans le corpus, visible à travers la bande verticale en Figure 2d.

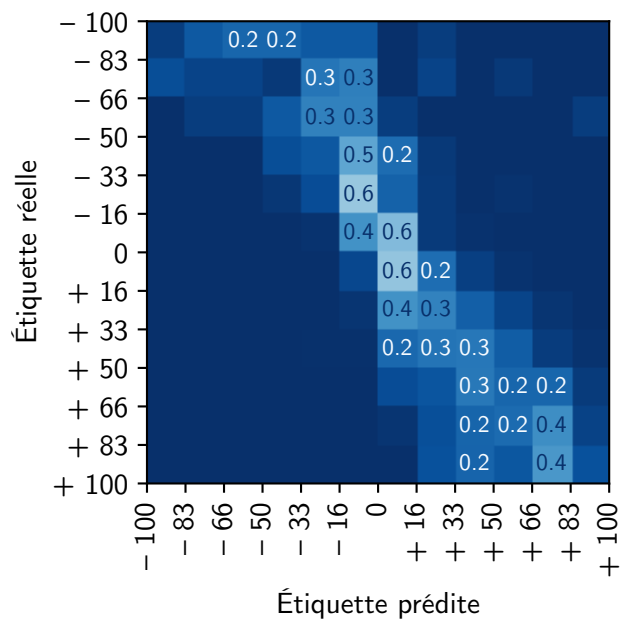
Ces résultats, bien qu’encourageants, sont encore trop approximatifs pour constituer une conclusion définitive à nos recherches. Nous avons entrepris des réflexions pour quantifier avec justesse les sorties des systèmes neuronaux. Parallèlement, un autre axe de travail porte sur l’exploration approfondie des critères influençant les décisions des modèles.

Métrique	Familiarité	Agréabilité	Utilité	Légitimité
Précision globale	95, 2	85.1	80.0	89.1
Précision +1	71, 4	78.6	82.3	77.4
Rappel +1	21, 6	55.2	71.8	49.7
Précision -1	80, 5	80.9	76.2	48.8
Rappel -1	69, 8	24.0	20.3	10.5
Précision 0	95, 9	86.2	78.7	90.4
Rappel 0	99, 1	95.7	89.2	97.6

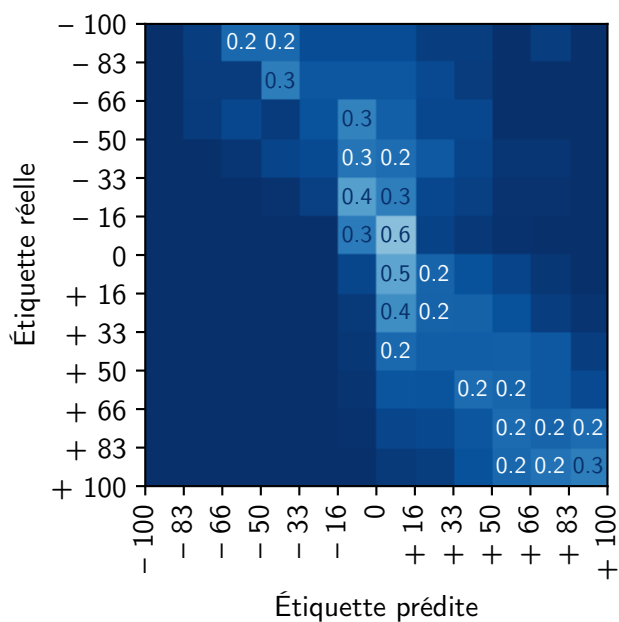
TABLE 4 – Modèle GTE affiné – Scores de précision et rappel, pour chaque dimension et valeur d’ E_{cog} , avec des seuils à -1/3 et +1/3 pour segmenter les valeurs.



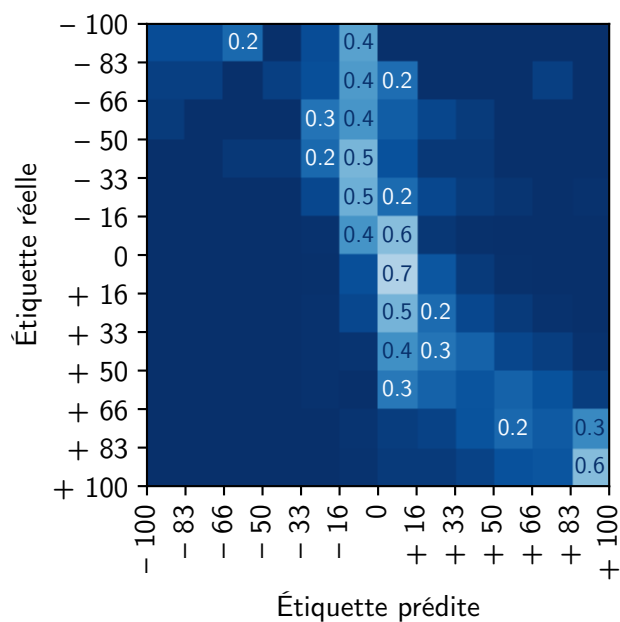
(a) Familiarité



(b) Agréabilité



(c) Utilité



(d) Légitimité



FIGURE 2 – Modèle GTE – dernière couche entraînée

5 Conclusion

Cet article propose plusieurs contributions dans le champ de l’annotation et de la modélisation des émotions. En premier lieu, le jeu d’étiquettes sur lequel nous nous appuyons est fondé sur une conception de l’émotion aujourd’hui peu exploitée, l’approche évaluative. Cette approche, qui complète les conceptions traditionnelles des émotions, permet de mieux prendre en compte la relation au stimulus – dans notre cas un concept innovant. De plus, en adoptant une démarche sans pré-filtrage et en recourant à six annotateurs pour traiter les mêmes unités, nous accédons à une large variabilité dans les jugements. Nous émettons l’hypothèse que cette variation observée est le reflet d’un gradient d’intensité dans l’expression de l’émotion : plus une dimension de l’émotion se manifeste clairement, plus elle aura tendance à être annotée et inversement. Pour tester cette hypothèse, nous entreprenons de modéliser le processus d’annotation à l’aide de modèles de langues affinés. Ces derniers montrent des résultats probants, indiquant à la fois que la variabilité des annotations est gouvernée par des tendances stables et que les modèles de langue sont en mesure de porter une conception évaluative des émotions.

Notre corpus présente néanmoins plusieurs limitations. Il repose exclusivement sur des discours écrits en réaction à un stimulus distant – une vidéo de concept innovant. Ce mode d’interaction peut constituer une barrière à l’engagement émotionnel, notamment lorsque les enjeux perçus sont faibles. En outre, bien que les objets physiques et les services représentent une part significative de notre environnement, d’autres sources potentielles d’émotion restent exclues. En particulier, ne sont pas pris en compte les rapports interpersonnels, qui constituent un aspect fondamental de l’émotion.

Nos travaux futurs viseront à approfondir la validation de notre hypothèse. L’opacité des réseaux de neurones constitue un enjeu majeur, rendant nécessaire le développement de méthodes algorithmiques permettant d’analyser les choix effectués par ces modèles et d’identifier les structures linguistiques sous-jacentes. En complément, nous envisageons un retour aux méthodologies linguistiques afin d’affiner l’analyse, en apportant une précision explicative que les outils statistiques ne permettent pas d’atteindre.

Références

- ALM C. O., ROTH D. & SPROAT R. (2005). Emotions from text : Machine learning for text-based emotion prediction. In R. MOONEY, C. BREW, L.-F. CHIEN & K. KIRCHHOFF, Éd., *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, p. 579–586, Vancouver, British Columbia, Canada : Association for Computational Linguistics.
- BARRETT L. F. (2018). *How emotions are made : The secret life of the brain*. Expert Thinking Series. London, England : Pan Books. original-date : 2017.
- BLANC N. & QUENETTE G. (2017). La production d’inférences émotionnelles entre 8 et 10 ans : quelle méthodologie pour quels résultats ? *Enfance*, 4(4), 503–511. Publisher : NecPlus Section : Sociologie, DOI : [10.3917/enf1.174.0503](https://doi.org/10.3917/enf1.174.0503).
- BOSTAN L.-A.-M. & KLINGER R. (2018). An analysis of annotated corpora for emotion classification in text. In E. M. BENDER, L. DERCZYNSKI & P. ISABELLE, Éd., *Proceedings of the 27th International Conference on Computational Linguistics*, p. 2104–2119, Santa Fe, New Mexico, USA : Association for Computational Linguistics.

- BUECHEL S. & HAHN U. (2017). EmoBank : Studying the impact of annotation perspective and representation format on dimensional emotion analysis. In M. LAPATA, P. BLUNSOM & A. KOLLER, Édts., *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics : Volume 2, Short Papers*, p. 578–585, Valencia, Spain : Association for Computational Linguistics.
- CIANCONE M., KERBOUA I., SCHAEFFER M. & SIBLINI W. (2024). MTEB-French : Resources for French sentence embedding evaluation and analysis. arXiv :2405.20468 [cs] version : 2, DOI : [10.48550/arXiv.2405.20468](https://doi.org/10.48550/arXiv.2405.20468).
- COWEN A. S. & KELTNER D. (2020). What the face displays : Mapping 28 emotions conveyed by naturalistic expression. *American Psychologist*, **75**(3), 349–364. Publisher : American Psychological Association, DOI : [10.1037/amp0000488](https://doi.org/10.1037/amp0000488).
- COWEN A. S. & KELTNER D. (2021). Semantic space theory : A computational approach to emotion. *Trends in Cognitive Sciences*, **25**(2), 124–136. Publisher : Elsevier, DOI : [10.1016/j.tics.2020.11.004](https://doi.org/10.1016/j.tics.2020.11.004).
- COWEN A. S., KELTNER D., SCHROFF F., JOU B., ADAM H. & PRASAD G. (2021). Sixteen facial expressions occur in similar contexts worldwide. *Nature*, **589**(7841), 251–257. Publisher : Nature Publishing Group, DOI : [10.1038/s41586-020-3037-7](https://doi.org/10.1038/s41586-020-3037-7).
- DAMASIO A. R. (1995). *L'erreur de Descartes : la raison des émotions*. Odile Jacob poches Sciences. Paris : Odile Jacob.
- DARWIN C. (1890). *The expression of the emotions in man and animals*. Cambridge, England : Cambridge University Press, 2 édition.
- DEMSZKY D., MOVSHOVITZ-ATTIAS D., KO J., COWEN A., NEMADE G. & RAVI S. (2020). GoEmotions : A dataset of fine-grained emotions. In D. JURAFSKY, J. CHAI, N. SCHLUTER & J. TETREAULT, Édts., *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, p. 4040–4054, Online : Association for Computational Linguistics. DOI : [10.18653/v1/2020.acl-main.372](https://doi.org/10.18653/v1/2020.acl-main.372).
- DUKES D., ABRAMS K., ADOLPHS R., AHMED M. E., BEATTY A., BERRIDGE K. C., BROOM-HALL S., BROSC T., CAMPOS J. J., CLAY Z., CLÉMENT F., CUNNINGHAM W. A., DAMASIO A., DAMASIO H., D'ARMS J., DAVIDSON J. W., DE GELDER B., DEONNA J., DE SOUSA R., EKMAN P., ELLSWORTH P. C., FEHR E., FISCHER A., FOOLEN A., FREVERT U., GRANDJEAN D., GRATCH J., GREENBERG L., GREENSPAN P., GROSS J. J., HALPERIN E., KAPPAS A., KELTNER D., KNUTSON B., KONSTAN D., KRET M. E., LEDOUX J. E., LERNER J. S., LEVENSON R. W., LOEWENSTEIN G., MANSTEAD A. S. R., MARONEY T. A., MOORS A., NIEDENTHAL P., PARKINSON B., PAVLIDIS I., PELACHAUD C., POLLAK S. D., POURTOIS G., ROETTGER-ROESSLER B., RUSSELL J. A., SAUTER D., SCARANTINO A., SCHERER K. R., STEARNS P., STETS J. E., TAPPOLET C., TERONI F., TSAI J., TURNER J., REEKUM C. V., VUILLEUMIER P., WHARTON T. & SANDER D. (2021). The rise of affectivism. *Nature Human Behaviour*, **5**(7), 816–820. Publisher : Nature Publishing Group, DOI : [10.1038/s41562-021-01130-8](https://doi.org/10.1038/s41562-021-01130-8).
- DUPRÉ D. (2016). *L'influence de produits innovants sur l'émotion des utilisateurs : une approche multi-componentielle*. These de doctorat, Université Grenoble Alpes (ComUE).
- EKMAN P. (1984). Expression and the nature of emotion. In K. R. SCHERER & P. EKMAN, Édts., *Approaches to emotion*, p. 319–344. New York : Psychology Press.
- EKMAN P. (1992). An argument for basic emotions. *Cognition and Emotion*, **6**(3-4), 169–200. Publisher : Routledge, DOI : [10.1080/02699939208411068](https://doi.org/10.1080/02699939208411068).
- EKMAN P. & CORDARO D. (2011). What is meant by calling emotions basic. *Emotion Review*, **3**(4), 364–370. Publisher : SAGE Publications, DOI : [10.1177/1754073911410740](https://doi.org/10.1177/1754073911410740).

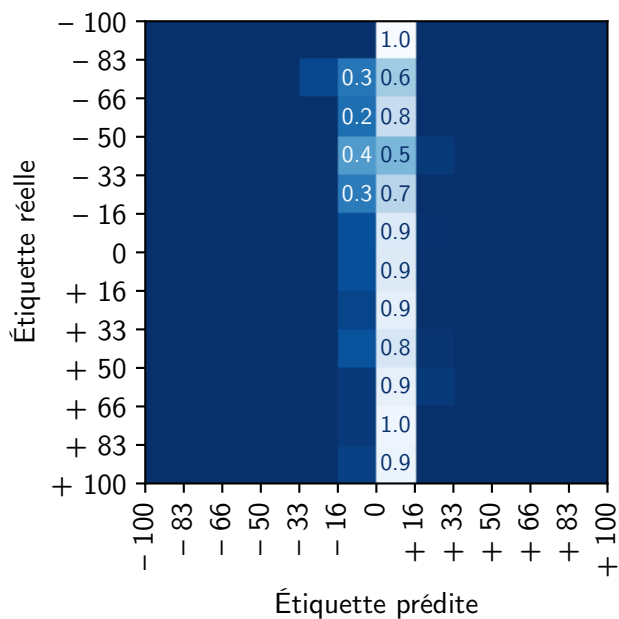
- EKMAN P., FRIESEN W. V., O'SULLIVAN M., CHAN A., DIACOYANNI-TARLATZIS I., HEIDER K., KRAUSE R., LECOMPTE W. A., PITCAIRN T., RICCI-BITTI P. E., SCHERER K., TOMITA M. & TZAVARAS A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology*, **53**(4), 712–717. Publisher : American Psychological Association, DOI : [10.1037/0022-3514.53.4.712](https://doi.org/10.1037/0022-3514.53.4.712).
- FOX A. S., LAPATE R. C., SHACKMAN A. J. & DAVIDSON R. J., Éd. (2018). *The nature of emotion : Fundamental questions*. Series in Affective Science. New York, NY, US : Oxford University Press, 2 édition.
- FRIJDA N. H. (1986). *The emotions*. The emotions. Paris, France : Editions de la Maison des Sciences de l'Homme. Pages : xii, 544.
- GOBIN P., CAMBLATS A. M., FAUROS W. & MATHEY S. (2017). Une base de l'émotionalité (valence, arousal, catégories) de 1286 mots français selon l'âge (EMA). *European Review of Applied Psychology*, **67**(1), 25–42. DOI : [10.1016/j.erap.2016.12.001](https://doi.org/10.1016/j.erap.2016.12.001).
- HOUWER J. D. & HERMANS D., Éd. (2010). *Cognition and emotion : Reviews of current research and theories*. London : Psychology Press. DOI : [10.4324/9780203853054](https://doi.org/10.4324/9780203853054).
- KELTNER D., SAUTER D., TRACY J. & COWEN A. (2019). Emotional expression : Advances in basic emotion theory. *Journal of Nonverbal Behavior*, **43**(2), 133–160. DOI : [10.1007/s10919-019-00293-3](https://doi.org/10.1007/s10919-019-00293-3).
- KRIPPENDORFF K. (2011). Computing Krippendorff's Alpha-Reliability. *Departmental Papers (ASC)*.
- LE H., VIAL L., FREJ J., SEGONNE V., COAVOUX M., LECOUTEUX B., ALLAUZEN A., CRABBÉ B., BESACIER L. & SCHWAB D. (2020). FlauBERT : Unsupervised language model pre-training for French. arXiv :1912.05372 [cs], DOI : [10.48550/arXiv.1912.05372](https://doi.org/10.48550/arXiv.1912.05372).
- LEMAIRE P. (2021). *Émotion et cognition*. Ouvertures psychologiques LMD. Louvain-la-Neuve : De Boeck supérieur.
- LI Y., SU H., SHEN X., LI W., CAO Z. & NIU S. (2017). DailyDialog : A manually labelled multi-turn dialogue dataset. In G. KONDRAK & T. WATANABE, Éd., *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1 : Long Papers)*, p. 986–995, Taipei, Taiwan : Asian Federation of Natural Language Processing.
- LIU C., OSAMA M. & DE ANDRADE A. (2019). DENS : A dataset for multi-class emotion analysis. In K. INUI, J. JIANG, V. NG & X. WAN, Éd., *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, p. 6293–6298, Hong Kong, China : Association for Computational Linguistics. DOI : [10.18653/v1/D19-1656](https://doi.org/10.18653/v1/D19-1656).
- LOESER F. (2019). *Modélisation probabiliste de l'influence des émotions sur l'acceptabilité des innovations*. Thèse de doctorat, Université Grenoble Alpes (ComUE).
- MASSON Z. & PARMENTIER G. (2023). Innover avec une communauté en ligne : les cas de Yoomaneo et d'EDF Pulse and You. In E. RUIZ, Éd., *Les capacités créatives des organisations ouvertes*, volume 2, p. 163–191. ISTE Editions.
- MICHELI R. (2014). *Les émotions dans les discours. Modèle d'analyse, perspectives empiriques*. De Boeck Supérieur. ISSN : 1374-089X, DOI : [10.3917/dbu.mchel.2014.01](https://doi.org/10.3917/dbu.mchel.2014.01).
- MOHAMMAD S. (2012). #Emotional tweets. In E. AGIRRE, J. BOS, M. DIAB, S. MANANDHAR, Y. MARTON & D. YURET, Éd., **SEM 2012 : The First Joint Conference on Lexical and Computational Semantics – Volume 1 : Proceedings of the main conference and the shared task, and Volume 2 : Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*, p. 246–255, Montréal, Canada : Association for Computational Linguistics.

- MOORS A. (2017). The integrated theory of emotional behavior follows a radically goal-directed approach. *Psychological Inquiry*, **28**(1), 68–75. Publisher : Routledge _eprint : <https://doi.org/10.1080/1047840X.2017.1275207>, DOI : [10.1080/1047840X.2017.1275207](https://doi.org/10.1080/1047840X.2017.1275207).
- NOBLET J. (2023). Mesurer l'accord inter-juge avec l'Alpha de Krippendorff : une étude des fonctions de différence. Publisher : Laboratoire LIDILEM (UGA) Published : Journées Linguistique de Corpus.
- NOBLET J. (2025). Le bruit dans la mesure de la composante cognitive de l'émotion pour l'évaluation de l'acceptabilité des innovations. *Corpus*, **26**. Publisher : Bases, corpus et langage - UMR 6039, DOI : [10.4000/1364u](https://doi.org/10.4000/1364u).
- NOBLET J. (2026). *Outils d'analyse sémantique pour l'autocodage qualitatif de corpus d'opinions d'utilisateurs selon des critères psycho-cognitifs*. Thèse de doctorat, Université Grenoble Alpes. Notes : À paraître.
- PLUTCHIK R. (1980). Chapter 1 - A general psychoevolutionary theory of emotion. In R. PLUTCHIK & H. KELLERMAN, Édts., *Theories of emotion*, p. 3–33. Academic Press. DOI : [10.1016/B978-0-12-558701-3.50007-7](https://doi.org/10.1016/B978-0-12-558701-3.50007-7).
- PLUTCHIK R. (2001). The nature of emotions : Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*, **89**(4), 344–350. Publisher : Sigma Xi, The Scientific Research Society.
- PREOȚIUC-PIETRO D., SCHWARTZ H. A., PARK G., EICHSTAEDT J., KERN M., UNGAR L. & SHULMAN E. (2016). Modelling valence and arousal in Facebook posts. In A. BALAHUR, E. VAN DER GOOT, P. VOSSEN & A. MONTOYO, Édts., *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, p. 9–15, San Diego, California : Association for Computational Linguistics. DOI : [10.18653/v1/W16-0404](https://doi.org/10.18653/v1/W16-0404).
- ROSEMAN I. J. (2013). Appraisal in the emotion system : Coherence in strategies for coping. *Emotion Review*, **5**(2), 141–149. Éditeur : SAGE Publications, DOI : [10.1177/1754073912469591](https://doi.org/10.1177/1754073912469591).
- RUSSELL J. A. (1994). Is there universal recognition of emotion from facial expression ? A review of the cross-cultural studies. *Psychological Bulletin*, **115**(1), 102–141. DOI : [10.1037/0033-2909.115.1.102](https://doi.org/10.1037/0033-2909.115.1.102).
- SBAI N. (2013). *The influence of specific emotions on consumer judgment and behavioural intention with respect to innovations*. These de doctorat, Grenoble.
- SCHERER K. R., SCHORR A. & JOHNSTONE T. (2001). *Appraisal processes in emotion : Theory, methods, research*. Oxford University Press.
- SIEGEL E. H., SANDS M. K., VAN DEN NOORTGATE W., CONDON P., CHANG Y., DY J., QUIGLEY K. S. & BARRETT L. F. (2018). Emotion fingerprints or emotion populations ? A meta-analytic investigation of autonomic features of emotion categories. *Psychological Bulletin*, **144**(4), 343–393. Publisher : American Psychological Association, DOI : [10.1037/bul0000128](https://doi.org/10.1037/bul0000128).
- STRAPPARAVA C. & MIHALCEA R. (2007). SemEval-2007 task 14 : Affective text. In E. AGIRRE, L. MÀRQUEZ & R. WICENTOWSKI, Édts., *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, p. 70–74, Prague, Czech Republic : Association for Computational Linguistics.
- SYSSAU A., YAKHLOUFI A., GIUDICELLI E., MONNIER C. & ANDERS R. (2021). FANCat : French affective norms for ten emotional categories. *Behavior Research Methods*, **53**(1), 447–465. DOI : [10.3758/s13428-020-01450-z](https://doi.org/10.3758/s13428-020-01450-z).
- TERRADE F., PASQUIER H., REERINCK-BOULANGER J., GUINGOUAIN G. & SOMAT A. (2009). L'acceptabilité sociale : la prise en compte des déterminants sociaux dans l'analyse de l'acceptabilité des systèmes technologiques. *Le travail humain*, **72**(4), 383–395. DOI : [10.3917/th.724.0383](https://doi.org/10.3917/th.724.0383).

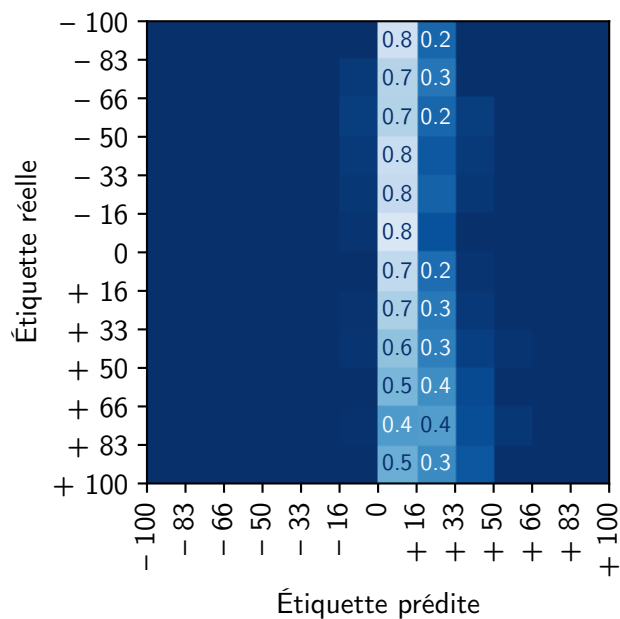
- TROIANO E., OBERLAENDER L. A. M., WEGGE M. & KLINGER R. (2022). x-enVENT : A corpus of event descriptions with experiencer-specific emotion and appraisal annotations. In N. CALZOLARI, F. BÉCHET, P. BLACHE, K. CHOUKRI, C. CIERI, T. DECLERCK, S. GOGGI, H. ISAHARA, B. MAEGAARD, J. MARIANI, H. MAZO, J. ODIJK & S. PIPERIDIS, Édts., *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, p. 1365–1375, Marseille, France : European Language Resources Association.
- VAN PELT C. & SOROKIN A. (2012). Designing a scalable crowdsourcing platform. In *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*, SIGMOD '12, p. 765–766, New York, NY, USA : Association for Computing Machinery. DOI : [10.1145/2213836.2213951](https://doi.org/10.1145/2213836.2213951).
- VASWANI A., SHAZEER N., PARMAR N., USZKOREIT J., JONES L., GOMEZ A. N., KAISER L. & POLOSUKHIN I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30 : Curran Associates, Inc.
- ZHANG X., ZHANG Y., LONG D., XIE W., DAI Z., TANG J., LIN H., YANG B., XIE P., HUANG F., ZHANG M., LI W. & ZHANG M. (2024). mGTE : Generalized long-context text representation and reranking models for multilingual text retrieval. DOI : [10.48550/arXiv.2407.19669](https://doi.org/10.48550/arXiv.2407.19669).
- ÉTIENNE A. (2023). *Analyse automatique des émotions dans les textes : contributions théoriques et applicatives dans le cadre de l'étude de la complexité des textes pour enfants*. phdthesis, Université de Nanterre - Paris X.
- ÉTIENNE A. (2024). Corpus TextToKids-Émotions.
- ÉTIENNE A., BATTISTELLI D. & LECORVÉ G. (2020). L'expression des émotions dans les textes pour enfants : constitution d'un corpus annoté. In C. BENZITOUN, C. BRAUD, L. HUBER, D. LANGLOIS, S. OUNI, S. POGODALLA & S. SCHNEIDER, Édts., *Actes de la 6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Volume 2 : Traitement Automatique des Langues Naturelles*, p. 205–212, Nancy, France : ATALA.
- ÖHMAN E., PÀMIES M., KAJAVA K. & TIEDEMANN J. (2020). XED : A multilingual dataset for sentiment analysis and emotion detection. In D. SCOTT, N. BEL & C. ZONG, Édts., *Proceedings of the 28th International Conference on Computational Linguistics*, p. 6542–6552, Barcelona, Spain (Online) : International Committee on Computational Linguistics. DOI : [10.18653/v1/2020.coling-main.575](https://doi.org/10.18653/v1/2020.coling-main.575).

Annexes

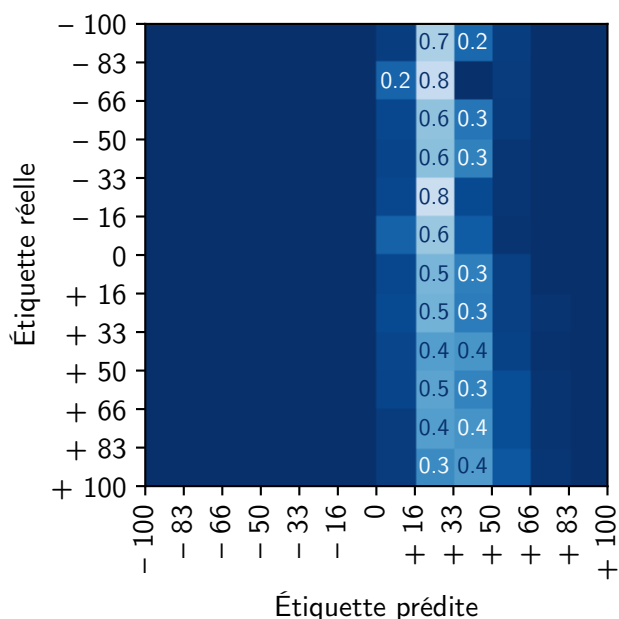
Nous présentons ici des figures complémentaires illustrant les performances des modèles Flaubert gelé et GTE gelé (voir Section 4).



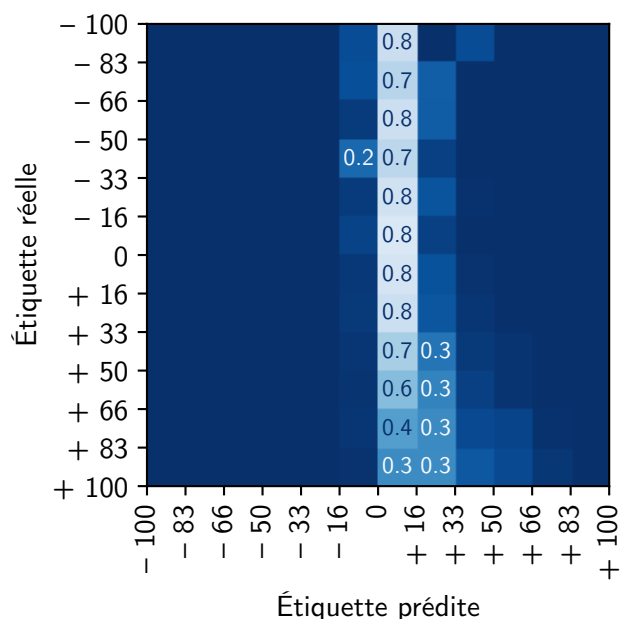
(a) Familiarité



(b) Agréabilité



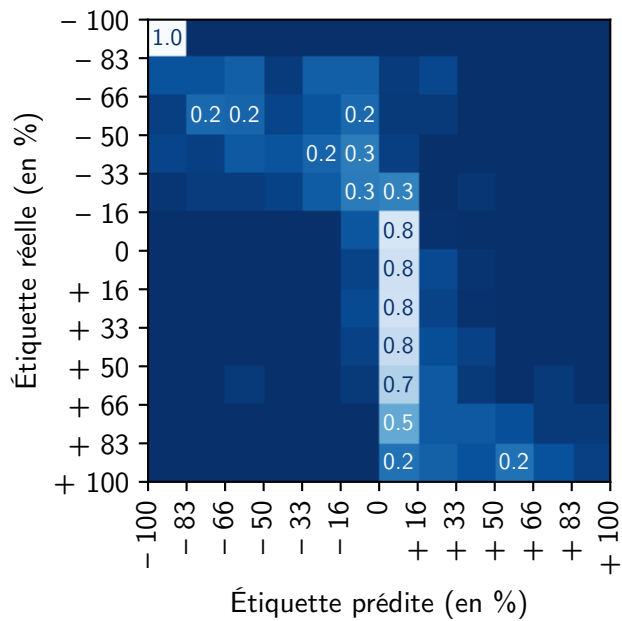
(c) Utilité



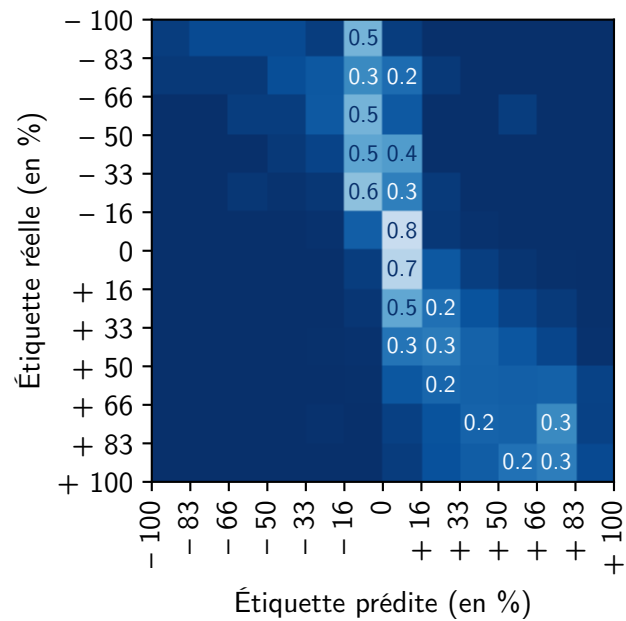
(d) Légitimité



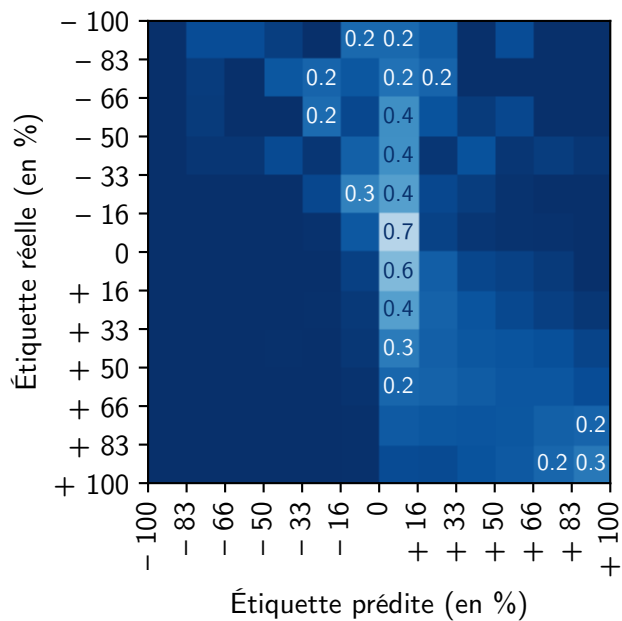
FIGURE 3 – Modèle Flaubert – modèle gelé



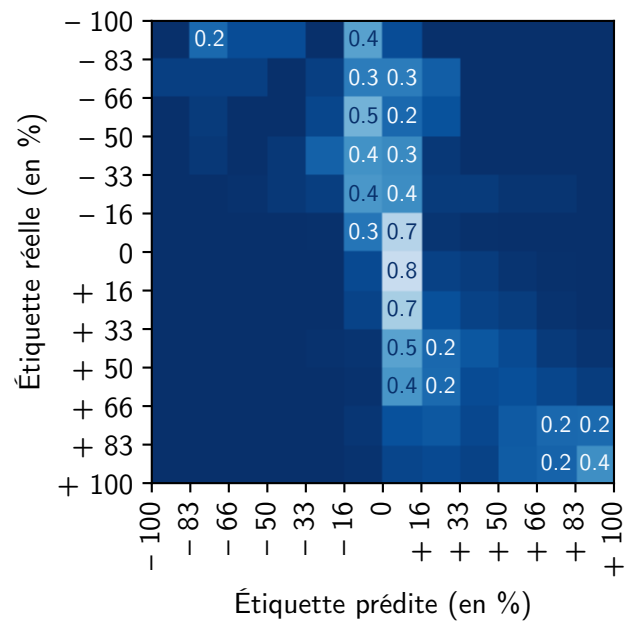
(a) Familiarité



(b) Agréabilité



(c) Utilité



(d) Légitimité



FIGURE 4 – Modèle GTE – modèle gelé