

# Lexical and Phrasal Analysis of Online Discourse of Type 2 Diabetes Patients based on Text-Mining

Moonl-Hyon Hwang\*, Jungsik Park\*\*

Dept. of Applied Physiology and Kinesiology, University of Florida, Gainesville, FL, USA\*

Dept. of English Language and Literature, Ajou University, Suwon, Korea\*\*

## 텍스트마이닝 기법을 이용한 제 2형 당뇨병환자 온라인 담론의 어휘 및 구문구조 분석

황문현\*, 박정식\*\*

플로리다대학교 응용생리학 및 운동과학과\*, 아주대학교 인문대학 영어영문학과\*\*

**Abstract** This paper has identified five major categories of the T2D patients' concerns based on an online forum where the patients voluntarily verbalized their naturally occurring emotional reactions and concerns related to T2D. We have emphasized the fact that the lexical and phrasal analysis brought to the forefront the prevailing negative reactions and desires for clear information, professional advice, and emotional support. This study used lexical and phrasal analysis based on text-mining tools to estimate the potential of using a large sample of patient conversation of a specific disease posted on the internet for clinical features and patients' emotions. As a result, the study showed that quantitative analysis based on text-mining is a viable method of generalizing the psychological concerns and features of T2D patients.

**Key Words** : Type 2 Diabetes, Psychological Condition, Text-Mining, Lexical and Phrasal Analysis, Online Discourse.

**요약** 본 연구는 질병과 관련한 온라인 포럼에서 추출한 언어 데이터를 통해 제 2형 당뇨병 환자의 질병에 대한 담론을 양적으로 분석하였다. 또한 환자 언어행위의 양적분석을 통해 환자들의 주요 관심사와 심리적 특징의 일반화가 가능한지에 대해 실증적으로 검증하였다. 분석방법으로는 기존의 인터뷰에 기반한 정성적 연구방법론과 달리 환자들의 담론 표본 전체를 파싱 (parsing)과 POS 태깅을 통해 언어학적으로 형태소 분류를 하였다. 주요 어휘빈도 추출과 N-gram을 통한 최빈도 구문구조 분석을 병행하여, 질병과 관련한 이슈의 주요 범주와 심리상태에 관한 언어적인 특징을 살펴보았다. 연구 결과 환자들의 자발적 대화는 주로 다이어트, 운동, 증상, 약물치료, 심리상태의 5가지 범주로 나타나고 있음을 확인하였고, 최빈도 구문구조 분석을 통해 질병치료와 식생활습관 개선 전반에 대한 부정적인 견해가 두드러진 것을 확인하였다. 결과적으로 의료진의 정확한 정보 전달과 전문가의 조언, 정서적 지원 등이 당뇨병환자에 대한 심리적 상태에 중요한 만큼 심리치료 서비스가 개선이 필요할 것으로 보인다. 이러한 결과는 기존의 의료제도 안에서의 환자의 관심사와 심리적 특징이 온라인 상에서도 적절하게 투영되고 있음을 시사한다.

**주제어** : 제 2형 당뇨, 심리상태, 텍스트마이닝, 어휘 및 구문분석, 온라인 담론분석

Received 16 April 2014, Revised 8 May 2014

Accepted 20 June 2014

Corresponding Author: Jungsik Park

(Dept. of English Language and Literature, Ajou University)

Email: jspark@ajou.ac.kr

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. INTRODUCTION

During the last couple of decades, Type 2 diabetes has been discussed regarding its etiology, pathogenesis, complications, social impacts and costs. Type 2 diabetes is a chronic and widespread disease that influences more than 300 million people world-wide and creates social and economic burdens for patients, families and national health care providers—with its cost of diagnosed diabetes in 2012 alone reaching \$245 billion [1]. Etiologically, diabetes has two broad but distinctive forms. Type 1 diabetes is an insulin-dependent diabetes mellitus and results from the functional loss of beta cells in a human pancreas. Insulin secretion is substantially limited, thus glucose molecules in the human blood stream are not likely to get into the cells. Type 2 diabetes is a non-insulin dependent diabetes mellitus and results from the increased resistance of insulin receptors when they combine with insulin despite normal or slightly less secretions of insulin. Compared to Type 1 diabetes, Type 2 diabetes possesses over 90% of total diabetic patients and is also known to be closely related to over-nutrition, obesity, and a sedentary lifestyle. Its chronic complications are major cardiovascular diseases, loss of visual acuity, and limb amputations, and there is no conclusive cure for this chronic disorder known so far [2], [3].

Despite medical and clinical studies with a wealth of accumulated data concerning the etiological and patho-generic development of type 2 diabetes (hereafter referred to as T2D), little research has been done on the T2D patients' online discourse on their physical conditions and psychological needs. Whereas other qualitative and ethnographic studies based on surveys and interviews have provided much insight into the experience the T2D patients undergo, this study attempts to identify and analyze the naturally occurring conversations where the patients voluntarily

express their own concerns and where they do not need to follow structured questionnaires, and by so doing, this study avoids the homogeneity issues in their answers and explanations.

With the advent of Internet forums and other SNS platforms, it has become relevant to analyze the discourse of the T2D patients: contemporary patients tend to find forums to share their experience with others who have the same disease and/or disorders, and numerous accounts of experience and medication specific to T2D are available. Traditionally, data sets such as testimonial accounts derived from patients had been the usual sources of the qualitative studies on illness experience. Though the traditional sources have been analyzed as evidence of common illness experience, they lack in quantity to generalize the clinical features of a disease and patients' discourse related to them, and the amount of information archived in the disease-specific platforms is not sufficiently manageable. Thus a different method of approach is required. Only recently, discourse analysts, health specialists, researchers and scholars have noticed the possibilities for mining those unstructured language data and extracting, patterning, and analyzing the ever-growing text data [4].

The objective of this study is to explore online discourse on type 2 diabetes and to identify patients' concerns and psychological needs in an online forum organized by T2D patients. Apart from medical treatment, T2D patients undergo a different experience and situation with physical and psychological characteristics with a marked difference from non-patients. Therefore, this study aims to represent those characteristics of T2D patients and identify some of the shared areas of their concerns and their attitudes towards their experience of T2D through lexical and phrasal analysis and language patterning.

## 2. METHOD

As known under the umbrella term ‘text-mining’ – that is, data-mining of unstructured language data – scholars from various disciplines, including health related professionals, have conducted identifying lexical, phrasal, or sentential patterns and thematic tendencies. They have suggested that massive amounts of data are more aptly studied by using text-mining methods than the qualitative approaches based on surveys and interviews [5], [6], [7]. The text-mining tasks such as text classification, clustering, word-frequency, and content analysis have been proven efficient at detecting the recurring patterns of clinical features and patients’ concerns [8], [9].

The present study observes the conversation in the form of ‘messages’ and ‘replies’ related to T2D carriers posted on the chosen forum – ‘the Global Diabetes Community’ where over 93,500 patients with diabetes and their family members share their experience, concerns, and information to better cope with the disease. By text-mining the discourse of the patients with T2D, our aim was first, to identify the recurring themes of T2D patients’ concerns, and second, to interpret the patients’ psychological conditions and welfare.

To study the current concerns of T2D patients and the characteristics of discourse on their diet and exercise, 400 patients’ conversations were extracted as samples from the T2D group site titled ‘the Global Diabetes Community’ at < <http://www.diabetes.co.uk/forum/category/type-2-diabetes.25/>> that provides a platform where various types of patients can share the experience of the disease. First, the unstructured data of conversation about T2D were collected through a crawling process. The collected words were then tokenized and parts of speech (POS) were tagged by Stanford Log-Linear Part of Speech Tagger. AntConc 3.2.4w was also used to arrange the words into the

categories of verbs, nouns, and adjectives and to perform N-gram analysis<sup>1)</sup>. Along with the retrieval and classification of the information, a lexical and phrasal analysis was conducted based on N-gram frequency ranking. In addition, Phrase Net visualization based on the web-based text-mining tool was used to check the phrase patterns and any two words that were connected with a conjunction<sup>2)</sup>.

## 3. RESULTS

This study identified T2D related words and phrases of high frequency by conducting POS tagging, and by so doing, aimed to generalize distinctive features of conversation in the T2D patients’ discourse. After data cleansing, which eliminated common words such as pronouns and personal proper nouns, we obtained 56,162 T2D related nouns. Among them, we chose 200 nouns with the highest frequency and identified five thematic categories manually: (1) Diet, (2) Exercise, (3) Symptoms, (4) Medication, and (5) Psychological Condition. For psychological condition, we then conducted additional analysis using N-gram analysis to identify the patients recurring patterns of phrase expression.

### 3.1 Diet

According to the analysis of noun frequency from the conversation of T2D patients, food consumption

1) AntConc is a Concordancer that provides such functions as concordance, collocation, word frequency and keyword listing. More information is found at <<http://www.antlab.sci.waseda.ac.jp>>.

2) ManyEyes is created by Fernanda Viégas and Martin Wattenberg. It offers a graphic representation of patterns and distinctions of a given dataset. Constructed on a public web site, it allows users to upload a set of data, visualize them, and gain new insights into patterns occurring in the data. Information about Phrase Net and other functions of ManyEyes are available at <<http://www-958.ibm.com/software/analytics/manyeyes/>>.



(184) is a representative cholesterol-lowering drug. Thus, the word, 'statin' indicates that it is also a frequently prescribed medication along with diabetic medications for T2D patients because they are likely to have higher cholesterol levels than other healthy people.

### 3.5 Psychological Condition

In the category of 'Psychological Condition,' 'shock'(125) was ranked first among the nouns related to the patients' sentiments, followed by 'worry' (145) and 'stress'(173). 'Shock' is often expressed in relation to the first diagnosis of T2D and to the reading taken after the medical treatment and exercise. For example, the patients tended to describe their experience with T2D diabetes, first when they are diagnosed as having Type 2 Diabetes (as in, "Yes it's a shock to find out you have diabetes," and, "For anyone who knows me that would be a hell of a shock."), and second after the reading the blood level (as in, "I thought that the exercise was a panacea. Then I tested one day last spring, and I got a shock," and, "2 hrs later and [after] a 30 min walk reading was 8.1 :shock:.")

Other key words related to the patients' psychological condition were 'stress' and 'worry' (as in, "My average readings are between 4.9 and 6 but last few days getting 7.2 why would that be I've not changed my diet. Only thing is stress levels up," and, "Older drugs such as gliclazide can help and they don't have the pancreatic cancer worries.").

Other keywords with high frequency were the terms associated with getting help and support such as 'advice'(29), 'forum' (36), 'information' (45) and 'help' (71), with variations such as 'question'(49), 'idea'(55), and 'answer'(180) (Appendix 1).

The five categories of Diet, Exercise, Symptoms, Medication, and Psychological Condition were approximately consistent in coverage when the 400 forum samples were presented using the Phrase Net

Visualization.

### 3.6 N-gram Analysis and Recurring Patterns

Though the patients' psychological condition was moderately reflected in the high frequency key nouns, we performed 3-gram analysis to see the phrase patterns. First, based on 300,496 phrase tokens, we extracted the most frequent 200 phrase types (that is, 8,286 phrase tokens). As the results show in (Appendix 2), 2,590 phrase tokens belong to the described psychological condition, comprising 31 percent of the all 8,286 phrase tokens. The percentage is estimated to be significant when considering the 3-gram phrases capture any three words that are consecutively arranged, including any daily expressions. Mostly, the feelings have to do with a possibility, security, capability, necessity, and a wish, which the patients are expected to perform or participate in, but feel unable to engage in active performance and participation. Based on the results, dominant psychological conditions are categorized into feeling of (a) Negation (for instance, 'I don't' or 'I didn't'), (b) Capability ('not able to' or 'I can't'), (c) Certainty ('if you have' or 'seem to be'), (d) Necessity ('I need to' or 'you need to'), (e) Duty ('I have to' or 'I had to'), and (f) Wish ('would like to' or 'I want to').

## 4. DISCUSSION

There are two major points noteworthy in the results of this analysis. First, we identified major psychological concerns and needs of the T2D patients; and second, online archives of the patients with the same disease can be a useful resource to analyze the patients' discourse.

Based on 200 nouns with the highest frequency from a total of 56,162 T2D related nouns, we identified five thematic categories of Diet, Exercise, Symptom,

Medication, and Psychological Condition, which are the major areas of the T2D patients' conversation. The category of 'Diet' is identified to be the most dominant concern of the T2B patients. The patients were most concerned with 'carbohydrates' often exemplified as 'pasta', 'rice', 'potato', and 'bread', which are known to be high in carbohydrates (Appendix 1, Figure 1). In the conversations, 'diet' was often associated with 'exercise,' most frequently in the form of 'diet and exercise' with some overlapping issues such as 'weight' and 'lifestyle' as shown in the Phrase Net Visualization (Figure 1).

T2D patients' concerns in the category of 'Symptom' are related to such body parts as 'hand,' 'foot,' 'eye,' 'knee,' and 'kidney.' The T2D patients also go through psychological needs for 'support,' 'advice,' and 'information': they often felt confusion with the information they have and desired to have more information from patients with the same experience, and needed clear and reliable information as pointed out by the qualitative studies done by Kinmonth and Goyder [10], [11]. The words such as 'anyone'(54) and 'someone'(62) also occurred frequently when T2D patients are in need of help or information; they often tried to find someone with the same experience or relevant information as in, "Can anyone help me break my vicious cycle?" "Has anyone got any diabetic friendly snacks?" "I hope someone comes along who has been where you are now to tell you how they coped," and, "Can someone out there who has had a diagnosis of neuropathy enlighten us to the first signs?" Their desire for getting information and finding a person with the same experience or relevant knowledge has to do with the lack of immediate solutions and answers outside the clinic or hospital.

The results imply that detailed and clear information provided in a timely manner is a crucial aspect of patients' need regarding health care service providers, and that digital forums, blogs, and other web-based

platforms are the sites where the patients circulate information, share experience, and ask for advice from T2D patients. Therefore, it can be suggested that the websites can be a primary and effective place where T2D patients can be educated about diabetes care, informed about different treatments and supported emotionally.

Regarding the patient's psychology and emotional needs, 'shock' appears to be the word that best describes the state of the patients. The finding is consistent with previous qualitative studies that found that 'shock' represented the primary and prevailing emotion when people were diagnosed with T2D [12], [13] and is also the sign for 'a degree or intensity of emotion with which an individual is unable to cope' [12].

The patients' general attitudes and emotions towards the experience of T2D were also identified with 3-gram analysis and classified into five categories of Negation, Capability, Necessity, Duty and Wish. As shown in the analysis, 'Negation' ('I don't,' 'I am not,' or 'I didn't') is the most frequent sentiment in the category of phrase analysis. The patients often expressed negative attitudes towards their condition related to T2D as in, "I am not satisfied with that anymore," and, "I don't find it easy to afford strips as I'm a pensioner," or they were in confusion or experienced a lack of knowledge as in, "I'm spiralling downwards and don't know how to stop," and, "I am not sure about a cure, I have read about cures, but doesn't it have something to do with the pancreas?"

The findings are also consistent with previous qualitative studies that showed that patients tend to have more negative attitudes and less ability to control [14], [15], [16]. For instance, Egede et al. argued that the T2B patients tend to show "lower self-care ability, lower self-care adherence, negative attitudes, and less perceived ability to control" [4].

Another dominant feeling that the patients undergo

is 'Capability' issue as in, "I can't exercise because of chronic arthritic pain, so any ideas would be really helpful," or, "Sadly diabetes can't be cured." As indicated in the (Appendix 3), besides 'Negation' and 'Capability,' expressions related to 'Certainty,' 'Necessity,' 'Duty' and 'Wish' comprise a large portion of the sentiment.

Though not included in the categories of 'Psychological Condition', 3-gram results also show that patients typically suffer from T2D for a considerable time. The duration phrases (560 phrase tokens, ranked second in the 3 gram analysis) such as 'I have been' and 'I have had' were distinctive in the results, and it indicates that T2D patients may have been in the status described for a period time and their symptoms may be described as chronic. The present perfect tense was often used when referring to the period of time from being diagnosed as having T2D as in, "I have been type 2 for over five years and at the moment I'm running at around 6.0," and, "I have been diabetic for 13 years and this year has been a hard lesson in relearning what I can and cannot eat," or to the patient's physical state such as obesity as in, "i have been overweight most of my life but being told i was diabetic sort of shocked me into doing something about it," and medical treatment as in, "I have been on 2 x metformin twice a day for 3 months." The chronic condition is reinforced when accompanied by the high frequency of the noun 'year' (11) suggesting that T2D is a long-lasting disease.

The nouns of highest frequency in the category of 'Psychological Condition' and additional 3-gram phrases indicate that the overall sentiment of the patients revolves around the sense of inability and instability. Despite the recognition of the need for life style change, this feeling of inability with the sense of depression and loss of control also permeated the online discourse of the T2D patients as previously pointed out by the qualitative studies of Gabbay et al. and Peyrot

et al. [17], [18].

Since the psychological conditions bordering on depression and compulsion such as overeating, impulsive food-seeking, and incessant checking blood glucose, it is of paramount importance that health professionals understand the ways to communicate the control issues with T2D patients because these feelings and behaviors can increase the risk of intensifying diabetes. The findings also support the previous studies that depression and compulsiveness in the patients' behavior can increase the risk of worsening diabetes and obesity [19], [20], [21], [22].

This study also demonstrates that it is possible to understand the interests and concerns of T2D patients' online discourse through text-mining technology. The results of the online discourse analysis showed the possibility of recognizing key words from a large database of texts. By understanding the major issues and concerns of the patients, medical professionals can increase their effectiveness when educating them. The preference for and dominance of online discourse are expected to continue or accelerate since T2D patients want to escape from pain and the feeling of inability, search for advice online, converse about medical treatment, and get support from others.

Despite the importance of quantitative online discourse analysis, there are some limitations to this study. First of all, though quantitative analysis based on keyword extraction and N-gram can help researchers deal with a large amount of information about a group of patients' online discourse, it is not definitive, and it is uncertain whether it may lead to clinical and diagnostic confirmation. Additionally, the sets of information on age, gender, occupation of patients were not specified, simplifying the complexity of individual disease patterns and processes. Further, the T2D discourse archived in the website that is intended to appeal to the people in the United Kingdom does not represent a guarantee of demographic coverage.

## 5. CONCLUSION

This paper has identified five major categories of the T2D patients' concerns based on an online forum where the patients voluntarily verbalized their naturally occurring emotional reactions and concerns related to T2D. We have emphasized the fact that the lexical and phrasal analysis brought to the forefront the prevailing negative reactions and desires for clear information, professional advice, and emotional support. This study used lexical and phrasal analysis based on text-mining tools to estimate the potential of using a large sample of patient conversation of a specific disease posted on the internet for clinical features and patients' emotions. As a result, the study showed that when compared to the previous qualitative research, quantitative analysis is a viable method of generalizing the psychological concerns and features from a large dataset of language.

Quantitative analysis of online discourse is considered to be more adequate in capturing the ever-flowing conversation of the T2D patients as there are a growing number of T2D patients who want to share their experience and information with others regardless of time and location. Though there are some limits in details and in the use of text-mining for diagnostic purposes, quantitative analysis is believed to complement qualitative and ethnographic methods to better understand the general patients' experience of T2D.

## REFERENCES

- [1] American Diabetes Association, Economic Costs of Diabetes in the U.S. in 2012, *Diabetes Care*, vol. 36, no. 4, pp. 1033-1046, 2013.
- [2] R. Klein, SE. Moss, BE. Klein, MD. Davis, and DL. DeMets, The Wisconsin Epidemiologic Study of Diabetic Retinopathy. XI. The Incidence of Macular Edema, *Ophthalmology*, vol. 96, pp. 1501 - 1510, 1989.
- [3] UK Prospective Diabetes Study Group, Tight Blood Pressure Control and Risk of Macrovascular and Microvascular Complications in Type 2 Diabetes, *British Medical Journal*, vol. 318, no.7175, p. 29, 1999.
- [4] A. Faulkner, *When the News is Bad: a Guide for Health Professionals on Breaking Bad News*, Stanley Thornes, Cheltenham, UK, 1998.
- [5] L. Fernandez-Luque, R. Karlsen, J. Bonander, Review of Extracting Information from the Social Web for Health Personalization, *Journal of Medical Internet Research*, vol. 13, no. 1, e15, 2011.
- [6] F. Greaves, D. Ramirez-Cano, C. Millett, A. Darzi, and L. Harnessing the Cloud of Patient Experience: Using Social Media to Detect Poor Quality Healthcare, *BMJ Quality & Safety*, vol. 22, no. 3, pp. 251-255, 2013.
- [7] U. Raja, T. Mitchell, T. Day, and JM. Hardin, Text-mining in Healthcare. Applications and Opportunities, *Journal of Healthcare Information Management*, vol. 22, no. 3, pp. 52-56, 2008.
- [8] Y. Lu, P. Zhang, J. Liu, J. Li, and S. Deng, Health-Related Hot Topic Detection in Online Communities Using Text Clustering, *PLoS ONE*, vol. 8, no. 2, e56221, 2013.
- [9] PR. Vasconcellos-Silva, D. Carvalho, C. and C. Lucena, Word Frequency and Content Analysis Approach to Identify Demand Patterns in a Virtual Community of Carriers of Hepatitis C, *Interactive Journal of Medical Research*, vol. 2, no. 2, e12, 2013.
- [10] EC. Goyder, PG. McNally, M. Drucquer, N. Spiers, and JL. Botha, Shifting of Care for Diabetes from Secondary to Primary Care 1990 - 1995: Review of General Practices, *Br Med J.*, vol. 316, pp. 1505 - 6, 1998.
- [11] AL. Kinmonth, E. Murphy, and T. Marteau, Diabetes and Its Care—What Do Patients Expect?, *J R Coll Gen Pract.*, vol. 39, pp. 324 - 7, 1989.
- [12] I. Jedlicka-Köhler, M. Götz, and I. Eichler, Parents'



- Recollection of the Initial Communication of the Diagnosis of Cystic Fibrosis, *Pediatrics*, vol. 97, pp. 204 - 209, 1996.
- [13] E. Peel, O Parry, M Douglas, and J Lawton, Diagnosis of type 2 diabetes: a qualitative analysis of patients' emotional reactions and views about information provision, *Patient Education and Counseling*, vol. 53, pp. 269 - 275, 2004.
- [14] L. Egede and C. Ellis, The Effects of Depression on Diabetes Knowledge, Diabetes Self-Management, and Perceived Control in Indigent Patients with Type 2 Diabetes, *Diabetes Technology & Therapeutics*, vol. 10, no. 3, pp. 213-219, 2008.
- [15] J. Kiecolt-Glaser, L. McGuire, T. Robles, and R. Glaser, Emotions, Morbidity, and Mortality: New Perspectives from Psychoneuroimmunology, *Annual Review of Psychology*, vol. 53, pp. 83-10, 2002..
- [16] LJ. Lange and JD. Piette, Perceived Health Status and Perceived Diabetes Control: Psychological Indicators and Accuracy, *J Psychosom Res.*, vol. 58, pp. 129 - 137, 2005.
- [17] R. Gabbay, I. Lendel, T. Saleem, G. Shaeffer, A. Adelman, D. Mauger, M. Collins, and R. Polomano, Nurse case management improves blood pressure, emotional distress and diabetes complication screening, *Diabetes Research and Clinical Practice*, vol. 71, pp. 28 - 35, 2006.
- [18] M. Peyrot and R. Rubin, Behavioral and Psychosocial Interventions in Diabetes, *Diabetes Care*, vol. 30, no. 10, pp. 2433-2440, 2007.
- [19] JV. Butler, JE. Whittington, AJ. Holland, H. Boer, D. Clark, and T. Webb, Prevalence of, and Risk Factors for, Physical Ill-health in People with Prader - Willi Syndrome: a Population-based Study, *Dev Med Child Neurol.*, vol. 44, no. 4, pp. 248 - 255, 2002.
- [20] AJ. Holland, JV. Butler, JE. Whittington, H. Boer, D. Clark, and T. Webb. Behavioural Phenotypes Associated with Specific Genetic Disorders: Evidence from a Population-based Study of People with Prader - Willi Syndrome, *Psychol Med.*, vol. 33, no. 1, pp. 141 - 153, 2003.
- [21] LE. Levitt Katz, S. Swami, M. Abraham, KM. Murphy, AF. Jawad, H. McKnight-Menci, and R. Berkowitz, Neuropsychiatric Disorders at the Presentation of Type 2 Diabetes Mellitus in Children, *Pediatric Diabetes*, vol. 6, pp. 84 - 89, 2005.
- [22] R. Rubin and M. Peyrot, Psychological Issues and Treatments for People with Diabetes, *Clin Psychol.*, vol. 57, pp. 457 - 478, 2001.

**황 문 현(Hwang, Moon Hyon)**



- 1999년 2월 : 고려대학교 체육교육과 (체육교육학사)
- 2001년 2월 : 고려대학교 대학원 체육학과(체육학석사)
- 2014년 8월 : 미국 University of Florida, Dept. of Applied Physiology and Kinesiology(이학박사 졸업예정)

· 관심분야 : 운동처방, 스포츠의학, 심혈관질환, 만성대사성질환

· E-Mail : mhhwang@ufl.edu

**박 정 식(Park, Jungsik)**



- 1994년 2월 : 서강대학교 영문학과 (문학사)
- 1999년 2월 : 미국 City University of New York at Queens 영문학과 (문학석사)
- 2006년 5월 : 미국 Texas A&M University 영문학과 (문학박사)
- 2008년 3월 ~ 현재 : 아주대학교

영문학과 부교수

· 관심분야 : 스토리텔링, 텍스트마이닝, 미국소설

· E-Mail : jspark@ajou.ac.kr

Appendix 1. Nouns with Highest Frequency

Rank	Frequency	Noun	Rank	Frequency	Noun	Rank	Frequency	Noun
1	1624	carb	36	238	forum	71	123	help
2	995	blood	37	237	morning	72	119	water
3	955	day	38	233	medication	73	116	everyone
4	874	test	39	215	bread	74	113	health
5	843	level	40	215	hypo	75	112	mg
6	837	type 2	41	213	tablet	76	111	dose
7	800	diet	42	200	anything	77	110	pasta
8	793	sugar	43	200	rrb-	78	107	moment
9	727	time	44	198	post	79	103	g
10	653	diabetes	45	197	information	80	102	walk
11	630	year	46	186	life	81	102	cholesterol
12	596	food	47	178	breakfast	82	102	hospital
13	573	doctor	48	176	vegetable	83	102	nothing
14	536	insulin	49	171	question	84	102	portion
15	517	meal	50	170	night	85	102	egg
16	502	weight	51	164	cd %	86	102	foot
17	444	problem	52	158	number	87	101	condition
18	434	reading	53	157	luck	88	98	part
19	434	month	54	156	anyone	89	97	appointment
20	403	glucose	55	152	idea	90	97	lunch
21	398	exercise	56	147	work	91	96	fat
22	394	gp	57	146	drug	92	96	home
23	359	metformin	58	145	im	93	96	pain
24	358	testing	59	144	potato	94	95	change
25	343	hour	60	143	stone	95	94	feeling
26	326	control	61	143	prescription	96	94	person
27	324	effect	62	142	someone	97	94	pressure
28	317	meter	63	140	diagnosis	98	91	reason
29	310	advice	64	138	today	99	89	bs
30	310	week	65	132	loss	100	89	chocolate
31	275	bg	66	130	everything	101	89	end
32	263	something	67	127	rice	102	89	fact
33	261	nurse	68	125	amount	103	89	family
34	260	body	69	124	fruit	104	89	intake
35	249	result	70	124	eye	105	89	reply

## Appendix 1. Nouns with Highest Frequency (continued)

Rank	Frequency	Noun	Rank	Frequency	Noun	Rank	Frequency	Noun
106	86	issue	138	62	support	170	44	neuropathy
107	84	difference	139	61	alcohol	171	44	situation
108	83	friend	140	61	one	172	44	soup
109	82	husband	141	60	energy	173	44	stress
110	82	gliclazide	142	58	lifestyle	174	44	wife
111	81	milk	143	58	money	175	43	knee
112	80	stuff	144	57	resistance	176	43	kidney
113	80	cake	145	57	worry	177	43	oil
114	79	cheese	146	56	stress	178	43	complication
115	78	news	147	55	fish	179	43	start
116	77	fasting	148	55	need	180	42	answer
117	77	meat	149	55	roll	181	42	cramp
118	76	lbs	150	54	check	182	41	hope
119	76	liver	151	54	toast	183	41	toe
120	76	surgery	152	54	website	184	41	statin
121	76	tea	153	53	clinic	185	41	target
122	75	evening	154	53	protein	186	41	thumbup
123	75	experience	155	52	drink	187	41	step
124	75	porridge	156	51	tomorrow	188	40	bacon
125	75	shock	157	50	carbing	189	40	hand
126	75	patient	158	50	lrb	190	40	diary
127	74	eating	159	50	practice	191	39	version
128	73	care	160	49	risk	192	39	consultant
129	73	stomach	161	48	cream	193	39	wheat
130	72	treatment	162	48	dinner	194	37	age
131	71	salad	163	48	disease	195	37	chance
132	69	heart	164	48	head	196	36	joint
133	68	place	165	48	nhs	197	36	knowledge
134	67	figure	166	47	monitor	198	36	salt
135	65	pancreas	167	47	regime	199	35	appetite
136	64	sleep	168	47	review	200	<null>	<null>
137	63	chicken	169	45	mile			

Appendix 2. 3-gram Analysis of Phrase Patterns

Rank	Freq	3-gram	Rank	Freq	3-gram	Rank	Freq	3-gram
1	296	I don t	34	52	all the time	67	40	don t think
2	161	I have been	35	51	I haven t	68	40	I think it
3	131	a lot of	36	51	seem to be	69	40	If you are
4	124	I was diagnosed	37	51	that I have	70	40	need to be
5	114	I ve been	38	50	you have to	71	39	and it s
6	105	you need to	39	49	I need to	72	39	seems to be
7	99	be able to	40	48	and I am	73	38	a day and
8	95	I m not	41	48	if you have	74	38	but it s
9	92	if you are	42	48	on this forum	75	38	I wouldn t
10	91	a couple of	43	48	one of the	76	38	in the UK
11	89	I can t	44	48	to be a	77	38	some of the
12	85	blood sugar levels	45	48	you have a	78	37	that I am
13	85	don t know	46	47	the amount of	79	37	they don t
14	83	I am not	47	46	I am a	80	36	I m sure
15	81	at the moment	48	46	I think I	81	36	need to test
16	80	you don t	49	46	if you can	82	36	t seem to
17	75	I have a	50	46	to the forum	83	36	the end of
18	74	I have to	51	45	I was told	84	35	blood glucose levels
19	70	I didn t	52	44	that you have	85	35	I used to
20	66	a bit of	53	44	when I was	86	35	I was on
21	62	in the morning	54	44	with type diabetes	87	35	most of the
22	59	don t have	55	43	have to be	88	35	my BG levels
23	59	I had a	56	43	I ve had	89	35	to see the
24	59	low carb diet	57	43	there is a	90	34	and I have
25	59	to have a	58	43	was diagnosed with	91	34	for a while
26	57	and I m	59	42	I m a	92	34	go back to
27	57	I have had	60	42	t want to	93	34	I know I
28	57	my blood sugar	61	42	to see if	94	34	I m on
29	55	as well as	62	41	I have lost	95	34	need to get
30	53	diet and exercise	63	41	in the s	96	34	times a day
31	53	I had to	64	41	put me on	97	34	to see what
32	53	that you are	65	40	a low carb	98	33	is a good
33	53	your blood sugar	66	40	diagnosed with type	99	33	it s not

## Appendix 3. Dominant Psychological Conditions

Category	# of tokens	Percentage	Sentence examples
Negation	1215	0.146633 (14.6%)	'I don't find it easy to afford strips as I'm a pensioner.' 'I'm spiralling downwards and don't know how to stop.'
Capability	345	0.041636	'I can't handle them at all at present.' 'like most type2 diabetics, you are not able to metabolise them efficiently.'
Necessity	288	0.034757	'We really need to know a bit more about your daily levels, fasting each morning, pre and post prandial levels, bedtime level.' 'It all takes time and we all so desperately need patience.'
Certainty	394	0.04755	'I am wondering if you have changed anything.' 'My trouble is controlling my overnight BS often waking with bs in the 8 to 10 range, have tried not eating before I go to bed and also eating before bed, both of which dont seem to make any difference.'
Duty	220	0.026551	'I have to increase my lipotar currently taking Aprovel 10mg.Really struggling to get control of blood sugars.' 'I have also struggled with high bp for many years and I have to take 5 different drugs each day for my bp.'
Wish	128	0.015448	'I only wish I had known about this earlier, could have saved myself a lot of grief.' 'As I said I want to be able to carry on with my life abeit at a slower pace without having any complications.'
Total	2590	0.312575	