



Université Paul Valéry - Montpellier 3

UFR 6 - Éducation & Sciences pour les LLASHS

Master MIASHS

## Prédiction des comportements suicidaires

### RENDU 4 - REVUE DE LA LITTÉRATURE



Équipe :

Lisa BÉTEILLE  
Matéo CALSACY  
Jean CHABANOL  
Anamé ROUMY  
Laura SÉNÉCAILLE  
Célia TEYSSIER

2021/2022

Encadrants :

Sandra BRINGAY  
Pierre LAFAYE de MICHEAUX  
Florian LOMBARDO

# LE SUICIDE

## A – INTRODUCTION

Une thèse de doctorat, par exemple, commence par dresser un état de toutes les recherches déjà faites dans le domaine. Cette recherche bibliographique permet au futur chercheur de mieux cerner son sujet et de se donner plus de chances de faire de réelles avancées, en évitant de refaire ce qui a déjà été fait, y compris des erreurs expérimentées par d'autres et décrites par la littérature.

C'est ici ce que nous allons essayer de réaliser, un état de l'art des recherches faites sur le suicide. Nous avons à notre disposition différents articles pour nous aider dans nos recherches (voir étape 2).

Tout d'abord, il faut savoir que le suicide fait davantage de victimes que la guerre ou que les catastrophes naturelles selon l'Organisation Mondiale de la Santé (1). En effet, il est la deuxième cause de décès dans le groupe d'âge des 15-29 ans. Il est frappant de constater qu'un quart de tous les suicides du monde se produisent dans deux pays seulement : la Chine et l'Inde, ce qui reflète la taille de leurs populations respectives. La Chine compte à elle seule 20 % des suicides mondiaux.

Les taux de suicide ont augmenté globalement de 60 % dans le monde sur les 45 dernières années. Jusqu'à tout récemment, le suicide était prédominant chez les personnes âgées mais il l'est maintenant chez les personnes plus jeunes (à la fois dans les pays développés et en développement). On peut noter aussi la prédominance des taux de suicide des hommes sur les femmes (excepté en Chine).

Chaque fois que des chiffres sur le suicide sont présentés, leur fiabilité est toujours remise en question selon l'argument que dans beaucoup de pays et pour plusieurs raisons le suicide est caché et que les chiffres réels doivent être beaucoup plus élevés. Ce point est effectivement reconnu par l'OMS, ce qui renforce la gravité des statistiques présentées.

Toujours selon l'Organisation Mondiale de la Santé (OMS), une personne met fin à ses jours toutes les 40 secondes. Le suicide est un problème majeur de santé publique avec au moins 800 000 personnes décédées chaque année dans le monde et au moins 20 fois ce nombre de tentatives suicidaire.

Les voies du suicide sont complexes, le suicide étant le produit final d'une interaction de risques biologiques, cliniques, psychologiques, sociaux, culturels et facteurs de protection. Bien que la connaissance des facteurs de risque de suicide ait considérablement augmenté au cours des dernières décennies, notre capacité à prédire le suicide n'est pas meilleure aujourd'hui qu'il y a 50 ans.

Il y a plusieurs raisons pour lesquelles le domaine du suicide reste un sujet sensible. La recherche n'a pas amélioré sa capacité prédictive, les candidats clés incluent la base basse des taux de comportement suicidaire, c'est-à-dire que les personnes prises en compte ne représentent qu'une petite partie de la population à évaluer (2).

De plus, jusqu'à récemment, il y avait une pénurie de cadres théoriques complets qui ont tenté de comprendre l'émergence des idées suicidaires et la transition de la pensée suicidaire à la tentative de suicide (ou à la mort par suicide).

C'est pourquoi l'OMS a pour objectif de tenter de réduire d'un tiers le taux de mortalité par suicide d'ici à 2030.

## **B - SYNTHÈSE DES ARTICLES**

L'une des conclusions les plus cohérentes en ce qui concerne l'épidémiologie du comportement suicidaire est sa répartition par sexe. Les hommes sont plus nombreux à se suicider que les femmes au niveau mondial, mais beaucoup plus de femmes adoptent un comportement suicidaire qui ne mène pas à la mort.

Plusieurs études suggèrent que les femmes sont plus susceptibles d'être confrontées à de nombreux événements à risque qui augmentent la présence d'appartenance contrariée et la perception d'être un fardeau, ainsi que de fortes dépressions (en effet les femmes sont plus susceptibles que les hommes de souffrir de dépression majeure).

De plus, les données montrent que les femmes classent le fait d'aider les autres, d'avoir une famille proche et d'être aimées par des êtres chers significativement plus haut que les hommes en tant que source de bonheur. Cela suggère que lorsque ces sources de bonheur sont absentes chez les femmes, elles sont particulièrement plus susceptibles de percevoir une appartenance contrariée et une perception d'être un fardeau, ce qui mène à une souffrance émotionnelle plus élevée que chez les hommes pour les mêmes situations.

Cependant, les hommes, en général, craignent moins de s'automutiler, à cause de leurs expériences plus nombreuses (notamment avec des bagarres, des expositions à des armes à feu, des sports violents...), et ont une tolérance à la douleur et à la peur plus élevée que les femmes. De ce fait, les hommes peuvent être plus capables de développer la capacité de mourir par suicide que les femmes.

La théorie interpersonnelle du comportement suicidaire permettrait d'expliquer la prévalence de comportements suicidaires. Mais, comme les conditions de cette théorie (l'appartenance contrariée, la perception d'être un fardeau et la capacité acquise à se donner la mort) sont extrêmement rares, la théorie est cohérente seulement à cause de la rareté du comportement suicidaire.

Jusqu'à présent, nous avons conclu que la présence simultanée des constructions de la théorie est suffisante mais pas nécessaire pour qu'un comportement suicidaire se produise.

D'après la taxonomie de Baechler, les comportements suicidaires semblent chercher à régler un problème, mais que le problème « résolu » par le suicide varie et aboutie à des types de suicides, qui ont chacun des origines étiologiques (8) différentes.

Types de suicides d'après Baechler :

- les suicides schappistes (fuite, deuil, châtement)
- les suicides agressifs (crime, vengeance, chantage, appel)
- les suicides oblatifs (sacrifice, passage)
- les suicides ludiques (ordalie, jeu)

L'une des principales tâches auxquelles font face les cliniciens travaillant avec des patients suicidaires, est l'évaluation du degré de risque auquel sont confrontés chaque patient. L'évaluation du risque fondée sur la théorie interpersonnelle (9) permettra une conceptualisation plus cliniquement utile de l'étiologie (étude des causes et facteurs d'une maladie) du suicide, car cette conceptualisation ne présume pas que l'évaluation du degré de risque de suicide des individus nécessite une mesure (ou une estimation) d'un grand nombre de facteurs de risque.

Les soins cliniques pour les patients suicidaires impliquent également un traitement visant à réduire le risque d'adopter un comportement suicidaire. Les campagnes de santé publique visent également à prévenir les comportements suicidaires en ciblant toutes les personnes présentant un risque élevé de développer des pensées suicidaires ou de se livrer à des comportements suicidaires (promotion de l'importance de maintenir les liens sociaux etc.)

Nous constatons que l'appartenance contrariée et la perception d'être un fardeau (ainsi que le désespoir concernant ces états) sont des facteurs dynamiques (c'est-à-dire qui changent fréquemment), tandis que la capacité acquise, une fois acquise, est relativement stable et immuable.

En théorie, les interventions des cliniciens qui s'attaquent directement ou indirectement à la perception d'être un fardeau et à l'appartenance contrariée devraient produire de meilleurs résultats chez les individus suicidaires. Alors que la capacité acquise à se donner la mort serait relativement difficile à aborder efficacement dans le traitement, puisqu'un thérapeute n'est pas en mesure de modifier les antécédents d'un patient. (4,5,6,7)

## C – METHODE

Nous allons étudier une méthode utilisée, plus particulièrement sur le texte C :

*Dévoilement du comportement suicidaire en ligne : que pouvons-nous apprendre sur la santé mentale des survivants du suicide de Reddit ?* (6)

Nous avons appris que le projet Durkheim a étudié l'activité en ligne et le contenu partagé par des groupes d'anciens combattants américains sur les réseaux sociaux afin d'identifier les signes de comportements malveillants.

Ils ont constaté que ceux qui se sont suicidés enregistrent fréquemment des comportements qui indiquent la peur, l'excitation et le délire avec une précision d'environ 65 %.

H. Sueki étudie un groupe de jeunes utilisateurs de Twitter afin d'évaluer la relation entre le suicide et les tweets relatifs au comportement suicidaire. Ils ont identifié les indicateurs les plus importants d'un futur suicide.

Les étapes réalisées pour obtenir des résultats sont :

#### Étape 1 - Acquisition des données

Tout d'abord ils ont récupéré tous les commentaires du fil '*Survivors*' sur Reddits avec Python. Sur Reddit, tous les commentaires sont organisés en un grand arbre de commentaires.

#### Étape 2 - Prétraitement des données

Ensuite ils ont supprimé les balises HTML, les espaces, les barres latérales et les retours chariot. On parle ici de nettoyage des données plus exactement. Puis ils ont utilisé des expressions régulières pour développer.

Ces étapes sont nécessaires pour marquer correctement les mots et identifier les négatifs.

#### Etape 3 - Extractions de caractéristiques

Puis il a fallu définir les caractéristiques utilisées dans le processus de classification.

- Trigrammes : Le lemme en lui-même n'a pas beaucoup de sens, il faut utiliser les trigrammes pour explorer les phrases et trouver des modèles en double.
- NLTK POS Tags : *tokenized* et *tagged* chaque mot avec NLTK POS tags.
- Balises POS personnalisées : Ils ont défini manuellement les balises dans le tableau. Toutes les balises sont auto-explicatives.
- Method list : Elle définit une liste de 34 méthodes de suicide.
- Liste POSEMO : Elle contient des expressions régulières pour détecter les émotions des personnes sur les réseaux sociaux.

#### Etape 4 – Classification

Enfin ils ont créé une application avec des algorithmes d'apprentissage automatique supervisés pour prédire et extraire automatiquement les méthodes de suicide au niveau des commentaires.

Chaque commentaire contenant au moins une annotation a été extraite.

Ils ont utilisé de 4 modèles : SVM, la régression logistique, SGD et Perceptron.

## D – CONCLUSION

Le suicide est l'un des problèmes de santé les plus courants et touchant la population mondiale, il est donc intéressant d'étudier l'épidémiologie du comportement suicidaire.

Tout d'abord, nos recherches nous ont montré une prévalence au suicide selon le sexe : au niveau mondial, les hommes sont plus nombreux à se suicider que les femmes, mais beaucoup plus de femmes adoptent un comportement suicidaire qui ne mène pas à la mort.

De la même manière, il existe également plusieurs facteurs liés au suicide. Selon la théorie interpersonnelle du comportement suicidaire proposée par Joiner, trois facteurs de risque provoqueraient le comportement suicidaire :

- la perception d'être un fardeau
- l'appartenance contrariée
- la capacité acquise à se donner la mort

Comme les typologies et facteurs de suicide, il existe aussi de nombreux modes de suicide. L'étude réalisée par Ashwin Karthik Ambalavana, Bilel Moulahib, Jérôme Azéc et Sandra Bringay, basée sur le traitement des caractéristiques linguistiques, lexicales et sémantiques de posts Reddit (site communautaire d'actualités sociales) pour classifier les méthodes de suicide, a mené à la conclusion suivante : la méthode de suicide la plus fréquente est la surconsommation de médicaments, suivie par la pendaison et la surconsommation d'alcool (coma éthylique).

D'autre part, il est important de noter qu'il faut différencier les personnes ayant uniquement des idées suicidaires sans passage à l'acte et celles allant jusqu'à la tentative de suicide. Effectivement, il existe une transition entre les idées et le geste suicidaire. Le modèle intégré motivationnel-volontaire (IMV) permet de décrire l'évolution et l'intensification des étapes du processus suicidaire avec les phases suivantes :

- phase pré-motivationnelle (troubles, environnement, événements négatifs)
- phase motivationnelle (ressentis négatifs, difficultés à faire face, sentiment d'enfermement, formation des idées suicidaires)
- phase volitionnelle (passage à l'acte)

Enfin, nous avons pu constater que de nombreuses études montrent que les personnes sont plus susceptibles de rechercher du soutien via l'online qu'en passant par des professionnels. Ainsi, l'évaluation des facteurs de risques suicidaires via les réseaux sociaux peut être utilisée pour prévenir le suicide et les idées suicidaires à ses premiers stades.

# LE SUICIDE DANS LES MÉDIAS SOCIAUX

Les objectifs de notre projet TER sont multiples mais la finalité du projet reste la prévention suicidaire. Nous allons nous appuyer sur les données des réseaux sociaux mis à notre disposition. Ces données proviennent de différents médias sociaux tels que Twitter, Wikipedia ou encore Reddits.

Twitter : Nous avons récupéré des tweets contenant le mot suicide. Ces personnes qui ont écrit ces tweets sont les plus aptes à tenter de se suicider bien que nous ne savons pas si ces personnes iront jusqu'au bout dans leurs actes au final. Avant de mettre fin à leurs jours, certaines personnes laissent des signes annonciateurs sur leur profil Twitter, qui pourrait laisser penser que ces personnes envisagent de se suicider. C'est pourquoi nous avons créé un algorithme pour trouver ces personnes-là, nous avons entraîné notre modèle sur des gens parlant du suicide ou en dépression tout comme sur des gens parlant du bonheur. Nous avons utilisé la technique de classification pour savoir si la personne derrière un tweet sélectionnée sera donc à risque ou non.

Notre algorithme pourra détecter un tweet à risque puis envoyer une notification avec comme but la mise en contact avec des groupes de survivants suicidaires pour apporter un soutien émotionnel. On cherche uniquement à repérer les éventuelles tentatives de suicides pour les signaler.

Wikipédia : Nous avons sélectionné des pages Wikipédia de personnes qui se sont suicidées. En effet sur ce média ce sont des biographies qui sont rédigées, ce qui permet de savoir que la personne s'est bien suicidée. Nous pouvons donc à partir de sa biographie, savoir son parcours de vie et l'étudier pour essayer de trouver des similarités avec d'autres personnes ayant eu recours au suicide. Les personnes regroupées sur Wikipédia sont souvent des personnes connues (à prendre en compte dans nos résultats). Nous avons aussi extrait des pages Wikipédia contenant les mots « happy » ou encore « sport » pour établir une comparaison avec les personnes suicidées et peut être prévoir si une personne encore en vie, présente sur le média Wikipédia pourrait ou non arriver au suicide.

Reddits : Nous avons à notre disposition des posts Reddits contenant le mot suicide. Nous pouvons trouver à la fois des gens parlant du thème du suicide comme trouver des gens en dépression qui sont amenés de mettre fin à leurs jours. Nous ne savons pas sur ce média si la personne qui a rédigé ce post s'est finalement suicidée ou non.

## OBJECTIFS DU TER

Notre projet pour cette année sera donc de mettre en place un algorithme qui identifie les textes présentant un risque de tentative suicidaire à partir de toutes les données dont on dispose (données tirées de Reddit, Wikipédia et Twitter). Pour cela nous utiliserons des techniques de classification. Cet algorithme devra prendre en compte toutes ces données, mais il aura une performance différente en fonction de la provenance de celles-ci. Cet algorithme devra, selon la source et le format des données, être le plus efficace possible. Nous avons plusieurs catégories de populations pour chaque média social, cela permet de comparer pour un même média différents types de personnes (pour entraîner notre modèle de classification).

Notre travail devra également s'appuyer sur les travaux des années précédentes. Dans un premier temps, les corpus de texte analysés feront référence uniquement au suicide et non au mal être d'une personne (texte traitant des idées noires par exemple). Mais il serait intéressant pour la finalité de pouvoir analyser également ces mal êtres. Pour réaliser cela nous utiliserons une méthode différente : le Deep Learning (10).

Cette année sera donc divisée en 2 parties distinctes :

Le 1er semestre, nous allons comparer les trois types de données sociales (Twitter, Wikipédia et Reddit).

Le 2ème semestre, nous allons étudier la transférabilité d'un modèle appris sur un type de données vers un autre.



## BIBLIOGRAPHIE

- (1) OMS communiqué de presse, le 09/09/2019  
<https://www.who.int/fr/news/item/09-09-2019-suicide-one-person-dies-every-40-seconds>
- (2) Conduites suicidaires  
<https://www.conduites-suicidaires.com/facteurs-de-risque/differents-facteurs/#introduction>
- (3) Les Suicides (1975)  
*Jean Baechler*
- (4) Texte A : Détection des idées suicidaires dans les forums de médias sociaux à l'aide de Deep Learning (2019)  
*Michael Mesfin Tadesse, Hongfei Lin, Bo Xu and Liang Yang*
- (5) Texte B : L'intégration motivationnelle/volontaire - modèle de comportement suicidaire (2018)  
*O'Connor RC, Kirtley OJ*
- (6) Texte C : Dévoilement du comportement suicidaire en ligne : que pouvons-nous apprendre sur la santé mentale des survivants du suicide de Reddit? (2019)  
*Ashwin Karthik Ambalavana, Bilel Moulahib, Jérôme Azéc, Sandra Bringay*
- (7) Texte D : La théorie interpersonnelle du suicide (2010)  
*Kimberly A. Van Orden, Tracy K. Witte, Kelly C. Cukrowicz, Scott Braithwaite, Edward A. Selby, Thomas E. Joiner Jr.*
- (8) Étiologique : Relatif à l'étiologie, à l'étude des causes des maladies  
*Le Robert*
- (9) La théorie interpersonnelle  
Selon Kiesler et Watkins (1989), la personnalité humaine peut être comprise comme étant un pattern d'interactions interpersonnelles récurrentes. De manière similaire, l'attachement peut être également appréhendé comme étant un pattern d'interactions interpersonnelles récurrentes.  
(<https://www.cairn.info/>)
- (10) Deep learning :  
C'est un type d'intelligence artificielle dérivé du machine learning où la machine est capable d'apprendre par elle-même, contrairement à la programmation où elle se contente d'exécuter à la lettre des règles prédéterminées.  
(<https://www.futura-sciences.com/>)