HOW DO I SAY "SAD?" BUILDING A DEPRESSION-LEXICON FOR *PSYCHOLOGIST IN A POCKET*

Roann Munoz Ramos[1], Paula Glenda Ferrer-Cheng[2],
Jó Ágila Bitsch[3] and Stephan Michael Jonas[1]
[1]Department of Medical Informatics, RWTH Aachen University Hospital (Germany)
[2]The Graduate School, University of Santo Tomas (Philippines)
[3]Communication and Distributed Systems, RWTH Aachen University (Germany)
[1]Department of Medical Informatics, RWTH Aachen University Hospital (Germany)

Language can be used as an indicator of psychological health. The study of language, particularly text analysis, assumes, among others, that mental states are reflected in the words we use and that certain language features can provide markers for mental illnesses. In studying depression and its characteristic symptoms, everyday language may reveal cognitive mechanisms, such as negative schemas and self-focus ruminations.

An emerging trend in health care, mobile health (mHealth) uses wireless (mobile) technology in medicine and public health. Extending its functions in the field of psychological health, mental mHealth augments clinical assessment via real-time measurements. It also reduces clients' recall bias and fear of being stigmatized. Although mental mHealth software application have been largely concentrated outside Asia, the popularity of mobile phones in the Philippines is perceived as a promising aspect in the development of mental mHealth applications, such as the *Psychologist in a Pocket* (PiaP) **(**Ramos, Winter, Smith & Bitsch, 2012)**.**

Despite the growing recognition of depression in the Philippines, apprehension surrounding mental illness prevents individuals from taking necessary steps towards depression assessment and treatment. To partially address this, *PiaP,* as an adjunct to initial psychological assessment, aids in screening depressed mood non-intrusively using text analysis technology on the smartphone device itself. Text inputs are gathered and logged in real time and compared against a predefined set of keywords indicative of depression. The language behavior is captured over a longer period of time, thereby creating a more complete picture of the user's experience and reducing recall bias.

This study presents the first part of our on-going validation of *PiaP*. In building the *PiaP* lexicon in English and in Filipino, we utilized two general approaches in order to represent the entire domain of depression and its 13 symptom categories based on DSM and ICD: top-down (deductive) and bottom-up (inductive). Words for analysis were derived from focus group discussions with university students, interviews with mental-health professionals and the review of psychological tests. In addition, we considered cultural idiosyncrasies in text writing (e.g., use of spelling variations/shortened words) and current expressive styles found in social media (e.g., use of emoticons and emojis). As a result, the lexicon of the current *PiaP* prototype is composed of over 11,400 main keywords and its derivatives and more than 800,000 spelling variations.

Keywords: depression, Psychologist in a Pocket, mHealth, lexicon development, text analysis

---

1. Introduction

Whether stating facts, asking questions or exchanging ideas, we use language as a means to communicate. Being social animals, language is instrumental in developing and maintaining relationships. In addition, language (and the words we employ) provides suggestions as to our educational background, social standing, various motivations and other (more or less) conspicuous personal circumstances.

From a psychological perspective, our use of everyday words reflects on-going thoughts and emotions, which we may or may not be fully aware of. Freud and psychoanalysis, for example, view that the contents of one's unconscious – the repository of repressed, distressing psychological contents and processes - are expressed in slips of the tongue (verbal slips or parapraxes). The aim of projective techniques, which require participants to, among others, fill in incomplete sentences, draw common objects or weave a story, is to externalize (project) unpleasant feelings or experiences. The use of vague stimuli in this technique allows the unburdening of thoughts or emotions that are difficult to articulate.

This perspective that language can be a diagnostic approach to gauge mental and emotional health has been further scrutinized using more precise approaches. For instance, one methodology is quantitative text analysis or the statistical analysis of word use. A specific example is word count strategies, involving simple word counts of standard grammatical units (e.g., personal pronouns) and psychologically derived linguistic dimensions (e.g., emotion words) (Pennebaker, Mehl & Niederhoffer, 2003).

Advancements and rapid changes brought about by technology have heralded greater scientific sophistication and yielded newer paradigms in the psychological analysis of word usage. One illustration is computerized text analysis, the assignment of scheme categories to text on pre-defined variables with the aid of computers. Programs that emerged in the 1960's, such as General Inquirer (Stone et al, 1966) and WORDS (Iker and Klein, 1974), were significant in providing the impetus towards this direction. The Linguistic Inquiry and Word Count (LIWC), which appeared in 2001, is a transparent text analysis program that counts words in psychologically meaningful categories (Tausczik & Pennebaker, 2010). Pedesis (Neuman et al., 2012), a system that screens for depression proactively and automatically in texts, identifies metaphors associated with depression. Another system, Emotex, identifies and classifies emotions in Twitter messages (Hasan, Rundensteiner & Agu, 2014).

1.1 Characteristics of Language among the Depressed

For the past years, there has been heightened awareness of the strain that mental illness brings – in terms of individual costs, social impact and economic liability. In addition, it affects physical health and can be more impairing than common chronic medical disorders (Kazdin & Rabbit, 2015; Druss et al., 2009). Depression, which is chronically disabling, is ranked as the third most common among mental and health diseases (World Federation for Mental Health, 2011) and is predicted to be the number one disability in 2030 (WHO, 2008).

Everyday language use among individuals with depression may serve as a marker of the illness. Beck's Cognitive Theory explains the role of negative schemas in the consistently pessimistic way depressed individuals view themselves. Weintraub (1981) notes that the frequency of first-person singular pronouns is associated with one's level of depression. Furthermore, the Self-Focus Model of Depression (Greenberg & Pyszczynski, 1986) emphasizes the inclination to ruminate and to construe events in terms of themselves. Such attention placed on one's self triggers negative self-schema. Individuals who are experiencing or had experienced depression may be identified through their preoccupation with negative thoughts and a heightened sense of self-awareness (Rude, Gortner, and Pennebaker, 2004). The dimensions of symptom-specificity in depression (affective, cognitive, somatic) are likewise reflected with the use of affective-, cognitive-and somatic-related words in natural spoken or written language (Vanheule, Desmet and Meganck, 2009).

Text, emails and social media, which have become communication staples, can provide us with an understanding of emotional expressions among depressed individuals in their written texts. Social media postings can be valuable measures of depression symptoms, such as high self-focus attention, heightened relational concerns, decreased social activity and raised negative affect, that signal its onset (De Choudhury et al., 2013). Analysis of texts can be derived from social network sites like Facebook and Twitter. Studies have shown that college students may reference their feelings of depression in their status updates in Facebook. In addition, the online reinforcements and comments to their status updates allow them to publicly discuss their experiences via Facebook (Moreno et al, 2011).

1.2 Mental Mobile Health

The pervasiveness of mobile technology opens an innovative avenue for the enhancement of health care practice. Mobile health capitalizes on mobile technology for more efficient health monitoring and generation of health data. Advantages of this approach are especially felt, for instance, when health practitioners have to remotely monitor patients' health and housebound patients to be in-touch with their clinicians.

Knowing the seriousness of depression and other mental illnesses, mobile technology can be tapped to address mental health concerns. Applying mobile health to mental health assessment may augment the conventional procedures of direct interactions between doctors and patients (e.g., face-to-face interviews, clinical observations). Responses to paper-and-pencil tests may be affected by recall errors and misperceptions. With this technique, greater objectivity is attained, since such errors and biases are reduced or eliminated.

1.3 Psychologist in a Pocket

Psychologist in a Pocket (PiaP) is an Android smartphone application (app) for screening depression symptoms via text analysis (Bitsch et al, 2015). It relies on Ecological Momentary Analysis (EMA) in capturing behavior – text inputs in electronic communications – at the moment of occurrence or in real-time. Advantages of adopting this approach include the ability to detect fleeting moods and to gather data in various settings passively. Most importantly, such an approach allows for privacy.

PiaP is not aimed at replacing human mental health professionals. Instead, it is designed as an adjunct in depression monitoring and assessment. To address privacy and security concerns, the app neither requires any Internet connectivity nor permission to access the Internet. PiaP data is gathered and evaluated only in one's local device. Unless given explicit consent, user input data will never be uploaded to a cloud service. Optional plug-ins may be installed only if the user wishes to inform a trusted third party of his/her mental health status.

2. Objectives

This paper focuses on the process of lexicon development used in PiaP. We describe the top-down and bottom-up approaches to build the corpus of words related to depression, in English and in Filipino, which serve as basis of our application's text analysis. We likewise considered the particular language culture of our target population (adolescents and young adults) in establishing our lexicon.

3. Methods

3.1 Use of Established Classification Systems

We adopted the symptom description of Depressive Episode based on the DSM-5 and ICD-10 as basis for PiaP's lexicon categories. All words were classified into at least one of the 13 symptom categories: Mood, Interest, Anxiety, Guilt & Self-esteem, Fatigue, Appetite & Weight, Concentration, Sleep, Psychomotor Agitation, Psychomotor Retardation, Histrionic Behavior, Alcohol & Substance Abuse, and Suicide.

3.2 Review of Depression Tests

We analyzed 18 commonly used depression tests (e.g., Beck's Depression Index (BDI)-II, Patient Health Questionnaire (PHQ)-9, Inventory to Diagnose Depression (IDD)). Frequencies for words, phrases and sentences from these sources were computed and assigned to a PiaP keyword category.

3.3 Cultural Considerations in Expressing Depression

To encompass a more accurate representation of the linguistic expressions among university students, we conducted seven FGD sessions (45-60 minutes each) with 76 students from Manila (mean age: 17.28 years; female: 61%). Based on BDI–II scores, their depression symptoms ranged from mild (42%) to severe (20%). Participants were randomly assigned to one session (10 – 11 participants per session). Discussion themes included: descriptions of personal experiences with depression, words to convey depression in mobile text inputs and in social media and recognition of depression in another's texts. They were also asked to select generally used emoticons or emojis to describe their experiences of depressive symptoms.

Filipino text writing culture is unique. Most especially among adolescents, it is typical to "shorten" or abbreviate texts. In addition, the practice of "textolog" and "tag-lish" – mixing Filipino/Tagalog and English words in informal verbal and written communications – is common. Recognizing this, we selected 328 students from various colleges in Metro Manila and in Central Luzon to provide at least three spelling variations of the keywords we compiled.

3.4 Consultations with Mental Health Professionals

We interviewed psychiatrists (2), clinical psychologists (2) and a support group facilitator on what typical words and phrases their patients and clients use to describe depression and their insights on having a mobile application for depression. We also presented to them the words we gathered with our FGD participants for validation.

4. Findings & Discussion

PiaP's lexicon (English and Tagalog) consists of 11,417 keywords related to depression (including phrases and their derivatives). Each was classified into one of the 13 categories of the PiaP, which are aligned with the symptoms of depressive episode as outlined in DSM-5 and ICD-10. Words are mostly classified under Mood (16%), followed by Appetite and Weight (14%) and Guilt & Self-Esteem (10%). As supported by the Self-Focus Model of Depression, we also observed that 1st person pronouns often accompanied the keywords, reflecting the self-referential tendency when participants describe experiences with depression.

Cultural specificity and generational attributes may have an impact on the differences in depression expression and experience. Since we also considered the texting background among Filipino adolescents, it was quite common among our participants to give more than three spelling variations to a particular keyword (e.g., "depressed" can be written in 6 different ways, including five spelling variations). In total, we collated and analyzed 823,869 spelling versions (keywords and personal pronouns). Aside from the lexicon, we included emoticons and emojis that our participants reported to use. Expressing feelings and body language using symbols and characters during text communication is widespread among young people and plays a dominant role in their social interchange as a way to compensate for the absence of face-to-face interactions during online communications. Adolescents tend to be adept at using emoticons. In a study of English language blogs, teen users were able to express intense and strong emotions using emoticons (Subrahmanyam & Smahel, 2010).

4.1 Review of Depression Tests

Words related to guilt and lowered self-esteem (33%) depressed mood (23%) and decreased interest (9%) were more frequently mentioned. Item statements were mostly written in the first-person point of view. We found almost no references to substance abuse in these tests. Substance abuse is generally regarded as a separate condition and, when combined with depression, comprises a dual diagnosis. Despite studies have shown a close link between the two conditions (e.g., Link et al, 1997), one does not directly cause the other.

4.2 FGD

Our participants (who were identified as experiencing depression to a certain degree) generally characterized their emotions as "sad," "unhappy" and "lonely" (27%). They also described themselves as "a loner" (9%). Another common description (14%) is having concentration problems (e.g., no focus, not myself, having far-away thoughts). Aside from the affective dimension, having negative thoughts are typically present among depressed individuals (Beck, 2011). Cognitive deficits, such as distractibility, are further exacerbated by the condition (e.g., Hammar & Ardal, 2009).

The use of social network sites (SNS), such as Facebook and Twitter, is popular among our participants to express depression (56%). This allows them to escape stigma and to subtly seek help from friends (9%). Expressing feelings and thoughts on suicide can be more "indirect" and "resigned," as represented in statements such as "God please take me." Further, our participants, despite being non-mental health practitioners, recognize depression in other's texts and social media through the presence of sad words (36%) and the abrupt change of topic or behavior (24%). As mentioned by De Choudhury et al. (2013), these can point out to the onset of depression.

4.3 Consultations with Mental Health Professionals

There was an agreement among the mental health professionals we interviewed that the PiaP lexicon is representative of the experience of depression among their clients and patients. They also shared their observations that personal descriptions of depression may differ from how items in standardized tests are formulated. Further, contemporary modes of communication, such as SNS and texting, provide an avenue for expressing emotions especially among adolescents and young adults. Lastly, mobile technology may help in reaching out to more individuals suffering from mental health illnesses. An application for depression may be useful in screening and monitoring emotions.

5. Conclusions & Future Research Directions

Mobile technology is changing the health care landscape. In the field of mental health, mobile technology provides an additional resource for earlier detection and screening for mental illness. In this work, we present the creation of a bilingual lexicon on depression based on the DSM-5 and the ICD-10 classification systems for the mobile application PiaP.

Initial results of PiaP are promising. The lexicon has undergone an initial validation phase and is currently being tested among university students. For future research directions, we will conduct additional studies regarding the lexicon's validity and possible extension to mobile application-based depression screening in terms of the dimensionalities in measuring depression, such as physical activity, sleep and voice features and markers. We plan to incorporate adjuncts to monitor physical activity. We are hopeful that research possibilities in the field of mobile health, such as the refinement in recognizing the more subtle signs of depression and other mental disorders, such as language use, will further advance.

While we did our best to address these issues in the study design, we likewise acknowledge that the novelty of this approach to mental health care leads to new legal and ethical questions. There is a need to establish more concrete and exhaustive guidelines to ensure patient safety for this technique to eventually become a part of the toolbox in the psychological community.

References:

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Virginia: American Psychiatric Publishing.

Beck, J. S. (2011). *Cognitive behavior therapy: Basics and beyond* (2nd Ed.). New York: Guilford Press.

Bitsch, JA, Ramos, R, Ix, T, Ferrer-Cheng, PG, & Wehrle, K. (2015). Psychologist in a Pocket: Towards depression screening on mobile phones. In B Blobel, M Linden & MU Ahmed, MU (Eds.). *pHealth 2015: Proceedings of the 12th International Conference for Wearable Micro and Nano Technologies for Personalized Health.*Amsterdam: IOS Press.

De Choudhury, M, Gamon, M, Counts, S & Horvitz, E. (2013). Predicting depression via social media. *9th International AAAI Conference on Web and Social Media Proceedings,* 1-10.

Druss, BG, Hwang, I, Petukhova, M, Sampson, NA, Wang, PS, & Kessler, RC. (2009). Impairment in role functioning in mental and chronic medical disorders in the US: Results from the National Comorbidity Survey Replication. *Molecular Psychiatry,14*, 728-737.

Greenberg, J., & Pyszczynski, T. (1986). Persistent high self-focus after failure and low self-focus after success: The depressive self-focusing style. *Journal of Personality and Social Psychology, 50 (5),* 1039-1044.

Hamar, A. & Ardal, G. (2009). Cognitive functioning in major depression – A summary. *Frontiers in Neuroscience,3*(26), 1-7.

Hasan, M, Rundensteiner, E, & Agu, E. (2014). EMOTEX: Detecting emotion in Twitter messages. *Social Computing - Academy of Science and Engineering (ASE),* 1-10.

Iker, HP & Klein, R. (1974). WORDS: A computer system for the analysis of content. *Behavior Research Methods & Instrumentation, 6*, 430-438.

Kazdin, AE & Rabbit, SM. (2015). Novel models for delivering mental health services and reducing the burdens of mental illness. *Clinical Psychological Science 1*(2), 170-190.

Link, B, Struening, E, Rahav, M, Phelan, JC, & Nuttbrock, L. (1997). On stigma and its consequences: Evidence from a longitudinal study of men with dual diagnoses of mental illness and substance abuse. *Journal of Social Health & Behavior, 38*(2), 177-190.

Moreno, MA, Jelenchick, LA, Egan, KG, Cox, E, Young, H, Gannon, KE & Becker, T. (2011). Feeling bad on Facebook: Depression disclosures by college students on a social networking site. *Depression and Anxiety, 28,* 447-455.

Neumann, Y, Cohen, Y, Assaf, D, & Kedma, G. (2012). Proactive screening for depression through metaphorical and automatic text analysis. *Artificial Intelligence in Medicine, 56,*19-25.

Pennebaker, JW, Mehl, MR, & Niederhoffer, KG. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology, 54*, 547-577.

Rude, SS, Gortner, EM & Pennebaker, JW. (2004). Language use of depressed and depression-vulnerable college students. *Cognition & Emotion,18*(8), 1121-1133.

Stone, JP, Dunphy, DC, Smith, MS, & Ogilvie, DM (1966*). General Inquirer: A computer approach to content analysis*. Massachusetts: MIT Press.

Subrahmanyan, K & Smahel, D. (2010). *Digital youth: The role of media in development*. New York: Springer.

Tausczik, YR & Pennebaker, JW. (2010). The psychological meaning of words: LIWC and computerized text analysis. *Journal of Language and Social Psychology, 29*(1), 24-54.

Vanheule, S, Desmet, M, & Meganck, R. (2009). What the heart thinks the tongue speaks: A study on depression and lexical choice. *Psychological Reports, 104*, 473-481.

Weintraub, W. (1981). *Verbal behavior: Adaptation and psychopathology*. New York: Springer.

World Federation for Mental Health. (2011). *The great push: Investing in mental health.*

World Health Organization. (2008). *Task shifting: Global recommendations and guidelines.* Geneva: WHO.

World Health Organization. (1992). *International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10*). Geneva: WHO.