A Corpus-based Study of Depressive Language in Online Teen Health Communications

Yu-Hua Chen and Chen Jin

University of Nottingham Ningbo China

Abstract

This paper presents a multidisciplinary study by using a corpus linguistic approach to investigate the topic of teen depression in an online discussion forum. The lexico-grammatical and semantic patterns of keywords in 129 online posts are explored, and five keywords (i.e. '*have*', '*feel*', '*know*', '*want*' and '*really*') are chosen for investigation. The results suggest that those posts are characterised by recurring expressions associated with intense emotions, which indicate this group's vulnerable mental state in relation to social contexts (e.g. family, school or relationship), and the semantic prosody of the text excerpts examined is predominantly negative (e.g. '*I feel so alone and angry*'). The findings shed light on the use of language expressions in a unique discourse of online health communications. (120 words)

Keywords: depression; online health communication; corpus approach; adolescents

1. Introduction

Online health forums have been increasingly popular for people seeking information or support about various stigmatizing medical conditions, such as depression, due to their easy access and anonymity (Ramirez-Esparza, Chung, Kacewicz & Pennebaker, 2008). Depression can be attributed to various factors, such as stress, anxiety, loneliness or problems in one's personal life, which in extreme cases can develop into severe conditions including suicide or self-harm behaviour (Harrington, 2001; Wilkinson, P., Kelvin, R., Roberts, C., Dubicka, B., & Goodyer, I., 2011; Wilson & Valstar., 2014). Depression concerns not only adults but also adolescents when they reach a particularly vulnerable developmental stage. According to Substance Abuse and Mental Health Services Administration (2014, p.37), 10.7 per cent of adolescents aged from 12 to 17 in the U.S (i.e. 2.6 million youths) suffered from 'at least one major depressive episode' in 2013, which was higher than the percentages in the previous decade (ranging from 7.9% to 9.1%). A similar statistic regarding the poor mental health of young people was reported in the U.K. According to the British Medical

Association (2006), about one in ten adolescents suffers from a psychological disorder. A more recent survey indicates that the number of young people going to A&E as a result of psychiatric conditions has increased significantly; in fact it more than doubled between 2010 and 2015, and self-harm, suicide or anorexia often seem to be associated with such psychiatric conditions among adolescent groups (British Medical Association, 2016).

Those statistics highlight the issue of mental health and its prevalence and severity among adolescents. Although some research has been conducted on similar mental health issues for adolescents in online discourses using corpus approaches (e.g. Adolphs et al. 2010; Harvey, 2013), very few studies has explored the language use of online teen health communications in terms of collocation, colligation, semantic preference and prosody (Sinclair1996; 2004). This study, therefore, aims to fill this gap and to identify lexico-grammatical and semantic patterns in the above online postings in a teen depression community by examining the most common lexical words which occur statistically more frequently in comparison with general English.

2. Literature review

2.1 Online health communications

An increasingly large number of studies have been carried out to analyze the text content or linguistic features (including functions, themes and styles) of online health communication. For example, drawing on sociolinguistics and medical sociology research, Angouria and Sanderson (2016) identify the functions of an online forum on Rheumatoid Arthritis and show how a 'collective identify' is constructed in this virtual community. To evaluate the predictive power of language use in a Web-based treatment course for young adults suffering from depressive symptoms, Van der Zanden, Curie, Van Londen, Kramer, Steen, & Cuijpers (2014) analysed the transcripts of chat sessions and found that the increasing use of 'discrepancy words' (e.g. '*should*', '*would*' or '*wish*') significantly correlated with improvements in depression.

In recent years, some studies have started to adopt corpus approaches in mental health related research as part of an integrated approach that combines corpus linguistic methods, and health-related contexts appear to 'offer new possibilities of data and theory building, as well as becoming a resource for practitioners themselves in clinical field settings' (Adolphs et al. 2004, p. 9). One such interdisciplinary project on online health communication was led by

Adolphs et al. (2010), where the language used by adolescents on a UK-based website, 'Teen Health Freak', was explored. The research team created a 2-million-word corpus of messages posted online from 2004 to 2009; the messages came from adolescents who submitted their questions in relation to health to an online GP (general practitioner) persona. Words and word clusters which occur statistically more frequent in their corpus in comparison with a reference corpus (i.e. keywords and key phrases defined in Corpus Linguistics) were identified as the main topics of the 113,480 messages in this longitudinal corpus, and these areas of concerns were: 1) sex, pregnancy and relationships, 2) sexual body parts, 3) body changes, 4) smoking, drugs and alcohol and 5) weight and eating. Based on this large corpus, Harvey and his associates further elaborated the topics of sex, sexual health, psychological distress (including suicide and self-harm) and depression in a variety of studies, which were foregrounded by corpus analysis as recurring themes in adolescent health communication (2007, 2008, 2012, 2013).

In a cross-linguistic study conducted by Ramirez-Esparza et al. (2008), computerized text analyses were used to compare language use in English and Spanish online forums concerning depression; the researchers focused on the identification of linguistic markers of depression and themes arising from those online posts. They found significantly higher use of first person pronouns and negative emotional words (e.g. *hurt*, *ugly*) in their data. In terms of themes, which were identified via a statistical approach called 'Meaning Extraction Method' (ibid, p.105), the main topics addressed in those posts were highlighted from the frequency wordlist (i.e. all the words occurring in the corpus), and they fall into a number of semantic clusters including treatment (e.g. *medication, therapy*), family (e.g. *mom*, *daughter*) and school (e.g. *university*, *college*).

As can be seen, various corpus approaches are utilized to identify themes and linguistic features in a collection of representative texts with a specific purpose, i.e. a corpus. Few studies, however, have reported corpus analyses beyond the single-word level of such online health communication. In rare cases where sample concordance lines are revealed, they are often used to exemplify connotations (e.g. the example of '*normal*' in Harvey et al. 2008, p.776) as opposed to data per se where systematic analyses can be drawn upon. Focusing on an online forum on teenage depression, the current study will therefore not only aim to identify recurrent themes in online health communication using keyword analysis but also to explore the lexico-grammatical and semantic profiles of keywords extracted by examination of sample concordance lines.

2.2 Corpus Approaches

A variety of corpus methods applied in the current study will be introduced in this section. This includes a frequency list and the keyword analysis used to identify recurring themes, and the framework of Unit of Meaning from Sinclair (1996; 2004) used to explore the linguistics features of online posts.

A frequency list contains all the words used in a corpus with information about the corresponding occurrence of each word. It can be used to quickly capture 'the main foci of a corpus', as evidenced in frequently recurring words (Baker, 2006, p.133). Frequency alone, however, is often not robust enough to represent the 'aboutness' of content in a corpus, and so statistical measures are therefore required, which leads to the introduction of keyword analysis. Retrieved by statistical analysis such as chi-square or log-likelihood tests, keywords are used to highlight the items in a target corpus which are statistically more or less frequent when compared to those in a reference corpus (Hunston, 2002, p. 68). Keywords which are content words (e.g. nouns, adjectives and lexical verbs), according to Bondi and Scott (2010), serve as the 'principle indicators' of the theme of a text. As for keywords which are grammatical words (e.g. pronouns, conjunctions, auxiliary verbs), these words can suggest the style of a text and also interact with the discourse of a text. As mentioned earlier, the themes in the online teen depression forum will be identified through keywords, and the focus here will therefore be on the use of content keywords as opposed to grammatical words. After the keywords are extracted, the use of the most frequent lexical words will be investigated at the lexico-grammatical and semantic levels, applying Sinclair's framework of analyzing units of meaning (1996; 2004) in order to gain insights into the linguistic features of the compiled corpus. This framework comprises four types of co-occurrence relations: colligation, collocation, semantic preference and semantic prosody. Colligation refers to patterned choices of grammatical collocates that co-occur with a search word or node word. Collocation refers to lexical items a node word tends to co-occur with, and one way of identifying collocation is via the examination of concordance lines (McEnery & Hardie, 2012, p. 126), which is also what the current study adopts here.

Semantic preference is identified as the semantic field of a set of lexical collocates of a node word. The notion of semantic prosody discussed by Sinclair is further clarified by Hunston and Thompson (2001) and refers to evaluative orientation in a context comprising adjacent words or phrases which tend to co-occur with the node word. Concordance lines are

inspected carefully by the research team to extract information regarding these four aspects. 'A concordance is a table of all of the occurrences of a linguistic item in a corpus' (Baker, 2006, p. 21), and it is a powerful tool when examining the linguistic context of language use, which will be demonstrated in the analysis section with examples.

The objectives of the study are twofold. The first is to identify themes and linguistic features in an online discussion forum concerning teenage depression, and the other is to explore the use of corpus methodology at the lexico-grammatical and semantic levels in the investigation of discourse in online health communications. The research questions are formulated as follows:

1. What are the recurring themes in an online discussion forum on teenage depression as revealed by keywords analysis?

2. What are the linguistic features of the most frequent lexical keywords explored in terms of collocation, colligation, semantic preference and prosody?

3. Methodology

3.1 Data collection

MedHelp (http://www.medhelp.org/) is an online health community based in the U.S. which allows users to post questions or comments about health and seek advice from other online members. The Teen Depression Community is one forum under MedHelp, and the webpage (http://www.medhelp.org/forums/Teen-Depression/show/185) has the following statement:

> *This patient support community is for discussions relating to teen depression, abuse, aggressive behavior, alcohol and drug abuse, anxiety, behavioral issues, fatigue, gastrointestinal problems, grief loss, parent issues, relationship problems, school issues, self-esteem, sexuality, sleep disorders, and step-families.*

One hundred and thirty posts from 1 June 2014 to 6 December 2015 were collected from the Teen Depression Community, covering about a period of around 18 months. One post during this period that was recognized by other users as being posted in the wrong community was removed. Otherwise, no duplicated posts are identified. All the final posts are put together as plain text for processing and analysis. The final corpus amounts to 25,806 words and 129 posts in total.

With regard to ethics or copyright of data collection, it is a thorny and controversial issue in corpus construction and has been much discussed (see, for example, Grinyer, 2007 or McEnery & Hardie, 2014). The rationale for collecting data online to build our own specialized corpus comes from arguably one of the largest corpora, Corpus of Contemporary American English (Davies, 2008–), where a massive amount of copyrighted material is collected. The reason why the use of text is allowed (at least under US Fair Use Law) is 'because of the limited 'Keyword in Context' (KWIC) displays', which is similar to the logic of 'snippet defence' used by Google that processes an enormous amount of text but only allows end users to access 'snippets' (ibid., for more details, see: http://corpus.byu.edu/copyright.asp). As our full corpus is not made available to the public, the current project follows the same rationale of 'snippet defence' because only a limited amount of data will be disclosed to the public in the format of KWIC displays in the paper. In terms of ethics, according to terms of use on MedHelp (http://www.medhelp.org/legal/terms_of_use), all user generated content in the public areas (including public forums) is non-confidential as specified in the statement 'By submitting communications or content to the Public Areas, you agree that such submission is non-confidential for all purposes'. As online health communications can involve highly sensitive information such as names or dates, all the data collected in this project still underwent a robust process of scrutiny and anonymization although it is believed that most of the users use nicknames/pseudonyms. User identifications and posting dates that accompany the posts are excluded. In other words, only non-identifiable information is included. In terms of the possibility of including vulnerable groups such as children (aged under 16), because it is impossible to determine the age of those anonymous online users, we paid particular attention to potentially age-sensitive content but did not discover any. In addition, the posts originated from a period of 1 June 2014 to 6 December 2015, at least one to three years before the paper was produced. It is believed that the data collected would not be as sensitive because they are not as timely as when they were posted although we do acknowledge the sensitivity of the topic of depression and the group involved.

3.2 Data processing

3.2.1 Spelling standardization

Although online posts on discussion forums fall into the written register, the language tends to have spoken features, including shortened forms (e.g. '*u*' instead of '*you*', '*kinda*' instead of '*kind of*'), typos (e.g. '*Im' instead of 'I'm*') or simply errors (e.g. '*side affect*' instead of

'*side effects*'). As the above spelling variations can impact on the results in keyword and frequency lists, standardization of spelling is carried out in order to improve the validity of follow-up analysis. Spelling errors are therefore corrected and spelling variations standardized with a combination method of manual correction and the automatic checking function from MS Office Word 2013. Concordance lines are checked whenever in doubt. The taxonomy of standardization (see Table 1) is adapted from a large research project mentioned earlier by Adolphs, Mullany, Smith, Harvey & Macfarlane (2010), where various types of spelling errors, abbreviations and acronyms extracted from their corpus of emails on teen health are documented. Note that the determination of standardization type can sometimes be rather arbitrary. For example, '*your*' instead of '*you're*' may fall into the category of either a typographic or a phonetic error, but as the main purpose of this procedure is to standardize the spelling to identify the themes and lexico-grammatical features of an online discourse via frequency and keywords, we consider that the categorization of such ambiguous cases has little impact on the results of follow-up analysis.

Table 1 Types of spelling standardization (adapted from Adolphs et al., 2010)

Type of standardization	Subtype	Examples from the current study	
		original	standardized
Typographic	Space deletion	*awhile*	*a while*
	Space insertion	*you self*	*yourself*
		mean time	*meantime*
	Missing apostrophe based on context	*Im*	*I'm*
		anothers	*another's*
	Redundant apostrophe based on context	*any of it's associated symptoms*	*any of its associated symptoms*
	Missing hyphen	*self harm*	*self-harm*
	Misspelling	*side affects*	*side effects*
Phonetic	Homophones substitution	*your*	*you're*
		too	*to*
Chat style	Letters or shortened forms for full words	*n*	*and*
		r	*are*
		u	*you*
		bf	*boyfriend*
		yrs	*years*
	Colloquial expressions	*kinda*	*kind of*
		wanna	*want to*
		cause	*because*
Emphasis	Use of capital letters	*I'm bleeding RIGHT NOW*	*I'm bleeding right now*
	Repeated alphabet	*sooooo*	*so*

3.2. Extraction of keyword and frequency lists

A corpus tool, WordSmith 5 (Scott, 2008), is utilized for data processing, and the Brown corpus, which contains 1,034,148 words, is chosen as the reference corpus. The Brown corpus is a collection of general American English which comprises text samples from a wide range of genres, such as news reports, novels, short stories and government documents (Francis & Kucera, 1979), and is therefore considered comparable with the variety of English used on Medhelp, which is a website based in the U.S. The other reason for choosing the Brown corpus is because it is a freely available corpus with a manageable size that can be easily processed with WordSmith tools. Note that the choice of a more general reference corpus might impact significantly on the extraction of keywords. Although other researchers seem to prefer general spoken English as a reference when comparing to a similar corpus of online health communication via emails or messages from teenagers who seek advice or information (e.g. Adolphs et al. 2010; Harvey 2008, 2012), we think it is also reasonable to use written English as the reference because the language used on an online forum often entails a mixture of both spoken and written features. We also acknowledge that the Brown Corpus is somewhat outdated. While other larger and newer corpora are available, however, they are either not easy to access or not possible to process with existing corpus tools. Yet our decision for selecting the Brown Corpus as the reference corpus is based on repeated experiments with other general corpora available in an online corpus tool, 'Compleat Lexical Tutor', developed by Tom Cobb (http://www.lextutor.ca/key/). Through a close inspection of different sets of keywords when compared with other general corpora and their corresponding concordance lines, we were satisfied with our choice of the Brown Corpus. Additional comparisons with the 520-million-word Corpus of Contemporary American English (Davies, 2015) or the 100-million-word British National Corpus (BNC Consortium, 2001), which includes both spoken and written registers, will also be presented where appropriate.

3.3 Procedures

First, the top 50 most frequent words and top 50 keywords (p<0.000001) are generated by Wordsmith 5 (see Appendices I & II).

For words which can serve as both grammatical and lexical words (e.g. '*have*' as in '*have had*' for the former and '*have depression*' for the latter), instances of these words functioning as lexical words will be counted manually by inspecting all concordance lines, and the calculated figures will be compared with other predominantly lexical words. Similarly, the grammatical functions of the top six lexical keywords will be identified by examining all the

concordance lines, and only corresponding occurrences of lexical meaning will be included manually.

When analyzing the top six lexical words from the keywords list, the first 30 concordance lines are examined for further analysis, these are randomly sorted by Wordsmith 5. The analysis will follow the sequence of colligation, collocation, semantic preferences and semantic prosody. The focus of analysis of these four aspects depends on specific node words. Dominant colligation patterns with a total hit of over five times are reported. In cases of collocation, the occurrences of lexical collocates are recorded, and those collocates are grouped semantically, with semantic fields which contain over three lexical collocates being reported. In terms of semantic prosody, 30 randomized concordance lines are grouped on the basis of positive, negative or neutral evaluation revealed by concordancing. All the semantic judgement tasks were completed by two researchers, and a third rater was involved in checking the semantic analysis, which also helped with final decisions in case of disagreement between the first two raters.

4. Analysis and findings

4.1 Keyword analysis

The top 50 extracted keywords are categorized into open class words and closed class words, some of which are further divided into subgroups depending on their semantics or parts of speech (POS) (see Tables 2 and 3). It is acknowledged that some of the groupings may be determined arbitrarily as meaning may depend on the context. For example, words such as '*like*', '*happy*', '*do*', '*have*', '*haven't*' are ambiguous unless presented in context (e.g. '*not happy*'). Speaking overall, the corpus appears to have an informal style, as expected, which can be identified from the use of shortened forms of ten grammatical words, such as '*I'm*', '*it's*', '*can't*', and three colloquial lexical expressions, including '*dad*' and '*mom/mum*'.

Table 2 Open Class Keywords

Semantic/POS Grouping		Open Class (n=freq.)
Affective (emotion/attitude)	negative	*depression (62), depressed (42), hate (36)*
	positive	*happy (44)*
Social	family	*dad (31), mom/mum (58), mum (19), parents (59)*
	others	*boyfriend (24), friends (91), school (92)*
Other verbs		*feel (141), help (83), know (136), get (86), going*

		(59), have (306), like (158), started (45), think (87), want (121)
Other nouns		porn (25), suicide (21)

Table 3 Closed Class Keywords

Grammatical Grouping		Closed Class
Pronoun	1st personal pronoun	I (2014), I'm (234), I've (88), me (428), my (706), myself (98)
	Others	it's (46), she's (16), someone (39)
Negation		can't (55), don't (188), didn't (46), doesn't (20), haven't (23), wasn't (16)
Others	Adv.	anymore (30), just (180), really (112)
	Conj.	because (124), but (311), so (173)
	Verbs	do (154)
	Others	about (139), am (146), lot (50)

As can be seen in Table 2, the content keywords provide a snapshot of the various themes, including the prevalence of affective expressions (e.g. '*feel*', '*depression*') and familial and social roles (e.g. '*mom*', '*friends*') in this online discourse of teen mental health. There are five keywords in the affective grouping, and two of them, '*depression*' and '*depressed*', correspond to the title of this forum, 'teen depression community'. Another seven keywords fall under the social grouping while six refer to the role of someone else (i.e. '*dad*', '*mom/mum*', '*parents*', '*boyfriend*' and '*friends*') by whom the adolescents are often surrounded. The remaining word in this semantic field, '*school*', suggests the social community that adolescents tend to socialize in. It is evident that these posts are highly concerned with affective conditions and social surroundings.

In terms of the grammatical keywords presented in Table 3, the dominance of first person pronouns in various forms with high frequencies (e.g. '*I*', '*I'm*', '*me*', '*myself*') is noteworthy. Together with a variety of negation forms (e.g. '*can't*', '*don't*', '*didn't*', '*wasn't*'), the grammatical keywords suggest a personal tone with a high degree of self-awareness and negative elements in those online posts.

4.2 Analysis of top five lexical keywords

The keyword list (Appendix II) shows that the top five most frequent lexical words are '*have*', '*feel*', '*know*', '*want*' and '*really*', with occurrences of 306, 141, 136, 121 and 112, respectively. Among the 306 instances of '*have*', 90 of them are identified in concordancing as grammatical words used in structures such as '*have been*', '*have done*', and thus they are eliminated from the concordance analysis. The final frequency of '*have*' functioning as a lexical word is therefore 216 times. Interestingly, four out of the five lexical words are verbs, the only exception is an adverb, '*really*'. For each of the lexical keywords, the first 30 concordance lines (randomly sorted) are manually examined for lexico-grammatical and semantic analysis. All the samples of concordance lines can be found in Appendix III.

4.2.1 *Have*

Colligation

From the samples of concordance lines, it is observed that among the left collocates of '*have*', the dominant colligation pattern can be summarized as 'pronoun + (lemma *DO/DON'T*) + *have*', and the first personal pronoun '*I*' occupies the position of pronoun 22 times in this structure. Occasionally there is the insertion of an adverb, '*probably/really*', or an auxiliary verb, '*do*', for emphasis between the pronoun and '*have*' (e.g. '*I really have no purpose in this world*', '*I do have two best friends*'). As for the sub-pattern 'pronoun +lemma *DON'T* + *have*', it has seven hits with '*I*' in the position of pronoun six times. For the right collocates of '*have*', the pattern '*have*+ noun' is dominant, with 28 hits, and occurrences include instances with nouns modified by adjectives. On the basis of the above analysis, the general colligation pattern of '*have*' can be summarized as 'pronoun + (lemma *DON'T*) + *have* + (adjective) + noun'.

Collocation and semantic preference

Considering the word class of '*have*' as a verb and the dominant colligation pattern of '*have* + noun' discussed above, the identification of collocations focuses on the noun collocates following the node word '*have*'. Noun collocates and their associated semantic fields are identified and listed in Table 4. After consulting the Oxford Dictionary online (Oxford Dictionary, 2015), it is found that the primary senses of '*have*' used here are 1) '*possess*', 2) '*experience*' and 3) '*suffer from illness/ ailment/ disability*'. It appears that '*have*' tends to associate with words which express affective and social connotations. In addition, seven

words in the semantic field of 'social' relate to social roles. The tendency of '*have*' to collocate with words in these two fields corresponds to the findings of 'aboutness' in the overall keyword analysis. Collocates in relation to both physical (e.g. '*cancer*') *and* mental health (e.g. '*depression*') are also found here.

Table 4 Semantic preference of '*have*'

Semantic field	Noun collocates		type	token
Affective	*heart, interest, support (2), trouble*		4	5
Social	Identity/relationship	*brother or sister, cousins, family, friend(s) (3), girls, parents*	6	8
	Life	*job, life, work experience*	3	3
Health	*cancer, depression, energy, panic attacks*		4	4
Others	*anything, anyone, bets, idea (2), night(s), purpose, record, sex*		8	10

Semantic prosody

The attitudinal meaning expressed in each of the sample concordance lines (see Fig. 1) is examined and listed in Table 5. Based on the evidence from concordancing, it can be seen that '*have*' is dominantly used in negative prosody, which often expresses disappointment with the status surrounding this teenage group, the struggles with the desire to possess something, and the suffering from undesirable situations or personal feelings. It is also noted that the colligation pattern of 'lemma *DON'T+ have*' contributes significantly to expressions of struggling and suffering.

Table 5 Semantic prosody of '*have*'

Attitude	Line numbers	Total hits
Positive	5,7,8,26,39,41	6
Negative	1,2,3,4,6,13,14,15,20,22,23,24,27,28,34,36,43,44	18
Neutral	10,12,21,25,31,42	6

Figure 1 Concordance lines of 'have'*

N	Concordance
1	slowly and they need to work fast. I have panic attacks at school, I feel as if
2	things and now I'm scared. I want to have my own life....without these
3	I don't want to hurt my baby and I don't have much support from family and very
4	his mind completely saying he doesn't have the energy or motivation and that
5	good girl who must be good at school have a good job. Everyone else keeps
6	no more jamming and its like i don't have interest in school at all like im
7	with everyone like i want to be alone. I have loving parents and i know they will
8	. Everyone else keeps an eye on me. I have the idea that i want to prove that i
10	course the answer is no but I still look. I have maybe 1 friend and no female
12	. the first thing i noticed was that I didn't have nearly as many as the girls around
13	just friend thirsty. i don't really feel like i have a best friend, i don't have anyone I
14	feel like i have a best friend, i don't have anyone I can genuinely trust or go
15	my mom waste money like that i didn't have the heart to tell him what i was
20	a cancerous tumor below her kidney. I have cancer in my veins. My grandma is
21	and that its time for her to meet god. I have 2 younger cousins, 3 mo apart. My
22	of being beaten was still with media have trouble sleeping because I dream
23	one in the area is boring me because I have zero work experience. I did receive
24	restrictive diets that now cause me to have bets or binge, eating disorder,
25	52 days in a single semester). If I don't have a bad night, I usually lay down
26	don't bother the baby or me anymore. I have my family support, but I feel that
27	It's about 2 am where I live, whenever I have bad nights I don't fall asleep till
28	happening during the school year and I have a bad attendance record (I missed
31	and forgot me? Her friends probably have no idea how close we were. Heck
34	again? Hi. I'm a 13 year old and i think i have depression get upset very easily
36	of the people I hang out with I do drugs, have sex, and drink. My parents are
39	so there are many kinds of people. I do have two best friends and some other
41	Hello my name is Lisa and I'm 16. I have a great family. We are in good
42	that I only live with my mom and I don't have any brothers or sisters. I honestly
43	was really talkative now I struggle to have anything to say in a group of friends
44	he knows how correct he is. I really have no purpose in this world. i don't

* Note that the instances of '*have*' functiong as grammatical words (e.g. '*have been*', '*have done*') are eliminated from the concordance lines. The sequence numbers above, therefore, have some omissions.

4.2.2 *Feel*

Colligation

The left collocates of '*feel*' shows that the dominant colligation pattern is 'pronoun + *feel*', which has 21 instances (including two instances of insertion '*just/only*'), with the first personal pronoun being '*I*' in all cases. Even in the remaining concordance lines, six out of nine instances still relate to how the authors of those posts talk about their own feelings (e.g. '*I can't feel happy*', '*my life feel so alone*', '*I don't feel a hundred percent fit with them*'). With regards to the colligation pattern of the right collocates, there are two primary structures: '*feel like* + clause/noun/pronouns' and '*feel* +adjective' are dominant. The former has nine occurrences with seven instances of '*feel like* + clause' (e.g. '*I feel like I can't breathe and I panic*) while the latter has 16 occurrences (e.g. '*why I feel so alone and angry a lot of the time*). Based on the above analysis, the dominant colligation pattern of '*feel*' can be summarized as either 'pronoun+ *feel like* + clause/noun/pronoun' or 'pronoun + *feel* + adjective'. It is noticeable that '*feel*' is integrated into '*feel like*' as a fixed phrase in use.

Collocation and semantic preference

Since '*feel*' functions as a verb in the corpus, a collocation search for '*feel*' will focus on the 13 adjective collocates following the node word '*feel*' (Table 6). According to the Oxford Dictionary (2015), the node word '*feel*' either carries the meaning of '*experiencing*' (e.g. '*I just feel different from everyone now*', '*sometimes I feel so alone*') or forms the fixed expression '*feel like*' (e.g. '*I feel like he's given up all hope on me*' or '*I feel like if the story is repeating*') in the concordance lines examined. The main semantic field of collocates appears to be affective, it mainly expresses a range of negative emotions in those posts. Note that the only two 'positive' collocates actually occur in the construction of negation: '*I don't feel welcome*' and '*I can't feel happy*'.

Table 6 Semantic preference of *'feel'*

Semantic field	Adjective collocates	type	token
Affective	*angry (2), alone (3), depressed, different, fit, happy, horrible, lonely (2), miserable, stupid, uncomfortable, welcome, worse (3)*	13	19

Semantic prosody

The semantic prosody of the concordance lines is presented in Table 7, and sample concordancing is presented in Figure 2. It can be seen that '*feel*' is almost always used in negative prosody, expressing a desperate emotional state, often without actually using the exact word '*depressed*', as in '*I feel lonely and sad I feel stupid ugly fat worthless and unhappy*', '*I feel that my entire world is falling apart*' or '*I have started to feel like nothing but a burden*'.

Table 7 Semantic prosody of *'feel'*

Attitude	Number(s) of concordance line(s)	Total hits
Positive		0
Negative	1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,17,18,21,22,23,24,25,26,27	24
Neutral	13,19,20,28,29,30	6

Figure 2 Concordance lines of '*feel*'

N	Concordance
1	fast. I have panic attacks at school, I feel as if I'm having a heart attack, but I
2	even after being on birth control, I don't feel welcome. We migrated to the US
3	they will do everything for me but i just feel different from everyone now. And i
4	like i don't know what to do with my life feel so alone that's why I'm trying to fit
5	year 10 and my birthday is 7 July 2000 I feel lonely and sad I feel stupid ugly fat
6	My dad thinks that I don't like him, and I feel like he's given up all hope on me.
7	feel stupid ugly fat worthless unhappy. I feel so horrible and lonely my brother
8	is 7 July 2000 I feel lonely and sad I feel stupid ugly fat worthless unhappy. I
9	herself now but I would love if she didn't feel this depression that she feels
10	anymore. I have my family support, but I feel that my entire world is falling apart
11	to see me until the baby was born. I feel miserable. I told him if that's what
12	. And I hate it. It ruins my day. I can't feel happy. How do I stop this? There
13	will be without a father just like me. I feel like if the story is repeating and
14	to feel like nothing but a burden. I feel like everything I do has a negative
15	months. The past week I have started to feel like nothing but a burden. I feel like
16	dad has proven her wrong, but at times I feel so lonely I turn to something to help
17	want me anymore and mentally I feel worse and worse every day. Suicide
18	what i was really feeling so i said no. i feel like I've missed my only shot at help
19	not go to school. That's not the case, I feel like if I go to school everyone there
20	old. I think that it traveled down to me. I feel like me, as a person, will never be
21	other close friends. But sometime I feel so alone...I don't think I belong with
22	everyone there hates me. I don't want to feel worse than I already do. Can
23	and laughing but then, just like that, I'll feel all depressed and angry. I'm not the
24	confidence. I just don't understand why I feel so alone and angry a lot of the time.
25	my mind goes straight to it and I feel like I can't breathe and I panic it
26	more confident with food now although I feel a bit uncomfortable eating around
27	I'm not as cool as them. And I don't feel a hundred per cent fit with them.
28	. I would rather do homeschooling but I feel like I would barely do that. I don't
29	live surrounded in bad. Although I only feel this way about myself and will
30	is "ew... jk ily" supposed to make me feel when the girl next to me was being

4.2.3 *Know*

Colligation

The dominant colligation pattern of '*know*' on the left is 'pronoun+ lemma *DON'T + know*'. It has 19 instances including cases where adverbs (such as '*always*' or '*really*') are inserted into the structure. Similar to the previous colligation patterns, 18 out of 19 concordance lines have the pronoun '*I*' forming the dominant expression '*I don't know*'. With regard to the

colligation pattern on the right, the common structures are '*know + what to do*' (e.g. '*I don't know what to do with my life*') and '*know + (if) clause*' (e.g. '*I know I need help*', '*I know I shouldn't*', '*I want to know if it's too late to heal this*), with six and eight hits, respectively. Nine of the remaining concordance lines have *wh-* clauses (starting with '*what/where/why/whether/who*') following '*know*', such as '*he doesn't know how offended I really get*' or '*I don't know where I should belong*'.

Collocation and semantic preferences

From the above colligation analysis and an inspection of sample concordance lines, lexical collocates that have a frequency exceeding five and directly relating to '*know*' are rare, except for six instances of the '*don't know what to do*' construction. The collocation pattern and semantic preference of '*know*' in concordance lines are therefore not easy to identify. According to the Oxford Dictionary (2015), the two primary meanings of '*know*' used in these concordance lines are '*being aware of*' (e.g. '*I also started cutting myself in August. I know I shouldn't*' or '*I know it was a long time ago*') and '*having knowledge*' (e.g. '*I don't know where I should belong*' or '*people just seem to hate me and I don't know why*'), with the latter accounting for nearly three-quarters of the sample (23 out of 30 lines). The numbers of corresponding concordance lines are listed in Table 8, and the sample concordance lines in Figure 3.

Table 8 Senses of *'know'*

Sense	Number(s) of concordance line(s)
'*be aware of*'	1,2,5,6,9,16
'*have knowledge*'	3,4,7,8,10,11,12,13,14,15,17,18,19,20,21,22,23,24,26,27,28,29,30

Figure 3 Concordance lines of '*know*'

N	Concordance
1	. I've been trying not to, but it's difficult. I know I need help, I just don't know how I
2	I also started cutting myself in August. I know I shouldn't, and I had stopped for
3	shape, and i want to stop this. I want to know if it's too late to heal this. I would
4	difficult. I know I need help, I just don't know how I could possibly tell anyone.
5	have a support group either. I don't even know what I'm asking but okay let me
6	care why did you write this?" I don't know really, if I get help that's fine but if
7	beginning of this school year! I just don't know what to do, I'm so stressed and
8	I feel like I would barely do that. I don't know. What should I do? I'm 13 years
9	teenager". i laugh but he doesn't know how offended i really get. then he
10	was adopted like everyone my relatives know it while i don't it pressures me in a
11	. I lost focus at everything like i don't know what to do with my life feel so
12	to be alone. I have loving parents and i know they will do everything for me but i
13	in my mind it seems that i don't know what to do. I became lost not
14	people just seem to hate me and I don't know why. Hi I'm Brittany barrett I'm a
15	her that I'm her son and I'll always know her son and her son only. I am a
16	more effin years till i graduate. I don't know i seemed lost. I lost focus at
17	understand this issue...I just don't know what to do. help? My boyfriend of
18	who I should be with at school. I don't know where I should belong. I don't
19	over think about everything. I don't really know like who I should be with at school.
20	to go back to my old self. I just don't know how and don't know if/when it will
21	know where I should belong. I don't know how to be social. My school is a
22	I still do it. Only a couple of people know I still do it, one is my 12 year old
23	. My mum knows I've tried it but I don't know whether she knows I still do it.
24	and I panic it might be dumb but I don't know what to do Hi. My name is
25	, it means a lot. Goodbye. I'm 18 and I know it was a long time ago but when I
26	old self. I just don't know how and don't know if/when it will ever happen. What
27	she sheltered me so much I didn't know what she was doing but I had a
28	just like their father. I was too young to know any of this and didn't find out till
29	to getting married having kids. I don't know what to do though because I am
30	. Things a child at that age should never know or hear. This in turn has caused

Semantic prosody

As aforementioned, the semantic prosody of '*know'* is doubly rated by the researchers, and the results are presented in Table 10. As can be seen, over half of the instances of '*know'* are used in negative prosody where the authors of those posts often find themselves in a clueless and helpless state, not having the necessary knowledge to cope with problematic situations (e.g. '*I don't know I seemed lost'* or '*I was too young to know any of this*'), which relates to the second dictionary meaning discussed previously. The colligation pattern '*don't know'*

plays a significant part in determining the semantic prosody here as this formulaic sequence accounts for two-thirds of the concordance lines (see Fig. 3).

Table 9 Semantic prosody of '*know*'

Attitude	Number(s) of concordance line(s)	Total hit
Positive	12	1
Negative	4,5,7,8,9,11,13,14,16,17,18,19,20,21,24,28,29,30	18
Neutral	1,2,3,6,10,15,22,23,25,26,27	11

4.2.4. *Want*

Colligation pattern

The primary colligation pattern of '*want*' on the left is 'pronoun + (lemma *DON'T*) + *want*'. The sub-pattern 'pronoun + lemma *DON'T* + *want*' has 11 hits, with the first personal pronoun '*I*' in nine of them. Another sub-pattern, 'pronoun + *want*', has 13 occurrences, and all the pronouns are '*I*'. As for the colligation pattern on the right, the structure '*want* + infinitive to verb' is dominant, with 22 occurrences. The colligation pattern can therefore be summarized as 'pronoun + (lemma *DON'T*) + *want* + infinitive to verb'.

Collocation and semantic preferences

The above colligation pattern stimulates further interest in exploring verb collocates in the structure of infinitive '*want* + infinitive *to do something*'. After the concordance lines are inspected and annotated, the verb collocates and their corresponding sematic fields are listed in Table 10. The verb '*want*' appears to be mainly used with the meaning of '*wish*' (Oxford Dictionary, 2015) in these collocation patterns. As the meanings of most verb collocates (such as '*take*', '*have*' or '*be*') depend on the multi-word expressions that they are part of, the whole expressions (such as '*do therapy*' or '*have my own life*') are included in Table 10. It should also be noted that there are three instances in the semantic field of harmful acts, '*cut*', '*hurt*' and '*self-harm*', and even the phrase '*get better*', which seems positive in appearance, but actually belongs to a longer negative statement, '*I don't <u>want to get better</u>*'.

Table 10 Semantic preference of '*want*'

Semantic field	Verb collocates	type	token
Health	*do therapy, get better, take medicine*	3	3
Status	*be alone/around/there/with, have my own*	7	7

	life, make it through, turn out this way		
Harmful acts	*cut, hurt (2), self-harm*	3	4
Other	*know, lie down, prove, quit, say, stop, tell, wake up*	8	8

Semantic prosody

Semantic prosody traits of the discourse are listed in Table 11. The node word '*want*' is primarily used in negative prosody that often suggests a struggle for love and attention (e.g. '*...like even my loved ones don't want me around anymore*', '*I want to tell my mum but I'm scared*'), desperate thoughts (e.g. '*I just want the pain to stop*', '*I want to cut so badly*') or fighting against those negative thoughts (e.g. '*how to cope with the feeling I get when I want to self-harm*' or '*I don't want to hurt myself*'). The mental states expressed through the use of '*want*' in those posts are highly conflictive.

Table 11 Semantic prosody of '*want*'

Attitude	Number(s) of line(s)	Total hits
Positive	7	1
Negative	3,4,5,6,8,9,10,11,12,13,14,15,16,17,19,20,21,22, 23,24,25,26,27,30	25
Neutral	1,2,18,28,29	5

Figure 4 Concordance lines of '*want*'

N	Concordance
1	it a couple times and she does not want to take medicines and I'm
2	and I'm assuming she doesn't want to do therapy. I don't know if there
3	me and like even my loved ones don't want me around anymore. My boyfriend
4	tired of getting along with everyone like i want to be alone. I have loving parents
5	many other things and now I'm scared. I want to have my own life....without
6	own life....without these thoughts.... I want to tell my mom but I'm scared
7	an eye on me. I have the idea that i want to prove that i am worth it to my
8	. My boyfriend of two years doesn't want me anymore and mentally I feel
9	but haven't told anyone because I don't want them thinking it's for attention. My
10	how to cope with the feeling I get when I want to self-harm or is there nothing I
11	or is there nothing I can do? I don't want to hurt myself but it seems that I
12	like if the story is repeating and didn't want to turn out this way. Hi, This is the
13	through the night. But sometimes I don't want to make it through, those nights
14	all she can say is "I'm sorry". I don't want pity, I don't want people to think of
15	is "I'm sorry". I don't want pity, I don't want people to think of me as any less
16	because I honestly don't care. I don't want to get better, I don't see the point
17	I do go its torture, I'm starting to not want to be around my friends because
18	to ask me about school and stuff. I just want to lie down. I wish I could just run
19	a solution and can't think of one. I just want the pain to stop.. This is going to
20	some of the people are awful there. I want to quit but I'm too anxious to quit
21	if I quit. I'm miserable and don't want to be there because it's destroying
22	is the only way out but I know I just want a solution and can't think of one. I
23	sleep too much or too little, and I never want to wake up. I've gained ten pounds
24	and I want to cut so badly but I don't want to hurt my baby and I don't have
25	if I'm having a heart attack, but I don't want to say anything....... Since I was 7,
26	and follow everything my parents want me to do, I'm never in the mood to
27	sides r just making things worse and I want to cut so badly but I don't want to
28	not it's in a very bad split shape, and i want to stop this. I want to know if it's
29	split shape, and i want to stop this. I want to know if it's too late to heal this.
30	my brother said no one is going to want to be with you when you're older

4.2.5 *Really*

Colligation

As '*really*' is used as an adverb, the search for a colligation pattern focuses on its modifiers. Through concordancing, it is found that the pattern '*really +verb*' has 20 hits. The verb collocates include infinitives, present and past tenses, as well as present and past participles. The remaining ten instances of '*really*' share the structure '*really* + adjective*'.

Collocation and Semantic Preferences

The above colligation analysis drives the search for collocates to focus on verb collocates and adjective collocates of '*really*'. The semantic fields of those collocates are listed in Table 12. According to the Oxford Dictionary (2015), the semantic analysis of those collocates shows that '*really*' is used with two main senses: 1) '*very*; *thoroughly*' (e.g. '*…he doesn't know how offended I really get*' or '*I'm just really tired of everything*'), and 2) '*emphasizing a statement or opinion*' (e.g. '*This life is really hard because…*' or '*I have a really strong self-harm addiction*'). Note that, as usual, sometimes the meaning of a collocate is not complete without the context; whole phrases therefore are included in Table 12 (e.g. '*committing suicide*' or '*draining me and dragging me*'). The node word '*really*' appears mostly to collocate with words in relation to affective senses, illustrating the intense emotions of the authors who wrote those posts.

Table 12 Semantic preference of *'really'*

Semantic field		Collocates	type	token
Affective	**Verb**	*appreciate, care/cared, committing suicide, draining me and dragging me, feel/feeling, get (offended), grieved, like, love, trusted, stand each other*	11	13
	Adj.	*angry, depressed, satisfied, tired*	4	4
Others	**Verb**	*had, helped, mean, need advice (2), tell (2)*	5	7
	Adj.	*bad, hard, nice, popular, strong, young*	6	6

Semantic prosody

Semantic prosody where '*really*' represents it is listed in Table 13, and sample concordance lines in Figure 5. As is often the case, the node word '*really*' is mainly used in negative prosody to refer to the unwillingness of an author to get involved in a situation (e.g. '*I don't really feel comfortable talking to her*') or the author's negative evaluation of a situation (e.g. '*I wouldn't really like to get into because…*'). A suicidal intention may also be observed, as in '*I make plans to die without really committing suicide'.* In contrast, in cases where positive prosody occurs, the act of seeking help, as in '*I really need advice*', or some positive aspect of a situation, as in '*some girls are really nice but I still don't fit in with them*', is also found, although it may not always have a positive impact.

Table 13 Semantic prosody of 'really'

Attitude	Number(s) of line(s)	Total hits
Positive	5,7,11,12,19,20,28	7
Negative	1,2,3,4,6,8,9,10,13,14,15,17,18,21,22,24,25, 27,29	19
Neutral	16,23,26,30	4

Figure 5 Concordance lines of 'really'

N	Concordance
1	we are not close at all (in fact we can't really stand each other), so i don't really
2	can't really stand each other), so i don't really feel comfortable talking to her. I
3	. Then i kind of realize that i never really had any. Every day i just lay here
4	I have often felt depressed, they don't really care about my feelings at all. And
5	me during some of the beatings. I would really appreciate help. I'm just so sad.
6	I don't know already what to do i am not really satisfied if a good thing happened
7	me into traffic. I know have someone I really love and I haven't cut in almost 9
8	old and I think am depressed I don't really tell how I feel because people
9	stress free talkative bubbly chick don't really tell about how I feel not even my
10	just to listen about how I feel. This life is really hard because even when am
11	up. One was still in jail, one became really popular, one became obsessed
12	for 3 years I had a therapist I loved and really trusted but I suffer from anxiety
13	Turkey to Sweden, he beat me. I have a really strong self harm addiction and
14	in grade 8 for something I wouldn't really like to get into cause its irrelevant
15	have the heart to tell him what i was really feeling so i said no. i feel like I've
16	of the twins I was carrying and I haven't really grieved or anything but it's hitting
17	us to be together an hour later. This is really draining me and dragging me into
18	i really get. then he asked me if i really was depressed and after his
19	take it seriously because in the past i'd really been helped and inspired by
20	of them are superficial. some girls are really nice but i still don't fit in with them.
21	but he doesn't know how offended i really get. then he asked me if i really
22	cut herself several times when she feels really bad about herself. She lacks
23	/uplifting words, thanks guys Hello all, I really need advice on what to do with
24	school and my father would get really angry when i wrote the letters
25	succeed. I make plans to die without really committing suicide and many
26	is unhealthy for me. Please help me. I really need advice. I hope I don't sound
27	I don't sound snippy or mean i'm just really tired of everything. My parents are
28	stopped. I wanted to see if my parents really cared for me.....during a track
29	if when I get older I might try it" I didn't really mean what was saying i don't
30	I am 16. I was exposed to porn at a really young age 11-13 I the first time I

5. Discussion and conclusion

From a semantic analysis of the top 50 keywords and the five most frequent lexical keywords, it can be summarized that the themes of the online posts in this 'teen depression community' are often expressed using frequently recurring language expressions associated with intense emotions of helplessness and desperation that indicate their vulnerable mental state in relation to social contexts (such as family, school, friend, relationship), which adolescents are often surrounded by. The feelings expressed are so powerful that, at times, they may escalate to self-reported intention or behaviour of self-harm, or they may even provoke suicidal thoughts. These themes suggest that the issues regarding mental health from this forum are perhaps primarily concerned with adolescents' struggles between themselves and the social circles surrounding them. This can be evidenced by the extremely high frequency of the first personal pronoun '*I*', with 2,014 occurrences, accounting for 7.8% of the whole corpus, when compared with 0.6% of the first personal pronoun '*I*' in the Brown reference corpus of general written English. The percentage of the first personal pronoun '*I*' accounts for only 0.3% of the spoken subcorpus and a fractional 0.007% of the written subcorpus of the 100-million-word British National Corpus (BNC Consortium, 2001). Two common colligation patterns '*I +have/feel/want*' and '*I +lemma DON'T + have/know/want*' may indicate a high degree of self-consciousness in these posts. This is corroborated by data collected in the corpus, where 127 out of 129 posts are found to be self-focused. Such a heightened level of self-focus is often associated with 'depressed pessimism' (Pyszczynski, Holt & Greenberg findings, 1987) or 'negative affect', which includes depression, anxiety or any negative mood (Mor & Winquist, 2002; Mor et al. 2010). Through a semantic analysis of concordance lines, it is not surprising to find that the majority of contexts where the top five lexical words occur tend to have a negative denotation, e.g. '*I don't want to get better', 'I just want a solution and can't think of one'*. There are also some disturbing and devastating statements, such as '*I feel nothing but a burden', 'I feel that my entire world is falling apart', 'I make plans to die without really committing suicide'*. Some signals which indicate attempts to make a cry for help, however, are also found, e.g. '*I really need advice*' or '*I would really appreciate help*', and the online discussion forum seems to provide a platform for such posters to share their conflicting emotions anonymously.

In terms of methodology, corpus approaches appear to be fairly effective for the identification of recurring themes (using keyword analysis) and linguistic features (using Sinclair's framework). Analysis at the single-word level (i.e. keywords and collocates), however, is

found to be insufficient, and often the bigger picture can only be revealed by lexico-grammatical and semantic analysis of concordance lines. For example, 18 out of the 30 sample concordance lines for the node word '*know*' point to the longer expression '*I don't know*', and nine out of 30 concordance lines for another node word, '*want*', also form a longer expression, '*I don't want*'. This is also true for the semantic analysis of collocations. Take the collocations of '*want*' for example. It collocates with a variety of delexical verbs, such as '*get*', '*take*', '*have*', which are semantically incomplete unless the whole phrases in use are provided, e.g. '*get better*', '*take medicine*' and '*have my own life*'. However, note that the semantic prosody of context can only be revealed when more context is disclosed, e.g. '*I don't want to get better*' or '*she does not want to take medicine*'. With regard to the linguistic style of these texts, the frequent use of shortened forms, colloquial expressions such as '*feel like*' or '*don't know what to do*', indicates the informality and chat style of these posts. It should also be noted that all the keywords identified and discussed are unavoidably influenced by the choice of a reference corpus, as mentioned before. As online forums are a unique genre with mixed features from both written and spoken registers (e.g. the look of narrative and greetings at the beginning for the former and some colloquial expressions for the latter), the adoption of a different reference corpus would most likely impact on the keywords extracted. In the experiments described earlier, it was, for example, found that using the spoken subcorpus of the BNC as the reference corpus with the online corpus tool Compleat Lexical Tutor highlighted a different set of keywords, where the top five lexical keywords are '*motivate*', '*weight*', '*freshman*', '*porn*' and '*program*'. Apparently, if those keywords were selected for discussion, this study would be pointing in a completely different direction from the current one. We argue that more research should address the issue of a reference corpus by comparing the results from the use of different sample texts, not just from the perspective of written and spoken registers, but also from, for example, online health communications with adults or face-to-face discussions between adolescents and healthcare professionals.

In terms of the limitations of this study, as the manual examination of concordance lines is very time-consuming, only the top 50 words and five most frequent lexical keywords are analyzed, including 30 concordance lines for each of the chosen keywords. It is also acknowledged that judgement of semantic groupings or prosody is unavoidably subjective, and at times the contextual information available from concordance lines or posts is limited for such decisions to be made. We, however, believe that the reliability of such tasks has been improved by the procedure of double-rating with two researchers, and complemented by a

third rater when needed. In addition, the corpus investigated is relatively small, and more data, covering a wider range of words and concordance lines, would certainly improve the validity of the analysis. It should also be noted that the term 'depression' used here is a blanket expression. As Harvey (2013: 172) has rightly pointed out, when an adolescent group mentions depression or being depressed, '*they will not necessarily be describing a clearly bounded condition, a unitary state of being*'. The analysis presented here, therefore, only refers to a self-reported mental state as opposed to a clinically diagnosed condition.

This paper presents a multidisciplinary study by using a corpus linguistic approach to investigate the topic of teen depression in an online forum for health communications. It is hoped that through such an interdisciplinary approach, researchers and practitioners can benefit from the synergy of linguistics and mental health communications, which otherwise might be deemed to be mutually irrelevant. The findings shed light on the use of language expressions in a unique discourse, on both lexico-grammatical and semantic levels, and it also demonstrates the effectiveness of corpus approaches in researching health communications. Moreover, it helps us to better understand the struggles and suffering of this vulnerable group, which are often otherwise concealed from their parents, teachers and friends.

References

Angouria, J. & T. Sanderson. (2016). 'You'll find lots of help here' unpacking the function of an online Rheumatoid Arthritis (RA) forum. *Language & Communication*, 46, 1–13.

Adolphs, S, Brown, B., Carter, R., Crawford, C. and Sahota, O. (2004). Applying corpus linguistics in a health care context. *Journal of Applied Linguistics*, 1(1), 9–28.

Adolphs S., L. Mullany, C. Smith, K. Harvey & A. Macfarlane. (2010). Health Communication and the Internet: An Analysis of Adolescent Language Use on the Teenage Health Freak Website. Funded by the *Economic and Social Research Council* (ESRC). Project No. RES-000-22-3448. Centre for Research Applied Linguistics, University of Nottingham, U.K. URL: http://www.nottingham.ac.uk/research/groups/cral/projects/thf/thfindex.aspx

Baker, P. (2006). *Using Corpora in Discourse Analysis*. London: Continuum.

British Medical Association. (2006). *Child and Adolescent Mental Health: A Guide for Healthcare Professionals*. BMA: London.

British Medical Association. (2016). Child and Young People's Mental Health. BMA: London.

Bondi M. & M. Scott (2010). *Keyness in Texts: Corpus Linguistic Investigations.* Amsterdam: John Benjamins.

Francis, W.N. & H. Kucera. (1979). *Brown Corpus Manual*. Providence: Brown University. URL: http://clu.uni.no/icame/brown/bcm.html (accessed 17 January 2017).

Grinyer, A. (2007). The ethics of Internet usage in health and personal narratives research. Social Research Update, 49. URL: http://sru.soc.surrey.ac.uk/SRU49.html (accessed 7 February 2017).

Harrington, R. (2001). Depression, suicide and deliberate self-harm in adolescence. *British Medical Bulletin*, 57 (1), 47–60.

Harvey, K., Brown, B., Crawford, P., Macfarlane, A., and McPherson, A. (2007). 'Am I normal?' Teenagers, sexual health and the internet. *Social Science & Medicine*, 65, 771–781.

Harvey, K., Churchill, D., Crawford, P., Brown, B., Mullany, L., Macfarlane, A. and McPherson, A. (2008). Health communication and adolescents: what do their emails tell us? *Family Practice*, 25(4), 304–311.

Harvey, K. & Brown, B. (2012). Health communication and psychological distress: Exploring the language of self-harm. *Canadian modern language review*, 68(3), 316–340.

Harvey, K. (2013). *Adolescent Health Communication: A Corpus and Discourse Approach*. London: Bloomsbury.

Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Hunston, S. & P. Thompson. (2001). *Evaluation in Text: Authorial Stance and the Construction of Discourse*. Oxford: Oxford University Press, 74-101.

McEnery, T. and Hardie, A (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.

Mor, N., Leah D. Doane, Emma K. Adam, Susan Mineka, Richard E. Zinbarg, James W. Griffith, Michelle G. Craske, Allison Waters, & Maria Nazarian. (2010). Within-person variations in self-focused attention and negative affect in depression and anxiety: A diary study. *Cognition and Emotion*. 24(1), 48–62.

Mor, N. & Winquist, J. (2002). Self-focused attention and negative affect: A meta-analysis. *Psychological Bulletin*, 128(4), 638–662.

Pyszczynski, T, Holt, K. & Greenberg, J. (1987). Depression, self-focused attention, and expectancies for positive and negative future life events for self and others. *Journal of Personality and Social Psychology*, 52(5), 994–1001.

Ramirez-Esparza N., C.K. Chung, E. Kacewicz & J.W. Pennebaker. (2008). The Psychology of Word Use in Depression Forums in English and in Spanish: Texting two text analytic approaches. *Proceedings of the International Conference on Weblogs and Social Media ICWSM*.

Sinclair, J. M. (2004). *Trust the Text: language, corpus and discourse*. London: Routledge.

Sinclair, J. M. (1996). The Search for the Units of Meaning. *Textus*, 9(1), 75–106.

Scott, M. (2008). *WordSmith Tools version 5*. Liverpool: Lexical Analysis Software.

Substance Abuse and Mental Health Services Administration. (2014). *Results from the 2013 National Survey on Drug Use and Health: Mental Health Findings*. *NSDUH Series H-49*. Rockville, MD: Substance Abuse and Mental Health Services Administration. http://www.samhsa.gov/data/sites/default/files/NSDUHmhfr2013/NSDUHmhfr2013.pdf

Van der Zanden, R., Curie, K., Van Londen, M., Kramer, J., Steen, G. and Cuijpers, P. (2014). Web-based depression treatment: Associations of clients′ word use with adherence and outcome. *Journal of Affective Disorders*, 160, 10–13.

Wilkinson, P., Kelvin, R., Roberts, C., Dubicka, B., & Goodyer, I. (2011). Clinical and psychosocial predictors of suicide attempts and nonsuicidal self-injury in the Adolescent Depression Antidepressants and Psychotherapy Trial (ADAPT). *American Journal of Psychiatry*, 168(5), 495–501.

Wilson M.L., S. Ali & M.F. Valstar. (2014). Finding Information about Mental Health in Microblogging Platforms: A Case Study of Depression. *Proceedings of the 5th Information Interaction in Context Symposium*, 8–17.

Web references

Davies, Mark. (2015). The Corpus of Contemporary American English (COCA): 520 million words, 1990–2015. Available at http://corpus.byu.edu/coca/ (accessed December 2015)

BNC Consortium. (2001). British National Corpus. Available at http://corpus.byu.edu/bnc/ (accessed in November 2016)

Cobb, Tom. (Unknown). Compleat Lexical Tutor. Available at http://www.lextutor.ca/key/ (accessed December 2016)

MedHelp. (2015). 'Teen Depression Community'. Available at http://www.medhelp.org/forums/Teen-Depression/show/185 (accessed November 2015)

Oxford Online Dictionary. (2015). Definitions of '*have*', '*feel*', '*know*', '*want*' and '*really*'. Available at http://www.oxforddictionaries.com/ (accessed December 2015)

Appendix I Frequency List

No.	Word	Freq.	%	Texts	%
1	I	2014	7.792609692	1	100
2	AND	979	3.787966728	1	100
3	TO	825	3.192106724	1	100
4	MY	706	2.731669664	1	100
5	A	494	1.911394835	1	100
6	THE	453	1.752756834	1	100
7	ME	428	1.656026363	1	100
8	OF	313	1.211066008	1	100
9	BUT	311	1.203327537	1	100
10	HAVE	306	1.183981419	1	100
11	THAT	277	1.071774006	1	100
12	IN	276	1.06790483	1	100
13	IT	276	1.06790483	1	100
14	I'M	234	0.905397534	1	100
15	#	220	0.851228476	1	100
16	WAS	212	0.820274711	1	100
17	FOR	190	0.735151887	1	100
18	IS	189	0.731282651	1	100
19	DON'T	188	0.727413416	1	100
20	THIS	181	0.700328887	1	100
21	WITH	181	0.700328887	1	100
22	JUST	180	0.696459651	1	100
23	SO	173	0.669375122	1	100
24	LIKE	158	0.611336827	1	100
25	DO	154	0.595859945	1	100
26	SHE	147	0.568775415	1	100
27	AM	146	0.56490618	1	100
28	FEEL	141	0.545560062	1	100
29	NOT	140	0.541690826	1	100
30	ABOUT	139	0.537821651	1	100
31	KNOW	136	0.526213944	1	100
32	OR	133	0.514606297	1	100
33	BE	126	0.487521768	1	100
34	ALL	125	0.483652532	1	100
35	BECAUSE	124	0.479783326	1	100
36	WHEN	122	0.472044885	1	100
37	WANT	121	0.46817565	1	100
38	AT	117	0.452698767	1	100
39	IF	116	0.448829561	1	100
40	HE	115	0.444960326	1	100
41	HER	115	0.444960326	1	100
42	WHAT	115	0.444960326	1	100
43	BEEN	113	0.437221915	1	100
44	ON	113	0.437221915	1	100
45	REALLY	112	0.433352679	1	100
46	AS	99	0.383052826	1	100
47	MYSELF	98	0.37918359	1	100
48	THEY	97	0.375314385	1	100
49	T	96	0.371445149	1	100
50	TIME	96	0.371445149	1	100

Appendix II Keyword List

No.	Key word	Freq.	%	RC. Freq.	RC. %	Keyness	P
1	I	2014	7.792609692	5931	0.573515594	6392.001465	3.11263E-21
2	MY	706	2.731669664	1320	0.127641305	2706.873779	4.18943E-20
3	I'M	234	0.905397534	0		1740.178833	1.61071E-19
4	DON'T	188	0.727413416	1		1385.334839	3.24191E-19
5	ME	428	1.656026363	1186	0.114683777	1376.587036	3.30569E-19
6	I'VE	88	0.340491384	0		653.9391479	3.35771E-18
7	AM	146	0.56490618	237	0.022917416	587.7803955	4.71061E-18
8	FEEL	141	0.545560062	216	0.020886758	579.6463013	4.92443E-18
9	MYSELF	98	0.37918359	129	0.012474037	424.1905212	1.34667E-17
10	JUST	180	0.696459651	872	0.08432062	418.1471863	1.41123E-17
11	CAN'T	55	0.212807119	0		408.6433105	1.52137E-17
12	MOM	58	0.224414781	4		401.4736633	1.61217E-17
13	T	96	0.371445149	146	0.014117902	395.5212402	1.69312E-17
14	WANT	121	0.46817565	329	0.031813629	391.5517273	1.75014E-17
15	REALLY	112	0.433352679	275	0.026591938	380.255127	1.92709E-17
16	DEPRESSION	62	0.239891663	24		360.0144348	2.30876E-17
17	FRIENDS	91	0.352099061	163	0.015761767	352.8357849	2.46825E-17
18	DIDN'T	46	0.177984133	0		341.7587585	2.74447E-17
19	IT'S	46	0.177984133	0		341.7587585	2.74447E-17
20	KNOW	136	0.526213944	690	0.066721588	305.8409119	3.98241E-17
21	DON	52	0.201199457	23		295.0198975	4.49851E-17
22	HAVE	306	1.183981419	3980	0.384857863	266.0133362	6.40547E-17
23	DEPRESSED	43	0.166376472	11		265.4140625	6.45536E-17
24	BUT	311	1.203327537	4382	0.423730463	238.874527	9.30307E-17
25	PARENTS	59	0.228284001	98		235.3410797	9.80151E-17
26	LIKE	158	0.611336827	1339	0.129478559	231.0153503	1.04614E-16
27	HELP	83	0.321145296	326	0.031523533	220.1791229	1.23964E-16
28	BECAUSE	124	0.479783326	883	0.085384294	213.5471649	1.38215E-16
29	ANYMORE	30	0.116076611	4		198.4349518	1.79932E-16
30	THINK	87	0.336622179	434	0.041966915	197.8430939	1.81889E-16
31	SCHOOL	92	0.355968267	518	0.050089542	191.7305298	2.03865E-16
32	PORN	25	0.096730508	0		185.7186279	2.29073E-16
33	BOYFRIEND	24	0.092861287	0		178.2889709	2.66312E-16
34	SO	173	0.669375122	2033	0.196586952	173.0852966	2.97335E-16
35	DAD	31	0.119945832	15		172.9514771	2.98195E-16
36	HAVEN'T	23	0.088992067	0		170.8593903	3.1208E-16
37	HATE	36	0.139291927	42		161.8514862	3.82801E-16
38	LOT	50	0.193461016	146	0.014117902	156.0735321	4.3977E-16
39	HAPPY	44	0.170245692	103		152.5487213	4.80175E-16
40	DOESN'T	20	0.077384405	0		148.5711212	5.31942E-16
41	DO	154	0.595859945	1863	0.180148289	148.0583649	5.39149E-16
42	MUM	19	0.073515184	1		133.2505646	8.18259E-16
43	SOMEONE	39	0.150899589	100		129.6795349	9.13341E-16
44	GET	86	0.332752943	752	0.072716862	121.6438217	1.1883E-15
45	ABOUT	139	0.537821651	1817	0.175700188	119.5367126	1.2783E-15
46	SHE'S	16	0.061907526	0		118.8544769	1.30938E-15
47	WASN'T	16	0.061907526	0		118.8544769	1.30938E-15
48	STARTED	45	0.174114913	194	0.018759403	112.6685028	1.64319E-15
49	GOING	59	0.228284001	404	0.039065976	105.0330124	2.23208E-15
50	SUICIDE	21	0.081253625	17		104.5817184	2.27517E-15