

# Temporal and Computerized Psycholinguistic Analysis of the Blog of a Chinese Adolescent Suicide

Tim M. H. Li<sup>1,2</sup>, Michael Chau<sup>3</sup>, Paul S. F. Yip<sup>1,2</sup>, and Paul W. C. Wong<sup>1,2</sup>

<sup>1</sup>Department of Social Work and Social Administration, The University of Hong Kong, China

<sup>2</sup>Centre for Suicide Research and Prevention, The University of Hong Kong, China

<sup>3</sup>School of Business, The University of Hong Kong, China

**Abstract.** *Background:* Text analysis of personal documents provides insight into the cognition of those who complete suicide. Many personal documents are digitalized and easily found on the Internet, which can be used to advance suicide research. *Aims:* (1) To examine the temporal relationships between posting intensity and language use to sketch the suicidal process of a young man on the basis of his blog entries. (2) To investigate whether digitalized personal documents and paper documents of suicide cases have similar or different language patterns. *Method:* Firstly, 193 blog entries of a 13-year-old boy posted during the year prior to his suicide were analyzed using the Chinese Linguistic Inquiry and Word Count (CLIWC) program. The temporal relationships between posting intensity and language use were illustrated by time series visualization. Secondly, the findings of this case study were compared with previous case studies from a systematic search of three Ovid databases. *Results:* Posting frequency and language use in the blog helped sketch the suicidal process of the young boy. In this case study, the ratio of positive to negative emotion words was associated with the posting trend. Progressive self-referencing appeared to be a primary predictive sign of suicide. However, the comparison did not show other clearly consistent patterns. *Conclusion:* Digitalized personal documents, when interpreted with other information of the individual, provide insight into the suicidal process of completed suicides. This study extends the findings of psycholinguistic analyses of suicides to the Chinese social context and online document form.

**Keywords:** blogs, Chinese, posting intensity, language use, suicidal process

The words that people use provide important psychological cues to their thought processes, emotional states, intentions, and motivations (Tausczik & Pennebaker, 2010). Text analysis of suicide notes has been one of the core methodologies for understanding the cognitive processes before suicide (Freuchen & Grøholt, 2013; Shneidman & Farberow, 1957). Some studies have shown that the wording in the final writings of people who completed suicide is predictable (Leenaars, 1988; Osgood & Walker, 1959); for instance, there were more first-person singular and death-related words in writings of poets who completed suicide than in writings by matched nonsuicidal poets (Stirman & Pennebaker, 2001). The findings may suggest that suicidal individuals are more self-focused and negative in response to emotional pain (Wolf, Sedway, Bulik, & Kordy, 2007).

One criticism of previous studies on writings of suicide cases is that the suicide notes were manually analyzed, which may lead to subjective interpretations of the material (Wong, Yeung, Chan, Yip, & Tang, 2009). Recently, computer technology has been employed to assist the analysis; for example, the Linguistic Inquiry and Word Count (LIWC) program (Pennebaker, Chung, Ireland, Gonzales, & Booth, 2007), a text analysis software program that provides over 70 psychologically meaningful language varia-

bles, such as emotion and self-referencing words, was used to analyze suicide notes (Pestian et al., 2012).

Suicide notes are typically brief and written moments before the suicide act (Lester & Linn, 1998). Hence, other studies have analyzed the personal documents of those completing suicide, which allows for a richer and longer observational timeframe of the cognitive process before suicide (Baddeley, Daniel, & Pennebaker, 2011; Barnes, Lalal-Solarin, & Lester, 2007; Fernández-Cabana, García-Caballero, Alves-Pérez, García-García, & Mateos, 2012; Lester, 2009; Pennebaker & Stone, 2004; Stirman & Pennebaker, 2001). In general, these studies found that the suicide victims tend to express more positive feelings in their later entries (Barnes et al., 2007; Lester, 2009; Lester & McSwain, 2011; Pennebaker & Stone, 2004) and this may be representing a sense of relief after they have decided to use suicide as a solution to their problems (Keith-Spiegel & Spiegel, 1967). Besides, they tend to write more about the self and less about others in later entries; for instance, people who completed suicide used more self-referencing words (e.g., I, my, me) and fewer references to others (e.g., we; Baddeley et al., 2011; Fernández-Cabana et al., 2012; Stirman & Pennebaker, 2001). The changes in language patterns found in the study of the diaries of suicide vic-

tims allow suicide researchers to identify both the distant and imminent issues that concerned the suicidal writers and may allow practitioners to develop an empirical-based understanding of the psychological changes of these individuals (Lester, 2009). The temporal relationships between behavioral and psychological factors in the personal documents of those who have died by suicide may provide insight into the trajectory of the suicidal process (Fortune, Stewart, Yadav, & Hawton, 2007).

In the Web 2.0 era, personal documents are often stored digitally (Fullwood, Sheehan, & Nicholls, 2009), and this provides a more easily analyzable and accessible source of information for text analysis of suicide (Barak & Miron, 2005). Blogging is a prevalent form of communication in expressing emotion and sharing information (Fullwood et al., 2009; Nardi, Schiano, Gumbrecht, & Swartz, 2004). The online disinhibition effect makes people feel more comfortable in revealing their feelings and thoughts on the Internet (Suler, 2004). Publicly accessible blogs (Fullwood et al., 2009) or Facebook (Ruder, Hatch, Ampanozi, Thali, & Fischer, 2011) provide new data sources for the study of suicidal behavior.

This is an exploratory case study using the blog entries of a 13-year-old Chinese youth who died by suicide. To be consistent with a unified classification method, the language patterns of the blog were analyzed using the Chinese version of LIWC (Huang et al., 2012). Most suicide studies have focused on English materials, while the Chinese population has not been well studied (Li, Cai, Graesser, & Duan, 2012; Tsai, Simeonova, & Watanabe, 2004). The primary aims of this study were to explore the temporal relationships between posting intensity and language use in an attempt to sketch the suicidal process of a young man and to investigate whether digitalized personal documents show similar patterns with those reported in previous text analysis studies of the written personal documents of suicide victims.

## Method

### Data Collection

The data source for this study was the publicly available blog entries of a suicide case found in a coroner's court file. To provide background on the suicide case, we will present some basic information collected from the coroner's court file. Ethics approval was obtained from the Human Research Ethics Committee for Nonclinical Faculties at The University of Hong Kong, China.

### The Background of Mr. A

Mr. A. was a 13-year-old male secondary school student who completed suicide by jumping from a building. He suffered from a stutter but did not appear to have received any psychiatric treatment previously. His family members

and best friend described him as an extreme, stubborn, socially isolated, and pessimistic person. Besides, his mother reported that he had been unhappy all of the time after starting his secondary education 1 year before his death. He disliked his school because he was teased and isolated by his classmates. He told his mother that he wished to study in another school. The deceased loved playing volleyball and was a member of the school volleyball team. He once failed to attend a training session and attributed that as the main reason for losing in a competition shortly before his death. This incident led to him being condemned and isolated by his team members.

On the day of his death, the deceased asked his teacher whether he could leave school early to attend a volleyball match; however, the teacher declined the request. After that, he was spotted crying in a classroom by another teacher. The coroner speculated that the deceased became emotionally unstable under the influence of alcohol, which might have contributed to his suicide by jumping from a building.

## On-Line Material and Data Analyses

There were three stages of analysis of the blog entries: (1) natural language processing, (2) trend of posting intensity and language use, and (3) correlation analyses and comparison with previous case studies.

### Natural Language Processing

The subject's 193 blog entries had been written during the year preceding his death and were collected after his suicide during the police investigation. The blog entries were segmented with the Institute of Computing Technology, Chinese Lexical Analysis System ICTCLAS software (Zhang, Yu, Xiong, & Liu, 2003). The Chinese version of the LIWC dictionary (CLIWC; Huang et al., 2012) was used to analyze the blog entries. The CLIWC has 72 categories of Chinese words, including linguistic and psychological word categories. The 193 parsed entries were scored in the 72 categories, including 11 Chinese-specific categories. With the exception of the total number of words, scores were all percentages of the total number of words.

### Trend of Posting Intensity and Language Use

The cumulative number of blog entries written by the deceased from time 0 (1 year before his death) up to time  $t$  (day of suicide  $t = 365$ ) was modeled by a counting process (Andersen, Borgan, Gill, & Keiding, 1993). The posting intensity function represented the expected daily number of blog entries while the language use represented the scores of CLIWC categories of the entries. The time-varying intensities were then estimated using a nonparametric curve estimation method, the local polynomial method (Chen, Chen, Gunnell, & Yip, 2013).

### Correlation Analyses and Comparison With Previous Case Studies

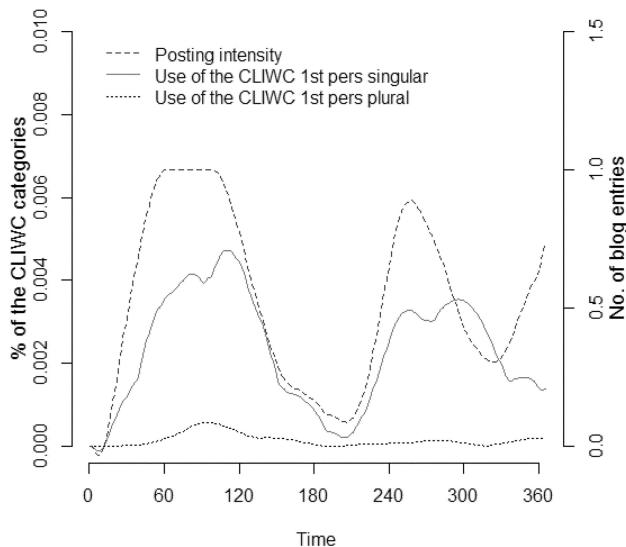
The 193 blog entries were analyzed from the first entry around 1 year before the subject's death to the last entry, written 1 day before his death. Pearson correlation analysis was used to examine the relationship between the sequence of the blog entries (numbered from 1 to 193) and each of the LIWC categories.

We searched for previous case studies using three Ovid databases: Journals@Ovid Full Text, MEDLINE, and Embase. The following keywords were used in the search strategy: "suicide" AND "LIWC." The search yielded 34 academic articles in April 2013. Five more articles were identified manually from the reference lists of the 34 articles and two more articles were identified through Google Scholar. Six duplicates were removed from the pool of studies. After reviewing the articles, ten were included as they (1) covered someone who completed suicide, (2) were case studies reported in English, and (3) analyzed the person's personal documents using LIWC. Two of them that were too brief for analysis were excluded, resulting in a total of eight case studies. The list of excluded studies is available from the corresponding author.

## Results

### Trend of Posting Intensity and Language Use

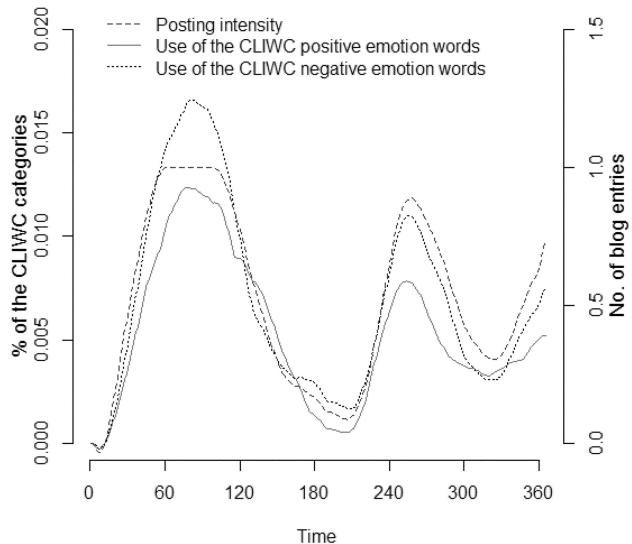
Figure 1 shows the trend of the posting intensity and the use of first-person singular and plural pronouns in the blog entries across time. There were two wave cycles or two peaks in the trend and the posting intensity increased in the last few weeks before death. The first peak appeared



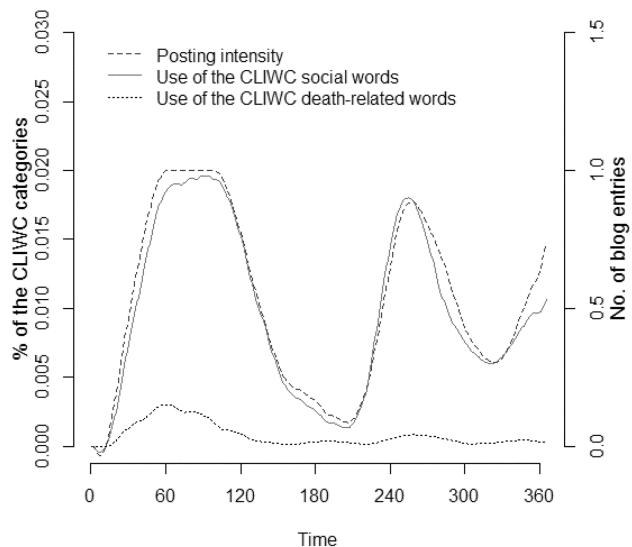
**Figure 1.** Trend of posting intensity (blog entries per day) and the use of first-person singular and plural pronouns.

at  $t = 60\text{--}120$ , the second was at  $t = 240$ , and the intensity then escalated from  $t = 320$ . The use of first-person singular pronouns followed the posting trend, whereas the use of first-person plural pronouns did not. In the last month, the use of first-person singular pronouns decreased but the posting intensity increased. Besides, the percentages of first-person plural pronouns in the blog entries were much lower than those of first-person singular pronouns.

The use of positive and negative emotion words in the blog entries was compared across time (see Figure 2). Percentages of negative emotion words in the blog entries were generally higher than those of positive emotion words across time. Specifically, a difference between the use of positive and negative emotion words was observed



**Figure 2.** Trend of posting intensity (blog entries per day) and the use of positive and negative emotion words.



**Figure 3.** Trend of posting intensity (blog entries per day) and the use of social and death-related words.

*Table 1.* Comparison of six major language patterns in nine suicide case studies

Case study	Age	Sex	Suicide means	Document type	No. of doc.	Period	No. of significant categories/ total	Self-reference	Religion	Death	Positive emotion	Negative emotion	Reported personal concerns
Katie (Pennebaker & Stone, 2004)	21	F	Hanging	Diary	n/a	6 months	8/72	↑	↑	↓	↑	↓	↑ (eating)
“Letter” (Barnes et al., 2007)	20	M	n/a	Letters	23	2 years	7/72	ns	↑	ns	↓	↑ (finances)	
Pavese (Lester, 2009)	41	M	Overdose	Diary	13	1 year	15/76	↑	ns	↑	ns	↓	↓ (occupations, schools, jobs)
Stephen (Lester, 2010)	18	M	Gunshot	Journals	7	6 months	4/74	ns	ns	ns	ns	ns	n/a
Teadale (Lester & McSwain, 2010)	48	F	Overdose	Poems	317	27 years	25/74	↓	↓	ns	↓	ns	n/a
Hellyer (Baddeley et al., 2011)	42	M	Gunshot	Letters, journals, and reports	11	7 years	n/a	↑	n/a	ns	↑	↑	n/a
Plath (Lester & McSwain, 2011)	30	F	Carbon monoxide poisoning	Poems	224	7 years	29/74	↑	ns	ns	↑	↑	n/a
Monroe (Fernández-Cabana et al., 2012)	36	F	Overdose	Personal notes, letters, and poems	50	1 month	1/74	ns	ns	ns	ns	ns	↑ (eating)
“Mr. A” (current)	13	M	Jumping	Blog	193	1 year	4/72	ns	ns	ns	ns	ns	(social)

Note. ↑ = increase in the category; ↓ = decrease in the category; doc. = documents; ns = change of the category was nonsignificant; n/a = information was not provided in the paper.

at the peaks of the two waves. When the posting intensity decreased, the percentages of positive emotion words became higher than those of negative emotion words ( $t = 120\text{--}160$  in the first wave and  $t = 300\text{--}330$  in the second wave). The ratio of positive to negative emotion words might be associated with the posting trend.

The use of social and death-related words in the blog entries was compared across time (see Figure 3). This comparison showed a difference between the two primary concerns of the deceased, namely, relationship with others and suicide intention. The use of social words closely followed the posting trend, whereas the use of death-related words did not. Furthermore, the percentages of social words in the blog entries were much higher than those of death-related words.

### Correlation Analyses and Comparison With Previous Case Studies

In the correlation analysis, four significant associations were identified at the 5% level of statistical significance or stronger. As the blog entries approached the day of death, there were more references to *hear* ( $r = .17, p = .017$ ) and *feel* ( $r = .15, p = .035$ ), and there were fewer tense markers ( $r = -.15, p = .037$ ) and present tense markers ( $r = -.15, p = .041$ ).

The findings of the comparisons in this study and eight previous case studies are shown in Table 1. Among the six aspects that previous studies have focused on (i.e., self-reference, religion, death, positive emotions, negative emotions, and personal concerns), there were no significant changes across the blog entries in the current study. The evidence accumulated from this and eight previous case studies indicates some consistent patterns of language use in the personal documents of suicide victims. Progressive self-reference appeared in five out of nine studies. Emotion changes were common (seven out of nine studies), but four out of nine studies showed increasing positive emotions (i.e., increased use of positive emotion words and/or decreased use of negative emotion words), and two studies showed increasing negative emotions over the last period before the suicides. The studies reporting personal concerns generally showed an increasing trend of personal concern words (three out of five studies). Only four out of eight studies demonstrated a significant change in death and/or religion-related words in their suicide cases, and these changes were not in a consistent direction.

## Discussion

### Trend of Posting Intensity and Language Use

Most previous case studies demonstrated the language patterns related to suicide using the relationship between the sequence of documents and the LIWC categories. It is reasonable to use the sequence of documents instead of

time because each piece of writing may be event-driven but not necessarily time-dependent. For example, a person responds to a letter only after receiving it. In this case, the time interval between letters does not depend only on the cognition of the writer, but also on other confounding factors such as delivery delay. Incorporating the time factor in the analysis may affect the findings (Barnes et al., 2007), but the time factor in some document types may be crucial in improving the analysis and in illustrating psychological changes. For instance, it is expected that individuals write a diary or blog daily to document their everyday lives. If a blogger suddenly posts more or less than his/her usual patterns, it may reflect some psychological and behavioral changes.

Counting pronouns in written documents can be an unobtrusive way of assessing underlying psychological processes (Tausczik & Pennebaker, 2010). Throughout the blog entries, the deceased used substantially more self-references than references to social integration, which is consistent with the findings from a previous study (Baddeley et al., 2011). The findings may also reveal that the deceased suffered from elevated interpersonal distress in his last year (Zimmermann, Wolf, Bock, Peham, & Benecke, 2013). In the last month, the decreased use of self-references, however, was opposite to the posting trend. The selection of the period for analysis may explain why the study of Teasdale demonstrated a downward trend of self-referencing (Lester & McSwain, 2010).

Although the deceased tended to express more negative emotions in his blog, he used more positive emotion words than negative emotion words when the posting intensity was lower. It seems the deceased posted less when he was happier. The use of social words in the blog also provided insight into the deceased's possible concerns and emotions. As the deceased had poor relationships with his classmates and was afraid of being isolated, he expressed social problems in his blog. Apart from personal concern, Chinese may use social words to convey important emotional meaning (Tsai et al., 2004). The deceased Chinese boy might have expressed his negative emotion implicitly.

### Correlation Analyses and Comparison With Previous Case Studies

Regardless of the suicide means and the sociodemographics of the suicide cases, the increased use of first-person singular pronouns prior to death was consistent in most studies. This increase fits with previous studies that associated this change with low social integration, which has been related to suicide according to different theories (Baddeley et al., 2011; Fernández-Cabana et al., 2012; Stirman & Pennebaker, 2001).

Although emotion change was a major psychological cue, the deceased in this study did not express such change before his death. Since previous studies analyzed documents written a long time before the suicide, emotion change may be explained by age (Pennebaker & Stone,

2003). The different levels of suicidal intentionality may influence emotion change (Fernández-Cabana et al., 2012). For example, depressed suicidal individuals may express stronger and longer emotion change than people who commit suicide impulsively (Baddeley et al., 2011).

Language patterns are unique in different suicide cases because of the individual's concerns and life circumstances (Baddeley et al., 2011). This may suggest that the individual's needs or desires were not reduced as they approached their death. On the contrary, few demonstrated a change in death and/or religion-related words traditionally regarded as closely related to suicide.

This study shows significant changes in Chinese-specific categories, such as the decreased use of present tense markers. This change suggests that the deceased might not have thought or lived in the here-and-now, but in the past (Tausczik & Pennebaker, 2010). However, tenses in Chinese are reflected by additional tense markers instead of changing verb form as in English. Language differences may affect the findings of studies conducted in different cultures.

More complex analyses and measures are necessary to understand the language patterns (Lester, 2009). Barnes and colleagues (2007) and Lester (2009) have suggested that psychological changes are long-term (over a few months prior to the death by suicide). This suggestion is reasonable because, for instance, the study of Plath (Lester & McSwain, 2011) did not find meaningful psychological changes over a short period (information listed in Table 1). In the same vein, Fernández-Cabana and colleagues (2012) and Lester (2010) have grouped and analyzed documents to understand different life periods of suicide victims. Moreover, two case studies have used a wide range of personal documents to examine the language patterns preceding suicide (Baddeley et al., 2011; Fernández-Cabana et al., 2012), while others have only used a particular type of document.

## Limitations

There are some limitations to this study. First, because of the anonymity of the Internet, the blogs written by people who completed suicide may not be easily identified if there is no extra information in the content about the writer's sociodemographic details. Large-scale analysis may therefore be difficult to conduct. In addition, the comparison of case studies is not comprehensive because we only compared six major categories of language patterns. Other psychological and linguistic categories may also be relevant to suicide. Moreover, the psychological interpretation depends only on the CLIWC categories but not on the meaning of the content.

## Research and Practical Implications

Our study extends the findings of psycholinguistic analysis of suicides to the online document form. One of the fundamental research questions about using publicly accessi-

ble digitalized personal documents in suicide research is whether the findings will be similar to or different from studies of written documents. The answer to this question has significant implications for research and practice because it may affect our current understanding of suicide, which is mainly based on studies using printed documents. The study may act as a bridge between technology- and suicide-related research to advance suicide prevention practices (Li, Chau, Wong, & Yip, 2012).

This study extends the findings of psycholinguistic analyses of suicides to the Chinese context and provides an overview of case studies. It has investigated a Chinese youth suicide case and compared it with suicide cases from other countries. Besides, this study analyzed Chinese material, whereas most previous studies have only analyzed English material. Although there are studies that have analyzed non-English diaries, the material was translated into English before the analyses (Lester, 2009). It is noteworthy that the writer's original psychological meaning and linguistic style may not be completely preserved after translation. A unique aspect of the current study is that we used blog entries that were written in Chinese and were analyzed by the Chinese version of the LIWC. Some language patterns related to suicide are similar in both Chinese and English, such as the use of first-person singular pronouns. However, the linguistic and psychological word uses in the English and Chinese languages are slightly different (Li, Cai, et al., 2012). Besides, Chinese-specific categories may further reveal psychological cues of suicide that cannot be used in other languages. Future research in different languages should therefore test whether there are language differences in the text analysis of suicides. Since the English version of LIWC has been translated into many different languages (Pennebaker et al., 2007), language research can now be easily conducted in different cultures. As Chinese becomes the most populous language in the world (One World Nations Online, 2013), there is a need for research on Chinese text analysis of suicides.

## Conclusion

Behavioral and psychological changes illustrated in the personal documents of people who complete suicide have been used to examine suicide thoughts. Posting frequency and language use in Internet blogs offer insight into the suicidal process and allow for the timely introduction of suicide prevention interventions. The Internet provides rich communication content, making it much easier to analyze posting and language patterns that may predict human behaviors such as suicide. It is believed that technology will advance our knowledge of suicide language and suicide prevention. The language patterns in the blog were also compared with previous studies so as to extend the research on the psycholinguistic analysis of suicide to the Chinese social context and online document form. The consistently increased usage of first-person singular pronouns suggests a consistent association between social iso-

lation and suicide. However, the comparison did not show other clearly consistent patterns.

### Acknowledgments

This research was supported in part by the following grants: General Research Fund (#742012), Research Grants Council, Hong Kong Government, and the Azalea (1972) Endowment Fund, The University of Hong Kong, China. Approval of sabbatical leave to PWCW was granted by the Human Resource Committee (FHRC) of the Faculty of Social Sciences, The University of Hong Kong. The authors thank the Coroner's Court and the Chinese University of Hong Kong for their assistance in data collection.

## References

- Andersen, P. K., Borgman, O., Gill, R. D., & Keiding, N. (1993). *Statistical models based on counting processes*. New York, NY: Springer-Verlag.
- Baddeley, J. L., Daniel, G. R., & Pennebaker, J. W. (2011). How Henry Hellyer's use of language foretold his suicide. *Crisis*, 32(5), 288–292.
- Barak, A., & Miron, O. (2005). Writing characteristics of suicidal people on the Internet: A psychological investigation of emerging social environments. *Suicide and Life-Threatening Behavior*, 35(5), 507–524.
- Barnes, D. H., Lawal-Solarin, F. W., & Lester, D. (2007). Letters from a suicide. *Death Studies*, 31(7), 671–678.
- Chen, Y.-Y., Chen, F., Gunnell, D., & Yip, P. S. F. (2013). The impact of media reporting on the emergence of charcoal burning suicide in Taiwan. *PLOS ONE*, 8(1), e55000.
- Fernández-Cabana, M., García-Caballero, A., Alves-Pérez, M., García-García, M., & Mateos, R. (2012). Suicidal traits in Marilyn Monroe's fragments: An LIWC analysis. *Crisis*, 34(2), 124–130.
- Fortune, S., Stewart, A., Yadav, V., & Hawton, K. (2007). Suicide in adolescents: Using life charts to understand the suicidal process. *Journal of Affective Disorders*, 100(1), 199–210.
- Freuchen, A., & Grøholt, B. (2013). Characteristics of suicide notes of children and young adolescents: An examination of the notes from suicide victims 15 years and younger. *Clinical Child Psychology and Psychiatry*. Advance online publication. doi: 10.1177/1359104513504312
- Fullwood, C., Sheehan, N., & Nicholls, W. (2009). Blog function revisited: A content analysis of MySpace blogs. *CyberPsychology & Behavior*, 12(6), 685–689.
- Huang, C. L., Chung, C. K., Hui, N., Lin, Y. C., Seih, Y. T., Chen, W. C., ... Pennebaker, J. W. (2012). The development of the Chinese Linguistic Inquiry and Word Count dictionary. *Chinese Journal of Psychology*, 54(2), 185–201.
- Keith-Spiegel, P., & Spiegel, D. E. (1967). Affective states of patients immediately preceding suicide. *Journal of Psychiatric Research*, 5(2), 89–93.
- Leenaars, A. A. (1988). Suicide notes: Predictive clues and patterns. New York, NY: Human Sciences Press.
- Lester, D. (2009). Learning about suicide from the diary of Cesare Pavese. *Crisis*, 30(4), 222–224.
- Lester, D. (2010). The final hours: A linguistic analysis of the final words of a suicide. *Psychological Reports*, 106(3), 791–797.
- Lester, D., & Linn, M. (1998). The content of suicide notes written by those using different methods for suicide. *Perceptual and Motor Skills*, 87(2), 722–722.
- Lester, D., & McSwain, S. (2010). Poems by a suicide: Sara Teasdale. *Psychological Reports*, 106(3), 811–812.
- Lester, D., & McSwain, S. (2011). A text analysis of the poems of Sylvia Plath. *Psychological Reports*, 109(1), 73–76.
- Li, H., Cai, Z., Graesser, A. C., & Duan, Y. (2012). A comparative study on English and Chinese word uses with LIWC. In G. M. Youngblood & P. M. McCarthy (Eds.), *Proceedings of the Twenty-Fifth International Florida Artificial Intelligence Research Society conference: Marco Island, Florida* (pp. 238–243). Palo Alto, CA: The AAAI Press.
- Li, T. M. H., Chau, M., Wong, P. W. C., & Yip, P. S. F. (2012). A hybrid system for online detection of emotional distress. In M. Chau, G. A. Wang, W. T. Yue, & H. Chen (Eds.) *Proceedings of the Pacific Asia Workshop on Intelligence and Security Informatics 2012: Kuala Lumpur, Malaysia* (pp. 73–80). Heidelberg, Germany: Springer.
- One World Nations Online. (2013). *Most widely spoken languages*. Retrieved from [http://www.nationsonline.org/oneworld/most\\_spoken\\_languages.htm](http://www.nationsonline.org/oneworld/most_spoken_languages.htm)
- Nardi, B. A., Schiano, D. J., Gumbrecht, M., & Swartz, L. (2004). Why we blog. *Communications of the ACM*, 47(12), 41–46.
- Osgood, C. E., & Walker, E. G. (1959). Motivation and language behavior: A content analysis of suicide notes. *The Journal of Abnormal and Social Psychology*, 59(1), 58–67.
- Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). The development and psychometric properties of LIWC2007. Austin, TX: LIWC.net.
- Pennebaker, J. W., & Stone, L. D. (2003). Words of wisdom: Language use over the life span. *Journal of Personality and Social Psychology*, 85(2), 291–301.
- Pennebaker, J. W., & Stone, L. D. (2004). What was she trying to say. In D. Lester (Ed.), *Katie's diary* (pp. 55–79). New York, NY: Brunner-Routledge.
- Pestian, J. P., Matykiewicz, P., Linn-Gust, M., South, B., Uzuner, O., Wiebe, J., ... Brew, C. (2012). Sentiment analysis of suicide notes: A shared task. *Biomedical Informatics Insights*, 5(Suppl. 1), 3.
- Ruder, T. D., Hatch, G. M., Ampanozi, G., Thali, M. J., & Fischer, N. (2011). Suicide announcement on Facebook. *Crisis*, 32(5), 280–282.
- Shneidman, E. S., & Farberow, N. L. (1957). *Clues to suicide* (Vol. 56981). New York, NY: McGraw-Hill Companies.
- Stirman, S. W., & Pennebaker, J. W. (2001). Word use in the poetry of suicidal and nonsuicidal poets. *Psychosomatic Medicine*, 63(4), 517–522.
- Suler, J. (2004). The online disinhibition effect. *CyberPsychology and Behavior*, 7(3), 321–326.
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1), 24–54.
- Tsai, J. L., Simeonova, D. I., & Watanabe, J. T. (2004). Somatic and social: Chinese Americans talk about emotion. *Personality and Social Psychology Bulletin*, 30(9), 1226–1238.
- Wolf, M., Sedway, J., Bulik, C. M., & Kordy, H. (2007). Linguistic analyses of natural written language: Unobtrusive assessment of cognitive style in eating disorders. *International Journal of Eating Disorders*, 40(8), 711–717.
- Wong, P. W. C., Yeung, A. W. M., Chan, W. S. C., Yip, P. S. F., & Tang, A. K. W. (2009). Suicide notes in Hong Kong in 2000. *Death Studies*, 33(4), 372–381.
- Zhang, H., Yu, H., Xiong, D., & Liu, Q. (2003). HHMM-based Chinese lexical analyzer ICTCLAS. *Chinese language processing*, 17, 184–187.
- Zimmermann, J., Wolf, M., Bock, A., Peham, D., & Benecke, C. (2013). The way we refer to ourselves reflects how we relate to others: Associations between first-person pronoun use and interpersonal problems. *Journal of Research in Personality*, 47(3), 218–225.

Received May 24, 2013

Revision received November 25, 2013

Accepted November 25, 2013

Published online April 6, 2014

#### About the authors

Tim Li is a PhD student at the Department of Social Work and Social Administration, The University of Hong Kong, China. He received MSc and BEng degrees in computer science from the University of Hong Kong. His current research interests include the Internet and mental health, E-learning, social networks, and text mining.

Michael Chau is Associate Professor in the School of Business at the University of Hong Kong, China. His current research interests include information retrieval, Web mining, data mining, social media, electronic commerce, and security informatics. He has been ranked as the 14th most productive researcher in information science (1998–2007).

Paul Yip is Director of the Centre for Suicide Research and Prevention and Professor at the Social Work and Social Administration Department, University of Hong Kong, China. He is the

recipient of The Stengel Research Award (2011) of the International Association of Suicide Prevention. His interests lie in adopting a public health approach for suicide prevention and population health studies.

Paul Wong is Assistant Professor at the Department of Social Work and Social Administration, The University of Hong Kong, China. He is also a National Representative of the International Association for Suicide Prevention (IASP). He is a clinical psychologist and his research interest includes the Internet and mental health, community-based suicide prevention, and suicide notes.

#### Paul W. C. Wong

Department of Social Work and Social Administration  
Faculty of Social Sciences  
The University of Hong Kong  
Rm 511, Building C  
The Centennial Campus, HKU  
Pokfulam Road, Hong Kong  
China  
Tel. +852 3917-5029  
Fax +852 2858-7604  
E-mail paulw@hku.hk