Letters and Languages Guide Volume: 5 Issue: 1 / 2026 /pp.140-153

Entre mots et machines : vers une modélisation automatique des émotions dans la langue

Between Words and Machines: Toward an Automatic Modeling of Emotions in Language

SOKRAF SAHAR
Docteur à la Faculté des Lettres et des Sciences Humaines de Sousse (Tunisie)
Sokraf.sahar@yahoo.com

Reçu: 27/03/2026	Accepté : 14/06/2026	Publié : 30/06/2026

Résumé

L'étude des émotions dans le langage écrit est centrale à l'interface entre linguistique et traitement automatique des langues (TAL), car l'expression affective repose sur des mécanismes complexes et hétérogènes. Les émotions se manifestent non seulement par des mots isolés, mais aussi par des unités polylexicales, des collocations, des constructions figées et des structures syntaxiques ou pragmatiques implicites. Cette diversité complique leur détection automatique et limite les approches basées sur le lexique ou les statistiques de surface. Ce travail adopte une approche interdisciplinaire, combinant méthodes statistiques, apprentissage automatique et connaissances linguistiques, pour mieux rendre compte de la polysémie affective, de l'intensité émotionnelle et des variations contextuelles. L'analyse montre le rôle central des unités polylexicales et des collocations affectives, ainsi que l'importance des constructions syntaxiques et

discursives dans l'expression implicite ou évaluative des émotions, soulignant la nécessité d'une articulation étroite entre linguistique et TAL.

Mots clés:

Émotions linguistiques, unités polylexicales, phraséologie, constructions syntaxiques de l'affect, collocations, traitement automatique des langues

Abstract

The study of emotions in written language is central at the interface of linguistics and natural language processing (NLP), as affective expression relies on complex and heterogeneous linguistic mechanisms. Emotions are conveyed not only through individual words but also via multiword units, collocations, fixed or semi-fixed constructions, and implicit syntactic and pragmatic structures. This diversity complicates automatic detection and limits approaches based solely on lexical or surface statistical cues. This work adopts an interdisciplinary approach, combining statistical methods, machine learning, and linguistic knowledge to better account for affective polysemy, emotional intensity, and contextual variations. The analysis highlights the central role of multiword units and affective collocations, as well as the importance of syntactic and discursive constructions in implicit or evaluative emotional expression, emphasizing the need for close integration of linguistics and NLP to develop systems sensitive to emotional nuances.

Keywords:

Linguistic emotions, Multiword units, Phraseology, Syntactic constructions of affect, Collocations, Natural language processing (NLP)

1. INTRODUCTION

L'étude des émotions dans le langage écrit occupe une place centrale à l'interface entre linguistique et traitement automatique des langues (TAL). En effet,

l'expression affective ne se réduit pas à la simple occurrence de lexèmes émotionnels : elle mobilise un ensemble complexe de mécanismes linguistiques impliquant des unités polylexicales, des constructions syntaxiques spécifiques et des stratégies discursives dépendantes du contexte (Novakova & Tutin, 2009 ; Tutin et al., 2006 ; Micheli, 2014).

Les travaux en linguistique ont montré que les émotions sont construites dans et par le langage. Il convient de noter aussi que les régularités phraséologiques et les unités polylexicales apparaissent comme des vecteurs privilégiés de la signification affective, permettant l'expression de nuances difficiles à capturer par des lexèmes isolés (Novakova & Tutin, 2009 ; Tutin et al., 2006). Les approches discursives, quant à elles, soulignent le caractère médiatisé et interprétatif de l'émotion, qui se construit dans l'interaction entre formes linguistiques et contexte (Micheli, 2014).

Dans le champ du TAL, les progrès récents ont permis le développement de systèmes capables de détecter et classifier automatiquement les émotions à partir de textes numériques. Les ressources telles que celles de Mohammad (2017), notamment sur l'intensité affective, ont amélioré la performance de ces systèmes. Cependant, la plupart des approches restent fondées sur des indices lexicaux ou des régularités statistiques de surface, ce qui limite leur capacité à saisir les formes implicites, métaphoriques ou contextuelles de l'expression émotionnelle.

En ce sens, la question centrale de cette étude est de déterminer comment modéliser automatiquement les émotions dans le langage écrit en tenant compte

de leur complexité lexicale, phraséologique, syntaxique et contextuelle, alors que les approches existantes se limitent à des indices de surface.

Pour répondre à cette problématique, l'étude adopte une démarche interdisciplinaire, articulant linguistique de corpus et méthodes de TAL.

À partir d'un corpus homogène de commentaires francophones issus des réseaux sociaux, elle s'attache à analyser en profondeur les mécanismes linguistiques de l'expression émotionnelle.

De ce fait, l'analyse se structure selon trois axes complémentaires :

1. La structuration lexicale, phraséologique et la conceptualisation des émotions, afin de comprendre comment les unités polylexicales et les schémas métaphoriques structurent le lexique affectif.
2. L'encodage syntaxique, discursif et pragmatique des émotions, pour saisir comment l'émotion est modulée et construite dans l'interaction linguistique et contextuelle.
3. Les implications pour la modélisation computationnelle, visant à explorer comment les connaissances linguistiques peuvent enrichir les systèmes automatiques et dépasser les limites des approches basées sur des indices lexicaux seuls.

Il est à souligner que cette structuration permet de rendre compte de la complexité multidimensionnelle des émotions, en articulant théorie linguistique et perspectives appliquées de TAL, tout en posant les bases pour des modèles automatiques plus nuancés et interprétatifs. Cette étude propose une modélisation linguistiquement informée et multi-niveaux de l'expression des émotions dans le

langage écrit, intégrant les dimensions lexicale, phraséologique, syntaxique et pragmatique, afin de dépasser les limites des approches fondées uniquement sur des indices de surface en traitement automatique des langues.

2. Méthodologie du corpus

Traditionnellement connu, le corpus a fait l'objet d'un protocole de constitution et de traitement visant à garantir la cohérence analytique des données. Il est constitué de quelques commentaires sélectionnés à partir de publications publiques selon des critères linguistiques précis, notamment la présence d'indices affectifs explicites ou implicites, ainsi que la diversité des formes d'expression émotionnelle.

Un prétraitement des données a été effectué, incluant la suppression des doublons, la normalisation partielle des variations graphiques et la conservation des marqueurs discursifs pertinents (ponctuation expressive, répétitions, marques d'oralité), essentiels pour l'analyse des émotions en contexte numérique.

L'analyse repose sur une annotation qualitative multi-niveaux. Chaque énoncé a été examiné selon plusieurs dimensions : (i) lexicale et phraséologique, avec une attention particulière portée aux unités polylexicales ; (ii) syntaxique, notamment à travers l'identification des constructions expérientielles ; (iii) discursive et pragmatique, incluant les phénomènes d'implicite, d'ironie et de modulation de l'intensité affective.

L'interprétation des données s'appuie également sur le co-texte et, lorsque cela est pertinent, sur le contexte interactionnel, afin de prendre en compte la variabilité sémantique et la polysémie affective des énoncés. Cette approche

permet d'identifier des régularités linguistiques fines tout en préservant la complexité des usages.

Enfin, cette méthodologie vise à articuler description linguistique et perspectives computationnelles, en fournissant des observations exploitables pour la modélisation automatique des émotions, notamment dans une perspective multi-niveaux intégrant les dimensions lexicales, syntaxiques et pragmatiques. L'annotation a été réalisée manuellement selon une grille d'analyse définie en amont, permettant d'identifier les types d'émotions, leur intensité, ainsi que les procédés linguistiques mobilisés. Afin d'assurer la cohérence de l'analyse, un protocole d'annotation explicite a été suivi, fondé sur les travaux existants en linguistique des émotions.

3. Structuration lexicale, phraséologique et conceptualisation des émotions

Pour appréhender l'expression des émotions dans le langage écrit, il est nécessaire de préciser les notions centrales mobilisées dans cette analyse. Tout d'abord, le lexique émotionnel désigne l'ensemble des unités linguistiques exprimant ou suggérant un état affectif, qu'elles soient explicites ou dépendantes du contexte (Novakova & Tutin, 2009). Ensuite, les unités polylexicales correspondent à des séquences stabilisées de plusieurs mots formant une unité de sens, relevant de la phraséologie (Tutin et *al.*, 2006). Les collocations renvoient à des associations préférentielles entre mots fondées sur l'usage, tandis que la conceptualisation des émotions concerne les mécanismes cognitifs structurant l'affect à travers le langage (Kövecses, 2000).

Dans ce cadre, l'analyse démontre que l'expression des émotions ne peut se limiter à des unités lexicales isolées. Elle repose sur une organisation systémique dans laquelle les régularités phraséologiques jouent un rôle structurant et dynamique. Comme le soulignent Tutin et *al.* (2006), les unités polylexicales constituent des vecteurs privilégiés de la signification affective, permettant l'expression de nuances fines. Elles apparaissent ainsi comme des unités intermédiaires entre lexique et discours, assurant à la fois stabilité et variabilité interprétative.

Par ailleurs, les émotions sont conceptualisées à travers des schémas métaphoriques récurrents. Des expressions telles que « avoir le cœur lourd » ou « exploser de colère » traduisent une représentation de l'émotion comme substance ou force interne (Kövecses, 2000). Cette structuration métaphorique constitue un principe organisateur du lexique affectif**,** révélant l'ancrage corporel des émotions et leur inscription dans des schémas expérientiels partagés.

De surcroît, le lexique émotionnel se déploie selon des dimensions continues, telles que la valence et l'intensité, ce qui rejoint les observations de Mohammad (2017). Cette approche graduelle remet en question les modèles catégoriels traditionnels et permet de concevoir les émotions comme des espaces sémantiques dynamiques et nuancés**.**

Enfin, la polysémie affective souligne que la valeur émotionnelle d'une unité linguistique est fortement contextuelle (Novakova et *al.*, 2013). Le sens affectif émerge ainsi comme un phénomène dynamique et interactionnel, dépendant des cooccurrences et des environnements discursifs. Cette perspective met en évidence

l'importance des usages situés pour la stabilisation des valeurs affectives et souligne la nécessité de dépasser une conception statique du lexique.

L'analyse des distributions lexicales dans les corpus numériques révèle également des schémas récurrents d'association entre émotions et domaines sémantiques expérimentiels (corps, espace, température), confirmant l'ancrage sensoriel de l'affect. Ces observations ouvrent la voie à une modélisation du lexique émotionnel fondée sur des réseaux de cooccurrence enrichis, tout en suggérant que les unités phraséologiques constituent des points d'ancrage privilégiés pour la détection automatique des émotions. Elles mettent enfin en évidence l'intérêt d'une articulation entre linguistique de corpus et sémantique cognitive, permettant une description plus fine et plus précise du lexique affectif.

Afin d'illustrer empiriquement ces phénomènes, le corpus de commentaires issus des réseaux sociaux fournit plusieurs exemples représentatifs. L'énoncé *« J'ai le cœur brisé depuis hier »* illustre une unité polylexicale reposant sur une métaphore conceptualisant l'émotion comme une atteinte physique du cœur, confirmant l'ancrage corporel des affects. De même, l'expression *« Ça me rend malade de voir ça »* met en évidence une collocation où l'émotion est assimilée à un état pathologique, révélant une structuration expérientielle du lexique émotionnel. Il convient de noter que ces exemples confirment que les émotions s'expriment à travers des configurations phraséologiques stabilisées, difficilement réductibles à des lexèmes isolés.

4. Encodage syntaxique, discursif et pragmatique de l'émotion

Pour analyser l'expression des émotions dans le langage écrit, il est nécessaire de prendre en compte les dimensions syntaxique, discursive et pragmatique : La syntaxe renvoie à l'organisation des relations entre les constituants d'un énoncé, tandis que le discours désigne le langage en contexte (Diwersy & Kraif, 2013). La pragmatique s'intéresse aux mécanismes d'interprétation dépendant de la situation d'énonciation, et l'implicite désigne les contenus inférés ou sous-entendus par le locuteur (Micheli, 2014).

D'ailleurs, l'analyse montre que les émotions sont souvent encodées dans des structures syntaxiques spécifiques, notamment les constructions expérientielles impliquant un stimulus et un expérimentateur (Goossens, 2005). Ces structures mettent en exergue le caractère relationnel et causal de l'émotion, où l'état affectif ne se comprend qu'en lien avec un objet, un événement ou un autre acteur du discours.

Signalons que les variations syntaxiques permettent également de moduler l'intensité émotionnelle à travers des procédés grammaticaux divers. L'emploi d'adverbes d'intensité, de passifs ou de constructions impersonnelles illustre la capacité de la syntaxe à nuancer et graduer les affects, confirmant que l'émotion s'inscrit dans un continuum plutôt que dans des catégories rigides.

Au niveau discursif, l'émotion est fréquemment implicite et contextuelle, nécessitant des inférences pour sa reconnaissance. Comme le souligne Micheli (2014), elle se construit dans l'interaction entre indices linguistiques et contexte, ce

qui met l'accent sur l'importance de considérer le texte non seulement comme une suite de mots, mais comme un processus communicationnel intégré.

De plus, les procédés pragmatiques tels que l'ironie, la litote ou l'hyperbole introduisent un écart entre la forme et l'interprétation**,** complexifiant la lecture automatique des émotions. Ces stratégies montrent que l'affect peut être atténué, amplifié ou détourné selon les intentions du locuteur, soulignant le rôle central des connaissances partagées et de la culture linguistique dans l'interprétation.

En outre, l'organisation informationnelle de l'énoncé, via la thématisation et la focalisation, oriente la saillance des contenus affectifs et influence leur interprétation. Les choix syntaxiques participent ainsi à une hiérarchisation des émotions dans le discours, renforçant leur visibilité ou, au contraire, leur subtilité. Par ailleurs, la dimension dialogique du langage, particulièrement dans les interactions numériques, favorise la co-construction des affects entre locuteurs, révélant la nature dynamique et interactionnelle de l'expression émotionnelle. Cette perspective invite à intégrer l'analyse des émotions dans une approche discursive élargie, attentive aux genres, aux contextes et aux pratiques communicationnelles. Nous illustrons concrètement ces mécanismes par des exemples analysés comme suit :

L'énoncé *« Ça m'énerve quand ils font ça »* constitue une construction expérientielle typique, dans laquelle un stimulus (*ils font ça*) provoque un état affectif chez un expérimentateur (*m'*), illustrant le caractère relationnel de l'émotion. Par ailleurs, l'expression *« Super, encore une panne… »* repose sur un procédé ironique : la polarité lexicale positive de *super* contraste avec une

interprétation négative dépendante du contexte, ce qui nécessite une inférence pragmatique. Ces exemples montrent que l'émotion est fréquemment implicite et construite à travers des indices syntaxiques et discursifs, confirmant la nécessité d'une analyse contextuelle.

5. Vers une modélisation computationnelle linguistiquement informée des émotions

Les approches classiques en traitement automatique des langues (TAL) reposent majoritairement sur des indices lexicaux et des modèles statistiques (Mohammad, 2017), ce qui limite leur capacité à traiter les expressions complexes, implicites ou métaphoriques. Bien que des modèles récents intègrent le contexte et la dynamique des phrases, ils restent souvent insuffisants face à la richesse linguistique des données, comme le soulignent Diwersy et al. (2014). L'intégration de corpus annotés et de connaissances linguistiques apparaît donc comme un facteur clé pour améliorer la précision et la subtilité des systèmes.

Une approche hybride et multi-niveaux semble dès lors nécessaire. Celle-ci combine la détection d'unités polylexicales, l'analyse des structures syntaxiques, l'identification des indices discursifs et la prise en compte des dimensions sémantiques continues (valence, intensité, contrôle). Cette intégration permet non seulement de mieux représenter la polysémie affective, mais aussi de saisir la variabilité contextuelle et interactionnelle des émotions.

Dans cette perspective, le développement de modèles computationnels linguistiquement informés constitue un enjeu central. L'utilisation de ressources structurées, telles que des bases phraséologiques, des annotations pragmatiques et

des schémas conceptuels, permet de créer des représentations plus fines et adaptées à la complexité de l'affect. De plus, l'évaluation de ces systèmes devrait dépasser les métriques classiques de précision et de rappel, en intégrant des critères qualitatifs liés à la finesse interprétative et à la capacité à détecter les émotions implicites.

Enfin, la collaboration entre linguistes et spécialistes du TAL s'avère indispensable pour concevoir des modèles vraiment sensibles à la complexité des émotions. Cette articulation interdisciplinaire favorise la prise en compte simultanée des aspects lexicaux, phraséologiques, syntaxiques et pragmatiques. Elle ouvre également des perspectives pour l'enrichissement des corpus annotés et la construction de modèles adaptatifs capables de généraliser à différents genres textuels et contextes culturels. Par ailleurs, l'intégration d'approches cognitives et distributionnelles pourrait permettre de mieux modéliser l'intensité, la valence et la dynamique temporelle des émotions, offrant ainsi une vision plus complète et réaliste des affects dans le langage. En somme, ces développements renforcent le potentiel des systèmes TAL pour des applications concrètes, allant de l'analyse des réseaux sociaux à l'assistance émotionnelle automatisée, tout en garantissant une interprétation linguistiquement et contextuellement informée.

Il convient de préciser que certains exemples issus du corpus mettent en évidence les limites des approches strictement lexicales et l'intérêt d'une modélisation multi-niveaux. L'énoncé *« Je suis mort de rire »* illustre une expression idiomatique dont la polarité globale est positive, malgré la présence du lexème *mort*, généralement associé à une valence négative, ce qui souligne

l'importance de la prise en compte des unités polylexicales. De même, l'expression *« Franchement, c'est pas mal… »* peut, selon le contexte, relever d'une appréciation modérée ou d'une critique atténuée (litote), ce qui met l'accent sur le rôle déterminant des indices pragmatiques et contextuels dans l'interprétation des émotions. Ces exemples confirment la nécessité d'intégrer des connaissances linguistiques fines dans les systèmes de TAL afin de dépasser les limites des approches fondées sur des indices de surface. Cette approche permet d'esquisser un cadre opérationnel pour l'intégration de connaissances linguistiques fines dans les systèmes de détection automatique des émotions, en améliorant leur capacité à traiter les expressions implicites, métaphoriques et contextuelles.

6. Conclusion:

En somme, cette analyse met en évidence que l'expression des émotions dans le langage écrit relève d'une complexité multidimensionnelle, où les affects émergent de l'interaction entre lexique, phraséologie, syntaxe et contexte. Les résultats montrent que les unités polylexicales, les structures syntaxiques et les mécanismes discursifs et pragmatiques jouent un rôle central dans la construction et l'interprétation des émotions, souvent implicites et contextuelles.

Ces constats soulignent les limites des approches actuelles du traitement automatique des langues fondées sur des indices de surface, et plaident en faveur d'une modélisation hybride et multi-niveaux intégrant des connaissances linguistiques fines ainsi que des dimensions sémantiques continues telles que la valence et l'intensité. Ainsi, une collaboration étroite entre linguistique et TAL apparaît-elle essentielle pour développer des systèmes plus sensibles à la

complexité des affects. Elle propose également un cadre analytique transférable aux approches computationnelles contemporaines.

Cette perspective ouvre la voie à des modèles plus nuancés, capables de mieux interpréter les émotions dans des contextes variés, tout en intégrant la dimension contextuelle, interactionnelle et interculturelle du langage.

7. Liste Bibliographique

- Diwersy, S., Goossens, V., Grutschus, A., & Kraif, **O.** (2014), « Traitement des lexies d'émotion dans les corpus », *Corpus*, 13.
- Diwersy, S., & Kraif**, O.** (2013), « Corpus et linguistique outillée : enjeux théoriques et méthodologiques », *Langages*, 191, 3–22.
- Goossens, V. (2005), *Les constructions syntaxiques de l'émotion*. Paris : CNRS Éditions.
- Kövecses, Z**.** (2000), *Metaphor and Emotion: Language, Culture, and Body in Human Feeling*. Cambridge : Cambridge University Press.
- Le Pesant, D. (2023), « Les verbes psychologiques et l'encodage linguistique de l'affect », In *Langue française*.
- Micheli, R. (2014), *Les émotions dans les discours*. Bruxelles : De Boeck.
- Mohammad, S. M. (2017), « Word Affect Intensities », In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
- Novakova, I., & Tutin, A. (2009), *Le lexique des émotions*. Grenoble : Presses Universitaires de Grenoble.
- Novakova, I., Tutin, A., & Grossmann, F. (2013), « Polysémie et émotions : approche lexico-sémantique ». *Langue française*, 178, 45–62.
- Tutin, A., Novakova, I., Grossmann, F., & Cavalla, C. (2006). « La phraséologie des émotions ». *Revue française de linguistique appliquée*, XI(1), 7–20.