

Sélection ordonnée de phrases associées aux symptômes de la dépression par classification zéro-coup

Yves Ferstler Catherine Lavoie Marie-Jean Meurs

Université du Québec à Montréal, Montréal, QC, Canada
ferstler.yves, lavoie.catherine.15}@courrier.uqam.ca,
meurs.marie-jean@uqam.ca

RÉSUMÉ

Cet article présente une méthode pour extraire d'un corpus les phrases les plus pertinentes pour répondre à un questionnaire d'auto-évaluation. Un modèle de classification zéro-coup évalue la similarité entre les phrases et les réponses du questionnaire. Les résultats obtenus par ce modèle frugal sont prometteurs par comparaison avec ceux d'autres grands modèles de langue.

ABSTRACT

Ordered selection of phrases associated with depression symptoms by zero-shot classification

This paper leverages a method for selecting in a corpus the most relevant sentences to answer a self-report questionnaire. A zero-shot classification model is used to evaluate the similarities between the sentences and the possible answers in the questionnaire. The results obtained by this frugal model are promising when compared to those of large language models.

MOTS-CLÉS : Classification zéro-coup, Extraction de phrases, Dépression.

KEYWORDS: Zero-shot classification, Sentences extraction, Depression.

1 Introduction

Selon l'Organisation mondiale de la Santé ([World Health Organization, 2025](#)), environ 5% des adultes dans le monde souffraient de dépression en 2023. Les symptômes de cette maladie sont liés par exemple à des sentiments de tristesse, d'irritation, à de la faible concentration et des changements du niveau d'appétit. Plusieurs questionnaires de référence existent pour dépister et mesurer la sévérité de ces symptômes. Les questionnaires d'auto-évaluations sont généralement composées de questions à choix multiples, à échelles graduées ou de questions ouvertes axées sur les comportements associés aux différents symptômes. Ces outils offrent un support au corps médical dans le processus diagnostique et permettent l'auto-évaluation de l'évolution des symptômes. La détection automatique de ces comportements constitue un domaine de recherche important en traitement automatique du langage naturelle (TALN), notamment grâce à la disponibilité de corpora de messages et de discussions entre internautes issus de réseaux sociaux. Actuellement, l'état de l'art de la détection de symptômes de la dépression repose principalement sur l'affinage de grands modèles de langue (LLM) pré-entraînés, qui remplace souvent d'autres méthodes performantes. Parmi ces méthodes, la classification zéro-coup (ZSC) présente un intérêt particulier, car elle permet de classer des données sans nécessiter l'affinage d'un LLM ou l'entraînement préalable d'un modèle dédié, ce qui constitue un avantage quand les données sont limitées. Les travaux présentés dans cet article ont pour objectif de déterminer, dans un corpus de phrases, quelles sont celles qui sont pertinentes par rapport à des états émotionnels associés aux symptômes de dépression et de les classer par ordre de pertinence décroissante. L'approche utilise un modèle ZSC, dont les résultats sont comparés à d'autres approches basées sur l'affinage de LLM.

2 État-de-l’art

Parmi les travaux récents en TALN centrés sur la santé mentale, [Karamat et al. \(2024\)](#) proposent une approche hybride permettant la détection de plusieurs comportements associés à des symptômes de troubles de santé mentale sur les réseaux sociaux. En utilisant deux modèles préentraînés, MentalBERT et MelBERT ([Ji et al., 2021](#); [Choi et al., 2021](#)), et en extrayant les caractéristiques avec un réseau de neurones convolutifs, leurs approches permettent une détection de comportements associés aux symptômes de la dépression, de l’anxiété, du trouble post-traumatique et du trouble de la personnalité limite. D’autres aspects de la santé mentale sont étudiés par [Ta et al. \(2022\)](#), comme la détection des symptômes associés à l’anxiété au sein de communautés en ligne en analysant les relations entre les personnes et les sujets abordés. Les méthodes pour détecter des comportements associés aux symptômes de la dépression sont variées. [Lim et al. \(2025\)](#); [Chen et al. \(2025\)](#) présentent des approches de conception de modèles capturant à la fois les caractéristiques audio d’un dialogue et sa transcription textuelle utilisant de XLM-RoBERTa ([Conneau et al., 2019](#)). De nouvelles approches utilisent des modèles accompagnés de base de connaissances pour améliorer la détection de comportements associés aux symptômes de la dépression. [Yan et al. \(2025\)](#) proposent une méthode détectant des émotions négatives, basée sur les symptômes de dépression définis dans le *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) ([Crocq et al., 2015](#)), et l’utilisation du transformeur BERT ([Devlin et al., 2018](#)). Ces travaux présentent de meilleurs résultats que d’autres approches n’utilisant pas de base de connaissance. [Su et al. \(2025\)](#) utilisent également le DSM-5 comme base de connaissance. Dans le cadre de la campagne d’évaluation eRisk 2024 ([Crestani et al., 2022](#)), les équipes participant à la tâche 1 ont travaillé sur l’évaluation de pertinence de phrases associées aux symptômes de la dépression. Parmi les résultats ([Parapar et al., 2024](#)), les approches ayant obtenu les meilleurs scores sont basées sur l’affinage de LLM ([Ang et al., 2024](#); [Bascuñana & Bedmar, 2024](#); [Barachanou et al., 2024](#)). Cependant, d’autres approches utilisant des modèles de langue telle que la classification zéro-coup (ZSC) montrent des résultats prometteurs ([Zhang et al., 2020](#); [Li et al., 2017](#)). [Leow et al. \(2025\)](#) présentent une comparaison entre l’utilisation de LLM ou de ZSC dans la détection de signes de la dépression. Les résultats indiquent que l’utilisation des LLM permet d’obtenir de meilleurs scores mais la méthode ZSC est plus frugale et aussi très performante.

3 Méthodologie

Pour déterminer quelles sont les phrases d’un corpus qui sont pertinentes par rapport à des symptômes de dépression, notre approche : (1) calcule, avec un modèle ZSC, un score de similitude entre phrases présentes dans le corpus et les énoncés pré-écrits d’un questionnaire d’auto-évaluation, (2) élimine les phrases ne contenant pas de pronoms personnels au singulier, (3) agrège les scores des phrases restantes en un score global qui permet de classer ces phrases par ordre de pertinence décroissante, (4) utilise la médiane des scores globaux pour définir un seuil au-delà duquel les phrases sont jugées pertinentes.

Ensemble de données et questionnaire de référence. Le corpus choisi dans les travaux présentés est le corpus d’évaluation de la tâche 1 de la campagne d’évaluation eRisk 2024 ([Crestani et al., 2022](#)). Il contient 11,8k phrases, manuellement annotées, issues d’un historique conversationnel de personnes utilisatrices de réseaux sociaux. Chaque phrase est enregistrée avec ses phrases précédente et suivante pour conserver son contexte. Un exemple fictif est présenté en annexe, figure 1. À partir des phrases manuellement annotées, deux ensembles de référence ont été créés : un contenant les phrases jugées pertinentes par consensus des personnes annotatrices et l’autre par la majorité de ces personnes. Le questionnaire d’auto-évaluation utilisé pour évaluer la pertinence entre les phrases et ses items est l’inventaire de dépression de Beck ([Beck et al., 1996](#)) 2ème édition (BDI-II). Ce

questionnaire est composé de 21 items associés à différents états émotionnels liés à des symptômes de la dépression. Chacun de ces items est qualifié par plusieurs énoncés pré-écrits reflétant la gravité de l'état considéré. Des exemples et leur traduction libre en français sont fournis en annexe (table 3).

Modèle d'évaluation de pertinence. Les directives d'annotation manuelle des phrases indiquent qu'une phrase pertinente doit donner de l'information sur l'état de la personne en rapport avec l'item d'auto-évaluation. Ainsi, une phrase exprimant du bien-être, de la joie, de l'appétit, ou d'autres manifestations positives est aussi pertinente qu'une phrase présentant des émotions ou des comportements associés aux symptômes de la dépression. Afin d'identifier les phrases pertinentes, le modèle ZSC doit les comparer à chacun des énoncés pré-écrits. Avant de calculer les similarités, un filtrage inspiré de [Ang et al. \(2024\)](#) est appliqué pour supprimer les phrases sans pronoms personnels au singulier. Ne se rapportant pas à la personne autrice de la publication, elles sont jugées non pertinentes pour le questionnaire. Le modèle ZSC choisi est bart-large-mnli ([Hugging Face, 2025](#)), pour sa légèreté et ses performances à l'état de l'art. Les scores ZSC des énoncés pré-écrits sont combinés pour produire deux versions de score global, l'une étant la somme des scores et l'autre le maximum. Le score global permet de classer les phrases et d'identifier celles qui sont les plus pertinentes pour chaque item. Un exemple de scores est présenté dans la table 1. La phrase « je vais bien, mais je n'ai juste pas envie de faire quelque chose » a été évaluée par le modèle ZSC sur les trois items présents dans la table 3 en annexe. Les résultats montrent que, pour les items 1 (tristesse) et 12 (perte d'intérêt), les étiquettes avec le score le plus haut sont « *I do not feel sad.* » et « *I am less interested in other people or things than before.* », ce qui est cohérent pour la phrase évaluée. Pour l'item 7 (dégoût de soi), les étiquettes obtiennent des scores plus faibles, ce qui est également cohérent, dans la mesure où aucun sentiment de dégoût de soi n'est exprimé.

Phrase		Je vais bien, mais je n'ai juste pas envie de faire quelque chose					
Scores		1er	2ème	3ème	4ème	somme	maximum
1.	r1 : 0,951	r2 : 0,000	r3 : 0,000	r4 : 0,000		0,951	0,951
7.	r1 : 0,114	r2 : 0,013	r4 : 0,000	r3 : 0,000		0,127	0,114
12.	r2 : 0,876	r3 : 0,354	r1 : 0,300	r4 : 0,038		1,568	0,876

TABLE 1 – Exemple de résultat obtenu par un modèle ZSC entre une phrase et 3 items du BDI-II caractérisés par les 4 réponses pré-écrites (r1 à r4)

4 Expériences et résultats

Le corpus de référence, utilisé pour l'évaluation, a été construit à partir des soumissions des participants à la tâche 1 de eRisk 2024 ([Parapar et al., 2024](#)). N'ayant pas accès à toutes les soumissions, il est impossible de reproduire les mêmes conditions d'évaluation. Nous avons donc évalué nos résultats en les comparant aux résultats fournis par les organisateurs pour les phrases du corpus de référence. Cette approche permet d'évaluer les capacités de classification de pertinence du modèle ZSC en utilisant les métriques classiques de précision, rappel et F-mesure. L'ordonnancement des phrases selon leur pertinence est mesuré par le gain cumulé normalisé actualisé (NDCG) qui, pour chaque item du BDI-II, additionne les scores de pertinence des phrases selon leur position dans le classement généré par le modèle, en accordant plus d'importance aux phrases placées en haut et normalise ensuite ce score en le divisant par le score idéal, obtenu avec un classement parfait.

Expériences. Les phrases du corpus de référence (manuellement étiquetées en deux versions : *majorité* et *consensus* entre annotations) sont filtrées puis évaluées par le modèle ZSC pour produire les deux versions de score global (somme et maximum). Un seuil est alors appliqué au score global des phrases pour distinguer celles qui sont pertinentes pour un item de celles qui ne le sont pas. Un

Évaluation sur le corpus de référence <i>majorité</i>			
Score global	Précision	Rappel	F-mesure
ZSC-somme	0,7883	0,6644	0,7210
ZSC-maximum	0,8151	0,6858	0,7449
Meilleurs scores par question ordonnée par F-mesure			
ZSC-maximum-question-14 (<i>Dévalorisation</i>)	0,9756	0,7751	0,8639
ZSC-somme-question-14 (<i>Dévalorisation</i>)	0,9707	0,7713	0,8596
Pires scores par question ordonnée par F-mesure			
ZSC-somme-question-13 (<i>Indécision</i>)	0,5847	0,5487	0,5661
ZSC-maximum-question-13 (<i>Indécision</i>)	0,5862	0,5519	0,5685
Évaluation sur le corpus de référence <i>consensus</i>			
Score global	Précision	Rappel	F-mesure
ZSC-somme	0,5631	0,7212	0,6300
ZSC-maximum	0,5862	0,7494	0,6600
Meilleurs scores par question ordonnés par F-mesure			
ZSC-maximum-question-14 (<i>Dévalorisation</i>)	0,9317	0,8603	0,8946
ZSC-somme-question-14 (<i>Dévalorisation</i>)	0,9121	0,8423	0,8758
Pires scores par question ordonnée par F-mesure			
ZSC-somme-question-2 (<i>Pessimisme</i>)	0,2846	0,6557	0,3970
ZSC-somme-question-13 (<i>Indécision</i>)	0,3356	0,5914	0,4282

TABLE 2 – Résultats de l’approche ZSC (somme et maximum) appliquée au corpus de référence *majorité* et *consensus* de la tâche 1 de eRisk 2024

seuil spécifique à chaque item est calculé en prenant la médiane des scores obtenus. Le code source est disponible sous licence libre [ici](#).

Résultats. Les résultats de classification de phrases obtenus par le modèle ZSC, selon les calculs de scores, sont présentés dans la table 2. Les approches proposées donnent des résultats intéressants, en particulier sur le corpus de référence *majorité*. Les résultats montrent que le calcul du score global par maximum est plus performant que le calcul utilisant la somme. ainsi les phrases sélectionnées par maximum pourraient potentiellement permettre de choisir une réponse parmi celles pré-écrites pour les items du BDI-II. Il sera donc intéressant d’évaluer la capacité du modèle à prédire directement les réponses au BDI-II à partir des phrases qu’il sélectionne dans les écrits d’une personne. Le modèle ZSC avec les deux options de calcul de score obtient un NDCG de 0,7 qui correspond à celui des systèmes présentés à eRisk 2024, confirmant la capacité d’ordonnancement de notre approche.

5 Conclusion

Ces travaux présentent une méthode basée sur un modèle ZSC pour identifier et classer la pertinence de phrases en fonction des items du BDI II (Beck *et al.*, 1996), questionnaire d’auto-évaluation d’états émotionnels spécifiques associées aux symptômes de la dépression qui est un des principaux outils utilisés pour le dépistage et la mesure de sévérité de ces symptômes chez les personnes adolescentes et adultes. Les résultats obtenus grâce au score maximum attribué aux phrases par le modèle sont prometteurs. Ils confirment qu’affiner un LLM n’est pas toujours indispensable quand un modèle plus frugal comme ZSC peut être mis en oeuvre.

Remerciements

Ces travaux ont été réalisés grâce aux ressources de calcul mises à notre disposition par [Calcul Québec](#) et [l'Alliance de recherche numérique du Canada](#), et grâce au soutien financier du Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG) [MJ Meurs, CRSNG à la découverte #06487-2017] et de la Chaire de recherche du Québec sur la découvrabilité des contenus scientifiques en français [MJ Meurs, [DOI 10.69777/358425](#)]. Nous souhaitons également remercier Yassine Chahdi pour sa participation enthousiaste aux réflexions et aux analyses post-évaluation.

Références

- ANG B. H., GOLLAPALLI S. D. & NG S.-K. (2024). NUS-IDS@ eRisk2024 : ranking sentences for depression symptoms using early maladaptive schemas and ensembles. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction : 15th International Conference of the CLEF Association, CLEF 2024, Grenoble, France, September 9–12, 2024, Proceedings, Part II*, p. 782–793.
- BARACHANOU A., TSALAKANIDOU F. & PAPADOPOULOS S. (2024). REBECCA at eRisk 2024 : search for symptoms of depression using sentence embeddings and prompt-based filtering. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction : 15th International Conference of the CLEF Association, CLEF 2024, Grenoble, France, September 9–12, 2024, Proceedings, Part II*, p. 794–802.
- BASCUÑANA A. & BEDMAR I. S. (2024). APB-UC3M at eRisk 2024 : natural language processing and deep learning for the early detection of mental disorders. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction : 15th International Conference of the CLEF Association, CLEF 2024, Grenoble, France, September 9–12, 2024, Proceedings, Part II*, p. 871–880.
- BECK A. T., STEER R. A. & BROWN G. K. (1996). Beck Depression Inventory (BDI-II). *Psychological assessment*, **10**.
- CHEN Z., WANG D., LOU L., ZHANG S., ZHAO X., JIANG S., YU J. & XIAO J. (2025). Text-guided multimodal depression detection via cross-modal feature reconstruction and decomposition. *Information Fusion*, **117**, 102861. DOI : <https://doi.org/10.1016/j.inffus.2024.102861>.
- CHOI M., LEE S., CHOI E., PARK H., LEE J., LEE D. & LEE J. (2021). MelBERT : Metaphor Detection via Contextualized Late Interaction using Metaphorical Identification Theories.
- CONNEAU A., KHANDLWAL K., GOYAL N., CHAUDHARY V., WENZKE G., GUZMÁN F., GRAVE E., OTT M., ZETTMLOYER L. & STOYANOV V. (2019). Unsupervised Cross-lingual Representation Learning at Scale. *CoRR*, **abs/1911.02116**.
- CRESTANI F., LOSADA D. & PARAPAR J. (2022). *Early Detection of Mental Health Disorders by Social Media Monitoring : The First Five Years of the eRisk Project*. Studies in Computational Intelligence. Springer International Publishing. DOI : [10.1007/978-3-031-04431-1](https://doi.org/10.1007/978-3-031-04431-1).
- CROCQ M.-A., GUELFY J. D., ASSOCIATION A. P. & FORCE A. P. A. D.-. T. (2015). *DSM-5 : manuel diagnostique et statistique des troubles mentaux (5e édition)*. Elsevier Masson.
- DEVLIN J., CHANG M., LEE K. & TOUTANOVA K. (2018). BERT : Pre-training of Deep Bidirectional Transformers for Language Understanding. *CoRR*, **abs/1810.04805**.
- DIAS G., Éd. (2015). *Actes de TALN 2015 (Traitement automatique des langues naturelles)*, Caen. ATALA, HULTECH.

- Hugging Face (2025). facebook/bart-large-mnli. <https://huggingface.co/facebook/bart-large-mnli> [Accessed : (03/05/2025)].
- JI S., ZHANG T., ANSARI L., FU J., TIWARI P. & CAMBRIA E. (2021). MentalBERT : Publicly Available Pretrained Language Models for Mental Healthcare.
- KARAMAT A., IMRAN M., YASEEN M. U., BUKHSH R., ASLAM S. & ASHRAF N. (2024). A Hybrid Transformer Architecture for Multiclass Mental Illness Prediction using Social Media Text. *IEEE Access*, p. 1–1. DOI : [10.1109/ACCESS.2024.3519308](https://doi.org/10.1109/ACCESS.2024.3519308).
- LEOW J. J. D., CHUA H. N., JASSER M. B., ISSA B. & WONG R. T. (2025). Comparison of Depression Detection Between LLMs and Zero-Shot Learning Using DAD Dataset. In *2025 21st IEEE International Colloquium on Signal Processing & Its Applications (CSPA)*, volume 21, p. 295–300. DOI : [10.1109/CSPA64953.2025.10933098](https://doi.org/10.1109/CSPA64953.2025.10933098).
- LI X., FANG M. & WU J. (2017). Zero-shot classification by transferring knowledge and preserving data structure. *Neurocomputing*, **238**, 76–83. DOI : <https://doi.org/10.1016/j.neucom.2017.01.038>.
- LIM E., JHON M., KIM J.-W., KIM S.-H., KIM S. & YANG H.-J. (2025). A lightweight approach based on cross-modality for depression detection. *Computers in Biology and Medicine*, **186**, 109618. DOI : <https://doi.org/10.1016/j.combiomed.2024.109618>.
- PARAPAR J., MARTÍN-RODILLA P., LOSADA D. E. & CRESTANI F. (2024). Overview of eRisk 2024 : Early Risk Prediction on the Internet. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction : 15th International Conference of the CLEF Association, CLEF 2024, Grenoble, France, September 9–12, 2024, Proceedings, Part II*, p. 73–92, Berlin, Heidelberg : Springer-Verlag. DOI : [10.1007/978-3-031-71908-0_4](https://doi.org/10.1007/978-3-031-71908-0_4).
- SU Y., ZHENG X., LU J., GONG Y., GAO Q., SHEN S. & LIU Q. (2025). Semantic distillation and enhanced diagnostic alignment : A novel approach for depression detection in social media. *Expert Systems with Applications*, **279**, 127346. DOI : <https://doi.org/10.1016/j.eswa.2025.127346>.
- TA N., LI K., YANG Y., JIAO F., TANG Z. & LI G. (2022). Evaluating Public Anxiety for Topic-Based Communities in Social Networks. *IEEE Transactions on Knowledge and Data Engineering*, **34**(3), 1191–1205. DOI : [10.1109/TKDE.2020.2989759](https://doi.org/10.1109/TKDE.2020.2989759).
- World Health Organization (2025). Depressive disorder (depression). <https://www.who.int/news-room/fact-sheets/detail/depression> [Accessed : (03/05/2025)].
- YAN Z., PENG F. & ZHANG D. (2025). DECEN : A deep learning model enhanced by depressive emotions for depression detection from social media content. *Decision Support Systems*, **191**, 114421. DOI : <https://doi.org/10.1016/j.dss.2025.114421>.
- ZHANG J., CHEN Y. & ZHAI Y. (2020). Zero-Shot Classification Based on Word Vector Enhancement and Distance Metric Learning. *IEEE Access*, **8**, 102292–102302. DOI : [10.1109/ACCESS.2020.2998495](https://doi.org/10.1109/ACCESS.2020.2998495).

Annexes

FIGURE 1 – Exemple fictif d'extrait du corpus de la tâche 1 d'eRisk 2024

```
<DOC>
  <DOCNO>5001_0_1</DOCNO>
  <PRE></PRE>
  <TEXT>Je vais bien, mais je n'ai juste pas envie de faire quelque chose</TEXT>
  <POST>Je pense que je vais passer ma journée chez moi</POST>
</DOC>
<DOC>
  <DOCNO>5001_0_2</DOCNO>
  <PRE>Je vais bien, mais je n'ai juste pas envie de faire quelque chose</PRE>
  <TEXT>Je pense que je vais passer ma journée chez moi</TEXT>
  <POST>Arrête de me dire ça, tu vas me faire culpabiliser</POST>
</DOC>
```

TABLE 3 – Exemples d'items du BDI-II avec leurs énoncés associés

Réponses Pré-écrites

1. Sadness – Tristesse

- r1. *I do not feel sad.* – Je ne me sens pas triste.
- r2. *I feel sad much of the time.* – Je me sens triste la plupart du temps.
- r3. *I am sad all the time.* – Je suis toujours triste.
- r4. *I am so sad all the time.* – Je suis tellement triste tout le temps.

7. Self-Dislike – Dégoût de soi

- r1. *I feel the same about myself as ever.* – Je me sens toujours aussi bien.
- r2. *I have lost confidence in myself.* – J'ai perdu confiance en moi.
- r3. *I am disappointed in myself.* – Je suis déçu par moi-même.
- r4. *I dislike myself.* – Je ne m'aime pas.

12. Loss of Interest – Perte d'intérêt

- r1. *I have not lost interest in other people or activities.* – Je ne me suis pas désintéressé des autres personnes ou des activités.
- r2. *I am less interested in other people or things than before.* – Je m'intéresse moins aux autres personnes ou aux choses qu'auparavant.
- r3. *I have lost most of my interest in other people or things.* – J'ai perdu la plupart de mon intérêt pour les autres personnes ou les choses.
- r4. *It's hard to get interested in anything.* – Il est difficile de s'intéresser à quoi que ce soit.