

---

# RU:TS

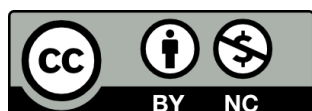
---



Review: Detecting depression in language

T.van Seggelen

Journal's Website: <http://ruts-journal.ruhosting.nl>



Published under CC-BY-NC  
licence Stichting RU:ts Student  
Linguistics Journal

# Review: Detecting depression in language

*T.van Seggelen, Language and Communication (Research), Radboud University, Nijmegen*

---

*Abstract:* Depression is a mental illness which causes a depressed mood for most of the day and/or diminishes interests and pleasures. Culture and language play a role in developing and experiencing depression as language makes it possible for someone to make sense of their experiences, while a depressed person has a disconnect between these emotional experiences and language. In this paper, it is found that depression alters a depressed person's language on different levels: on the level of phonology, vocabulary, and grammatical structure. On the level of phonology, a depressed person takes longer to utter sentences, takes more pauses and repeats more words. Depressed people also tend to use more 'I' in their writing and more absolute words like 'always'. Their sentence structure is atypical too as they write longer and more descriptive texts with shorter sentences, use more inversion and ellipses, repeat words more often and use more figurative language. However, most of the studies have focused on written text only, so there is a need to extend this research to naturalistic communication. And as most studies are conducted in English speaking countries, there is also a need for future research to look at the language use of depressed people in other languages as the signs may be different than in English before any findings can be implemented in therapeutic settings to detect possible depression earlier on and to treat it better.

---

**Keywords:** Depression, Language use, Lexicon, Phonology, Style

## 1. Introduction

Depression can be described by citing the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders' (DSM-5) diagnostic criteria (American Psychiatric Association, 2013): someone has to experience a depressed mood for most of the day and/or diminished interest or pleasure accompanied with four other symptoms to receive this diagnosis.

Depression can, however, be described differently by different people. In the first place, laymen sometimes see depression as something different as one would expect it to be based on the official criteria stated by the DSM-5 (Robertson, Venter & Botha, 2005). They underlined White's notion that experiences of depression by 'ordinary' people is far from uniform (1982, in: Robertson et al., 2005). In the second place, there exist cultural differences in the way depression is perceived. Shao, Doucet, and Caruso (2014), for example, established that emotions in general have both universal and culture-specific aspects. They found that emotion perception is the more universal domain, whereas emotion understanding and regulation are culture-specific.

So, it was indicated that not everyone talks the same about depression. Even though this is interesting on its own, the aim of the current paper is to look at how people with depression talk themselves. Based on the *structural equation model* of Şimşek (2013) – which states that personal reflection and self-rumination relate to language use and to developing depression – it is plausible that people with depression use language differently. He based his

theory on Bucci's Multiple Code Theory which states that the possibility to connect language and psychological experiences makes it possible for someone to make sense of these experiences (1984). In this model, the ability to describe these states with words (referential activity) is linked to depression, as depressed people have a disconnect between the emotional experiences and language. In Şimşek's model it is assumed that there is a gap between what is/can be expressed and the actual feeling, and that the size of this gap differs between people. By letting 459 people fill in numerous questionnaires, he tested his model and found that a big gap between experience and language will lead to a lower level of self-concept, which in turn leads to more self-reflection and self-rumination. The latter, which is a form of unhealthy recurrent thinking, can lead to depression, while self-reflection, a healthy aspect of self-consciousness, lowers the chance on depression. When a person has a high need for absolute truth - as is the case in many depressed people - self-reflection also has a negative effect on depression and makes it worse.

Şimşek (2013), thus, stated that language and depression have influence on one another. In this paper, I will focus on how depression can influence a depressed person's language use. I tried to include research from multiple languages, but all but one of the studies were conducted in English. This means that this review mostly focuses on English language features. Firstly, I will explore how depression can influence someone's phonology - especially focusing on prosody as this is mostly stereotypically seen as affected by depression. Secondly, I will discuss the influence on depressed people's lexicon. Lastly, I will talk about the style of their language. I will end this paper with some suggestions for future research and for practical usage in therapy practices.

## 2. Method

### 2.1 Search strategy

In November 2018, 23 papers that were published between 2000 and 2018 were identified on the topic of 'language of depression' using the search engines RUQuest and Google Scholar. The limit was set on 2000 to exclude the more outdated research, yet still include the relevant first studies that were conducted using the Linguistic Inquiry Word Count (LIWC; Pennebaker, Francis & Booth, 2001). To search for these papers, the terms 'language of depression', 'language + depression', 'grammar + depression', 'depressive speech', and 'phonology + depression' were used. Furthermore, papers that were referred to in the papers that were selected were used. Two papers on phonology were added that were outside of the publication date limits, because these were pioneering this kind of research and could not be kept out.

### 2.2 Data selection

From the papers that were found, studies about how other people talk about depression were omitted out as this paper focuses on the language use of depressed people themselves. Another four papers were excluded as there were only theoretical possibilities stated in these instead of empirical research. One more paper was not used as it was a master's thesis instead of a peer-reviewed paper. In the end, a total of 15 empirical and review papers were included in this study.

### 2.3 Procedure

The papers were ordered by the linguistic level they focused on (phonology, lexicon, style). Some of the papers fitted only in one of the levels, while others were used in multiple levels as they talked, for instance, about the lexicon and style. These papers were summarised individually, after which the results and specific methods were compared within the linguistic level it was ordered in. Within each of the three linguistic levels, an overview of the most important findings was made, after which an overall conclusion about ‘the language of depression’ was drawn.

## 3. Results

### 3.1 Phonology

Kristeva (1989) already said that the speech of depressed people is repetitive and monotonous. They utter interrupted sentences and when “that frugal musicality becomes exhausted in its turn [...] the melancholy person appears to stop cognizing as well as uttering” (33). More recent studies showed that indeed 20% of depressed patients showed signs of catatonia, which is a state of psycho-motor immobility. These people tend to stop talking, start to stutter, or start repeating what has just been said by someone else (Marvel & Paradiso, 2004).

In this paragraph, I will focus on the phonology of depressed speech only, instead of also focusing on people that completely stop talking. Already in 1921, Kraepelin noted that depressed people tend to speak in a low voice, that they talk slowly and hesitatingly, sometimes becoming mute in the middle of a sentence, monotonously, and sometimes stuttering. Empirical research followed when Kuny and Stassen (1993) observed 30 depressed patients during their recovery from depression. In their study, they assessed the patients’ language six times during two weeks and compared these findings with the language of 192 healthy people. They found that a speaker’s loudness and the variation in loudness over time, and the speaker’s voice timbre correlated highly with depression.

Results of Alpert, Pouget, and Silva (2001) were partially similar. Based on the free speech of a group of 22 elderly participants with major depressive disorder and 19 healthy controls, they observed that depressed patients showed less prosody than healthy subjects and that abnormal pauses in their utterances decreased when these patients were treated for their depression via medication. They, thus, concluded that temporal changes in speech, such as pausing, reflect the depressed state whereas prosodic features (e.g. stress patterns) seem to reflect a depressed trait: agitated, which is characterised by heightened psychomotor activity, or retarded, which is characterised by slow thinking and behaviour, depression.

The previously mentioned papers were all about depressed speech in relatively small samples of clinical patients. However, Low et al. (2011) investigated the acoustic correlates of depression in a sample of 68 depressed and 71 non-depressed adolescents in naturalistic interaction with their parents. By measuring the number of additional harmonics due to air flow in the vocal tract, as well the prosodic, cepstral, spectral, and glottal features, they found that the air flow in the vocal tract predicted depression quite well: accuracy ranging between 81%-87% for males and 72%-79% for fe-

males. By combining glottal features with prosodic and spectral features, the accuracy was slightly less accurate, but still it predicted the depression well: 67%-69% in males, 70%-75% in females. They explain this difference in speech by saying that speech production systems show physical manifestations of psychological difficulties. In other words: depression may have a significant effect on the vocal cord function, making it more difficult to produce sound by interaction with structures in the vocal tract.

### 3.2 *Lexicon*

One of the first papers written about the 'language of depression' in written texts is Stirman and Pennebaker's 'Word use in the poetry of suicidal and non-suicidal poets' (2001). Based on the notion that the frequency of word use in written text can be used as an indicator of psychological state, they tried to identify predictors of depression in these poets' poems as measured by these poets having committed suicide. Furthermore, they tried to link their findings to the two most prominent models about suicide: Durkheim's social integration/disengagement model (1951, in: Stirman & Pennebaker, 2001) and the hopelessness model of suicide. The first model states that a suicidal person has failed to integrate in society and becomes detached from social life. In other words, these people detach from the source of their pain and this way withdraw from social life and become more self-oriented. The hopelessness suggests that suicide takes place during extended periods of sadness and desperation in which an individual also has the tendency to think in absolutes leading to the conclusion that suicide is the only option. In case if the Durkheim's model is right, the authors expected that the suicidal poets would use more self-references and less references to others in their work, whereas if the second model is right, they would use more negative emotion words, such as anger, sadness, fewer positive emotion words, such as happiness, gratitude, and more references to death. For these purposes, the authors analysed 156 poems from well-known poets who committed suicide from their early, middle and late work without looking at the subject of the poems beforehand. They compared these poems with 135 poems from non-suicidal poets who were matched in nationality, era, education and sex with one of the suicidal poets. They found that suicidal poets did use more first-person singular pronouns, but did not use less references to others or communication words like 'talk'. Furthermore, the non-suicidal and suicidal poets did not differ in their use of positive or negative emotion words, even though suicidal poets mentioned death more often. Moreover, they did not find any phase effects, i.e. suicidal poets did not use more self-references or references of death in their later work than in the beginning. Because of these results, it seemed like neither of the models fit quite right, as the suicidal poets were not less positive or more negative, and they did not show any more social withdraw during their careers either.

Another artistic form of language use is the writing of song lyrics. To study depressive language clues in this form, Lightman, McCarthy, Dufty and McNamara (2007) analysed the lyrics of 8 suicidal and 8 non-suicidal songwriters matched on genre of music, date of birth, degree of fame, nationality, status of addiction and mental health and education. They used 35 songs from each artist and analysed the self- and other-references, emotion words, references to time, communication words and death-themed words. Another thing they

measured was the concreteness of the language used based on the word concreteness index in the Coh-Metrix from Graesser et al. (2004; in Lightman et al., 2007), as they predicted that suicidal singers would be less concrete due to the detachment from society as the Durkheim's model stated. They found that the suicidal artists were less concrete and that they sang more about the future, or at least they used more future-tense verbs. This was not predicted, because it is believed that suicidal people tend to be preoccupied with the interminability of the present without thinking about the future. The expectation that suicidal artists use more self-references and less other-references was not fulfilled as there was only a non-significant trend present. Moreover, contrary to the hypothesis, they found that the non-suicidal singers used more death-themed words than the suicidal ones. They explain this discrepancy by using the writing as something therapeutic: the non-suicidal artists were matched to the suicidal ones on the basis of their mental health and as writing down traumatic events or negative emotions can be cathartic; the non-suicidal artists could have been depressed too, but dealing with this by writing. Another reason why the results of this study might be different from that of Stirman and Pennebaker (2004) is because the process of writing poetry and writing song lyrics is different: many poets work independently, while songwriters have band members for example. Secondly, irony is often prevalent in music (e.g. positive lyrics on melancholic melodies) but computational tools like the LIWC cannot recognise irony, because it only focuses on text.

One could say that poetry as well as the closely related genre of music lyrics and depression are closely linked already, since suicide rates among poets are higher than among authors of other literary forms or the general public (Jamison, 1993). Thus, they are not a very representative medium for testing the depressed language features. That is one of the reasons why Rude, Gortner and Pennebaker (2004) looked at the language use of depressed and depressed-vulnerable college students. Another problem this study sorted was measuring depression by suicide. These are two different concepts, as depressed people do not have to commit suicide and suicidal people do not have to be depressed. Rude et al. (2004), thus, let 31 currently-depressed, 26 formerly-depressed, and 67 never-depressed participants write an essay about their deepest thoughts and feelings about starting college. They based their hypothesis that depressed and formerly-depressed students would use more first-person pronouns on Pyszczynski and Greenberg's theory that depressed individuals tend to think a great deal about themselves (1987, in: Rude et al., 2004) and also included the formerly-depressed participants because Beck's idea that depressive schema's or habits may be latent (1967, in: Rude et al., 2004). And indeed, they found that depressed students did use more negative emotion words and first-person singular words, on the other hand the formerly-depressed students did not differ from those that never suffered from depression. They explained this may be because of self-awareness of these formerly-depressed students, so they analysed the essays again by splitting them in three parts, assuming the old habits of depressed language would 'slip in' later on. Consistent with this idea, they found that the formerly-depressed students' use of 'I' increased across their essay, and that this increase was greater than that of never-depressed students.

In 2010, Rodriguez, Holleran and Mehl (2010) let laymen assess subclini-

cal depression in others by reading personal diaries or blogs. By reading the self-descriptions from depressed people, the laymen had to say if the person who had written the text had depression on a 7-point Likert scale, after which the texts were also analysed automatically by the LIWC (Pennebaker, Francis & Booth, 2001) to see if there were actually any thematic or linguistic differences between depressed and not-depressed writing. They found that in both genres the laymen achieved high and comparable levels of accuracy, meaning that depression is recognised in written text by these laymen. The cues they used to distinguish text from depressed people from non-depressed people were highly correlated with the cues the writers themselves (sub)consciously used. This means that laymen spot depression in text not only by explicit notions of sadness, but also implicit notions of verb tenses. Moreover, they found differences in how the depression was worded in each genre. In the diaries, people tended to use more words related to sadness, cognitive mechanisms indicated by words such as 'know', and metaphysical references like using the word 'religion'. In blogs, people tended to use more swear words or references to sleep as an indicator of depression. Rodriguez et al. (2010) then conclude that laymen seem to be implicitly aware of these context-specific differences by which they can spot depression in texts.

As mentioned earlier, depression cannot only lead to a high self-absorption, but also to thinking in absolutes (see Şimşek (2013) and the hopelessness model of suicide). Al-Mosaiwi and Johnstone (2018) focused on the use of these absolutist words instead of the references to the self in internet forums. They predicted that there would be a higher percentage of absolutist words in anxiety, depression, and suicidal ideation forum groups than in general forums and that the percentage in suicidal ideation forums would also be higher than in both anxiety and depression forums. They collected 30,000 words from 63 forums. They then made two lists of words to look for in the texts even though they only reported the outcomes for the second list: a list of extreme words like 'very' and of absolute words like 'always'. They found that indeed the percentage of absolutist words in anxiety, depression, and suicidal ideation forums was significantly greater than in the general forums and that the percentage of absolutist words in suicidal ideation forums was greater than in the anxiety or depression forum. Because the researchers were afraid they could have been measuring psychological distress instead of the use of absolutists, they did a second study in which they compared four forums of mental health conditions: two of which are known to not use absolutist words – namely post-traumatic stress disorder and schizophrenia – and two which are – bipolar disorder and eating disorder. They found that even though all groups contained many negative emotion terms, only the bipolar disorder group and eating disorder group contained absolutist thinking, indicating that they indeed measured absolute thinking instead of psychological stress.

### 3.3 *Style*

The previous paragraph focused on the lexicon, but there are also indicators that a depressed person's writing style differs from non-depressed people. Eichstaedt et al. (2018), for example, looked if Facebook language could predict depression in medical records. They found that the word use of depressed

people was the same as what could be found in the articles described in the last paragraph: these people were talking about emotions like sadness, loneliness, and hostility, were prone to rumination, and had increased self-references. They also found something non-lexical related: post length and frequency. These two variables were as predictive of depression as demographic characteristics which already had a fair accuracy. They concluded that word count across posts was 1,424 words higher for users who developed depression. So, posting more and longer texts on Facebook could be an indication of depression.

This is not only the case on Facebook. Smirnova et al. (2018) found that people with mild depression wrote longer and more descriptive styled texts than healthy individuals when asked to write an essay about their life. Moreover, they used more informal and figurative language, more inversions and lexical repetitions, omitted words more often, and even though their essays were longer overall, their individual sentences were shorter. When comparing participants with normal sadness to the healthy participants and depressed participants, it was shown that their language use was in between the two groups. For example, they omitted more words than the healthy group, but less than the depressed group. However, it is important to note that this last study was conducted in Russian while the other ones were done in English.

#### 4. Conclusion

First and foremost, it can be concluded that there actually exists some sort of ‘language of depression’. Previous studies have at least found that there are language features that are typical of the language of depressed people, yet differ from the language use of non-depressed people. These differences are seen in phonology, vocabulary, and style. As far as phonology is concerned, one could notice that people with depression take longer to utter sentences, take more pauses during speech, and that they repeat words more often than healthy people. Moreover, the stress pattern in a depressed person’s speech seems atypical from normal speech. Unfortunately, it is not clear yet how depression affects the speech production system exactly.

In their vocabulary, depressed persons also differ from their peers. This holds to some degree for writing in the artistic sense like poetry – even though here suicidality was more of a dividing factor – but definitely for writing essays or forum entries. Depressed people tend to use more ‘I’ in their writing, as well as absolute words like ‘always’. This indicates that a depressed person is more self-absorbed and thinks a lot about how they are feeling and what they are, but also that they tend to think in absolutists. Important to note is that even though a person is recovering or already recovered officially, these habits of self-references and using absolutist words can still hold. This indicates that the mindset of a formerly depressed person still differs from that of a healthy person even though the evidence of this can be masked by a greater self-awareness in this group.

Lastly, the grammatical structure of sentences in the language of depressed people seems different than that of non-depressed people. When looking at natural language of depressed people, one still finds atypical grammatical structures. Depressed people tend to post more and longer texts on Facebook for example and also write longer and more descriptive texts when writing essays.



Moreover, their sentences are shorter, their language more figurative, their word order is inversed, and they repeat words more often or omit them altogether.

## 5. Discussion

When concluding that there must be something like ‘the language of depression’, it is worth noting that most of the studies cited in this paper are conducted in English speaking countries or online where English was the language that was used. The only exception is the article from Smirnova et al. (2018), which is Russian of origin. Moreover, all of these studies were done with the Western definition of depression in mind. Therefore, these findings may not be universal, because cultures differ from each other in how they see depression and how they treat it. It is even the case that someone’s culture moderates the effect of emotions on developing depression as Chan et al. (2016) found. This is why it is important to continue doing research on linguistic differences of depressed and non-depressed people who speak other languages than English or who live in different cultures.

Another important shortcoming in the research is that it is mostly conducted on (artistic) written texts. If we want to use these linguistic variables that are different in depressed language versus non-depressed language, we ought to do studies of language use in naturalistic communication. In the paragraph about phonology it is already shown that this is possible, for instance, by measuring air flow and prosody to show differences between depressed and non-depressed language.

Moreover, future research ought to clearly distinguish suicide and depression. Two studies in this review used suicide commitment as a measure for depression, but having depression does not mean wanting to commit suicide, and vice versa. Furthermore, the hopelessness model of suicide might not be the same as a hopelessness model of depression, as Alloy, Abramson, Metalsky, and Hartlage (1988) indicated.

Still, when more is known about the language use of depressed people, one could start to think on how to implement it in the setting of therapy. By teaching professional counsellors to look for specific signs in language use, diagnosing depression could become easier; especially in people who cannot concentrate enough to fill in the depression questionnaires.

All in all, the research papers mentioned in this paper indicate that depressed people differ in their language from healthy people; specifically, on the following linguistic levels: phonology, vocabulary, and grammatical style. However, more research is needed on this topic. Are the differences in depressed and non-depressed languages dependent on the specific language and/or culture? Are these differences dependent on the format in which the language is used? Only when there is more clarity in these subjects, we could start to implement the findings in a therapeutic setting or use them as signs for early intervention.

## References

- Al-Mosaiwi, M. & Johnstone, T. (2018). In an absolute state: Elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. *Clinical Psychological Science*, 6(4), 529-542.
- Alpert, M., Pouget, E.R., & Silva, R.R. (2001). Reflections of depression in acoustic measures of the patient's speech. *Journal of Affective Disorders*, 66(1), 59-69.
- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Washington, DC: Author.
- Angst, J., Gamma, A., Benazzi, F., Ajdacic, V., & Rössler, W. (2009). Does psychomotor agitation in major depressive episodes indicate bipolarity? *European Archives of Psychiatry and Clinical Neuroscience*, 259(1), 55-63.
- Beck, J.S. (2011). *Cognitive Behaviour Therapy: Basics and Beyond* (2<sup>nd</sup> ed.). New York: The Guilford Press.
- Brandt, M.E. & Boucher, J.D. (1986). Concepts of depression in emotion lexicons of eight cultures. *International Journal of Intercultural Relations*, 10, 31-346.
- Eichstaedt, J.C., Smith, R.J., Merchant, R.M., Ungar, L.H., Crutchley, P., Preotiuc-Pietro, D., Asch, D.A., & Schwartz, H.A. (2018). Facebook language predicts depression in medical records. *Proceedings of the National Academy of Sciences of the United States of America*, 115(44), 11203-11208.
- Fleck, D.W. (2003). *A Grammar of Matses* (Doctoral dissertation). Retrieved from Rice University electronic theses and dissertations (12051).
- Halfin, A. (2007). Depression: The benefits of early and appropriate treatment. *The American Journal of Managed Care*, 16(4), S92-97.
- Jamison, K. (1993). *Touched with Fire: Manic-Depressive Illness and the Artistic Temperament*. New York: Free Press.
- Kraepelin, E. (1921). Manic depressive insanity and paranoia. *The Journal of Nervous and Mental Disease*, 53(4), 350.
- Kristeva, J. (1989). *Black Sun: Depression and Melancholia* (L.S. Roudiez, Trans.). New York: Columbia University Press.
- Kuny, S. & Stassen, H.H. (1993). Speaking behaviour and voice sound characteristics in depressive patients during recovery. *Journal of Psychiatric Research*, 27(3), 289-307.
- Lightman, E.J., McCarthy, P.M., Dufty, D.F. & McNamara, D.S. (2007). Using computational text analysis tools to compare the lyrics of suicidal and non-suicidal songwriters. *Proceedings of the Annual Meeting of the Cognitive Science*, 29, 1217-1222.
- Low, L.S.A., Maddage, N.C., Lech, M., Sheeber, L.B., & Allen, N.B. (2011). Detection of clinical depression in adolescents' speech during family interactions. *IEEE Transactions on Bio-Medical Engineering*, 58(3), 574-586.
- Marvel, C.L. & Paradiso, S. (2004). Cognitive and neurological impairment in mood disorders. *The Psychiatric Clinics of North America*, 27(1), 19-36, vii-viii.
- Mehl, M.R. & Pennebaker, J.W. (2003). The sounds of social life: A psychometric analysis of students' daily social environments and natural conversations. *Journal of Personality and Social Psychology*, 84(4), 857-870.
- Pennebaker, J.W., Francis, M.E. & Booth, R.J. (2001). *Linguistic Inquiry and Word Count*. Mahwah: Erlbaum Publishers.
- Robertson, A., Venter, C. & Botha, K. (2005). Narratives of depression. *South African Journal of Psychology*, 53(2), 331-345.
- Rodriguez, A.J., Holleran, S.E. & Mehl, M.R. (2010). Reading between the lines: The lay assessment of subclinical depression from written self-descriptions. *Journal of*

- Personality*, 78(2), 575-598.
- Rude, S.S., Gortner, E.M. & Pennebaker, J.W. (2004). Language use of depressed and depression-vulnerable college students. *Cognition and Emotion*, 18(8), 1121-1133.
- Shao, B. Doucet, L. & Caruso, D.R., (2014). Universality versus cultural specificity of three emotion domains: Some evidence based on the cascading model of emotional intelligence. *Journal of Cross-Cultural Psychology*, 46(2), 229-251.
- Şimşek, Ö.F. (2013). The relationship between language use and depression: Illuminating the importance of self-reflection, self-rumination, and the need for absolute truth. *The Journal of General Psychology*, 140(1), 29-44.
- Smirnova, D., Cumming, P., Sloeva, E., Kuvshinova, N., Romanov, D., & Nosachev, G. (2018). Language patterns discriminate mild depression from normal sadness and euthymic state. *Frontiers in Psychiatry*, 9, 105.
- Solomon, A. (2001). *The Noonday Demon: An Anatomy of Depression*. New York: Scribner.
- Stirman, S.W. & Pennebaker, J.W. (2001). Word use in the poetry of suicidal and non-suicidal poets. *Psychosomatic Medicine*, 63, 517-522.

## **Acknowledgements**

The researcher thanks prof. dr. J.M. Sanders and prof. dr. W.P.M.S. Spooren for the feedback and their guidance during the period of writing this paper. Context: Term Paper LET-REMA-LC104-2018-PER\_2V.