NLP Applied to Online Suicide Intention Detection

Mathieu Guidère

NLP Applied to Online Suicide Intention Detection

Guidere M.[1], Fluhr C., Rossi A., Wang Z.[2]

[1]National Institute of Health and Medical Research (INSERM), Paris, FRANCE
[2]GEOLSemantics, Paris, FRANCE

Abstract

Combining linguistic data and behavioral sciences, we use NLP to implement machine learning and inspect social media in a suicide surveillance system. We have conducted prospective studies to understand the linguistic expressions, and discourse features of suicidal subjects. The goal of this research was to build a machine learning processing using the linguistic characteristics. To achieve this, we applied machine learning classifiers on linguistic data captured from heterogeneous sources (blogs, websites, forums, social networks, etc.). The captured data were then used for training machine on information extraction in order to identify linguistic markers of suicide. In this paper, we provide an overview of the automated tracking and monitoring system for suicidal ideation and risk, which draws on predictive linguistics methods and techniques, based on a large sample of suicide messages posted online.

Key words: predictive linguistics, machine learning, suicide surveillance, online media, monitoring system, demo.

Introduction

Suicidal behavior takes over a million lives worldwide every year. Non-fatal suicidal behavior is estimated to be 25 to 50 times more common. So finding ways to identify risky behavior is a key public health goal, but the existing predictors simply do not work well, especially in identifying short-term risk (Reardon, 2013). The poor performance of predictors may relate to how well they are identified and tracked. That is why we propose a new method for monitoring suicidal behavior based on predictive linguistics and natural language processing (Guidere, 2019).

Predictive Linguistics (PL) is an emerging discipline that studies the interrelation between linguistic markers and action plans considered as an intentional, conscious and subjectively meaningful activity. Studying prediction in language aims at forecasting what would happen in the future based on a rigorous analysis of linguistic data and markers in the present (Guidere, 2015). A primary concern of PL is to analyze the way language may predict human actions and explain the relationship between time and behavior within language, notably how a future action is expressed in present language. Experimental methods indicate that combining statistic word processing and automatic discourse analysis, under certain conditions, enable to predict upcoming actions and mind states. The combination of predictive linguistics and predictive modeling enables us to recognize patterns and trends within data.

Suicidal Post Monitoring

In order to analyze the data, we use semantic similarity which is a metric defined over a set of posts, where the idea of distance between them is based on the likeness of their meaning or semantic content. Computationally, semantic similarity is estimated by using ontologies to define the distance between concepts, and can be extended to key words, or categories of words, that can reveal something about predictiveness.

For the monitoring of suicide posts online, the system was trained on several "Suicide Forums" available in French. These forums contain a huge number of suicide posts and personal messages posted online by people who think about suicide or/and who attempted suicide. These forums include "I would like to die" (J'ai envie de mourir[1]),

[1] See the forum: https://www.filsantejeunes.com/forum/sante/j-ai-envie-de-mourir-t291598.html

"Get by or Die" (S'en sortir ou mourir[2]), "I think of suicide" (Je pense au suicide[3]).

Based on these forums, the "French Suicide Posts" (FSP database) has been built automatically from a collection of suicide messages posted on the World Wide Web. The database contains over 1000 posts, written by people who attempted or completed suicide. These posts were identified based on linguistic markers with an emotional connection to the subject of suicide (such as sorrow, blame, guilt, hopelessness, etc.).

The point of this was to train the machine at the way a suicide post is written and the way people talk about the posts on the internet forums. As a matter of fact, some linguistic markers are found to be common across most of the suicide posts that are distinctly about referencing the future, including the discovery of the suicide. They may be further distinguished by their use of specific discourse markers.

Investigating Suicidal Thoughts and Intentions

Suicidal thoughts, or suicidal ideation, means thinking about suicide. The "FSP" database shows that these thoughts can be identified in discourse when some people express feelings such as: "Feel hopeless" (se sentir désespéré); "Feel intolerable emotional pain" (ressentir une souffrance émotionnelle intolérable); "Feel mood swings" (ressentir des changements d'humeur); "Feel unable to experience pleasurable emotions" (être incapable de ressentir du plaisir); "Feel very lonely" (se sentir très seul); "Feel very anxious" (se sentir angoissé); "Feel depressed" (se sentir déprimé).

The suicidal intention means planning suicide but does not include the final act of suicide. It can be identified in discourse when some people talk about: "suicide and dying or death" (le suicide, la mort, mourir); "being a burden to others" (être un fardeau pour les autres); "revenge, guilt, or shame" (la vengeance, la culpabilité, ou la honte); "Express regret about being alive or ever having been born" (exprimer le regret d'être en vie ou meme d'être né); "Say goodbye to others as if it were the last time" (dire adieu comme si c'était la dernière fois).

The intention can range from a detailed plan to a fleeting consideration of a suicidal attempt. It can be identified when some people: give things away; get suddenly their affairs in order; get hold of medications, or dangerous substances; start drinking or consuming drugs or more alcohol than usual; engage in risky behavior.

Based on these observations, predictive linguistics attempts to extract patterns from the available data:

1) Classification patterns: When the person feels hopeless, anxious, lonely, and moody, he is probably suffering from depression;

2) Associative patterns: When the person talks about "burden, revenge, guilt, shame, remorse", he is likely to move to "violence, dying, or death";

3) Sequential patterns: After feeling lonely, the persons will live in increased isolation, then get their affairs in order, start drinking or consuming drugs then they will engage in risky behavior.

Linguistic Markers of Suicide Posts

When it comes to the pursuit of finding out what people write about, predictive linguistic monitoring is a better method to prevent the risk for suicidal behavior because it is based on the assumption that there is a cognitive feature present, since the suicidal people are using similar structure and language, although they can have no awareness that they are doing so.

For example, when it comes to the expression of "love" (amour), the first person pronoun is always present: "I / je", "my / mon", "am / suis" and "me / moi" (Self-referring words). The pronoun "you / vous" and the word "love / amour" might indicate the writers' concern with their recipient: "I love you and hate myself for doing this" (je vous

[2] See the forum: https://www.filsantejeunes.com/forum/sante/s-en-sortir-ou-mourir-t290303.html
[3] See the forum: https://www.filsantejeunes.com/forum/sante/suicide-t291473.html

aime et me déteste de faire ceci); "I love you and always will" (je vous aime et vous aimerais toujours); "I send you my fond love" (je vous adresse mon amour profond); "Tell them how I love them" (dis-leur combien je les aime); "Remember I will always love you" (souviens-toi que je t'aimerais toujours).

This expression of love can be stressed with an intensifier such as "All": "I love you all" (je vous aime tous); "I love you all more than you'll ever" (je vous aime plus que vous ne le pensez); "All I can say I love you to bits" (tout ce que je peux dire est que je vous aime énormément); "All my love to all" (tout mon amour à vous tous).

As we can see, concern with the "Self" (I), the "Recipient" (You) and "Affection" (Love) are typical of most of the suicide posts, and thus may be a predictor of the suicidal behavior.

There are other linguistic features present in the suicide posts such as the expression of "forgiveness" (pardon): "Please forgive me, do not want to be burden" (s'il vous plaît, pardonnez-moi, je ne voudrais pas être un fardeau); "Please forgive me and don't think I'm a coward" (s'il vous plaît, pardonnez-moi et ne croyez pas que je suis un lâche); "Please forgive me for deceiving you all" (s'il vous plaît, pardonnez-moi de vous avoir tous déçus); "Please forgive me, I don't want to live anymore" (s'il vous plaît, pardonnez-moi, je ne veux plus vivre davantage); "Please forgive me, I can't stand anymore" (s'il vous plaît, pardonnez-moi, je n'en peux plus du tout); "Please forgive me for letting you all" (s'il vous plaît, pardonnez-moi de vous avoir laissé tomber); "Darling, please forgive me, I am too much in pain" (ma chérie, s'il te plaît, pardonne-moi, je souffre trop).

The expression of "sorrow" (désolation) is also typical of the suicide posts: "I'm sorry for the trouble I've caused" (je suis désolé pour le souci que j'ai cause) ; "I'm sorry, I really am" (je suis désolé, vraiment désolé); "I'm sorry, I'm so very very sorry!" (je suis désolé, vraiment vraiment désolé); "I'm sorry, you have all tried so hard" (je suis désolé mais vous avez tout essayé); "Sorry for letting you down" (désolé de vous laisser tomber ainsi); "I'm so sorry" (je suis vraiment désolé); "I'm very sorry for doing this" (je suis vraiment désolé de faire ça); "Again I'm sorry" (encore une fois, je suis désolé).

The predictive expressions of intention include terms about wanting, desire and aspiration: "I hope you will understand me" (j'espère que vous me comprendrez); "I hope you will remember me sometimes" (j'espère que vous vous rappellerez de moi parfois); "I hope you won't forget me for a little time" (j'espère que vous ne m'oublierez pas de sitôt); "I hope in time you will forgive me" (j'espère que vous me pardonnez avec le temps); "I hope you can survive what I have done" (j'espère que vous survivrez à ce que j'ai fait).

This expression of "wish" (espoir) can be stressed with different intensifiers such as: "I sincerely hope the future will hold better time" (j'espère vraiment que l'avenir sera meilleur); "I really do hope that you find a better man" (j'espère vraiment que tu trouveras un homme meilleur); "I wish them all the best in the future" (je leur souhaite à tous le meilleur pour l'avenir); "I wish you all the happiness in the world" (je vous souhaite tout le bonheur du monde).

As we can see, there is a potential for a group of predictive linguistic markers about "future" which could include the word "will" (vouloir), "hope" (espérer), "wish" (souhaiter), associated with this type of structure: "I shall no longer" (je ne serais plus), "shall no longer be" (ce ne sera plus), "no longer be here" (ne sera plus ici).

Further, there is the potential for a group of predictive linguistic markers about negativism and pessimism. This would include the word "can't" (peux pas): "I can't handle it no more" (je ne peux plus le supporter); "I can't see any other way out" (je ne vois plus d'autre issue); "I can't seem to understand anything" (je ne semble plus rien comprendre); "I can't take the waiting any more" (je n'en peux plus d'attendre); "I can't come to terms with it, can't" (je ne peux pas m'y faire, je ne peux pas); "I can't face it any more, can't" (je ne peux plus l'affronter, je ne peux plus).

What this shows is that using "negativism" and "pessimism" structures is indicative of a text's probably being a suicide post. The presence of these linguistic markers as a frequently-occurring terms merely means that suicide posts include a large quantity of words and lemmas such as "not" (pas), "nothing" (rien), "none" (aucun) and "no one" (personne).

Using Linguistic Markers for Online Monitoring

The use of the above mentioned linguistic markers within the semantic engine enables to detect many suicide messages on Twitter. Here are some samples[4]:

2019-10-28,13:47:06,"Paris, Madrid",***,***,***, ''de toute facon ma vie a tjrs été dla merde jperds tjrs les prsn qui comptent le plus pour moi la jevais me suicider,,, (*anyway my life has always been shit i always lose people that mean the most to me i will kill myself*)

2019-10-27,23:46:50,"Paris, Madrid",***,***,***, "Je te le dit maitenant , moi j'ai déjà essayer de me suicider , avec un couteau , bon à cause de cette merde j'ai une cicatrice au bras ,et sa sert a rien hein de se suicider , j'y est réfléchi longtemps très longtemps, au bous du compte sa sert a rien". (*I tell you now, I already tried to kill myself, with a knife, well because of this shit I have a scar on my arm, and it is useless to commit suicide, I think about it for a long time very long, in the end it is useless*)

Messages on Twitter (tweets) have some special features:

1) a tweet has a strict constraint in content size: consisting of up to 280 characters (Perez, 2017);

2) a text of a tweet has a large spelling variable, such as numerous abbreviations and, misspellings, and emojis.

Therefore analyzing such text needs an analysis of the forum on which it is posted as a general context for the processing of the meaning.

Our system uses a 2-steps approach to detect suicidal ideation markers: 1) a global linguistic analysis (GLA) which allows simplifications in the variety of possible structures of the language and a better disambiguation and relevance; and 2) an ontology-based knowledge extraction (OKE).

The global linguistic analysis is domain-free. It aims to represent a text with lemmas and binary relations. For example, the sentence "It's killing me" (ça me tue) will be analysed as the following:

Words: "it" (ça), "kill" (tuer), "me" (me), and

Relations: "Subject-Verb kill" (Subject-Verb tuer), "Verb-Object kill me" (Verb-Object tuer me).

In addition, the analysis takes into account modalities and negations, to catch the feelings expressed in the text, and the object of these feelings.

As shown in the following samples (1) and (2), modalities and negations are essential to detect alarming messages:

(1) "I don't feel well." (Je ne me sens pas bien.)

(2) "I want to die." (Je veux mourir.)

Using such deep linguistic analysis allows to avoid some ambiguities and misunderstandings: for example, in sentences like "I'm going to kill myself" (je vais me tuer) instead of "I will kill myself overworking" (je vais me tuer au travail).

But processing messages such as tweets implies new difficulties: dealing with the numerous abbreviations and misspellings. For this purpose, we use specific dictionaries for the most common abbreviations and specific idioms, and a phonetizer for the misspellings, a specific algorithm to cut concatenated words, and specific rules to capture syntactic structures particular to such messages. For example:

(3) "Chui dég. Jvai me tué. " (Im sik'nd gona kil mi self: I am sickened, I am going to kill myself.)

The second level of analysis, ontology-based knowledge extraction (OKE), is domain dependent. A domain ontology is created with a "Formal Concept Analysis methodology" (Ganter & Wille, 1999) in order to represent the criteria that indicate suicidal thoughts. This step is critical because it will determine the final result, i. e. alerting, and correctly visualizing concepts contained in a message.

[4] The demo is available online on this link: http://geoldemo.com:8080/demohealth/

For the demonstration purposes, we focused on the representation of “Discomfort” and “Suicidal Thought” concepts. But our algorithm also takes into account family and professional status, medical situation, etc. We represent also the negative and first person markers within the posted messages.

Based on this ontology, we wrote semantic rules to catch the text content. These rules are defined on semantic basis and not only with string characters. Therefore the ontology enables to group together expressions in a same semantic concept, aiming to visualize the situation and its criticality, and then to use probabilities to give an alert. The warning system is based on the presence of concepts like “Discomfort”, “Suicidal Thought”, “Family”, “Health”, “Love”, and “Violent Act”. But expressions like this are frequently used in an ironic or metaphoric way on tweets: “It’s raining, I’m going to kill myself.” (Il pleut, je vais me tuer). To try to catch this reality, we ponderate the weight with the length (number of words) of the message.

The linguistic markers of suicidal ideation are represented in rules as the following:

(4) Subjective_1st_person_personnal_pronoun “se tuer” ø

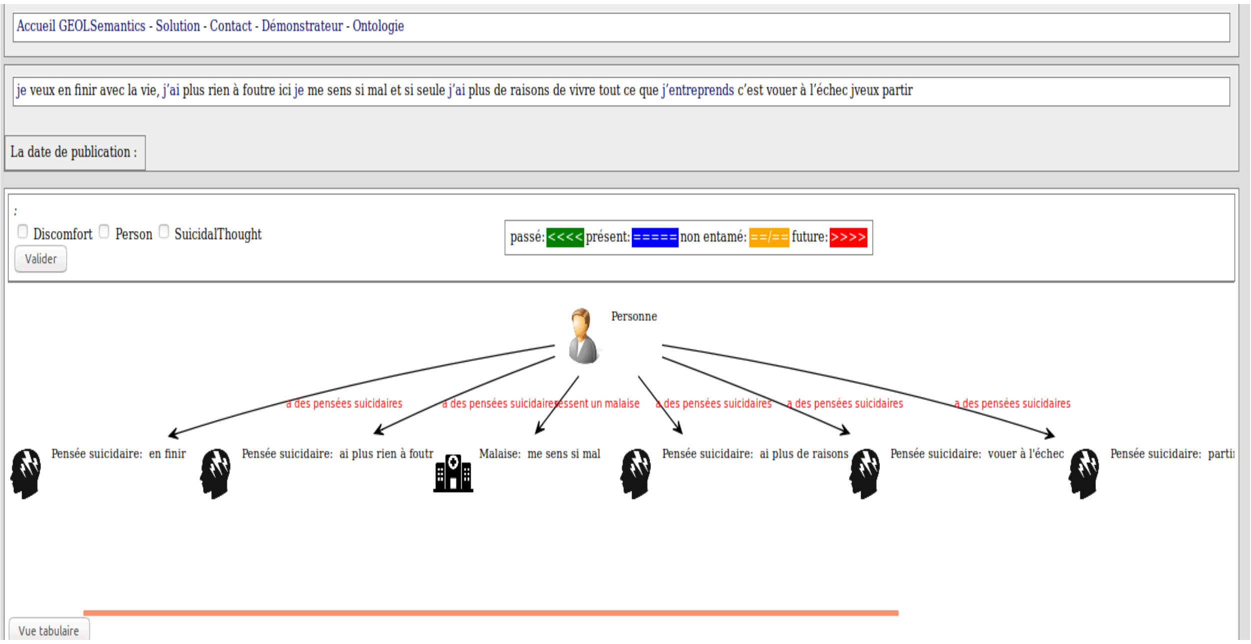
This rules captures “Je veux me tuer” (I’m going to kill myself), but they don’t capture “ça me tue” (it’s killing me) nor “je me tue à la tâche” (I am killing myself to the task).

The result of these 2 steps is a set of semantic concepts extracted from a text. To interpret them, we implement 2 complementary tools:

1) an early-warning system, to report to experts messages with alarming content,
2) a graph representing the content of a message, in a mind map style.

Our current interface enables us to show all the concepts in a message. We work now on improving this visualization to adapt it to the experts’ point of view.

An intuitive way of visualizing the predictive linguistic markers is by gathering terms which are closely related and spacing wider apart the ones which are distantly related. This is also common in practice for cognitive mapping (mind maps and concept maps).


Accueil GEOLSemantics - Solution - Contact - Démonstrateur - Ontologie

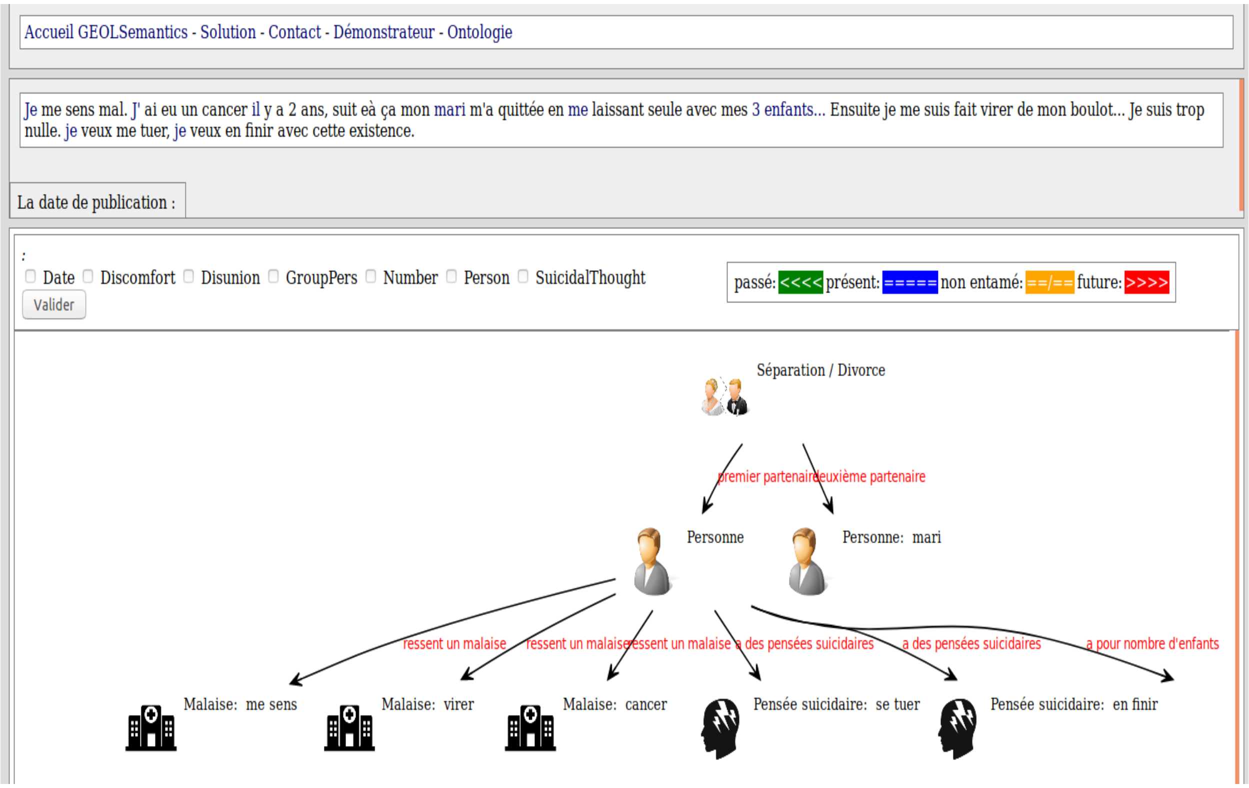
je veux en finir avec la vie, j’ai plus rien à foutre ici je me sens si mal et si seule j’ai plus de raisons de vivre tout ce que j’entreprends c’est vouer à l’échec jveux partir

La date de publication :

:

□ Discomfort □ Person □ SuicidalThought
Valider
passé: <<<< présent: ===== non entamé: ==/== future: >>>>
Personne
a des pensées suicidaires
a des pensées suicidaires
ressent un malaise
a des pensées suicidaires
a des pensées suicidaires
a des pensées suicidaires
Pensée suicidaire: en finir
Pensée suicidaire: ai plus rien à foutr
Malaise: me sens si mal
Pensée suicidaire: ai plus de raisons
Pensée suicidaire: vouer à l'échec
Pensée suicidaire: partir
Vue tabulaire

Figure 1. Graph visualization of a tweet with a suicidal thought.


Accueil GEOLSemantics - Solution - Contact - Démonstrateur - Ontologie
:
□ Date □ Discomfort □ Disunion □ GroupPers □ Number □ Person □ SuicidalThought
Valider
passé: <<<< présent: ===== non entamé: ==/== future: >>>>
Séparation / Divorce
premier partenaire
deuxième partenaire
Personne
Personne: mari
ressent un malaise
ressent un malaise
ressent un malaise
a des pensées suicidaires
a des pensées suicidaires
a pour nombre d'enfants
Malaise: me sens
Malaise: virer
Malaise: cancer
Pensée suicidaire: se tuer
Pensée suicidaire: en finir

Je me sens mal. J' ai eu un cancer il y a 2 ans, suit eà ça mon mari m'a quittée en me laissant seule avec mes 3 enfants... Ensuite je me suis fait virer de mon boulot... Je suis trop nulle. je veux me tuer, je veux en finir avec cette existence.

La date de publication :

Figure 2. Graph visualization of a tweet with a suicidal intention.

Evaluation

As our solution is an online alert system for suicide ideation expressed on Twitter, the difficulty of the evaluation is to collect a significant corpus. Indeed, the access and validity of real suicidal messages on Twitter is restrained. Our system is trained with significant idioms that express discomfort and suicidal thoughts from the FSP database. We've made requests on Twitter with these markers, and we've created a corpus of 4792 posts. The corpus is not larger, because we have to manually validate the results, then verifying which messages might express suicidal thoughts.

We have applied our algorithm to the corpus and compared the results automatically obtained with the reference. See details in the table 1.

Table 1: Evaluation of suicidal thought detection with precision and recall.

Precision	**Recall**
0.55	0.93

We obtained a very good recall score, but the precision score is lower than we expected. First, the language used in tweets is very hard to analyze. We have to improve the analysis to resolve the numerous mistakes they contain (verb conjugation, misspellings,. . .), like "a"/"à", "tué"/"tuer". Second, idioms that express discomfort or suicidal thought are commonly used in an exaggerated manner, for example : "jcomprends rien à mon dm de maths jv me suicider" (I don't understand my math homework I will kill myself). Here the context is sufficient to exclude the possibility of a suicidal ideation, but it's not always the case, as we explain below. Third, because of constraint in content size, a tweet focuses on only one topic. This is a specialty that we have to take into account in our next improvement. Our algorithm is enforced by adding negative criteria to ignore ironic messages, for example when the post talk about sport : "J'veux me suicider c'est quoi ce match???" (I will kill myself what's this match???).

The second point is that sometimes, posts are too short to give an opinion. That's the case for tweets containing only a suicidal expression. These are not isolated instances, they represent 15 % of our corpora : "je vais me tuer" (I will kill myself). The study of other messages sent by the same person will help to take a decision. This is the context of short posts sent on Twitter, where suicidal idioms are often used in a metaphoric way, which explains the lower precision we obtained for now.

Conclusion

Online suicide posts can contain many things with some of them distinctive from other texts. They are using the future, and may be further identified by their greater than normal use of pronouns, negativism and pessimism markers, and intensifiers.

Some of the linguistic markers may contain explicit expressions of love, regret, blame and guilt. They are also likely to contain mentions of events and memories.

Future work could look at transcriptions of suicide notes made on audio or video tapes. These might contain interesting differences from written suicide posts. It could also investigate correlations with non-linguistic variables by looking at the distribution of any linguistic findings between various combinations of suicide post-writers' styles, ages, reasons, etc. Predictive linguistics' software would make it possible to include many such attributes as a matter of course.

For the specific case of a suicide surveillance system based on forum posts, the future works will concentrate in studying messages send by a same person, to raise the doubt about its intentions, and improve visualization interface to show the post content.

References

Blaszczak-Boxe, A. Bullying linked to suicidal behavior in adolescents. *Live Science Retrieved*, March 13, 2014.

Courtet, P., Gottesman, F., Jollant, and Gould, T.D. The neuroscience of suicidal behaviors: what can we expect from endophenotype strategies? *Translational Psychiatry*, 2011.

Farzindar A. and Roche M. Les défis de l'analyse des réseaux sociaux pour le traitement automatique des langues. *Revue TAL-Traitement Automatique des Langues*, 54(3) :7–16, 2013.

Ganter, B. B., & Wille, R. (1999). *Formal concept analysis, mathematical foundation*. Berlin: Springer Verlag.

Guidère M. *La linguistique prédictive : de la cognition à l'action*, Paris: Editions L'Harmattan, 2015.

Guidère M., *Towards an automated surveillance system for suicide prevention*, 4th Mental Health for All, Canadian Association of Mental Health Conference, Toronto, Canada, 23-25 September, 2019, Session / Workshop H6, p.76.

Howard N. & Guidère M. *LXIO The Mood Detection Robopsych*. *The Brain Sciences Journal*, 1(1), 98-109, 2012.

Kontopoulos Efstratios, Berberidis Christos, Dergiades Theologos, BassiliadesNick, *Ontology-based sentiment analysis of twitter posts. Expert Systems with Applications*, Volume 40, Issue 10, August 2013, Pages 4065-4074

Reardon, S. Suicidal behaviour is a disease, psychiatrists argue. *New Scientist*. 2013.

Zachary C. Steinert-Threlkeld. *Twitter as Data. Elements in Quantitative and Computational Methods for the Social Sciences*. Cambridge University Press, 2018.

Patents:

LXIO: Predictive Linguistic-Based Mood Detection Robopsych.
https://patentscope.wipo.int/search/fr/detail.jsf?docId=US73511719&_cid=P12-JZZ9W7-76086-1

GEOL: Linguistic Analysis-Based Geopositiong System.
https://patentscope.wipo.int/search/fr/detail.jsf?docId=WO2011012834&_cid=P12-JZZ9W7-76086-1