A NOVEL THEORY OF SUPPORT IN SOCIAL MEDIA DISCOURSE (WITH ARTIFICIAL INTELLIGENCE AND LINGUISTICS ANALYSIS)

By Bazil Stanley Solomon

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF MASTER OF PHILOSOPHY

AT OXFORD BROOKES UNIVERSITY, OXFORD, UK

UNDER THE SUPERVISION OF

Prof. NIGEL CROOK, Dr ALON LISCHINSKY, AND Dr KENNETH BONESS

2020

Declaration

This thesis is my work and does not include outcomes from work done in collaboration with others except where explicitly shown. This work brings together areas of both Artificial Intelligence and Linguistics. This thesis does not exceed the regulation length of between 50,000 words of the social sciences and 40,000 words prescribed by the computing sciences, including tables and footnotes. However, the thesis also contains about 29000 words of anonymised examples of text-based data and ethics correspondence.

Bazil Stanley Solomon

ABSTRACT

This thesis aims to inform the way that people are directly affected by various issues and conditions and how they can support each other on social media. It explores their utilisation of novel salient high-frequency and diverse device-enabled discourse categories of purpose and content. The predominate analysis is from computing and artificial intelligence. It is with an aspiration for the novel methods and novel theory to find its place in corpus linguistics—subtle linguistic analysis guides the exploratory research. An investigation of sophisticated patterns of support alludes to an architecture for social media discourse. The primary patterns can include device-enabled discourse categories of purpose and content. My thesis calls this vose and nent. Advice with stance-taking offers an entry into the large-scale analysis. It forms part of a multitude of online topics and discourses from diabetes advice to raising charity money. The thesis focuses mainly on domain-specific targets of diabetes, blood, and pumps from many others ranging from targets of usernames, child, school, greetings, places, time, to years. The thesis proposes a novel theory of support constructs in social media discourse. The thesis makes a methodological contribution that seeks to combine Artificial Intelligence (AI) computational analyses with corpus linguistics (CL) and qualitative SFL analysis. It is carried out on a big large-scale dataset of 218,068, anonymised Facebook Diabetes UK posts and 16,137 anonymous diabetes-related users of the platform. AI with Latent Dirichlet Allocation (LDA) topic-modelling probabilistic modelling, automated content analysis, an annotation for entity recognition and Discourse Analysis (DA) is utilised with consideration of its limitations but to analyse the corpora for potential patterns. An adapted anonymisation process is used on the data to meet the ongoing challenges of online ethical research requirements.

People living with diabetes employ high-frequency patterns of relevant device-enabled categories of purpose and content with for example linguistic forms of advice with stance-taking, diabetes, blood, pumps, to humour/sarcasm or questioning or raising charity funds in their support of themselves and each other amongst other interactional ways. They may not respond directly to each other in consecutive posts but to many other similar posts with similar targets. The focus on particularly advice with stance-taking and domain-specific targets helps to place the novel device-enabled discourse categories in context. These can be in a broader context of power and solidarity, demonstrating social relations concerning risk and trust. Hence, the uncertainty and the variation of effect displayed when sharing information for support.

The support trends from a probabilistic model of social media support are confirmed with log-likelihood, precision measures and a multi-method approach.

The implications of the new theory are aimed at healthcare communicators to work with organisations. To help their social media users support each other by understanding a peer-focused view of chronic illness support. Corpus linguistics may benefit from the use of combined AI and DA approaches to anonymised large-scale online data.

This thesis also offers preliminary work for support-bots to be programmed to utilise the language patterns to support people who need them automatically. The bots may be able to have conversations instantaneously with many people but to do so in natural ways.

Originality Statement

I hereby declare that this submission is my work and does not, to the best of my knowledge, contain material that has been previously published or written by another person.

It does not contain substantial amounts of material that have been accepted for the award of any other degree or diploma at Oxford Brookes University or any other educational institution, except where due acknowledgement is made within the thesis.

Any contribution made to the research by others with whom I have worked with at Oxford Brookes University or elsewhere is explicitly acknowledged within the thesis.

I also declare that the intellectual content of this thesis is the product of my work, except to the extent that assistance from others in the project's design, conception, style, presentation and linguistic expression is acknowledged.

Bazil Stanley Solomon

2020

Copyright

Anaconda©

AntConc©

Facebook©

Flickr©

LIWC©

Microsoft Excel©

MeaningCloud©

Python©

R Programming©

Spyder©

Twitter©

Acknowledgements

First and foremost, I would like to thank my supervisors, Prof. Nigel Crook, Dr Alon Lischinsky and Dr Kenneth Boness. During this work, they have supported me academically and taught me many useful lessons, which I will utilise throughout my research and teaching career. They have been a constant source of inspiration and have provided me with daily access to their communications, computing and linguistic expertise, and have set thought-provoking challenges for me. For this, I would like to acknowledge how deeply indebted I am to all of them.

I would also like to thank Artie the Artificial Intelligence Lab 'Robot', who gave me pause to think when he shook my hand and greeted me several days ago.

My research benefited from conversing with Corpus Linguistics, DA and AI. I would like to thank Mrs Jill Organ, Mr Jim Pye, Dr Tjeerd olde Scheper and the Oxford Brookes University Ethics Council for their guidance and help with university policies, procedures, ethics and IT skills. We were also afforded some social coffee breaks and a set of brilliant courses and researcher training.

I would also like to thank the British Computer Science Society (BCS) especially Geoff Hunt, the secretary of the Wiltshire branch; the secretary of the Swindon Diabetes Group, Mathew Spencer, and committee member Stephan McMahon.

I would also like to thank my families in the UK, the Netherlands, New Zealand and South Africa who have kept me going mainly my daughter Amelie Solomon, my wife Dr Sharon Jacob, and my late grandmother, Mrs Sylvia Solomon. There were many ways and trips in order to get me away from the computer. There were many hot meals during some of the freezing UK winters, and my friends always kept me in good cheer.

Oxford Brookes University, the Department of Computing and Communication Technologies and the Department of History, Philosophy and Religion have funded this work.

Abbreviations and Acronyms

AI: Artificial Intelligence

BERA: British Educational Research Association

BNC: British Natural Corpus

BPA: British Psychological Association

CL: Corpus Linguistics

DA: Discourse Analysis

EU: European Union

FDP: Facebook Diabetes UK page

FDUK: Facebook Diabetes UK Page Corpus Data

GDPR: General Data Protection Regulation

HRA: Health Research Authority

ICO: Information Commissioner's Office

LDA: Latent Dirichlet Allocation

LIWC: Linguistic Inquiry and Word Count

MRC: Medical Research Council

NER: Named Entity Recognition

NHS: National Health Service

NLP: Natural Language Processing

NMF: Non-Negative Matrix Factorisation

Peer: A Facebook Diabetes UK User

SFL: Systemic Functional Linguistics

SMCW: Social Media Collaborative Work

SSMD: Support in Social Media Discourse

UK: United Kingdom

UREC: Oxford Brookes University Ethics Committee

Abstracts, publications, workshops and presentations

1. APCLC 2020: Asia Pacific Corpus Linguistics Conference 2020, Yonsei University, Seoul, South Korea, June 23-27, 2020
   A Novel Discourse Analyses Method with SFL and Appraisal Framework in a Corpus Linguistics Approach with Artificial Intelligence on Large-Scale Data: Towards a Novel Theory of Support Constructs in Social Media Discourse (SSMD)
2. The 10th International Corpus Linguistics Conference (CL2019) at Cardiff University, United Kingdom, Workshop Monday 22nd July 2019: How to Use Artificial Intelligence with Topic Modelling, as part of Corpus Linguistics and as a Discourse Analysis Method
   A Novel Discourse Analyses Method with SFL and Appraisal Framework in a Corpus Linguistics Approach with Artificial Intelligence on Large-Scale Data: Towards a Novel Theory of Support Constructs in Social Media Discourse (SSMD)
3. 17th Corpus Linguistics in the South, 24th November 2018, Combining Qualitative and Quantitative Frameworks in Corpus Linguistics, University of Roehampton London
   Combining Quantitative Artificial Intelligence Topic-Modelling on Large-Scale Social Media Data with Qualitative Systemic Functional Linguistics and Appraisal Model Frameworks in Corpus Linguistics: Towards a Novel Theory of Support Constructs in Social Media Discourse (SSMD)
4. HSC Workshop 21st November 2018, Abstract and Presentation, Workshop of the British Association for Applied Linguistics (BAAL), Health & Science

Communication Special Interest Group (HSC SIG), “Mixing it up: Multi-methods, media and modalities”, #healthsci18, Wednesday 21st November 2018, Hosted by the School of Education, Communication & Society, King’s College London

A Critical Use of Discourse Analyses and Artificial Intelligence on Large-Scale Social Media Data towards a Novel Theory of Support Constructs in Social Media Discourse (SSMD)

5. LIP sessions, Language Ideology and Power Research Group, Abstract and Presentation, October 2018, Lancaster University County South D72, Department of Linguistics and English Language:

A Novel Theory of Support Constructs in Social Media Discourse (SSMD)

Table of contents

List of figures

List of tables

1 INTRODUCTION

The thesis seeks to contribute to a deeper understanding of the problem of support patterns in social media discourse for diabetes support. The challenge is to discover patterns of online support, so that analysts may understand how the peers post about their diabetes. The research is predominately from a computing and artificial intelligence (AI) perspective with guidance from linguistics. AI and topic modelling alone can produce too broad base categorisations of text data. What are the potential sophisticated linguistic concepts that underly these topic modelling patterns? However, AI can help reduce the big data and large-scale text corpus from many potential linguistic dimensions to a few high-frequency ones. In counteracting these potential platforms, analytic tools and their difficulties, the thesis will bring artificial intelligence and a necessary linguistic analysis together. Such an approach may help to resolve these challenges but primarily help to develop a greater awareness of the inherent limitations. In overcoming the challenges, such an approach may firstly find a place in corpus linguistics and secondly a useful theory of support. Searching for an 'architecture' of these types of discourses is analogous to the search for an architecture of sentences. People use sentences daily, and take it for granted, producing them in a certain way. For instance, a sentence may consist of verbs 'doing words' and nouns 'the entities'. People use posts and discourses in social media, and they can benefit from a comprehensive and subtle study.

In 2020, the Corona Virus Pandemic caused the United Kingdom government similar to many countries across the world to make their citizens go into a state of lockdown. The lockdown was where the government told the main population of people to stay at home in isolation for several weeks to prevent the spread of the virus. Key workers such as NHS staff, public transport and other essential services could carry on but with safeguarding measures in place. Non-infected and non-vulnerable people could only go out of their homes for essential groceries, medicines, and for a daily exercise but all with social distancing. People formed many more large and significant data interactions in online social media groups. These are to communicate and support each other throughout the national and global health crisis emergency. The frequency of the use of COVID-19 terms and topics increased significantly on social media. Understanding technologies and human behaviour are at the forefront of a national and global response. There are research and development with, for example, the NHS contact-tracing Mobile smartphone App (Ferretti et al., 2020). It builds a memory of proximity contacts and immediately notifies contacts of positive cases. However, they argue that it could achieve epidemic control if used by millions of people in the national UK population together with ethical requirements for an intervention of this kind. Besides, by targeting recommendations to only those at risk, epidemics could be contained without need for mass quarantines ('lock-downs') that may be harmful to society.

Social media offers large-scale human interaction and Big Data available for investigation on human behaviour. Social media is hugely popular. For example, Facebook is continually evolving to attract new members and to keep existing account holders engaged on the thousands of different pages and groups. The research work in the thesis investigates the Support on Facebook Diabetes UK before the COVID-19 outbreak. It perhaps may be of use to help illuminate how people went about coping. This additional support via technologies could be significant. They have the potential to help peers cope. They may develop contacts with each other. The thesis is vital in discovering high-frequency online behaviour patterns. If these pertain to sophisticated support patterns, then they could be useful. They may be about primarily specific and salient ways in the discourses that people go about talking to each other. They can communicate about online diabetes support with linguistics forms of advice. They can take a position concerning their care. These stances can be on many topics from diabetes, blood, pumps, humour/sarcasm or questioning or to raise charity funds. There are many other interactional activities amongst peers. The thesis places emphasis on how people share the purpose and content of their discourses. These potential support patterns may be regarded as straightforward but are of importance in critical social media online groups. People can benefit from this additional support in newer technologies.

The thesis studies the Social Media Diabetes UK support group Facebook Diabetes UK (2015-2020). It is for an understanding of what makes peers interact with a post. Such an understanding may help to promote their organisational aims. Diabetes UK posts online content in different content formats, for example, text, images, video, links, polls, and quizzes. The posts can have a content focus to give the post meaning. It is for, example, with the apparent sharing of Diabetes related content, and communication. It can be for charity, business, support, information, promotion, humour/sarcasm and special offers. It can be for building a peer community that support each other. Diabetes UK can develop an audience and encourage people to follow or 'like' the group page through the creation and use of engaging content. They may do this by using keywords. They may link relevant medical sites to individuals. They may share content posted by others. Diabetes UK can employ direct and indirect advertising and news feed. People may create posts, reply or comment to other people's posts or like or share or do all of these to posts. It does not appear always to follow a linear consecutive posting pattern.

The subtly of insights from the field of linguistics guides the thesis with questions such as, is there an architecture for online support discourses? Can online discourses be organised into different categories analogous to the architecture of sentences? The structure of sentences is involved, as proposed by Lehmann (1987). A well-established view is that sentences are made up of, for instance, verbs and nouns. Linguists are continually exploring the architecture of discourses analogous to the study of sentences (Crystal, 1997).

The thesis shows social media discourse, for instance, may contain remarkably sophisticated high-frequency stance-taking with diverse affective stance and low epistemic stance. It is not necessarily in consecutive posts but somewhat indirectly in usage across many posts about the same target. People can post support about similar targets at anytime and anywhere in the discourse. It is to anyone or the entire group at once in a post. The 'Social Media Diabetes Discourse' is potentially another 'peer'. Post can concern advice with stance-taking with a target of diabetes or blood sugar levels or insulin pumps. People can tend to use support patterns in social media diabetes discourse. The delicate patterns show a tendency, e.g., for an indirect manner of interaction for advice. However, there can be many other direct interactional ways of support, such as greetings, humour/sarcasm.

A possible way of looking at closely related posts with high-frequency patterns would be to look at the post with similar high-frequency discourse devices. It would be to look at mainly similar targets such as diabetes. A comparison made to compare consecutive and non-consecutive posts. However, it is to look at the many random samples of posts for that target. It can focus on stance-taking related to that topic. Using a methodology to find similar stances in consecutive and non-consecutive posts can also help to check if posts respond directly to each other or the discourse. The research shows that people may not follow a linear conversational approach. Instead, they may follow a social media type non-linear conversational approach. It would seem to be an online community of sharing and posting at different times and days or linearly in response to the many diverse posts of many different people. The corpus is rich, and the Discourses and topics covered in the diabetes-related posts are diverse. They can cover for, example, diabetes, humour/sarcasm or questioning or raising charity funds or healthcare, blood glucose levels, advice or on the politics of the NHS.

Diabetes-related support posts may phrase in a particular way—the site promotes as a peer support site. The thesis focuses on and suggests it may hold clues as to what is the higher frequency posts or patterns of online behaviour. It can offer insights when thousands of people attempt to help each other online. People can express complex attitudes, opinions and sentiments in their online attempts to support each other. What people say and what people mean are just as contentious in online communications as they are in face-to-face communications. There is much complexity to be understood in both instances. The thesis is interested in considering if the discourse shows evidence of consecutive and non-consecutive posts in Facebook diabetes discourse that contain high-frequency categories. It is for devices for discourse purposes (voses) as well as the content (nents). Though the focus is on the advice with stance-taking in the many varied posts, there are potentially many other interactional activities amongst peers. These are a useful guide into the discourses. It is for a placing of the novel discourse devices of the thesis into the context of online support. The thesis shall show that they can be relevant for an understanding of an online corpus. The thesis shall argue for and offer evidence from rigorous systematic data analyses. What people talk about maybe straightforward online. However, it can become more apparent when the posts are investigated in terms of finer-grained purposes and content devices across the many diverse, thousands of posts and discourses. The thesis goes even further to show how people go about posting and using high-frequency words. These pertain to high-frequency devices to achieve purposes in the discourse. It is is a delicate and complex matter. The research focuses on the subtlety of advice with stance-taking. However, it gives a cursory treatment to other linguistic forms. The discourses can contain, e.g. humour/sarcasm or questioning or hope or raising charity funds. Of primary importance in the thesis are the novel particular high-frequency *device-enabled discourse purpose categories* (topics) and *device-enabled discourse content categories* (TOPICS) in combination. I call these voses and nents.

Brownson and Heisler (2009) have established that a primary element in the effective management of diabetes is in support for patients to self-manage. They argue that barriers to care exist for both patients and healthcare systems. Integrating peers into diabetes care has great potential for improving diabetes outcomes worldwide. *Support* is from peers who can offer emotional, social, and practical assistance for people to do the things they need to do to stay healthy.

Dennis (2003) took a critical look at the web and email-based programs and support. He considered the importance of establishing exactly who is the target audience, what these groups need, and especially what can feasibly meet the needs of these groups. He defines peer support in a broader context. It is as 'the provision of emotional, appraisal, and informational assistance by a created social network member who possesses experiential knowledge of a specific behaviour or stressor'. Peers may have similar characteristics as the target population in order to address a health-related issue of a potential or stressed focal person.

Researchers like Mead at al. (2001) define peer support as 'a system of giving and receiving help founded on key principles of respect, shared responsibility, and agreement on what is helpful'. My thesis likewise suggests support to be dynamic and construct that people can work together to achieve.

Solomon (2004) helps to establish the critical argument of my thesis. She defines peer support as occurring when people provide knowledge, experience, emotional, social, or practical help to each other. It may either offline or online with internet support groups. It is by offering, for example, information and support that can include advice, guidance, and feedback. She suggests that peer support is distinct from other forms of social support. In that, the locus of support is a peer – a person who is similar in fundamental ways to the receiver of the support where they perceive their relationship as one of equality. She argues that a peer can offer support through relevant experience. He or she "*has been there, done that and can relate to others who are now in a comparable situation*".

A theoretical issue that dominates the field is concerned about the identification of support. How exactly are people able to create constructs recognised as support? An excellent place to start is to look at 'advice' which my thesis places importance. Existing research recognises the critical role played by the advice in support (Morrow, 2012).

The thesis incorporates the ideas as mentioned earlier when developing a comprehensive notion of online support. The thesis notion of an online peer is developed as '*participants dealing with the illness directly either by having the illness or by helping others who have the illness together with the entirety of the 'social media diabetes discourse' acting as a peer* '. In the pages that follow, support with, for example, advice can be looked at primarily from a study of language, discourse-analytical together with a sociological viewpoint.

A significant challenge for this thesis is to identify from a social media support group data what 'support' entails. It is how to analyse large-scale online data in social media support discourse. My thesis prescribes how Simmon (1969) used the idea of a 'pattern' in his work with natural language about question-answering systems. It is for my thesis to identify patterns in social media support discourse. He defines a pattern as 'a statement that can be composed of words or sequence of words that can match any event of the keyword or sequence of words without any concern about what happened before or followed them'. He thinks about the AI conversation machine Eliza developed by Weizenbaum (1966) when giving an example of a pattern. The pattern (0 you are 0) which then is used to match any event of the sequence of words 'you are' without any concern about what happened before or is followed by them.

My thesis is about a predominate utilisation of computing and artificial intelligence analyses with guidance from the field of linguistics. The methods can find a place in corpus linguistics. Blei et al. (2003) describe '*what people converse about*' in the sense that a social media post contain many topics. For them, each social media post can have different rankings of topics by comparing the topics in the post to other all other posts, be it consecutive-linear or non-consecutive-linear postings for that particular social media discourse. This AI machine learning understanding of topics in the text, for example, a Facebook post, may not be equivalent to Halliday's (1967) understanding of a 'topic'. Halliday considers a topic to be the 'first expression of any sentence'. What exactly is the topic produced with AI text analysis? So, both the first idea expressed in a sentence or post or the highest frequency theme in many posts are significant. Latent Dirichlet allocations (LDA) 'topics' of the 'sentences' in posts are, however, about the interactions of many individuals in many discourses with many different interactions in overall communications. They may indicate what is remarkable about a corpus of conversations. The exciting thing for my thesis is that LDA can cluster similar posts under a set of top trigrams. So, in effect, the highest frequency top trigram maybe what all those posts are inherently about, even though the human eye would not see the posts as immediately similar.

The focus on particularly advice with stance-taking and domain-specific targets helps to place my thesis novel device-enabled discourse categories in context.

The capitalised 'TOPIC' *'nents' will be used in the thesis to describe the device-enabled discourse content.*

The 'Target' *will be used to describe the entities mentioned in the posts.*

The lower-case 'topic' *'voses' will be used to describe the purpose of the conversations*.

Both Blei's and Halliday's notions can help to define 'topic', 'TOPIC' and 'target' as used in my thesis. They are essential in understanding what people converse about and how they purposefully do this.

The thesis gives the terms 'topic' and 'TOPIC' more precision. They are related to fulfilling another objective of the research, i.e., to use LDA for large-scale data analysis of text data and, critically, to think about what LDA offers. The primary objective is to develop and propose a theory of support in social media discourse (SSMD). The following terms are of primary importance in my thesis are used throughout the thesis: '*device-enabled discourse purpose categories*', or 'topics' and '*device-enabled discourse content categories*' or TOPICs. These will be used when referring to discourse devices. They can make up the 'purpose' and the content of the many conversations by many different people in the corpus, respectively. The thesis focuses mainly on domain-specific targets of diabetes, blood, and pumps from many others ranging from targets of usernames, child, school, type, greetings, places, time, to years. They are explained and analysed in more detail throughout this thesis.

This thesis suggests that the '*device-enabled discourse purpose categories* are essential. They can be a leading factor for discourse-based conversation support systems. For example, social media diabetes support discourse. They can show trends where people together are purposely creating constructs of online support. These devices may perhaps be some of the constructs that provide meaning in online support. They can tend to contain the building blocks, for example of an important element, i.e. of 'advice' with stance-taking. It will be elaborated throughout the thesis with other topics such as humour/sarcasm or questioning or hope or raising charity funds.

This study provides new insights into the online support of people with chronic illness via a Facebook social media diabetes page. It looks at the discourses about peer support with numerous participants on the Facebook Diabetes UK page Facebook (2015, 2016, 2017, 2018).

Diabetes UK believes that people living with lifelong illnesses can utilise social media to help each other cope with the condition and its associated issues. They attempt to explain what diabetes is on their webpage: 'Diabetes is a serious, lifelong condition where your blood glucose level is too high'. They explain that there are two main types of diabetes: Type 1 and Type 2. These are different conditions, but they are both dangerous, and there are some other rarer types of diabetes too. They state that when you have Type 1 diabetes, you cannot make any insulin at all, but if you have Type 2 diabetes, then there are some differences, as the webpage explains: 'The insulin you make either can't work effectively, or you can't produce enough of it. In both types of diabetes, because glucose can't get into your cells, it begins to build up in your blood. And too much glucose in your blood causes a lot of different problems'.

Chronic illness presents many daily issues and conditions. These are added to the normal daily living ones that everyone deals with from time to time. Diabetes can present some common and highly specific challenges (Barrera et al., 2002; Shaw and Johnson, 2011; Diabetes UK, 2015, 2018). Diabetes UK state some of the symptoms: ‘like … feeling very tired…. may also lose weight, get infections… slow healing wounds…. If your child has been diagnosed with diabetes, you may be feeling overwhelmed and worried. Diabetes UK aims to provide you with the information to help you and your child manage their diabetes together. See below for help and advice, from dealing with diagnosis and finding suitable childcare to speaking to teenagers’.

Diabetes UK highlights the importance of managing the condition, as there can be many different high-risk complications, ‘Over a long period of time, high glucose levels in your blood can seriously damage your heart, your eyes, your feet and your kidneys’. They believe that with the right treatment and care, people can live a healthy life. There was less risk that someone was looking after their treatment will experience these potentially life-threatening complications.

My thesis seeks to investigate this type of online ‘management of diabetes’ by participants/peers. It will attempt to find out the support patterns for their prevailing conditions and issues. Some of the important ones suggested by (Harris et al., 2013; Karami et al., 2018; Diabetes UK, 2015, 2018) tend to be about managing blood glucose levels with medication, automatic medical insulin-measuring devices, diet, and looking after children with diabetes. My thesis needs to consider the prevalence of some of these essential conditions and issues. It is to find out who and what is being ‘supported’? It can be done by analysing an example of online discourse about diabetes.

1.1 Motivation

The field of linguistics concerning the study of such support groups on social media has also become a subject of increasing academic interest in the past decades. It concerns the development and usage of social media technologies such as Facebook or Twitter.

It has been proposed that conversations on social media may be analysed as a discourse (Koteyko, 2006). They can be naturally occurring conversations that are instances of language in use. Koteyko also sees the language in use as more than utterances, as they take place between individuals and are both interactive and can be sequential. Secondly, my thesis seeks to investigate the crucial people who are living with diabetes. It is for their online behaviours as they engage in conversation with other members of their social media communities. The latter share the same challenges or medical conditions. It proposes that their behaviour connects to the technology that they use.

Very little literature covers the methodological approach taken in this study. It is a mixed-methodology study based primarily on Artificial Intelligence (AI), automated computer analysis and secondly DA of large-scale corpora. Such analysis may together provide some of the more complicated aspects of large-scale communication and produce a broader context (social relations) of any discourse. My thesis as a type of analysis is described by very few researchers (e.g., Hashmi, 2012). My thesis looks for examples of the subtler aspects of advice and at the broader aspects of people that share stance-resources and objects as a way into the analysis of a large corpus of text data.

1.2 Context

There is limited literature on peer usage of Facebook to foster support for chronic illness. The corpus offers potentially linguistic forms of advice with stance-taking and others such as humour/sarcasm or questioning or hope or raising charity funds for investigation. Al Mamun et al. (2015) studied 'open' Facebook groups. It is for groups that were entirely accessible to the public. They showed that the user could raise awareness amongst the other platform users about chronic illness, for example. Researchers have established that people diagnosed with a chronic illness such as diabetes can actively seek health information on the internet. It is precisely in interactive venues such as social media forums (Barrera et al., 2002; Shaw and Johnson, 2011; Harris et al., 2013). These articles provide valuable insight into 'support', conducted by researchers on healthcare communication.

Until the widespread availability of social media, diabetes health information was available online on static lists with no interactivity and engagement (Kaufman, 2010). The interactive nature of social media can provide ongoing support for people with diabetes who are faced with managing their disease daily. Shaw and Johnson's (2011) article provides valuable insight, providing an example of a blog post to show this point. The blogger explains, '*...because diabetes is every day.... It is not a disease that you can manage by simply popping a pill and seeing your doctor once or twice a year. this disease, as a whole, requires thought and care every day...*'

As is well known, social media technology has developed from Web 1.0. It was based mainly on discussion boards, to Web 2.0 (Tannen and Trester, 2013). The phenomenon has been widely researched. Web 1.0 consisted of static websites compared to the interactive content of Web 2.0 websites. Earlier forms of online health support and the extensive body of research into them have informed the present study.

Social media can be seen to be related to earlier Web 1.0 phenomena, such as discussion boards. Kaplan and Haenlein (2010) define social media as internet-based applications that facilitate the creation and exchange of user-generated content based on Web 2.0. They use Web 2.0 (O'Reilly, 2009) to describe how end-users started to employ the World Wide Web as a platform.

There is a broader history of online work suggested, starting with CSCW, which combines the cooperative work of individuals through networking, hardware, and software (Greif, 1988). Solomon et al. (2011, 2012) have proposed that the use of social media collaborative work (SMCW) is more complicated than the 'cooperative' nature of CSCW. CSCW can usually involve a predetermined, relatively small group of employees working under the direction of a manager to complete a well-defined project within a limited timescale. Recent developments in CSCW are concerned with integrating social media for general purposes (Al Mamun et al., 2015). My thesis suggests that the full potential of SMCW needs to be fully utilised, and the constructs for online support need to be identified. The full power of mass work collaborations offers ways of doing distinct types of collaborative work. It can be for how people utilise language to support each other online. A study of this phenomenon is critical.

NHS Diabetes (2015, 2016) recommends the use of the Diabetes UK social media (Facebook 2015, 2016) platform as a 'useful link' on its website. Users of these sites have many confusing things to deal with (at least to the layperson), often being guided by the notions in how they see others use the platform or from the guidance of FDP (2017).

People share diverse posts, especially if they pertain to the complexity of diabetes support on Facebook. However, it is not easy to define 'support'. There does not seem to be any definition of support that can account for all possible scenarios in people's many similar or differing posts.

Facebook's (2017) vision statement states that '*Facebook's mission is to give people the power to share and make the world more open and connected*'. People can report anything that does not meet an acceptable standard of posting, e.g., posting illegal and offensive content, to FDP's administration or Facebook management. The usage of this available source of large-scale data for research or business activities is undergoing much criticism. New regulations are being developed to control the fair usage of this large-scale data and to protect the users and participants in these platforms and online communities (ICO, 2018).

Individuals or organisations can take part in or set up their page on social media in order to help each other (for example, the Diabetes UK Facebook page, 2015). Organisations like Diabetes UK have launched social media sites to help people who are living with diabetes (Diabetes UK Facebook, 2015, 2018). Diabetes UK is a non-profit organisation based in London. Their mission statement is that they are '*the leading UK charity that cares for, connects with, and campaigns on behalf of all people affected by diabetes*'.

They argue that people in need of help with issues can communicate their concerns about on their site. It is for example, on how to use the latest meters that measure blood glucose readings. Peers can use text in posts about their readings. These are amongst many other things to do.

Their idea is that people can share their blood glucose levels and experiences with others online. It is in order to improve management of their often-common conditions and issues. They can then set up many different pages for others to 'like'. This page and the 'like' function allow numerous people to become part of the conversations on a page. People can ‘post’ as part of conversations. A post is more than writing something on the page, as the expectation is that someone will respond to it. Posting, commenting and replying can be similar to a natural conversation where people engage apparently ‘directly’ with each other. Facebook pages can be about any TOPICs or topics ranging from politics to healthcare, to healthcare support systems. Facebook describes the steps involved ‘to like other pages’ that other people have set up or that a group of individuals or an organisation has set up.

Diabetes is an example of a lifelong illness. It is primarily about the body not being able to control its blood glucose levels. It is usually done by the human body naturally in healthy people throughout their lifetimes (NHS Diabetes, 2016). NHS Diabetes indicate that people living with diabetes must manage their blood glucose levels in many ways. They recommend that when illnesses last a lifetime (i.e., a chronic illness), there are many things that individuals must manage daily throughout their lives. They highlight the need for standards of care, as proposed by the National Institute for Health and Care Excellence (NICE, 2017). Different people will struggle with fluctuating glucose concentrations. It is as their natural way of keeping their blood glucose at an optimal level declines as their pancreas functions deteriorate or they develop a resistance to insulin (Diabetes UK Facebook, 2015, 2018). Diabetes UK insists that where their levels are so extreme that they can result in death, users should direct them back to their GPs. These examples give an idea of the expectations of support but not necessarily what people are doing to support each other online.

This research focuses on social media, and Facebook in particular, and so cannot extrapolate the findings to all other platforms. Diabetes UK (2015, 2018) utilises many technologies covered in healthcare communications research over the past years from message boards, online forums, blogs, websites, and social media sites:

‘Online Communities, Get involved, support each other. Our online communities let you chat, find support, and discuss issues with other supporters, as well as discover more about our campaigns, information, and activities. We want our supporters to be able to talk to each other and us online in an environment that is informative, supportive, engaging, and – most importantly – safe. Our biggest communities can be found on Facebook and Twitter’ (Diabetes UK, 2015, 2018) and cannot give medical advice.

FDP (2015, 2018) argues that a population in the absence of the availability of a 24/7 medical specialist can utilise social media online communities for non-clinical help. Specialist if always available, would converse with them directly about clinical issues and non-emergency conditions and issues. Their vision is that online sites where people can converse with each other about their health problems become an essential part of the users’ treatment. The FDP (2015, 2018) contains many posts from people who share information about, for example, their blood glucose levels online with each other. They can get a sense that they are going to be all right: ‘*…yes i need it as i have just come out of the hospital… with yet another high blood sugar… went in it was approximately forty…i am out now… so going to take it easy…no drinking!...*’

This verbatim example of a post can be about maintaining proper blood glucose levels. It can be related to diet. It is how to use the meters regularly where the person needs to take a sample of blood from their fingers. They insert it into the meter to get a blood plasma reading, or as many peers refer to it in the corpus, ‘blood sugar’ or ‘glucose readings’.

Furthermore, NHS Diabetes (2015, 2018) states that the meter must be able to give a reading of the sugar content in a person’s blood. The person can use their posts to ask a question about their blood sugar levels if it is too high or too low. They may have many negative symptoms, such as extreme thirst or dizziness.

If the reading is at the upper or lower limit of levels determined by a medical specialist from the NHS (2017), then they can be at severe risk of fatal complications. The FDP would then instruct people who are at immediate risk to go to straight to A&E.

My thesis suggests that individuals use social media to seek guidance. It can be for the many different things through the many ways that different people on the site can help each other (Diabetes UK Facebook, 2015, 2018). In the corpus of my thesis finds that people do this multitude of things with many types of conversations.

Jones (2013) argues that people do think about the text they use to communicate on social media platforms. On the FDP (2015, 2018), people use text to converse about their everyday needs as they can have many issues or conditions to resolve. They can also talk about broader issues about their illness but hold specific positions on the topics. They must find ways to communicate with numerous people in order to gain support.

1.3 Research questions

Petyko (2017) has studied another online phenomenon, i.e., trolls and political blogs with the use of corpus linguistics (CL). He wanted to determine the common linguistic patterns attributed to trolls. His approach can be useful for knowing about common linguistic patterns attributed to peers and support in the research. His ideas can be adapted to my thesis and the research questions. It is to be operationalised in specific ways concerning Halliday's (1985, 1994) SFL framework. It is, for instance, advice with stance-taking in the methodology and analytical framework (see Chapter 3).

Another good example to guide the research is given by that of Miguel et al. (2017). They have used appraisal theory in their research. It is to study a corpus of emotive and evaluative language. It is to identify the hidden ideologies reflected in the discursive features of their narratives. They have researched, similarly, by using SFL and the research of prominent linguistic researchers in their field (e.g., Martin and White, 2005; Fuoli and Hommerberg, 2015). Their research stance is similar to my research stance, in that '*all forms of speaking and writing are inherently emotional and evaluative*'. Evaluation is 'context-dependent at any level with a local co-text having a cultural context' Evaluative meaning is 'prosodic with discourse semantics and is related to meaning as text' (Martin and White, 2005, p. 9; Miguel et al., 2017).

The critical work of Partington (2008) on how to conduct CL research offers my thesis ways to do contextualization of the research questions. It helps to operationalise the questions. The text-data and discourse used in this study should give evidence to any claims made by the research. A general guiding question is: what is it that the research is looking for mostly? The research is thought to be exploratory. It is generally about how the authors/peers construct the discourse for support. It is for example, how they formulate advice-seeking and advice-giving together with stance-taking.

Partington (2008) explains that there are typical questions that can be answered by corpora. Where P is a discourse participant (or an institution), and G is a political goal, he looks at:

1) how P achieves G with language;

2) what this tells us about P;

3) in comparative studies, how P1 and P2 differ in their use of language. Does this say anything about their different principles and aims?

It involves the search for 'non-obvious meaning'. There would be little point in involving corpus techniques to uncover meanings which were readily available to traditional types of DA. Partington also refers to Sinclair's (2004, p.185) similar argument: '*we should not get caught in using corpora just to tell you more about what you already know*'.

Meaning in a language is theorised as always and simultaneously *ideational, interpersonal, and textual* (Halliday, 1985, 1994; Martin, 1992; Halliday and Matthiessen, 2004). They argue in an *ideational* manner, that language is about something. In essence, it functions to construe different or similar kinds of experiences. They have proposed that *interpersonally*, language negotiates social relations; it enacts roles and functions to share values. *Textually*, language presents the flow of information, and it distributes meanings in ways that are functional to modes of interaction. They have established that language can be viewed from semantic perspectives. It can be as a semiotic system, which represents the full *meaning potential* available to speakers or writers.

My thesis uses what Hoffmann and Bublitz (2018) refer to as the most comprehensive theorised model of evaluative language. It has been used across a wide range of contexts, and by researchers from both qualitative and corpus-based traditions: namely, the appraisal framework (Martin and White, 2005). This framework was developed within SFL to account for how evaluative language functions within situational and cultural contexts. Other prominent social media scholars, such as Zappavigna (2011, 2012) use SFL as an approach. My thesis can deal with the shortcomings of understanding how people create meaning in their online interactions, if at all. In discussions of voice and stance, the research can primarily be involved in interpersonal meaning.

In realising 'APPRAISAL', my thesis focuses on exploring realisations of interpersonal meaning at any or all strata of language in choices in lexico-grammar (Halliday, 1985, 1994; Halliday and Matthiessen, 1999) and discourse semantics (Martin and Rose, 2007; Martin, 1992). The explanations and descriptions in this chapter are primarily focused on interpersonal meaning in discourse semantics. This area of more critical theory is referred to as the system of APPRAISAL (Martin and White, 2005).

Online healthcare communications research has suggested that online healthcare support writing is a highly social practice. It is one that needs the careful positioning of claims within communities of intelligent people against a background of prior views and voices (Tannen and Trester, 2013; Sidnell and Stivers, 2014). Linguists working within functional perspectives, furthermore, have shown how this positioning work is accomplished in texts. It is through specific linguistic resources for expressing attitude and affects. It is tuning epistemic commitment up and down, and opening up discursive space for the reader (Hyland, 2005; Hood, 2004, 2007, 2010, 2011). In contrast to this growing body of research, it is less well understood whether peer writing is more highly regarded when it projects these stances and reader-positioning meanings in nuanced ways. That is when other peer-readers evaluate peer writing, for example, as displaying or showing a lack of uncertainty or sentiment about advice-seeking or advice-giving. How, if at all, are these judgements related to linguistic expressions of stance in posts by peers?

It is also worthwhile defining the notion of social media, as used in my thesis. Likewise, to the other notions, as mentioned above, the development of other terms in my thesis is with no claim to their completeness. Hoffmann and Bublitz's (2018) definition of social media is broad. It 'refers to (the totality of) digitally mediated and internet-based platforms which are interactively used by individual and collective participants. It is to exchange, share and edit self-and-other-generated textual and audio-visual messages'.

If the above research is considered then when peers/writers construct their diverse posts, they may create meaning, with a purpose and content for the discourse. These purposes may be seen through what the research calls (and of primary importance in my thesis) '*device-enabled discourse purpose categories*. The content can be obtained from what the research calls '*device-enabled discourse content categories*'. These devices may be thought of as analogous to communication acts that use high-frequency phrases, idioms or clichés. However, the devices have importance placed on linguistic forms such as 'advice' to help to categorise the discourse. The devices are used for particular purposes and context-building of a target domain and audience. People partaking in the discourse can also use entities/targets in the conversations that are about the domain, i.e., diabetes and medication. My thesis can focus mainly on domain-specific targets of, for example, diabetes, blood, and pumps from many others ranging from targets of usernames, places, time, to years. It is also crucial, therefore, to consider what forms of evaluation are present. What are the stances taken toward the entities or targets in consecutive and non-consecutive posts across the entirety of the corpus? My research considers entities/targets and TOPICs in the sense of Halliday (1967).

The established research suggests that LDA from AI, machine learning, and topic-modelling (Blei et al., 2003) can be employed in the first instance to find latent patterns and trends in the data. The research analyses are done with LDA, with keeping stopwords to identify '*device-enabled discourse purpose categories*'. This step is crucial: as the purpose of the discourse may help to identify salient patterns such as advice and stance-taking. The research analyses are done with LDA with removed stopwords to identify '*device-enabled discourse content categories*'. This approach helps to give a wider context to the purposes of the discourse. These devices are referred to as 'topics' and 'TOPICs' in the thesis and are explained further in Chapter 2 (Literature Review). They are a critical way of interpreting LDA topic, e.g., trigram patterns formulated by the model for 218,068 posts, as proposed by this thesis, but may help to understand online discourse.

Higgins (2012) has employed similar research, that is proposed by aspects of my thesis, on language attitudes, including stance-taking. An overall finding of his data is that when the participants expressed stances about their language use, they articulated the lowest degrees of affect, judgement, and appreciation. However, when people produce stance attributions for others, they tend to use more evaluative language. Evaluations were the most active when the speakers responded to stances produced by others. My thesis uses the linguistic theory of SFL as a way to understand the peer's support practice. People's attitudes about the information or opinions that they share in their support practice may be understood from their stance-taking towards the targets for that information or opinion.

A critical understanding and usage of artificial intelligence topic-modelling, automated content analysis, the annotation for entity recognition, sentiment analysis, dictionaries, and DA may help with the overall analysis.

The analysis of an open social media organisation's chronic illness support page may inform the research questions.

A combination of quantitative and qualitative approaches was used in the data analysis. For example, LDA and SFL usage for data analysis, as in the manner of Louvigné and Rubens (2016), can inform the research.

My thesis uses a database of FDP (2015) posts. It is made up of text that is collected into a corpus for the research. The corpus has Diabetes UK, the one-owner person referred to in this thesis as an organisation, owner or duk or expert team, and they posted 8,317 posts or 4% of the total. There are 16,137 peers, and they posted 218,068 posts or 96% of the total. Therefore, how peers engage in support constructs with each other is significant. Data for this study were collected using R programming to access the FDP posts and to download the data. Facebook provides a token to access their data, and their permission was obtained together with meeting UREC requirements about ethical research. The posts are from peers who are expressing diverse support needs. Peers expressed a need for support together in an open diabetes organisation's Facebook page.

The research considers that people can, for example, use 'topics', e.g., those related to advice-seeking, advice-giving with stance-taking behaviours, as ways to go about having conversations with each other. They attempt to co-construct meaningful support through their use of language in very particular ways.

The purpose of a primarily LDA-based model, therefore, is to find the distinct latent 'topics' trends or related purposes and TOPICS or content categories of support. The LDA algorithm is based on the estimation of topic distribution over the dataset (Section 2.5).

The different focus of my research is on the users and their potential utilisation of language patterns for support across the entire corpus. The specific questions which drive the research are:

RQ1: What attitudes, opinions, and sentiments are people expressing, about their conditions and issues in Facebook Diabetes UK posts?

RQ2: How do people express their attitudes, opinions, and sentiments about their conditions and issues in Facebook Diabetes UK posts?

The operationalised questions are:

OQ 1.1: What are the discourse purposes? What are the particular word trigram choices used by people, often together, to express what the discourse purposes are in their posts on social media for supporting people with chronic illness?

This question considers, for example, the advice, events, humour/sarcasm, questioning, emotion, hope, and charity in a post, topic and discourses. They may contain stance-taker and stance-taking resources that have been drawn upon but to what discourse purposes, or, as the thesis calls them: for what 'topics'.

OQ 1.2: About what is the discourse? What frequently used content word trigrams relate to the discourse contents?

This question considers, for example, the advice, events, humour/sarcasm, questioning, emotion, hope, and charity in a post, topic and discourses. They may contain stance-taking, and related content words for support concerning other posts, or, as the thesis calls them: for what 'TOPICs'.

OQ 1.3: What is its primary target? What nouns/entities is discussed in these posts?

This question considers the stance-taking and related target (noun/entity) object being discussed ranging from diabetes, advice, events, humour/sarcasm, questioning, emotion, hope, to charity, across consecutive and non-consecutive posts.

OQ 2.1: What is the poster's stance about the certainty of their information? How certain is the person in these posts and do people express these directly to each other in consecutive or in non-consecutive posts?

This question considers the epistemic stance (certainty) in the giving of information in the speech acts of, for example, advice-seeking and advice-giving in consecutive and non-consecutive posts.

OQ 2.2: What is the poster's stance concerning their feelings about the information and how do they feel in these posts and do people express these directly to each other in consecutive or in non-consecutive posts?

This question considers the affective stance (attitude) in consecutive and non-consecutive posts.

1.4 Answering the research questions

Analysing patterns of potential behaviour on social media can help to answer the research questions. It includes a study of a corpus consisting of 218,068 FDP (2015) posts. These have been posted by 16,137 users/peers and the Diabetes UK charity. Together, these posts provide a corpus of text that is useful for answering the research questions.

The research questions are operationalised later in the methodology and analytical framework chapter and provide ways to go about answering the questions. A summary of the novel DA method for answering the research questions mentioned in section 1.3 is outlined as follows.

1. The vital factors involved in collecting a corpus of diabetes posts are initially considered. The difficulties involved in doing so on a large-scale corpus are made clear with an awareness of their limitations. The crucial ethical implications are

stated and followed with state-of-the-art ethical processes for online research. The novel DA method begins with a focus on the research questions. The corpus, therefore, needs to be suitable for answering them. It contains about 96% peer posting of the total post, in a top social media, in a Facebook Diabetes UK support group page.

2. The corpus needs to be annotated. Since it is a large corpus of 218,068 posts, 6,960,998 tokens and 64,904-word types, automatic software procedures are preferable to using purely manual annotation. The annotation is also done in cycles. Fuoli (2018) describes the methodological rigour in annotating data for APPRAISAL. These procedures are used to determine salient features in the corpus, which are further explained in points 3–9.
3. Artificial Intelligence LDA topic-modelling is predominantly used to automatically identify the high-frequency top topic trigram patterns and features of the large corpus of 218,068 posts and 6,960,998 tokens and 64,904-word types a Facebook Diabetes UK corpus data (FDUK). All posts, both consecutive and non-consecutive, are analysed. The results are critically considered so that the quantitative analysis of patterns needs to fit within a Linguistic qualitative analysis. The AI machine-learning results need to be looked into and not merely to be gathered. LDA helps to identify the top 'topics' with their top trigram words. A novel approach in the research is to include stopwords at this stage of analysis. The resulting LDA topic model is suggested to have *discourse purpose words.* My thesis calls this '*device-enabled discourse purpose categories*' or 'topics'. The posts that were clustered into these 'topics' of high-frequency totalled 92,254. The remaining 133,106 posts were not clustered into these topics of high frequency. It suggests that they may be more crucial and are used across the corpus by

participants to construct support. The devices are possible purposes of and resources for support constructs.

4. Typical AI LDA topic-modelling is also used, and for identifying high-frequency top topic trigram patterns and the stopwords are removed in the analysis. State-of-the-art research suggests that this approach ultimately provides 'content' related words. However, this thesis is critical of this approach as it defines the *discourse content words* as '*device-enabled discourse content categories*' or TOPICs. This approach within the novel DA method provides the familiar category-related LDA topic model of the discourse. It covers the salient content words (*discourse content words*) used across the corpus by participants. They represent possible content and objects in the support constructs.
5. The process involves a research choice in line with artificial intelligence text analysis processes of bringing the large corpus down in size as the analyses process to one of the potential high-frequency patterns. After the LDA analysis shown above and the creation of the 2 LDA Probabilistic models, LDA '*device-enabled discourse purpose categories*' or topic numbers 0, 24, 34, 13 and 33 are selected randomly for further analysis and gives 92,254 posts. The ones with consecutive post numbers are then selected and put into a table (see Appendix 7-1) for further analysis. It results in 73 posts and their substantial quantity of text, and consecutive posts numbers. Thus, a smaller analysis sample will ultimately have consecutive posts and non-consecutive posts that could exhibit similar behaviours.
6. The decision to study the predominant and salient features meant using conservative 500 topic features. The posts that were thus clustered into these 'topics' of high-frequency totalled 92,254. The remaining 133,106 posts were not clustered into these topics of high frequency. It is essential to consider what is left

out of the analysis and what is included. LDA models are approached critically in my thesis.

7. These randomly selected topic numbers in point 5 also have the LDA '*device-enabled discourse content categories*' TOPICs assigned to each post from the above analysis. They are also put in the table next to each post identified in point 5.
8. MeaningCloud (2017) automated software is used to find the nouns/entities (targets), and this is assigned to the posts from point 5 in the table. The thesis focuses mainly on domain-specific targets of diabetes, blood, pumps and medication from many others ranging from targets of usernames, places, time, to years. It results in a sample of 73 consecutive posts with these targets from the LDA random five topics and 92, 254 posts. So, a new sample of posts for analysis will be about the target and contain ultimately potentially similar posts that are both consecutive and non-consecutive posts. The idea is to consider any potential high-frequency patterns for analysis but for posts that may have people/peers in a 'conversation' or as closely related posts to each other. It is in keeping with the AI analysis of all posts for patterns.
9. The MeaningCloud automated software is used to find global positive and negative sentiments for each post. It is assigned to the post from point 5 in the table.
10. A Linguistic Inquiry and Word Count (LIWC) (2016) is used to find values for healthcare, positive emotion, negative emotion, and certainty. The values are assigned to the posts from point 5 in the table.
11. Linguistic forms ranging from Advice, Humour/sarcasm to Questioning categories are identified by a manual comparison to labels in the well-established literature and assigned manually in each of the posts from point 5. This analysis is done in

comparison to state-of-the-art-research. LDA topics and the posts in the table are used to help make sense of each other in the analysis. Other discourses and ways that peers interact are also identified.

12. The analyses, as mentioned earlier, are then integrated to help identify the stance for each advice-related post in the table. Crucially the stance can relate to a noun/entity (target). The '*device-enabled discourse purpose category*' and the '*device-enabled discourse content categories*' are also shown to be resources for advice with stance-taking. There are many other linguistic forms such as humour/sarcasm or questioning or hope or raising charity funds. They are also shown to be related to the entity of healthcare. The stance is shown to be made up of an affective stance which can be related to a positive or negative emotion. They are shown to have a global positive or a Negative Sentiment value in the posts. The stance is shown to be made up of an epistemic stance. It can be shown to be related to Certainty values in the posts.
13. The above steps (1–11) are then integrated into a complete analysis of the discourse. The novel DA method shows sophisticated linguistic features such as advice with stance-taking. It is essential, for example, in the analysis to concentrate on who is involved, the resources, 'topics' (purposes), TOPICs (content), objects (targets), relationships (peer-to-peer), and the history of the stance.
14. The above analyses help to show patterns of support. They are evaluated further with DA, which is informed by SFL appraisal analysis. For example, they can suggest elements of solidarity and risk mitigation in the giving and receiving of advice.

15. Precision testing is used to help validate the results of the LDA. A table is created to offer comparisons of the LDA, MeaningCloud and LIWC results. It is in order to see if the integrated analyses of different individual procedures make sense.

1.5 Contributions

My AI and corpus-assisted study contribute to the academic study of two under-researched aspects of support in social media discourse, on a theory of support and, on AI and linguistic analysis of large-scale data.

1. A novel theory of support is proposed that places importance on high-frequency patterns of device-enabled discourse purpose (vose) and content (nent) categories. The theory focuses on linguistic forms for, example, advice with stance-taking in peer support on Facebook Diabetes discourses. They can exist amongst a multitude of linguistics forms ranging from humour/sarcasm or questioning or hope to charity or many different peer interactional forms.
2. A novel DA method is proposed concerning the analysis of large-scale corpora for high-frequency support patterns attributed to the peers. It is predominately carried out with artificial intelligence and automated software analysis. It is guided by SFL APPRAISAL analysis. The method would require more substantial uptake in the DA field to become of significance in DA potentially.

Of primary importance in my thesis are '*device-enabled discourse purpose categories*' *and* '*device-enabled discourse content categories*'.

The benefits are for corpus linguists and healthcare communicators seeking to understand support in large-scale diabetes online social media support.

Firstly, the thesis contribution will add to the field of CL. It is by showing how a large-scale corpus can be collected and then analysed for high-frequency patterns, and automatically 'annotated' with an example of the FDUK Facebook (2015). The use of voses and nents from an LDA model of analysis to understand the context of the diabetes corpus is novel in the DA. My thesis is about collecting and analysing data from Diabetes UK's Facebook page Facebook (2015). The investigation seeks to utilise current AI ways of collecting and analysing a large-scale number of posts on the FDP. It is about confronting the limitations and raising awareness of a way forward for AU and Linguistic analyses. The vast amounts of unstructured text and linguistic features make data analysis a challenging task. The goal is to combine primarily AI (Blei, 2012; Lee and Seung, 1999) approach to DA (Thornbury, 2010) to identify salient features in the text.

Secondly, another contribution is the development of a novel theory of support. It is by attending to the high-frequency sophisticated support patterns in the use of, for example, primarily voses and nents and targets, emphasising advice with stance-taking, and appraisal resources. The theory points to the broader use of linguistic forms and interactional types but focuses on primarily high-frequency types. It is to pay attention to more global patterns of the textual organisation by which for instance appraisal values are sequenced and made to interact in order to effect rhetorical outcomes. LIFT Psychology argues that people can feel psychologically vulnerable to these challenges in their lives NHS (2017). So, finding, for example, the sentiment in a post can be useful to gain an understanding of the broader emotional context of the support. The theory would need future refinement from its broad approach but is falsifiable. My thesis claims that *support in social media conversations is the result of language-based high-frequency patterns of device-enabled discourse categories of purpose 'voses' and content 'nents'. It contains linguistic forms such as advice with stance-taking. It contains many other interactional activities amongst peers carried* *out by people in their meaningful shared interactions that influence real outcomes about people's concerns and issues. There is remarkably a high-frequency stance-taking with diverse affective stance and low epistemic stance but not necessarily in consecutive posts but rather indirectly in usage across many non-consecutive posts about the same target. People can post support at any time and not always in a linear-consecutive order. The theory may be falsified if counterexamples of such discourses in support become available.*

The aim is for a novel theory of support constructs in social media discourse.

1.6 Thesis structure and objectives

The chapters are summarised below.

- Chapter 2 – The Literature Review and Conceptual Overview. This chapter presents a review of the relevant literature of the primary research about healthcare communications, support, advice, SFL, and the meaning of a text is crucial. It contains a cursory look into humour/sarcasm or questioning or hope or charity. Of primary importance are AI, automated content analysis and annotation for entity recognition, sentiment and dictionaries. The critical linguistic research on language patterns about support in social media is also reviewed. The methodological approach of using CL with machine learning and automated content analysis and annotation for entity recognition and sentiment and dictionaries is carefully considered, together with DA via the best available practices.
- Chapter 3 – The Methodology and Analytical Frameworks. The research questions are refined and operationalised for the analysis of the FDUK (2015). The crucial understanding of the limitations of an AI and linguistics analysis approach is confronted. The importance of the ethics processes throughout the research is

shown together with a particular process as per the OBU Ethics Committee and critical research, and the General Data Protection Regulation (GDPR). The data analyses are shown along with the use of CL with different tools such as LDA machine learning, automated content analysis, the annotation for entity recognition and sentiment, and the DA approach.

The MeaningCloud and LIWC results are used in the mixed-approach analysis. The qualitative analyses are carried out on the quantitative results. The quantitative results are obtained automatically by the software and aid the manual identification of advice, humour/sarcasm, questioning, hope, and charity features. There are comparisons of results to the state-of-the-art research in the diabetes domain. LDA precision testing is used for validation. A descriptive statistic of the corpus is also presented.

- Chapter 4 – Data Analysis and Findings. This chapter describes the data analysis and findings and validation. It addresses each of the research questions in turn by answering the research questions about support patterns of the users. Machine learning with LDA, automated content analysis, the annotation for entity recognition and sentiment and DA are used to deal with the significant social media data. LDA is used as a new way of keeping the stopwords and by removing stopwords to produce high frequency '*device-enabled discourse purpose categories*' and '*device-enabled discourse content categories*', respectively. How and what people are conversing about in their support of themselves and each other is also of importance. Advice with stance-taking, humour/sarcasm, questioning, hope, and charity features are identified in the corpus. For a similar stance object in a post, people can mostly have different types of affective and similar types of epistemic stances. They can employ some particular '*device-enabled discourse* *purpose categories*' or resources and context of support together with '*device-enabled discourse content categories*' or resources and content of support. The qualitative analyses are carried out on the quantitative results.

- Chapter 5 – A Proposed Theory of Support in Social Media Discourse is proposed. It will be useful for understanding peer support. The theory is critically evaluated with an awareness of its limitations.
- Chapter 6 – The Conclusion evaluates the findings in terms of answering the research questions. It points out the relevance and contributions of this thesis. An exploration is made about the impact of the research on healthcare communication, and future platform development with AI support-bots.

2 LITERATURE REVIEW AND CONCEPTUAL OVERVIEW

2.1 Introduction

This chapter reviews the state-of-the-art studies on the high-frequency topics and topic-clustering analysis with artificial intelligence (AI) machine learning and topic-modelling in automating large-scale data analysis. It investigates automated content analysis, the annotation for entity recognition and sentiment for advice, stance-taking, with a cursory look into humour/sarcasm, questioning, hope, and charity and dictionaries. It reviews the well-established fields of discourse analyses (DA), applied linguistics for health communication, support, advice, sentiment, stance-taking, SFL, and meaning concerning support on social media platforms.

A seminal study in this area is the work of Harvey and Koteyko (2013, p.185). They suggest that peer support is connected to advising, stance-taking and meaning (this is explained in detail in the next section). They use the example '*…instead of briefly outlining her position on the topic and giving direct advice through the use of imperatives……. participant 15 chooses to state her views indirectly and takes time to elaborate on her background and experiences…creating common ground between herself and the advice-seeker*'. Another excellent example of a linguistic approach can be found in an earlier study by Pennebaker and Davison (1997). They looked at the use of particular words in similar types of online discourse. More recently, Sidnell and Stiver's (2014) study on social media and healthcare communications has highlighted online healthcare and DA studies in particular.

Computational analyses (Blei et al., 2003) can offer a way to analysis high frequencies in linguistic data. It is for some timely and online-generated large-scale data: of 218,068, anonymised Facebook Diabetes UK (2015) posts and 16,137 diabetes-related users. AI is a sophisticated approach that includes latent Dirichlet allocation (LDA) topic modelling, automated content analysis and annotation for entity recognition. The humanities fields, where questions about texts are essential, is an ideal place to test AI topic-modelling and can benefit from interdisciplinary collaborations from AI and linguistics and DA.

2.2 Online peer support and advice

Lehmann (1987) and (Crystal, 1997) consider rightfully that the grammar of a language and the architecture of discourse are problematical. Together with the continual work of other linguists in this chapter, the thesis explores concepts ranging from sentences, topics, discourse grammar to the sociolinguistic view emphasising the purpose of discourse. There are potentially many other interactional activities amongst peers in online support. However, my thesis focuses on linguistic forms such as the advice in order to place potential discourse categories in the context of any potential high-frequency support pattern. However, the field of advice is not easily understood. In Locher and Limberg's (2012) discussion on advice, they criticise how some authors have understood support practices. For example, they argue for the firmly embedded linguistic form of advice-giving in the study of pragmatics - a study of language in use. Previous research by Verschueren (2009) has established that '*pragmatics is a general functional perspective on any aspect of language, i.e. as an approach to language which takes into account the full complexity of its cognitive, social, and cultural or meaningful functioning in the lives of human beings*'. Locher and Limberg (2012) have demonstrated that an understanding of this definition should be as broad as possible. In that, it does not preclude a reliable choice of methodology nor a set of questions to investigate language use.

It is now very well established from a variety of studies that advice – a seemingly straightforward and everyday act – is significant in people's lives. It can be revealed to be complicated in scope and has versatility and importance for day-to-day living (Locher and Limberg, 2012).

Gumperz (1992) also suggests shortcomings in studies on advice. The lens 'is not only on the speech act of rendering advice but, in an attempt to grasp the practice of advice-seeking and advice-giving more globally, the speech activity is crucial. So, from an advantage of the studies mentioned above, this thesis can show how the interactants orient themselves towards advice-seeking and advice-giving. The data is to explore for the constituents of the speech events under analysis.

2.2.1 Support content and purpose

A frequent criticism of much of the research on conversations concerns a general lack of a solid linguistics definition of '*what people utter, talk, or converse about*'. Researchers have not treated the notion of topics in conversations in much detail. Reinhart (1981) has proposed that *what people converse about* is a 'topic'. Her definition is useful to help examine the results gained from data analysis as regards this notion of 'topic'. A broader understanding of what the topic of a given sentence is can be determined. It is by its context of utterance (what is it about?) and by its linguistic structure (what are the language features?)'.

She uses this combination of understanding of a topic to build a notion of pragmatic about-ness. It connects a sentence to the notion of a topic. It is part of a more substantial way of understanding pragmatics, sentences and topics (Grice, 1975).

There are limits to taking for the concepts used in the research. The use of terminology from linguistics shows how challenging everyday notions of language are. Any analysis of support and conversations employed it carefully.

This thesis agrees with Reinhart's (1981) way of thinking about a topic to help counteract the tendency shown above. However, it is too broad for the needs of this thesis.

The established AI and machine-learning LDA (Blei et al., 2003) is used to find topics of high-frequency across a corpus. It is still not easy to say what precisely a topic is. What is it that LDA data analyses result mean even though Blei calls it a topic? The findings need to be further interpreted. From computational linguistics, Blei's (2012) definition of a 'topic' can help in how to find a 'topic' in the many social media posts. He utilises a quantitative stance and tries to offer a more precise definition of a topic, which he regards as high frequencies of words or social media posts. It is 'a particular distribution over a fixed vocabulary'. He gives examples of the genetics topic, which has words about genetics with high probability versus the evolutionary biology topic, which has words about evolutionary biology with high probability. However, he uses the term 'topic', because terms that frequently occur together tend to be about the same subject.

What is the relationship to the TOPIC or target of a post in the sense of Halliday (1967)? I use a TOPIC as content and target as an entity concerning contents in a post. It is different from the 'topic' in the review mentioned above. How should we visualise and navigate the structure in these LDA 'topics'? What do the topics and document representations tell the researcher about the texts? The humanities fields, where questions about texts are essential, is an ideal place to test topic-modelling and can benefit from interdisciplinary collaborations with computer scientists and statisticians. A new understanding can fill this gap by combining AI and DA and using it to analyse social media text data.

Therefore, for Blei (2012), the assumption is that LDA can find possibly latent words or n-grams. These topics, e.g., high-frequency trigrams that possibly generated the corpus and posts. N-grams share terms frequently occurring together. An n-gram can be a pattern of a certain number of words that mostly occur together across a discourse (Petyko, 2017). Blei considers n-grams in helping to think about posts, sentences, and topics of potentially large-scale corpora. Calculating high frequencies of topics in large numbers of sentences is difficult and analysing n-grams can shorten this computational complexity. An analysis of trigrams in a large-scale corpus may be more manageable. However, it would provide more information about the corpus than only looking at high-frequency words.

In considering the issues as mentioned above, the research needs to think about the LDA model findings. Does it help identify a way of organising the shared terms that are frequently occurring together into a TOPIC? By a macro organisation? It needs to describe the content of the LDA 'topics'. May it be something different from the prevailing notions of the topic? The notion of LDA's 'topic' or Halliday's (1967) 'topic' is not straightforward for my thesis.

The notion of conversation, though ubiquitous, can be more challenging to define. However, Reinhart (1981) establishes 'conversation' to be a way of mediating and building relations with others. It can be a way for people to represent or define themselves. To the people having daily 'ordinary' conversations, a description of a conversation may be unnecessary. It could be difficult to contemplate this aspect while conversing. For the research, the question of what precisely constitutes a conversation is an important issue when considering posts, topics and online support.

If individuals do think about what a conversation is, it may be in generically vague terms. It may be assumed to be as any other form of spoken discourse (Goffman, 1981). There could be other helpful norms for defining the conversation. Halliday (1978, p.140) refers to conversations as *ordinary everyday interactions*. Wilson (1989) criticises this definition for its lack of specificity, believing that it needs to be made more transparent. Even though regarding conversation as an unremarkable phenomenon, it is both a widespread and commonplace occurrence. The literature suggests that the notion of conversation may gain a more precise definition when considered in exceptional cases, potentially in posts, topics and online support.

2.2.2 Importance of advice and questioning patterns

Sillence (2013) suggests that the employment of advice strategy while aiming to get a response is to its degree of honesty and clearness. He argues that the level of directness needs examination because of the tension that exists between showing support and appearing to force help on people. The research can benefit from an understanding of social relations based on power or solidarity.

Goldsmith's (2000) typology of *advice asking* can make the thesis study more accurate and comprehensive. He shows that there are different themes that people use to call for advice. He considers the asking for advice to be also related to asking questions for advice. Sillence (2013) uses and adds a fifth advice-guidance pattern (same boat no.5, e.g., '*is anyone in the same boat as me?*') to Goldsmith's typology as shown in Table 2-1.

Table 2-1: Types of advice solicitation: Goldsmith's (2000) typology

No.	Advice pattern	Description frequency	My thesis advice-asking labelling keys
1	Request for advice	Explicit solicitation of advice using the following phrases: (a) '*I need your advice*'; (b) '*What should I do?*'; and (c) '*Should I do X?*'	AA_R
2	Request for opinion or information	Questions such as '*What do you think?*' or '*What do you think of X?*' that can often generate advice responses even though they may be ambiguous about whether the poster wants to solve a problem or obtain emotional support.	AA_{OI}
3	Problem disclosure	Also, potentially ambiguous, as it can be interpreted as a *request for advice, sympathy, or solidarity*.	AA_P
4	The announcement of a plan of action	The poster may receive advice after announcing their intentions.	AA_A
5	Anyone in the same boat? Sillence (2013)	The poster asks specifically to hear from anyone in the same boat as themselves or those who are going through the same experience.	AA_B

Sillence (2013) argues that the way of giving advice can be vital if the giver must succeed in passing on his or her way of thinking to others. He shows that advice-giving can be related to the directness of producing their answers.

There are several developed taxonomies for advice, and different methods to classify advice. It is unclear whether general findings are applicable to current social media online support.

Crystal (2001, p.81) uses '*It wouldn't hurt if you had your child evaluated*', as an example of advice to mothers in responses to vignettes from a US teaching context. Advice-givers can frequently attempt to establish a sense of rapport or solidarity with the advice-seeker (DeCapua and Dunham, 1993; DeCapua and Huber, 1993, 1995; Locher, 2006; Morrow, 2012). Such a rapport expressed through statements of empathy, such as '*I understand your frustration*'. The rapport established by acknowledging the feelings of the advice-seeker with, for example, '*I know it is frustrating to deal with such behaviour.*'

Kouper (2010) focuses on showing the patterns that can help to identify *advice-giving*. He argues that people searching for or giving advice must make choices about how they exchange advice and how this affects the beliefs of the community. He suggests the development of social relations through the giving and taking of advice. There are different classes of advice reported by Kouper (2010) and discussed by Sillence (2013). The different classes are shown below in Table 2-2.

Table 2-2: Levels of the directness of advice: Kouper's (2010) categories

No.	**Type of advice**	**Description**	**Advice giving labelling keys**
1	Direct advice	Any comment that included imperatives or the modal verb '*should.*'	AG_D
2	Hedged advice	Any comment that contained explicit hedges or hedging devices, e.g., '*I think*', '*It seems*', or '*Why don't you?*'	AG_H
3	Indirect advice	Any comment that lacked explicit or hedged advice but had enough information to act upon it, for example, '*Here's one possibility*' or '*There are some options.*'	AG_I
4	Description of personal experience	An account of how the person dealt with the situation the advice-seeker had described.	AG_E

People may need to prove legitimacy in order to get a response to advice-seeking or advice-giving, and they may need to convey authority to give advice. However, in doing so, they must overcome the danger of being challenged online (Harvey and Koteyko, 2013; Tannen and Trester, 2013; Sidnell and Stivers, 2014). Researchers have explained how people influence each other to promote their views (Dwyer, 2007; Park et al., 2011). They suggest that the individuals in support networks need to work through the risk of asking for, or employing, any advice given to them.

Considering trust issues in advice, people may be critical in their responses. People also need to trust the sensibility and trust that the replies given are correct. People may need to employ a core of advice strategies to convince the other users but to expect still to be challenged on their posts.

It is not always easy to show which posts are questions or answers in-social-media. Miles and Winget's (2010) study explicitly targets finding questions in a microblogging type of social media. They utilised some rules that gave up generous samples from social media. For example, a tweet from Twitter can be a question if it matches some types of patterns. It has a question mark that is not part of an URL.

2.3 Discourse analysis of online support, e.g., advice, humour/sarcasm, questioning, hope, and charity

Lupton (1992) has defined DA as being composed of two main dimensions: textual and contextual. Textual dimensions can account for the structures of discourse. In contrast, contextual dimensions relate the structural descriptions to various properties of the social, political or cultural context in which they take place. Harvey and Koteyko (2013, pp. 165–187) argue that textual is more concerned with such microelements of discourse as the use of grammar and rhetorical devices. For example, metaphor, syntax, sound forms, the plain meaning and content matter of words, and sentences of text or talk.

There are essential strands of research that have explored how people communicate via computers and the internet, such as computer-mediated communication (CMC). Theories from social psychology can have their limitations in explaining CMC. Sproull and Kiesler's (1986) article provides a contrary insight into online support. It is with their reduced social cues approach. They contend that CMC can give participants relative anonymity. People probably engage with each other in uninhibited antisocial behaviour.

There is also the social identity de-individuation model of Spears and Lea (1994). It is about norms and social perspectives of online behaviour and which can be valuable for the understanding of online support.

Van Dijk (1990b) has used multidisciplinary DA to examine communication processes in their social, political, and cultural dimensions, with a particular interest in *theory formation*. It can be a limitation to focus mostly on linguistic features (Van Dijk, 1990a; Lupton,1992).

Irwin (1989) describes the agenda for health communication as including action to apply more interpretive models of communication processes to research. The agenda exists to supplement or replace quantitative methodologies with more humanistic, qualitative methodologies, bringing the field in line with developments in mass communication.

Moreover, it is difficult to separate the notions of, e.g., 'topics', 'TOPICS', and 'conversation' from particular social practices, discourses and domains. Van Leeuwen (2008) suggests that social practices are 'socially accepted' models of how people should conduct social activities. They are observable by identifying the social actors or selected participants within a discourse. Therefore, these ideas applied to the support conversations and its discourse as a social practice may see the utilisation of many social actors in action. It is to recontextualise meaning by way of the deployment of representations through self-expression for a diabetes discourse. It can, in a much wider context, signify the perceived cognitions of the intended 'hearers' or audience, i.e., other peers.

This thesis agrees with Harvey and Koteyko (2013, pp.165–187). They consider some of the theories mentioned above to be a negative view of communication. They can also be mechanistic in not allowing for the fluidity and dynamism of participant identities. People do have norms of online behaviour.

Androutsopoulos (2006) explains that the earlier literature on computer-mediated communication became too focused on prioritising technological means. It is as the sole explanatory factor for linguistic practices, which resulted in what he terms '*computer determinism*'. The research looks at a *complementary* impact of technology on language use. However, it is aware of the technological determinism that is implicit in much of the early work. The contributions to this theme extend a growing body of research, inspired by sociolinguistics and discourse analysts. They aim at demythologising the alleged homogeneity. They highlight the social diversity of language use in CMC.

In studying the actual uses of modern technologies with its large-scale data, the research can subscribe to a primarily social view of CMC (Lamrichs and Molder, 2003, p. 18). Furthermore, by taking a predominately-linguistic perspective, the research can look through a 'prism of language' and make the research judgement on observations of language use. Herring (2004) terms this perspective as computer-mediated DA. He draws on the study of spoken and written communication. These range from conversational analysis, pragmatics, discursive psychology, to critical DA. The research likewise cannot only look at separate statements by individual participants. It considers the interaction with other members of online groups. It can illuminate different actions towards which participants' written descriptions can orientate.

Antconc (Laurence, 2014) is usable in more extensive analyses for each word versus the keyness of that word. It can be compared to its use in other well-known corpora, for example, the British National Corpus (BNC). The BNC contains text ranging from the newspapers to the internet.

Schiffrin et al. (2003, pp. 197–214) looked at discourse studies and the theories of conversations and engaged with a DA that is interdisciplinary. They use a vital study of '*the apology*', a study in pragmatics or conversation analysis, and they consider it from a perspective of sociolinguistics. They argue that 'apology', more than most speech acts, places psychological burdens on its maker. It is less severe on its maker than on its recipient. There may be unambiguous forms but rarely so. This thesis will use their idea and argue throughout that 'support' is similarly complex and places psychological burdens on the peer.

Morrow (2012) finds that assessment and advice can make up the majority of discursive moves in a particular practice studied. The salience and importance of assessment moves are explained for his example site on divorce as being contributed to by peers who share personal experiences.

The main strength of Halliday's (1977, 1985, 1992, 2004) work is that he focuses on the relative frequencies of choices made in the use of language. It helps to explain the paths to be taken by participants given the available resources in a system of language. He views language as a system and as functional. He is concerned with '*what for*' i.e., how language in some particular social practices, namely systemic functional linguistics (SFL). SFL aligns with the functional tradition in linguistics manifested in the Prague school (Jakobson, 1971). It arose out of the linguistic school known as Firthian systematics (Firth, 1971).

One criticism of much of the literature on online support can be about the lack of investigating pragmatics and discourse semantics. These are essential to understanding why people behave with language use in a particular way. Halliday's (1985) interpersonal meta-function of language can help focus on tenor, which is about the speaker/writer persona, social distance, and relative social status. The speaker/writer persona can help to focus on the stance of a person in their expressions. They are in an attitude that is positive, neutral or negative in their use of language. In SFL, tenor refers explicitly to the participants in discourse and their relationships to each other and their *purposes*.

It is difficult to understand what is meant by the notion of 'online social support'. However, the research can illuminate areas such as advice, sentiment, and stance-taking. Posting on social media, where people use text to converse with each other, is more complicated than perhaps is recognised for 'support'. For example, researchers have looked at electronic health communication in general and peer-to-peer interactions and particularly about support practices (Sidnell and Stivers, 2014; Harvey and Koteyko, 2013, pp. 165–187; Tannen and Trester, 2013). They follow Mickelson's research (1997, p. 157) and argue that such interactions are typically a locus of social support. They define it in the words of Mickelson as a '*transaction of empathy and concern, information and advice, or tangible aid, i.e., goods and services, between two or more individuals*'. Verbal or nonverbal behaviours to seek or provide help. This definition of support is broad enough for the research. However, it can help to identify specific salient language features for a particular domain and a particular discourse.

2.4 Qualitative

High-frequency stance-taking with affective stance and epistemic stance may be available in consecutive and non-consecutive posts of peers supporting each other about similar targets. The question it raises is: do peers use the corpus similar to a conversation with responses to each other's turn-taking or not. The other question is, do they use the corpus as a way to 'talk' to many others and gain responses from many others? Is it a non-consecutive posting in the corpus about similar targets. Stance-taking and appraisal analysis from linguistics can help with the computational and artificial intelligence analysis. It can check if peers instead respond indirectly to each other across many posts about the same target.

Aikhenvald (2004) shows that an epistemic stance in languages such as English is not grammaticalised in the traditional sense. It can include elements such as the use of adverbs and particles: for example, '*supposedly*'. Included are extensions of other grammatical categories like mood and modality, such as '*should*' and '*if-then*' constructs. They may also include reported speech and complement clauses that express perception and cognition. Evaluation is through the appraisal system, an approach to evaluation based on SFL (Martin and White, 2005). The research can also focus on the appraisal subsystem of attitude, which helps the research to see how speakers attach an intersubjective value or assessment to participants and processes. Following the work of Martin and White (2005), the attitude expressed through affect, judgement, and appreciation sub-systems of attitude.

The SFL showed by Halliday (2004) inform appraisal theory. Epistemic stance-taking, affective stance-taking and appraisal are informed by the highly relevant literature (Biber and Finnigen, 1998; Biber et al., 1999). An epistemic stance can be related to the degree of certainty concerning the object of discussion. In contrast, an affective stance related to emotions, feelings and sentiment about the object of discussion.

2.4.1 A stance-taking framework for advice

Stance can derive from social actors with consequences (Du Bois, 2007). People's lives impacted by the stances they and others take. Du Bois established that in many cases, the current stance act 'resonates both formally and functionally with a stance taken in prior discourse'. It is important to note that *the value of any stance utterance tends to be shaped by its framing through the collaborative acts of co-participants in dialogic interaction*.

Fundamental research from Du Bois (2007) has established that at least three things needed in knowing about a given occasion of stance-taking. It is beyond what may be overtly present in the words and structures of the stance sentence itself: (1) Who is the stance-taker? (2) What is the object of stance? (3) What stance is the stance-taker responding to most?

Who is the stance-taker? In conversation, participants can place importance on who says what and monitor it accordingly (Du Bois, 2007). A participant can perform this monitoring by noting or transcribing an utterance or turn with a label indicating the identity of the speaker. In any utterance involving different speakers, the identity of the speaker indicated is at the beginning of each turn.

What is the object of stance can be crucial in making sense of any given stance. The other participants or observers may need to know who is speaking. They also need to know what others are speaking about (Du Bois, 2007). Emphasis placed, amongst other things, on knowing the referential object or target toward which stance. It helps in the process of identification of targets. It is applicable to the thesis for healthcare studies. It can be about what is claimed by a speaker to be important in the management of any healthcare condition.

Finally, what stance the stance-taker is responding to is vital to the study. It is for knowing the identity of the stance-taker and the object of stance, which are essential for understanding. Participants or observers may be on the uncertain ground until the knowledge of what the previous stance the current stance formulated a response to 'counter stance' (Du Bois, 2007). It is essential to know the originator of why any particular stance. It is why at any particular moment in time and why in some particular terms. He suggests that in answering these questions, it is necessary to monitor the dialogic and sequential shape of the ongoing exchange of stance and counter stance.

What is particularly important is a consideration of the alignment given previously (for example *I agree, I agree with you*). Du Bois argued that any particular sentence could be grammatically complete. Incomplete as stances, and so, therefore, they are pragmatically incomplete. ‘People do not agree in the abstract; they agree with someone about something’. He uses the example that while a sentence containing the phrase *agree with you* foregrounds or profiles the dimension of alignment almost exclusively. In terms of its interpretation, it must still indexically incorporate a prior stance content, including the relevant object of stance. He suggests that, in general, the relevant stance content will be locatable in the prior discourse. It can help to specify what specific stance people agreed.

My thesis has problematised the view of *online support, advice, stance-taking, certainty* and *attitudes* in online interaction. The chapter now moves on to further examine possible frameworks for studying social interaction in online support settings.

Researchers have not treated advice and stance-taking in much detail for SSMD. There are more recent arguments against a one-dimensional view on ‘support’, including a study about meaning. This view has been established by Roos (2003), who describes the meaning-making process that takes place in a community of practice. Wenger’s (1998, p. 45) concept of a community of practice is useful where communities develop over time through shared enterprise pursuits.

Harvey and Koteyko (2013, pp. 165–187) have adapted Wenger's 'community of practice' to explore the phenomena of online groups centred around discussions on health and illness-related topics.

Previous studies of support have not dealt with what Fernback (2007, p. 66) proposes and what the research calls a *stance.* It can be adopted. An internet-based discussion forum should be seen and analysed primarily as a '*mutable construct, determined by social actors who create meaning about it*'. People negotiate personal meaning in this online discursive community. It can also be about unique stance-taking, but also about support and norms.

Chiluwa and Ifukor (2015) adopt the appraisal framework and focus on writers' affect and judgement, which reflect in the attitude of those who, for example, tweeted about the #BringBackOurGirls campaign and posted comments on Facebook. Peers can also use the hashtag # to share common topics and TOPICS and targets to gain information and support this way. They may build a community that may achieve a particular aim for a particular social media discourse. These studies, however, do not entirely attempt to understand how the many participants in that particular online community create a purpose or context for that particular discourse.

Furthermore, an appraisal framework can be adopted from SFL (Halliday, 1985, 1994) and can focus on the social function of language. Text expressed it. It is not only as a means through which the speaker or writer expresses their feelings and takes particular stances but also when they engage with socially determined positions of value. They can thereby align or dis-align themselves with the social subjects who hold to these similar positions.

Furthermore, Berman et al. (2002, p. 1) argue that the term STANCE is used in the discourse literature in diverse ways. Higgins (2012) provides a conceptual framework for attitudes to stance-taking.

As Du Bois (2007, p. 163) explains, '*in taking a stance, the stance-taker 1) evaluates an object; 2) positions a subject (self and others), and 3) aligns with other subjects*'. Speakers take positions mainly by evaluating another person's stance and other objects, including language.

The scope of research on support, stance and advice have been relatively narrow, being primarily concerned with advice. However, Hyland (2005) thinks about advice and stance together and explains stance as an 'expression of a textual "voice" or a community recognised personality'. It is an attitudinal dimension and includes features that refer to the ways writers present themselves and convey their judgements, opinions, and commitments. 'It is the ways that writers intrude to stamp their authority onto their arguments or step back and disguise their involvement'.

Hyland (2005) argues that it is difficult to account for the relatively small number of suggestions in written feedback practices. He uses the example of teachers. If teachers were seeking to effect change less directly through the speech acts of '*praise*' and '*criticism*'. In this example, these acts are not advice, and their different illocutionary points mean they are typically understood differently.

The above sections and other research suggests that the outcome of an utterance and its pragmatic meaning developed. It can be by conducting a study of advice with stance-taking, but more needs to be done in studying this field. Appraisal systems can inform the interpretation of evaluation in the text (Martin and White, 2005, p. 210). Martin and White's type of analysis of attitude and engagement and graduation is helpful. It is to develop an account of how, by such interactions, text constructs a model of the putative addressee and position the author concerning that addressee.

Attitude and stance-taking may need to be further related to the field of online chronic illness SSMD. Barton and Lee (2013) show that expressing stance in online writing spaces are about people expressing their opinions and attitudes about something or someone. Martin and White (2005) have done a similar analysis of interpersonal, political texts.

Herring's (2004) classification of features that characterise an online group as a community includes identity, sociability, and support. Their research focuses on support, i.e., '*the demonstration of solidarity and reciprocity and following pre-existing and emerging social participation guidelines, members can then maintain the group*'. Stance-taking can be used to achieve these. Harvey and Koteyko (2013, pp. 165–187) show that people outline their positions. This idea of stance-taking can be applied in this thesis and then utilised in the research about understanding 'topics' and 'TOPIC'. It can include investigations about stance-taking: for example, stance-taking can be used together with particular types of advice. Such as direct advice using the imperatives '*do this…but not that…*'. These may be a potentially viable action for online participants in a non-face-to-face online setting.

A claim can be checked whether people can work together with each other online to achieve such aims. Stommel and Koole (2010) propose that such practical applications that are informed by DA can be used to help participants align with the expectations of a particular online community. Such a study may help to prevent social isolation of people dealing with chronic illness.

My argument follows that of Zappavigna (2012), in that posting on Facebook is 'rarely about presenting bald facts or narrating activity'. The interpersonal meaning is essential in understanding posting where people can take a stance. Evaluations in the online group can become a domain of interpersonal meaning which can happen when language is used to express attitudes and to adopt stances about other text. The studies above are part of a significant body of research on CL and evaluation. They can influence the direction of the thesis.

An understanding of emotional language can benefit from sentiment or opinion-mining studies (Whitelaw et al., 2005; Gruzd et al., 2011). Martin and White (2005) argue that emotional reaction develops earlier as people socialise into culture and institutions.

Feelings become institutionalised as ethics and morality, forming JUDGEMENT systems, which people construe as rules and regulations on behaviour. 'Feeling' is also institutionalised as aesthetics and value, forming appreciation systems with which people generate assessments based on each other's reactions to phenomena. Simply described, 'affect' is about expressing emotion, 'judgement' is about assessing behaviour, and 'appreciation' is about estimating value. This thesis can look at the appraisal system of Martin and White (2005). IT IS from the aspect of attitude which covers 'affect' and its polarity and in the manner of Zappavigna (2012, pp. 52–70) who uses several different types of tables in her analysis.

2.4.2 Objects of advice, events, humour/sarcasm, questioning, emotion, hope, and charity in support patterns

Some of the gaps in support from Section 2.2.2 may be filled by investigating linguistic forms of advice with stance-taking humour/sarcasm, events, questioning, emotion, hope, and charity. The thesis aims to combine the crucial researchers of advice work with that of stance-taking (Du Bois, 2007) and look into the many others.

This view is necessary for this thesis when considering healthcare support discourse. Hall (1997) argues that people's opinions, preferences and facts all play a role in the aim of their discourse. Hall views conversations as having the potential to define the way individual objects are represented in, for example, for thinking about, practising, and for studying around any social practice.

There is a need to consider the social relations of solidarity about risk and trust (Brown and Gilman, 1960; Alter, 2010). Many other aspects of social relations for healthcare and their interpretations have been questioned above. The researchers above may help inform the research with consideration about solidarity or friendship, which is about people's sense of closeness, and familiarity with each other.

Jones (2013) suggests that 'risk' in healthcare communication is not only about managing the risk of getting more severe conditions or declining health. It can be about the danger of giving the wrong type of support or advice. He considers that in any support practice, the risk of giving advice can be mitigated. It is done by, for instance, sharing the advice strategies with others that have more power (e.g., clinical experts).

Jones (2013) argues that trust in advice is more likely to be gained in social relations of solidarity. It can say more about why people use language with 'topics' and 'TOPICs' and targets and for instance, advice with stance-taking devices for '*the what*' of their conversations. It is in '*how they*' converse with each other. It can help this thesis to show thinking, focus on the possible language patterns in support, and find explanations for any discoverable patterns. Why are particular patterns meaningful on these platforms?

An explanation for any main patterns can be understood by utilising Suler's (2004) idea. People connect to communities based on their shared interests and that they express them on social media. Suler's (2004) disinhibition or boldness effect can help to explain why people trust each other enough to share their direct experiences in the first place. Zappavigna (2011) calls this pattern of sharing 'ambient affiliation', where the users of a platform perform these interactions publicly on social media.

People may employ mood-related words in their conversations. Fraser (1999) argues about pragmatics as being 'everything else' a sentence must contain. These include mood markers and can be, for instance, about the declarative structure of the sentence. It can also be about the varying length and complexity of lexical expressions.

People are known to receive help from face-to-face and private doctor-patient conversations in healthcare support practices (Hunt and Harvey, 2012). Suler (2004) argues that people can take advantage of non-face-to-face meetings. There may be fewer social restrictions and perceived limited fears of sharing personal information with others in a similar predicament. The non-face-to-face community can develop an understanding of their conditions and issues with particular linguistic devices, such as topics and conversations (Zappavigna, 2011). For her, people tend to develop 'normal' behaviours or values of the social media group during their regular interactions with them. It is worth focusing on trust and solidarity in an analysis of support as people can agree as to what is acceptable and what is not suitable for support.

Biber and Finegan (1989, p. 124) define stance as 'the lexical and grammatical expression of attitudes, feelings, judgements, or commitment concerning the propositional content of a message'. It can include adverbs, verbs, and adjectives, which mark affect, certainty, doubt, hedges, emphasis, possibility, necessity, and prediction. Ihara (2006) argues that in any communication encounter, speakers or writers may communicate information in words. They also convey their attitudes, emotions, feelings, moods or dispositions. Non-referential information, such as mood and feeling, is as vital as referential information. It is because together, they enable the hearer or reader to interpret the message correctly. It is to find the intention of the speaker or writer. It is to evaluate the position and proposition conveyed in the message'. She uses the term 'affect' rather than emotion when referring to such non-referential information. She argues that in the 'course of communication, it is essential to understand the affective stance of the speaker. It is also necessary to understand the content of the message. It can help the listener to interpret the intention of the speaker and to evaluate the proposition conveyed.

Furthermore, the above can be expressed as epistemic and affective stances. The epistemic stance is a 'socially recognised disposition'. In contrast, the affective stance is a 'socially recognised feeling, attitude, mood or degree of emotional intensity' (Ochs, 1990, p. 2). As a theoretical concept, 'stance' has been described as evaluation (Hunston, 2002), appraisal (Martin, 2000; Martin and White, 2005) or attitude (Halliday, 1985, 1994).

Besides, lexis, grammar and other meaning-making resources can be used to express attitudes, feelings, beliefs, evaluations, judgements, and commitment towards a *precise target* (Biber and Finegan, 1989; Du Bois, 2007; Englebretson, 2007; Kiesling, 2011). This target can be the interlocutor, the person represented in discourse, or ideas represented in discourse and other texts. Stance components are stance-taker, stance object, i.e., the topic under discussion; and resources, e.g., evaluative lexis, modal verbs, punctuation, typography, different languages and addressee(s) (Barton and Lee, 2013). His ideas of the distinct kinds of stance are as follows:

- Epistemic stance, i.e., knowledge, beliefs, certainty, doubt, actuality, precision, limitation (e.g., *definitely, I know, I doubt, in fact, possibly, might, must, it seems that*). It can include the source of knowledge or perspective from which information is given (e.g., *according to, rumour has it*).

- Attitudinal stance, i.e., feelings, personal perspectives (adverbials such as *ironically* and *fortunately*, verbs such as *fear* and *love*, adjectives such as *happy* and *angry*). These can include stylistic speaker/writer comments on communication itself (e.g., *honestly, with all due respect*).

The research focuses on certainty and attitude as people sharing information about chronic illness may have to deal with advice-seeking and advice-giving.

Blei et al. (2003) offer an expansion of these notions by suggesting that people can search for shared topics. They can seek to understand and explain the meaning of, for example, a single word or n-gram. People express themselves through topics in the conversation.

Support is hard to define, and the notion of support is developed throughout this thesis. It can be a single detailed knowledge of a domain of social practice.

The literature review focuses on language-based advice with stance-taking. It is amongst, for example, others such as humour/sarcasm, questioning, emotion, hope, and charity of individual. It is to understand how and what people converse about (Hall, 1997). People may converse about specific 'topics', 'TOPICs' and targets using specific linguistic devices to support each other. Insights from Hall (1997) and Foucault (1980) can help when thinking about 'topics' and 'TOPICs' and targets. The organisation or users may employ these. It can help to expand on the general conception of the 'subject'. It is at the core of any sentence or post or topic or discourse.

From their analysis, the idea of 'subject' is an individual who is full of awareness. He or she is regarded as a self-ruling and stable 'something', the 'core' of the self, and the independent, real source of action and meaning. For them, an analysis must try to discover several ways many people may be referred to in conversations.

Du Gay et al. (1997) can help in showing that there can be many ‘topics’ in a conversation. They draw differences between a static, and an energetic or a combination of the static and energetic employment of words. There are mostly some significant combinations of words about many objects to sustain and to develop in a conversation.

Key researchers in the field, Harvey and Koteyko (2013); Hunt and Koteyko (2015) and Hunt et al. (2015) have suggested some topics and linguistic devices concerning online support for healthcare (shown in Table 2-3).

Table 2-3: Examples of Support categorisations

Topic Sub-Categories	Description
Organisational Events	Duk holds events for a person with diabetes
Charity	Duk holds events to raise charity and also for research
Advice	Duk owners and users offer advice
Knowledge Topics on Diabetes	Posters posts about medical conditions, e.g., blood glucose levels
Emotional Support issues	Posters look (for) and gain a desire to do something and may feel better

2.5 Quantitative

Data from several studies suggest that CL is a much broader concept that can be applied to many more aspects of linguistic enquiry (e.g., Koteyko, 2006). Furthermore, corpus linguists can include sociolinguistic data in their studies (e.g., McEnery and Wilson, 2001, pp. 115–117; Baker, 2010; Gabrielatos et al., 2010; Torgersen et al., 2011). The seminal work of Hardt-Mautner (1995) employed critical discourse analysis (CDA) and CL to study language usage with insights from Fairclough’s (1992) research on DA. He also discovered how particular discourses are related to how people and cultures interact.

A significant and growing body of literature shows the many investigations with corpora that can be used in many types of research. It is spreading to research in other disciplines (e.g., McEnery and Wilson, 2001). It has been demonstrated by Zappavigna (2012) that the field of social media referred to with the umbrella term CMC encompasses a wide range of perspectives on language. It is across disciplines, from information systems to linguistics. He highlights the issues of building social media corpora. These can show that web corpora are like traditional corpora. There are bound by issues of representativeness, balance, and comparability. These researchers offer strategies to counteract the problems mentioned above (e.g., on the cleaning of the text). These methods can be applied to the thesis in a careful analysis of the Diabetes UK (2015) corpus.

Zappavigna (2012) uses a methodology for Twitter social media analysis for analysing micro-post that is a form of corpus-based DA applying SFL. SFL posits language as a meaning-making resource. It is a theory tailored to answering questions about how meanings work within some particular contexts in which they are made, and in this sense, are 'functional'. It is distinct among linguistic theories, as SFL can help to both develop a theory about the social process and a description of language patterns.

For Baker et al. (2008), CL is not any single method. It employs a collection of different methods. They can be related to being performed on large-scale collections of electronically stored, naturally occurring texts, or instances of people employing language.

They show that many CL methods are quantitative and make use of statistical tests, which are calculated by computer software. However, Baker et al. (2008) understand that most CL methods need considerable human input, which often includes qualitative analysis, such as examining keyness or concordance lines. Therefore, Baker et al. (2008) attempt to avoid describing DA and CL as different 'methods' but instead sometimes refer to 'methods traditionally adopted by DA practitioners or by corpus linguists'.

Pulman (2017) has recently established the importance of n-grams higher than one. So, a simple newspaper headline with for example '*sales fall but profits rise*' means something very different from '*sales rise but profits fall*' (Pulman, 2017). Pulman is motivated by a linguistics analysis approach to the text via computational means.

The opposite approach is to treat the text as unstructured bags of words, – 'chuck the words in a bag, and ignore the order in which they occur'. The grammatical structure can be significant. Pulman (2017) gives an example of a word like '*kill*'. Which on its own, is negative, and a word like '*bacteria*', which again is negative on its own. However, '*killing bacteria*' is positive in most contexts most of the time. If it is '*fail to kill bacteria*' then that becomes negative again. If it is '*never fail to kill bacteria*' then that is positive again. Therefore, this is an example of four words (unigram) on their own that are negative, but in combination (4-gram) can give an overall positive message. If the analyses do not pay attention to the grammatical structure in which words occur, it could be that the analyses may be incorrect.

The number of iterations can counter the non-deterministic aspect of LDA. It is until LDA settles on a suitable number of topics and the words that belong to each topic is important. In recent years the LDA approach to text and DA has been challenged by the work of some linguistics researchers. The results of topics models cannot be taken for granted. This thesis can also challenge any assumptions of what exactly is meant by an LDA topic.

The idea of latent 'topics', in LDA (Blei et al., 2003) is that each word or n-gram within a post can be modelled as number-based vectors. These vectors can describe the posts' distribution over the topics, as a vector space model. Their basic idea is to describe a post as a mixture of different 'topics'. For them, a 'topic' is merely a collection of words that often happen with each other. They suggest that people converse about many different issues in each post.

The LDA model is convenient to calculate posts that are the same as another post and then do clustering analysis. It employs matrices as a mathematical calculation by the algorithm to work out the clusters. The LDA is a probabilistic and generative model. Norvig (2011, 2016) explains that a probabilistic model specifies a probability distribution. Blei et al. (2003) suggest that this distribution is over possible values of random variables, rather than a strict deterministic relationship. LDA can '*allow for sets of instances of observing topics in posts to be explainable by unseen groups*'. His algorithm helps to explain why some parts of the dataset are the same. If there are instances of observing certain words as collected in different posts, then the model guesses that each post is a mixture of specific topics. One of the post's topics causes each word created in the post.

AI Non-Negative Matrix Factorisation (NMF) (Lee and Seung, 1999) can also be used to model a corpus. They suggest that NMF results can look like LDA and argue that LDA in comparison is related to the study of how likely or unlikely topics are to happen. It is a model capable of expressing doubt about the placement of themes across posts and for the assignment of words to topics. However, their NMF is a predetermined calculating algorithm, which arrives at a single model of the corpus (Dariah, 2015, 2016).

Moreover, LDA (Blei, 2012) and NMF (Lee and Seung, 1999) can help with dimension reduction and clustering for topic identification. They both suggest that LDA and NMF can arrive in the same way of looking at a corpus of topics. They suggest the need for the use of hyperparameters settings to tune the model, e.g., with the use of stop words, n-grams, alphas, betas, and validation (Glossary 7-7).

2.5.1 Artificial intelligence topic-modelling

Kokkinakis and Malm (2013), for example, have used available topic-modelling software. They use empirical data from the content of the Swedish literature bank. In contrast, distant reading encapsulates quantitative methods, in which the reality of the text undergoes a process of deliberate reduction and abstraction (Moretti, 2005; Jockers, 2013). In the latter view, understanding literature is not carried out by studying individual texts, but by aggregating and analysing massive amounts of data. This understanding enables experimentation and exploration of new corpora uses and development that otherwise would be challenging to conduct. For such purposes, several available techniques can be applied, one popular technique being topic-modelling (Wallach, 2006; Brett, 2012; Graham et al., 2012).

Schiffrin et al. (2003, pp.197-214) argue for a critical look at the autonomous treatment of some aspects of language (e.g., syntax, or phonetics). Discourse cannot be satisfactorily analysed without contextual or methodological assertions

Therefore, machine learning can be used to model a corpus, and it can be done with for, example, either NMF (Lee and Seung, 1999; Scikit-learn, 2015) or LDA (Blei, 2012). Both methods can help with dimension reduction and clustering for topic identification. Blei et al. (2003) argue that probabilistic modelling can give a language for expressing assumptions about data with algorithms and for computing within the limits of those assumptions. LDA, which is a probabilistic model of texts, makes two assumptions as follows:

1. There are a fixed number of patterns of word use, groups of terms that tend to occur together in documents. They can be known as topics.
2. Each document in the corpus exhibits topics to varying degrees.

The exact number of topics to be found in a corpus with LDA is challenging to work out but has some mathematical means of estimation. Blei (2012) suggests that in LDA, the number of topics can be between 50 and 150; however, the optimal number usually depends on the size of the dataset or the researcher's knowledge of the domain. Similarly, the decision is the same for finding too few LDA topics or too many, which could lead to some nonsensical results or increase the complexity of human interpretation.

Therefore, topic-modelling is about organising, understanding, and summarising text documents, and there can be hidden topical patterns. Blei's topic is a group of words from the collection of text and their respective documents that can offer some description of the information in the document collection.

2.5.2 Automated content analysis and annotation for entity recognition and sentiment analysis

Hanks (2004) argues that evidence from large-scale corpora shows striking patterns of word use in natural language. The details of which are only now beginning to be adequately recognised and studied.

LIWC (2016) is used for healthcare studies and also has dictionaries such as a health-related dictionary that can help find whether a topic contains words associated with health. Karyotis et al.'s (2017) analysis included LIWC to find health-related topics. LIWC is used as a linguistic analysis tool to reveal thoughts, feelings, personality, and motivations in a corpus, as discussed by Karami and Zhou (2014a, 2014b, 2015). Karyotis et al. (2017) argue that 'understanding and classifying emotions is a very complex and delicate task, still under debate among psychologists'.

Karami et al.'s (2018) article provide a valuable insight where they have used LDA and LIWC together to show that from 4.5 million tweets, the LDA found 425 topics. They then used LIWC to filter the detected 425 LDA topics and found 222 health-related topics. They then used topic content analysis and aimed at an objective interpretative approach before using a lexicon-based approach to analyse the content of the topics. The lexicon-based approach uses dictionaries to disclose the semantic orientation of words in-a-topic.

Pennebaker and Davison (1997) have indicated that illness representations can be a central issue for health psychologists because patient's lay models may direct particular attitudes. Patients may develop attitudes and expectations by talking with others, hearing their stories, and comparing them with their own experiences. Patients can seek out from others how to comprehend dealing with a chronic illness.

Pennebaker and Davison (1997), in their seminal study, used LIWC to study the chronic illness of diabetes, capturing emotion words and cognitive words in people's online postings. They found that diabetes patients carry a 'strident tone. Their many exchanges could indicate an extensive understanding of their disorder, sensitivity to criticism or disagreement and occasionally outright hostile tones, towards each other'. They found that 'there is an undercurrent of emotional volatility'. Other LIWC studies making use of the automated nature of the software tools have investigated the emotional expression among breast cancer patients (e.g., Alpers, 2005; Lieberman and Goldstein, 2006).

Abdallah et al. (2016) have highlighted the difficulties with extracting structured information from unstructured text. An understanding of the difficulties is vital for qualitative data analysis. Leveraging natural language processing (NLP) techniques for qualitative data analysis can speed up the annotation process. It may allow for large-scale analysis and provide more insights into the text. A crucial step in gaining insights from the text is Named Entity Recognition (NER). These are the target words in the text. A significant challenge of the NLP and NER process is the domain diversity in qualitative data. The represented text varies according to its domain in numerous aspects, including taxonomies, length, formality, and format. Many approaches and systems have been developed for the NER task. Abdallah et al. (2016) have looked into MeaningCloud (2017). It gives an advanced opinion mining functionality. It extracts both a globally aggregated polarity of the text and more in-depth analysis. It gives a sentence-level breakdown of the polarity, extracting entities and concepts and the sentiment associated to each of them. It can also be used in a hybrid manner, with machine learning and dictionaries for entity extractions. There is a range of studies that have used and evaluated MeaningCloud (e.g., Dale, 2015; Segura-Bedmar et al., 2015).

Rodrigues et al. (2016) have shown that sentiment analysis methods can be used to automatically detect the positive, neutral, or negative mood of cancer patients. It is done by analysing their messages in online communities. It is not easy to develop or use existing sentiment analysis tools for text analytics or, as argued for earlier, advice or stance-taking.

Rodrigues et al. (2016) have compared different tools and have also developed their own. They obtained different collections of posts from two cancer communities on Facebook. MeaningCloud identifies the polarity of the sentiment of a text. It identifies the target of the sentiment and the theme of the text. Segura-Bedmar et al. (2015) have used MeaningCloud in their methods as a system for detecting drug effects from user posts extracted from a Spanish health forum.

With AI and machine learning (Lee and Seung, 1999; Blei, 2012), an AI-based automatic unsupervised modelling of large amounts of text-based data can be inferred into topics. LIWC uses dictionaries based on studies from psychology. LIWC (2016) has employed this type of 'sentiment feeling' analysis. They count particular words and identify them as part of predefined mood categories. They suggest that this approach can identify the mood in any text. MeaningCloud (2017) can use the given dictionaries or user-made dictionaries and ontologies in the English language. There are two types of sentiment analysis: global sentiment analysis, topics sentiment analysis, and aspect-based sentiment analysis. MeaningCloud's (2017) core engine with sentiment information, allows it to extract a sentiment analysis at every level. MeaningCloud (2017) can obtain a global polarity of the text, or it can go in deeper and see the polarity expressed in each one of the sentences that make up the text. MeaningCloud also offers the possibility of combining this analysis with the topics extraction feature, allowing analysts to obtain the polarity associated with the entities and the concepts in the text. This approach is usually referred to as aspect-level sentiment analysis.

2.5.3 Ethics for online research

It was frustrating to present the anonymised text data in the thesis and not the actual text data as collected, but ethics is crucial. The data is presented after the analysis. For instance, with removed identifiers from a much smaller sample that can make it harder to use the Facebook post sample to find the individual that made the post and wrote that text. Narayanan and Shmatikov (2008, 2009) demonstrate that when working with social media data, anonymising data is complicated. They show that anonymisation procedures are still evolving for aggregated or Big Data. It is difficult to anonymise units of data extracts, for example, for tweets from Twitter, when these are reproduced in publications and during presentations. 'Related to concerns over identity breaches is the risk of harm that researchers potentially place on their research subjects' (Association of Internet Research, 2012)

Facebook's (2017) mission statement is 'to give people the power to share and make the world more open and connected'. Facebook has its standard terms of service, but the issues of privacy, anonymity and confidentiality are problematic, even within the network of Facebook users (Grimmelman, 2009; Zimmer, 2010).

This research adopts amongst others, the essential ethics work of social media researchers as suggested by Oxford Brookes University Ethics Committee (UREC) (Appendix 7-6). It is together with UREC's ethical expectations. It also includes the ethics reference paper (Townsend and Wallace, 2016) as recommended by Diabetes UK (Appendix Figure 7-1).

Townsend and Wallace's (2016) framework is widely used in this research. It has also helped develop my understanding of ethics for research in a dynamic and technologically challenging environment like social media.

The use of ethics in this thesis covers adult subjects and their openly available online information, as described by the British Psychological Association (BPA, 2013, 2015). The issues, as mentioned earlier, profoundly influence the thesis. As an example, it can make sure there is no direct contact with the people (online posters) on Facebook that share their personal medical information.

2.6 Conclusion

This section has reviewed the critical aspects of online support. These are topics, TOPICS, targets, linguistic forms such as advice with stance-taking, events, humour/sarcasm, questioning, emotion, hope, and charity. These can be made through the employment of text, e.g., sentences in posts, topics and social media discourse. This text or sentences may have sophisticated linguistic devices in need of further study. The conversations that make up the support can be about the advice with stance-taking advice. It is amongst many others from events, humour/sarcasm, questioning, emotion, hope, to charity, of the people taking part. They can concern the 'topic'/purpose and 'TOPIC'/content of the discourses. This thesis has elaborated on the concept of device-enabled discourse category.

The pragmatic meanings are as important as the semantic meanings in the discourse. The task is to find the salient elements that help construct the meanings in support practice. The support patterns may be constructed by a significant amount of people taking part.

These patterns can be related to social relations of power, and solidarity. For example, it makes overcoming the risk of bad advice. It contributes towards building trust in giving advice and asking for advice whenever possible.

The research suggests that SFL and the Appraisal model can contribute a theoretical framework towards developing research questions and filling the gap in the literature. A theory is needed to help describe and from the evidence, salient and of non-obvious linguistic patterns of support. It can become a theory of peer support in social media discourse.

A combined approach to DA is explored to fill the gaps in the literature. Corpus linguistics with SFL, DA, AI, LDA, automated content analysis and annotation for entity recognition and sentiment for advice with the stance-taking approach is explored. *The literature shows that both pattern and purpose/content is needed to understand what is going on in a corpus*. It means that the gap in the knowledge may be filled.

An approach can take advantage of primarily employing AI machine learning to go about automatically analysing the large-scale corpora for language-based features. LIWC and MeaningCloud can also be employed as contrastive methods but more importantly, to help identify the elements of any sophisticated linguistic devices. The literature shows that there is not a readily available AI that can identify sophisticated linguistic devices like stance-taking. Elements of stance-taking may be identified. These may be integrated and could lead to identifying trends of advice with stance-taking.

CL-based approaches - a multi-dimensional approach can also help to confirm the findings. Qualitative analysis can be carried out on the quantitative results. The quantitative results are obtained automatically by AI software and enable automatic annotation on large-scale text data. They can also aid the manual identification of, for example, advice features. The comparisons of results are made to state-of-the-art research in the diabetes domain.

AI and software-automated analysis have limitations for the identification of sophisticated linguistic features. Discourse analysis can benefit from AI and software-automated analysis.

The use of an ethics process in collecting, anonymising, processing and reporting on findings for online data research is prioritised very highly in this thesis. The people who share their personal and medical information online are central to the research.

3 METHODOLOGY AND ANALYTICAL FRAMEWORKS

3.1 Introduction

This chapter discusses the specific aspects of the multi-method approaches to conducting the research and analysis. It consists of CL with quantitative approaches that concern high posts and topic frequencies. The qualitative approaches can help to give context and meaning to the findings. The multi-method approaches include predominantly AI with machine-learning LDA topic-modelling, automated entity recognition and sentiment analyses. Linguistic approaches that guide the thesis include DA with systemic functional linguistics (SFL) and appraisal model analyses. It is to identify device-enabled discourse categories of purpose and content and linguistic forms such as advice with stance-taking, events, humour/sarcasm, questioning, emotion, hope, and charity in Facebook posts and many other interactional activities amongst peers. For example, to investigate high-frequency stance-taking, affective stance and epistemic stance. It is crucial whether they exist in consecutive or non-consecutive posts. Is support somewhat indirectly in usage across many posts about the same target throughout the corpus rather than primarily in a linear conversational manner?

These multi-methods can provide a systematic analyses approach, but there are limitations to the study's exploratory methodology. It includes the difficulty of employing features of Advice or APPRAISAL theory in automated analyses. As an example, in manual coding of the data into broad categories of advice-seeking and advice-giving and the many other interactional activities amongst peers. The development of my theory of support involves using high-frequency potential linguistic patterns. It involves bringing a large-scale corpus of text into a small random number of topics and posts. They may best represent social media, diabetes UK Facebook corpus.

The quantitative approaches include some of the most popular tools used to assess text data. They include AI LDA, entity recognition and sentiment analysis with dictionaries via the MeaningCloud and LIWC software applications. Quantitative research is generally associated with the positivist paradigm, used to gather high-frequency data that contains potential patterns.

Mixed CL methods are to find elements of linguistic features that frequently used.

An advantage of using automated computer analysis is that it allows for the identification of stance-related features in the text that is part of rich data. It can also identify many items from lexis and grammar to other meaning-making resources in the text. These resources can express attitudes, feelings, beliefs, evaluations, judgements and commitments.

The thesis is guided overall by the idea of stance components: stance-taker, stance object or 'topic' and 'TOPIC' and target. These may be found in the support conversations and can be understood in the broader context and pertain to specific resources. They can include (for example) evaluative lexis, modal verbs, punctuation, typography or different languages, and addressees (Barton and Lee, 2013).

3.2 A description of the research methods

Analysis of discourse is a complex undertaking. The process of evaluation in discourse studies is context dependent (Fuoli and Hommerberg, 2015), and the process of annotating a discourse is not straightforward (Fuoli, 2018). It is achievable by applying a broad set of categories, which can include an appraisal framework. It means obtaining a process that can use concrete instances of text. It to obtain information that can develop and renew a model of what a corpus is about, but progressively. Fuoli (2018) argues that manual annotation is an essential part of the process of building theory. However, Fuoli and Hommerberg (2015) argue that manual annotation should facilitate comprehensive and detailed corpus analysis that would not be possible with only automated techniques.

For example, Mohammad et al. (2016) have shown that a dataset consisting of the tweet target consecutive posts annotated for both stance and sentiment. However, they show that 'while knowing the sentiment expressed by a tweet is beneficial for stance classification, it alone is not sufficient'. To automatically detect stance-taking has its problems. Is the author of the text in favour of, against, or neutral towards a proposition or 'target'. The target may, for example, be a person, an organisation, a government policy, a movement or a product.

Mohammad et al. (2016) look critically at the concept of stance-taking in any discourse. They argue that 'we can often detect from a person's utterances whether he/she is in favour of or against a given target entity – their stance towards the target'. It does not preclude a person who may express both stances towards a target by using both negative and positive language.

The thesis uses insights from the research steps and procedures of critical researchers to weave together quantitative and qualitative approaches (Miles and Huberman, 1994; Hardt-Mautner, 1995; Blankenship, 2010). They suggest not only the use of quantitative but also combined approaches that involve qualitative data analysis. Hardt-Mautner (1995) expands on this idea and shows that corpus analysis is a cyclical process. Looking at anything found about a corpus may be seen against other more extensive patterns inherent in a corpus. The thesis's study of language with CL and DA is conducted under the philosophical positivistic umbrella (e.g., Miles and Huberman, 1994). It sets up a quantitative high-frequency analysis. These analyses are part of the CL approach, which is a study of language approach since the data are of a linguistic type.

Primarily AI machine learning is used to develop a topic model of corpora. An automatic machine-based analysis can deal with the vital data much better than a long manual human-based approach. LDA for topic detection is conducted over the entire corpora. It offers an unsupervised way of finding latent patterns. Topic-modelling needs to be evaluated against claims of linguistic pattern discovery. There is a significance in the usage of factors such as topics. There is cross-checking of topics found against the posts clustered within those topics. It suggests a need for a further DA of the topics and constituent posts found by the process.

The high-frequency LDA topics are validated with many iterations and precision measures and compared to the academic research on the specific domain of Diabetes Support. Summaries of the research mixed-methods steps are given in Figures 3-1 and 3-2.

Figure 3-1: An overall diagrammatic summary of the methods.

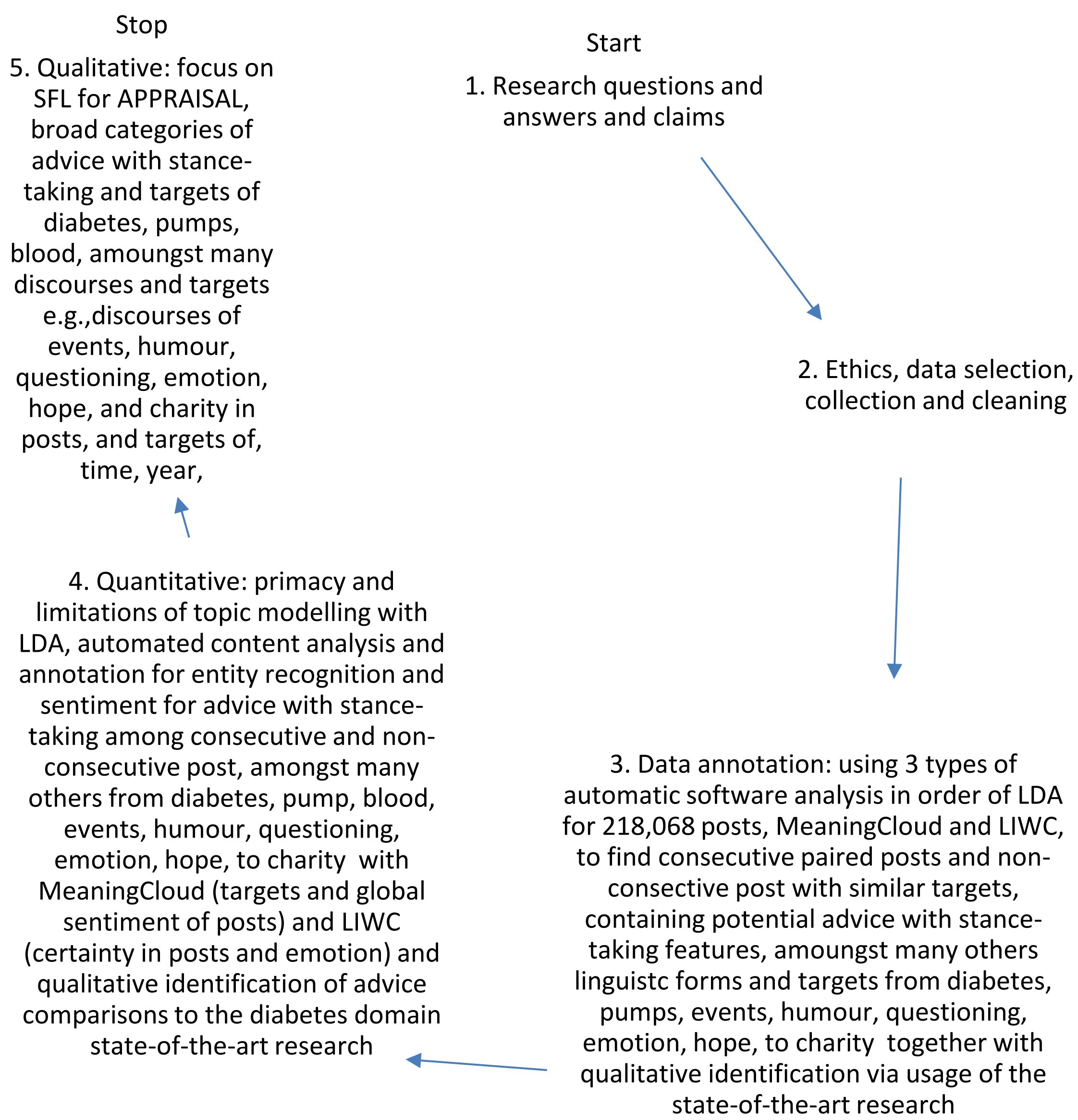


Figure 3-2: A step-by-step diagrammatic summary of the overall methods.

1. The factors involved in collecting a corpus of diabetes posts are initially considered. The difficulties involved in doing so on a large-scale corpus are made clear together with its inherent limitations. The ethical implications are also crucial and are stated and followed with state-of-the-art ethical processes for online research. The novel DA method begins with a focus on the research questions. The corpus, therefore, needs to be suitable for answering them. It contains about 96% peer posting of the total post, in a top social media, in a Facebook Diabetes UK support group page.
2. The corpus needs to be annotated. Since it is a large corpus of 218,068 posts, 6,960,998 tokens and 64,904-word types, automatic software procedures are preferable to using purely manual annotation. The annotation is also done in cycles. Fuoli (2018) describes the methodological rigour in annotating data for APPRAISAL. These procedures are used to determine salient features in the corpus and with an awareness of its limitations. It is further explained in points 3–9.
3. Artificial Intelligence LDA topic-modelling is predominantly used to automatically identify the high-frequency top topic trigram patterns and features of the vast corpus of 218,068 posts and 6,960,998 tokens and 64,904-word types a Facebook Diabetes UK corpus data (FDUK). All posts, both consecutive and non-consecutive, are analysed. The results are critically considered so that the quantitative analysis of patterns needs to fit within a Linguistic qualitative analysis.

The AI machine-learning results need to be looked into and not merely to be gathered. LDA helps to identify the top 'topics' with their top trigram words. A novel approach in the research is to include stopwords at this stage of analysis. The resulting LDA topic model is suggested to have *discourse purpose words.* My thesis calls this '*device-enabled discourse purpose categories*' or 'topics'. The posts that were clustered into these 'topics' of high-frequency totalled 92,254. The remaining 133,106 posts were not clustered into these topics of high frequency. It suggests that they may be more crucial and are used across the corpus by participants to construct support. The devices are possible purposes of and resources for support constructs.
4. Typical AI LDA topic-modelling is also used, and for identifying high-frequency top topic trigram patterns and the stopwords are removed in the analysis. State-of-the-art research suggests that this approach ultimately provides 'content' related words. However, this thesis is critical of this approach as it defines the *discourse content words* as '*device-enabled discourse content categories*' or TOPICs. This approach within the novel DA method provides the familiar category related LDA topic model of the discourse. It covers the salient content words (*discourse content words*) used across the corpus by participants. They represent possible content and objects in the support constructs.
5. The process involves a research choice in line with artificial intelligence text analysis processes of bringing the large corpus down in size as the analyses process to one of the potential high-frequency patterns. After the LDA analysis shown above and the creation of the 2 LDA Probalistic models, LDA '*device-enabled discourse purpose categories*' or topic numbers 0, 24,

	34, 13 and 33 are selected randomly for further analysis and gives 92,254 posts. The ones with consecutive post numbers are then selected and put into a table (see Appendix 7-1) for further analysis. It results in 73 posts and their substantial quantity of text, consecutive posts numbers. The smaller analysis sample will ultimately have consecutive posts and non-consecutive posts that could exhibit similar behaviours.
6.	The decision to study the predominant and salient features meant using conservative 500 topic features. The posts that were thus clustered into these 'topics' of high-frequency totalled 92,254. The remaining 133,106 posts were not clustered into these topics of high frequency. It is essential to consider what is left out of the analysis and what is included. LDA models are approached critically in my thesis.
7.	These randomly selected topic numbers in point 5 also have the LDA '*device-enabled discourse content categories*' TOPICs assigned to each post from the above analysis. They are also put in the table next to each post identified in point 5.
8.	MeaningCloud (2017) automated software is used to find the nouns/entities (targets), and this is assigned to the posts from point 5 in the table. The thesis focuses mainly on domain-specific targets of diabetes, blood, pumps and medication from many others ranging from targets of usernames, places, time, to years. It results in a sample of 73 consecutive posts with these targets from the LDA random five topics and 92, 254 posts. So, a new sample of posts for analysis will be about the target and contain ultimately potentially similar posts that are both consecutive and non-consecutive posts. The idea is to consider any potential high-frequency patterns for analysis but for posts

	that may have people/peers in a 'conversation' or as closely related posts to each other. It is in keeping with the AI analysis of all posts for patterns.
9.	The MeaningCloud automated software is used to find global positive and negative sentiments for each post. It is assigned to the post from point 5 in the table.
10.	A Linguistic Inquiry and Word Count (LIWC) (2016) is used to find values for healthcare, positive emotion, negative emotion, and certainty. The values are assigned to the posts from point 5 in the table.
11.	Linguistic forms ranging from Advice, Humour/sarcasm to Questioning categories are identified by a manual comparison to labels in the well-established literature and assigned manually in each of the posts from point 5. This analysis is done in comparison to state-of-the-art-research. LDA topics and the posts in the table are used to help make sense of each other in the analysis. Other discourses and ways that peers interact are also identified.
12.	The analyses, as mentioned earlier, are then integrated to help identify the stance for each advice-related post in the table. Crucially the stance can relate to a noun/entity (target). The '*device-enabled discourse purpose category*' and the '*device-enabled discourse content categories*' are also shown to be resources for advice with stance-taking. There are many other linguistic forms such as humour/sarcasm or questioning or hope or raising charity funds. They are also shown to be related to the entity of healthcare. The stance is shown to be made up of an affective stance which can be related to a positive or negative emotion. They are shown to have a global positive or a Negative Sentiment value in the posts. The stance is shown to be made

up of an epistemic stance. It can be shown to be related to Certainty values in the posts.
13. The above steps (1–11) are then integrated into a complete analysis of the discourse. The novel DA method shows sophisticated linguistic features such as advice with stance-taking. It is essential, for example, in the analysis to concentrate on who is involved, the resources, ‘topics’ (purposes), TOPICs (content), objects (targets), relationships (peer-to-peer), and the history of the stance.
14. The above analyses help to show patterns of support. They are evaluated further with DA, which is informed by SFL appraisal analysis. For example, they can suggest elements of solidarity and risk mitigation in the giving and receiving of advice.
15. Precision testing is used to help validate the results of the LDA. A table is created to offer comparisons of the LDA, MeaningCloud and LIWC results. It is in order to see if the integrated analyses of different individual procedures make sense.

3.3 Research questions and operationalised questions

In developing the general research questions into operationalised ones, the use of algorithms and coding, e.g., Python programming to carry out the analysis, are essential considerations. For example, the use of topic-modelling and Blei’s (2012) LDA algorithm may lead to identification and testing of potential latent patterns in a corpus. The idea is to find and identify a potential language-based pattern of support.

The general research questions are operationalised with insights from the literature and to fill the gaps in the knowledge: from CL, AI, online ethics, healthcare communications, CMC, and a potential DA of the corpora. They can be employed to show possible patterns of support.

The research questions, as mentioned earlier, are as follows.

RQ1: What attitudes, opinions, and sentiments are people expressing, about their conditions and issues in Facebook Diabetes UK posts?

RQ2: How do people express their attitudes, opinions, and sentiments about their conditions and issues in Facebook Diabetes UK posts?

The literature review and the methodology inform the operationalised research questions. They are developed throughout this chapter and are set out below:

OQ 1.1: What are the discourse purposes? What are the particular word trigram choices used by people, often together, to express what the discourse purposes are in their posts on social media for supporting people with chronic illness?

This question considers, for example, the advice, events, humour/sarcasm, questioning, emotion, hope, and charity in a post, topic and discourses. They may contain stance-taker and stance-taking resources that have been drawn upon but to what discourse purposes, or, as the thesis calls them: for what 'topics'.

OQ 1.2: About what is the discourse? What frequently used content word trigrams relate to the discourse contents?

This question considers, for example, the advice, events, humour/sarcasm, questioning, emotion, hope, and charity in a post, topic and the discourses. They may contain stance-taking, and related content words for support concerning other posts, or, as the thesis calls them: for what 'TOPICs'.

OQ 1.3: What is its primary target? What nouns/entities is discussed in these posts?

This question considers the stance-taking and related target (noun/entity) object being discussed ranging from diabetes, advice, events, humour/sarcasm, questioning, emotion, hope, to charity, across consecutive and non-consecutive posts.

OQ 2.1: What is the poster's stance about the certainty of their information? How certain is the person in these posts and do people express these directly to each other in consecutive or in non-consecutive posts?

This question considers the epistemic stance (certainty) in the giving of information in the speech acts of, for example, advice-seeking and advice-giving in consecutive and non-consecutive posts.

OQ 2.2: What is the poster's stance concerning their feelings about the information and how do they feel in these posts and do people express these directly to each other in consecutive or in non-consecutive posts?

This question considers the affective stance (attitude) in consecutive and non-consecutive posts.

3.4 Ethics

1. *OEQ1. How best can a Facebook post or large numbers of posts be collected, processed, saved and shared in academic writing without breaching online ethical guidelines?*

A focus of this thesis is to make sure that the data processing used in this research is lawful. The UREC research organisation do specify a statutory basis for data processing (Appendix 7-6). For example, for NHS data, a researcher should also know this basis because approval bodies like NHS Digital (2018) will ask researchers to specify it. The data involved in this research is social media and FDP data about diabetes from the users of their Facebook page and is carried out with the organisation's permission (Appendix 7-6). However, even so, it needs careful handling in order to ensure adequate anonymisation of this online, 'public' freely shared health-related data.

Generally, safeguards apply widely to research with personal data. They need to include obtaining UREC 'research ethics committee' (Appendix 7.6) approval, and only for processing personal data that is deemed necessary. This process is known as data minimisation.

The ethics process of this thesis draws on the theoretical framework provided by Townsend and Wallace (2016) and also insights from other prominent researchers. The framework is used as a guideline rather than as fixed rules in the research. Principles need to remain flexible. It is in order to respond to the varied contexts in which social media data is found. It considers the platform used, the target population, the topic, the methodology used, and the type of data collected (e.g., text, images or video).

Facebook (2017) says about their data policy: '*We transfer information to vendors, service providers, and other partners who globally support our business, such as providing technical infrastructure services, analysing how our Services are used, measuring the effectiveness of ads and services, providing customer service, facilitating payments, or conducting academic research and surveys. These partners must adhere to strict confidentiality obligations in a way that is consistent with this Data Policy and the agreements we enter into with them*' (Appendix 7.5).

The next task involved working through the framework and in determining whether the data the research needs to access is public. It was decided that it is public. Nevertheless, permission was still sought and obtained from Diabetes UK (Figure 7-

1). The data is about people sharing their personal experiences about their illness with others on a social media page. These need to be addressed to proceed.

UREC ethics have insisted that the research must consider the anonymisation of posts and content versus research that shows content. It is also to cite other studies and good practice and must be subject to data protection law. The study presented in this thesis, having been checked by the UREC committee, has been shown *to involve the automated, non-consensual collection of identifiable, non-anonymous data from participants discussing physical and emotional difficulties arising from a long-term illness on a publicly accessible webpage*. A clear statement of UREC's ethical position and that of the Information Officer is provided in the Appendix (Table 7-24). The Information Officer at OBU also investigated whether the GDPR applied to this thesis. It was highlighted that two key areas where needed, and these are covered in my research ethics process:

1) There needs to be a 'Lawful Basis' to process the FDP data. Although there are currently concerns how social media data is processed, the thesis 'is for genuine scientific research under Article 89 of the GDPR'. The thesis has taken the necessary steps to anonymise and apply the principle of data minimisation. The thesis has taken the necessary steps to safeguard the rights and freedoms of subjects by contacting Diabetes UK and seeking their consent and advice. This process means that the thesis has a lawful basis to proceed.

2) Transparency of processing: 'Under normal circumstances, you would have to display a privacy notice to your participants. However, as you are anonymising the quotes and to locate and provide such notice via social media would take "disproportionate effort" for your scientific research (Article 14(5) (b)), you are exempt from doing so.'

3) The rest of the GDPR principles are pretty much covered in the ethics approval. ‘Therefore, the study’s institutional ethics panel and the examples of state-of-the-art research to support the statement, and existing studies do provide a precedent for this type of research.’

2. *OEQ2. How can Facebook posts or large numbers of posts be adequately anonymised so that they do not match the original post and therefore cannot be used to identify the individual or individuals who created the posts quickly?*

In light of the above, I have gone through the data and have ensured that the posters are anonymised. The content has been ‘rearranged’ into trends, ‘topics’, ‘TOPICs’, targets and for example, linguistic forms such as advice with stance-taking. It is in order to ensure that the data reported on in the thesis cannot be easily traced through search engines or via algorithms back to the individuals who posted them on Facebook. Therefore, in my research context, the broader trends of the actual data that exist on the platform are reported. Samples of specific evidence for the arguments in the thesis are provided. ICO’s (2018) key anonymisation techniques are used. It is with data masking, partial data removal and data quarantining, pseudonymisation, aggregation, derived data items and banding. LDA analysis results in the form of patterns from the 218,068 posts are presented in my thesis. I have used parts of 73 posts and their substantial quantity of text samples of consecutive and non-consecutive posts. It is with anonymisation of the parts of the postings. It is for presenting them in the thesis and did this once the analysis of the actual text had been carried out and correctly saved. The trends and patterns and sampled anonymised posts can now be made available to a broader audience of this thesis and for future publications.

Anonymisation so that it can be used as examples in the thesis and future publications.

This process means that an exact match to a post may not be made back to the original posts in the data on the Facebook webpage. Also, a systematic process is needed to change the posts into an acceptable form for sharing in the thesis and future publications. Therefore, all individual names were removed from the posts. Besides all instances referring to gender identification were removed as this information was not needed in the research: for example, concerning marriage ‘he’, ‘she’ was replaced with ‘partner’.

Particular items that could identify an individual such as addresses, where a person works, where they may be going on holiday were removed from the posts. Also, spelling mistakes, for example, ‘*dieing of kidney failier*’ becomes ‘*dying of kidney failure*’, so they were corrected in the parts of the posts. They can therefore also make it more challenging to find the posts on the webpage. ‘Blood sugar’ was changed to ‘blood glucose levels’. Words that could be represented by a single word were changed to a single word, e.g., ‘*hes got to see the diabetic doctor*’ becomes ‘*appointment with a diabetes doctor*’. Acronyms or abbreviations or incomplete aspects of words are changed to full words, e.g., ‘net’ becomes ‘internet’, ‘carbs’ becomes ‘carbohydrates’. Where numbers had been used, they were changed to words and approximated so ‘48 years’ would be changed to ‘about fifty years’. ‘My daughter’ or ‘my son’ was replaced with ‘my child’. Where a sentence is shortened, the full sentence is written out. An emoji is replaced with the ‘emotion word’. The text was changed to small letters.

After these steps were carried out on a post, the different parts that represent the salient linguistic features were further separated by using ‘…’ ‘And’ to represent ‘and’ or ‘,’ or connectors. It is so that it can be difficult for a machine to put them together and find a matching post on any internet webpage. Therefore, an example post would look something like: 86, '*...just came out of hospital…developed keto-acidosis and kidney infection…my blood glucose level was approximately fifty…. great...happy…*'

Seventy-three posts and their substantial quantity of text were randomly selected for further analysis from a total of 218,068 posts that were utilised in the research. This anonymisation would also mean that data usage is minimised and is only for scientific purposes (GDPR, 2018).

Parts of the 73 randomly sampled full posts is shown in the thesis. Two hundred eighteen thousand sixty-eight posts were used in the entirety of the research. They are not shown in my thesis except for their topics, trends and patterns. The parts are used to show trends of linguistic features in this thesis. Spelling mistakes were corrected to make the posts non-searchable back to the Facebook Diabetes UK webpage. The clear statement on this thesis's ethical position is given above, with details from the UREC (Appendix 7.5). Existing studies in this section provide a precedent for the thesis.

3.5 Data collection

As per Blankenship's (2010) research process, Step 7: Collect Data, the data for the investigation was collected from the FDP. The data collected ranges from February 2008 to July 2015. A total of 226,385 posts were obtained for analysis (as shown in Table 3-1). The next section discusses the descriptive statistics of the corpus.

Table 3-1: Total posts and poster types

Total posts and poster types		
Raw data	226,385 posts	
Cleaned data	225,360 posts	
Unique user/peer persons	16,137	
Unique user/peer person total posts	218,068	96%
Owner/organisation/duk person	1	
Owner/organisation/duk person posts	8,317	4%
Total cleaned posts for all persons	226,385	
Total person tokens	16,138	

These posts are made up of initial posts, and comments or replies (referred to as just posts in the research) and they form the central part of the discourse organisation on Facebook. The text exists within the posts form the corpus, and it is these that are used in the analysis. A data decision is about finding a useful corpus and comparing different corpora. For example, the research corpus selected here contains posts, comments, and replies to many people about online support.

The software used for collecting the data and for the analysis includes R Programming and Python with Anaconda and Spyder. These are employed together with text-mining algorithm-computer codes for topic extraction, e.g., LDA, in this study. Secondary software such as LIWC, MeaningCloud, and MS Excel is used for automated content analysis and automated annotation, as well as for entity recognition. The primary and secondary software are used together for linguistic forms such as questioning, advice and sentiment, certainty and stance-taking analysis.

This thesis uses the study of language approach (e.g., Biber, 1991; Hyland, 2005; Hood, 2011; Harvey and Koteyko, 2013). Corpus selection in the thesis is also about using machine learning that allows for a large-scale selection of data for analysis.

The FDUK selected for the research is the current latest corpus and procedures. These steps for data analysis have decisions about defining what data must be gathered and why the data are essential. Principal researcher's works play a part in my decisions on how to go about finding, selecting, and cleaning the data.

3.5.1 Descriptive statistics

The descriptive statistics about the FDUK are provided in this section. More details about the process are given in order for the reader to understand why the particular reference corpus has been chosen. A description of the corpus is shown in Table 3-1. There are total 16,137 users/peers who posted 218,068 posts, 96% of the total. There is just one 'owner', who is referred to in this thesis as an organisation, owner, duk or expert team and they were responsible for the remaining 4% of the posts – 8,317 in total. The duk poster can consist of different professionals and charities who post under the duk label. Therefore, it is vital to find out what 96% of posts (users) support practices are about, and for them to be related to the thesis and research question.

LIWC is another software program that is employed to compare the findings with the stopwords that are not used. An LIWC count of the non-cleansed data gives 6,960,998 tokens. LIWC shows 64,904-word types and 36,151 critical words.

Antconc (Laurence, 2014) can also be utilised to check the frequencies. The linguistics analysis software program (Laurence, 2014) gave 69,304-word types and 35,413 keywords with no stopwords being used. For the non-cleansed corpus, there are 64,904-word types and 36,151 keywords. The total number of word tokens for the cleansed corpus is 6,826,614, and 6,898,948-word tokens for the non-cleansed corpus.

The thesis's usage of software, (Python, R Programming, MeaningCloud and LIWC as multiple uses of software types) can help in confirming close sets of number descriptors about the utilisation of words in the corpus. The software is utilised for the n-grams, keyness and colocation. The n-gram, keyness and colocation are considered about the study conducted by Petyko, (2017). The multiple uses of software can help with an overall analysis of the data, as shown in Chapter 4. Different software is needed for various types of language-based questions, but each use of the software is understood by its capabilities (e.g., for entity recognition or 'topics').

3.5.2 Data cleaning

Another not so easy decision is that of cleaning the corpus data (Caren, 2015), such as removing punctuation. Punctuation is relevant in any discourse but is removed from the process of topic-modelling, unlike LIWC and MeaningCloud, where it is retained. Researchers like Caren (2015) use widely known procedures that can be employed in machine learning and the analysis of cleansed data through R Programming and Python. A comparison of the cleansed and non-cleansed data must be employed to help to make sense of the differences between the two datasets.

Widely used software such as Spyder, with Anaconda and Python, MS Excel and R programming is used to collect, clean and analyse the data. All the information that has been gathered from Facebook with R programming is cleaned and stored in chronological order into an MS Excel file. The data includes the posts and its text.

The posts are cleaned with R programming in this thesis, which is used to remove, for instance, URLs. Another example is to change the text into lower-case letters. It can also be used to keep the question marks (?) by changing all question marks to the word "questionmark". The collected data gives 225,360 posts that consist of text that was also cleansed (for example, removing vacant posts). However, the stopwords were kept for finding ‘device-enabled discourse purpose categories. It is to find the content word and operative type words in the corpus (e.g., about questioning, possessive nouns, *wh* question-related words, and ‘I’, and ‘you’). A comparison is provided for when stop words are used to analyse the post.

Stopwords from R programming ‘English’ are employed. Particular stopwords are taken out from the ‘English’ list that relates to the patterns of questioning and advice as documented by key researchers (See literature review). So that these stopwords are kept ‘ask’, ‘asking’, ‘help’, ‘would’, ‘wonder’, ‘think’, ‘what’, ‘when’, ‘where’, ‘which’, ‘who’, ‘whom’, ‘whose’, ‘why’, ‘how’, ‘can’, ‘cannot’, ‘cant’, ‘do’, ‘donot’, ‘if’, and ‘so’. This type of LDA analysis can yield trigrams that have more content words from the domain and possible trigrams related to questioning or advice requests. These stopwords also kept for finding ‘*device-enabled discourse content categories*’, and so we are left with mostly context words.

3.5.3 Data annotation

Fuoli and Hommerberg (2015); and Fuoli (2018) show that annotation is iterative and progressive. The thesis applies sets of broad categories included in the appraisal framework to the annotation of concrete instances of text. The research can obtain information that can be used to develop and refine a model. This process is used for collecting data, and then to go through the phases of corpus expansion, transcription, mark-up and annotation iteratively, until the final annotation. Categories of interest are decided on and checked in the data.

Endres and Whitlock (2017) have shown some limitations to the coding process. They argue that how the data is collected and analysed can cause problems for this process of data collection. There are large-scale data and information on these forums, and researchers cannot obtain it very rapidly. To sift through this large-scale data and code it using a manual scheme can be a very lengthy process. However, it is also a lengthy process even for a small amount of data. There is a considerable time involved in coding and analysing any form of data in order to arrive at any meaningful conclusions. Computer software can quickly and automatically code text information, such as using the software program LIWC (Tausczik and Pennebaker, 2010). However, using software to code has its own set of limitations too. These types of software can automatically categorise words and classify text. For example, it may categorise words and code for the emotional tone of a post. Tausczik and Pennebaker (2010) also make the point that coding is a nuanced process. A computer program for text analysis, while very good for some things, cannot always handle complexity as competently as a human coder. By their natures, human coders take a fundamentally different approach to coding than computer software.

Even among multiple human coders, there can still be expectations for some variation and disagreement. 'Interrater reliability needs to be high enough to declare specific rules and requirements for putting each specific code in a specific case'. The idea is to create a coding manual for coding the data. However, there is still a degree of subjectivity in coding that cannot always be addressed.

Abdallah et al. (2016) also provide a way to use software to carry out annotation in an 'automatic way'. The first step for the qualitative analysis of the unstructured text is often to annotate the data to identify 'named entities' NER (Nadeau and Sekine, 2007). It is one of the primary tasks in NLP. It refers to techniques that are used to locate and classify atomic elements in text into predefined categories, such as the names of persons, organisations, locations or expressions of time (Beheshti et al., 2013).

Data annotation is not easy, and approaches from Fuoli (2015, 2018) are used to annotate the post iteratively but in a broad manner with an awareness of the limitations. Consideration is given to potential key elements in sophisticated linguistics forms:

1. The shared communication acts in consecutive and non-consecutive posts;
2. The context and purpose sharing words amongst posts ('topics');
3. The content or TOPIC (taken according to Halliday) may be about 'what is talked about'.
4. The targets/nouns/entities in the posts may be the objects of that talk;
5. The certainty in the posts;
6. The positive and negative emotions and global sentiment available in posts.

In the sections that follow, it will be argued that quantitative and qualitative approaches and the use of primarily machine learning to analyse the corpus can help to answer the research questions.

3.6 Quantitative and qualitative analysis

Data annotation is not easy. Approaches from Fuoli (2015, 2018) are used to annotate the posts iteratively. Software processes help in discovering, e.g., 'topics' or entities to focus on in many posts versus looking at the individual posts within a 'topic'. As an example, the annotation is done by LDA 'topics' by it given top trigram patterns, for instance, wherein similar post are clustered together. The labels are the same trigrams since the trigram may contain some easily identifiable aspects of the diabetes domain.

The next step is to find out the content of the posts, and LDA is used, again without stopwords, to reveal the high-frequency 'TOPICS'. These are groups of content words that together may form, for instance, a trigram, i.e., '*device-enabled discourse content categories*'.

In keeping with AI topic modelling and data and dimension reduction from Big Data to sensibly sized samples for further human analyses are selected. Out of the fifty high-frequency topics, five high-frequency LDA device-enabled discourse purpose category 'topics' are then randomly selected for further analysis. These give topics 0, 13, 24, 31 and 34 are they are analysed further. They contain together 92,254 out of the total 218,068 posts. Their LDA 'TOPIC' '*device-enabled discourse content categories*' is also utilised in a comprehensive analysis.

The next stage is the gathering of the randomly selected 'device-enabled discourse purpose categories' threads. This process was done by selecting only posts that are closely related to the post-id-number and time order of the randomly selected posts in the 'topics' mentioned above. This combined analysis yielded seventy-three posts and their substantial quantity of text for further analysis (Appendix 7.1., Table 7-6).

These seventy-three posts and their substantial quantity of text are then analysed using MeaningCloud with 'TOPICs' extraction for extracting the different elements present in them. The focus was to get the objects of any advice and stance-taking so that the primary entities in a post could be found, (e.g., nouns such as 'diabetes' or 'medication').

MeaningCloud provides sentiment analysis, and as the 73 posts contain text, they are analysed to determine if they express positive or negative or neutral sentiment. The local polarity of the different sentences in the text is identified, and the relationship between them is evaluated, resulting in a global polarity value for the whole text. This analysis gives a negative, neutral or positive global sentiment value towards the many entities or targets that are mentioned in the posts.

These seventy-three posts and their substantial quantity of text are then assigned healthcare, certainty and emotion values from the analysis by LIWC. LIWC is a dictionary-based tool with input from the field of psychology. Behaviour is related to classes of words for that behaviour. Since the study is interested in emotions, the positive or neutral or negative emotion values for the posts are used. These values are guidelines to be used carefully with further qualitative analysis. Since the study was interested in the healthcare, certainty and emotion displayed in a post, the healthcare, certainty and emotion values are used, and the values can be zero, low or high.

The approaches are validated with LDA validation strategies. They are validated by comparing the findings of the quantitative and qualitative LDA and DA language-based investigation to those of the LIWC (2012, 2016) and MeaningCloud (2018). LIWC (2012, 2016) is a type of software where dictionaries of terms are employed to compare against words utilised in a corpus. LIWC can offer some insights from psychology. Groups of words found in a corpus can be related to behavioural patterns in people. As an example, the use of the word ‘cried’ can belong to the set of five-word categories that include sadness: a negative emotion with a feeling of love, hate and guilt; an overall affect; a verb; and a past-tense verb LIWC (LIWC.net, accessed 2016).

LIWC has an internal validation strategy. So if, for example, the word 'cried' is found in the Facebook corpora posts, there is an increment of each of these five sub-dictionary scale scores. It can be compared to a base standard for general text. These different checks are then used to compare the findings of the corpora.

Therefore, this thesis suggests that LDA, LIWC and MeaningCloud can be used to interpret posts for advice-seeking, advice-giving, affective stance and epistemic stance. For example, particular words used for epistemic stance can show how 'certain' people are about the advice they are sharing in the posts. LIWC can be used to help interpret the epistemic stance. The LIWC 'certainty' behaviour considers words such as '*always*' (Appendix 7.4).

In this section, the quantitative and qualitative literature is further examined by combining both LDA, LIWC, MeaningCloud and DA (SFL). These are approaches to implement a combined approach that responds to the research questions. Krishnamurthy (1996) suggests that analysts need to guard against looking at the corpus compared to looking for the things expected to be found or compared to looking for things not supposedly discoverable in the corpus.

Norvig (2011, 2016) suggests that for researchers in AI with machine learning (using statistical approaches) that try to produce behaviour that mimics something in the world, they need to understand the meaning of that behaviour. Norvig reconciles Chomsky's (1969) doubts about patterns as he suggests that having a context can counteract the problems about the sense of language patterns versus a general theory of language use.

For example, if CDA (Fairclough, 1995; Baker, 2006) is only used, then it can help to show what the patterns found mean about the social power relations (context).

However, the usage of SFL can also help to show the stance-taking in seeking or giving advice systems.

However, the use of machine learning from AI is part of the more extensive debate about finding patterns automatically in large-scale corpora with AI versus finding the core truths about language and meaning (Chomsky, 1969, Norvig, 2011 2016). Applying Norvig's way of thinking, for example, machine learning can help to find instances of 4gram words used mostly in some large-scale corpora of text. For example, a 4gram '*how can I find*' may be found and is used by most people in some corpora. However, Norvig cites Chomsky and argues that there is no explanation given as to why that pattern should be employed rather than any other pattern. He argues that patterns should also reveal the context of the employment of language by people for some specific social activity.

The thesis proposes that Blei et al.'s (2003) LDA model may be applied to the diabetes corpora to model the posts and their 'topics' or 'TOPICs'. It can indicate '*about how and what*' people converse about and the possible links to for, example, linguistic forms of advice, stance-taking and support. It uses AI machine-learning and employs topic-modelling as well as a computational-math-based model that is related to the computational linguistics study of language. Blei et al.'s (2003) machine learning with LDA topic model is about the employment of a probabilistic model of language. It is to automatically find topic patterns of common usage (see Chapter 2 – Literature Review) in corpora. The LDA approach gives an entry point into a large-scale corpus and can be used to begin analysis.

In applying Blei's (2012) way of thinking, a quantitative LDA model can make sure that the research will report on the features that have high frequency across all the posts. He would see a post as containing many topics, in the common usage of this term for LDA analysis. People converse with each other about many things at once, although mostly about one topic.

Quantitative approaches also need to show, to what extent, the corpus did or did not confirm the researchers' expectations (Lee and Seung, 1999; Kireyev et al., 2009; Ramage et al., 2010; Blei, 2012; Caren, 2015; Srinivasan, 2015). Different quantitative approaches can help remove researcher bias.

A probabilistic LDA model (Blei, 2012) can be compared to a deterministic NMF model (Lee and Seung, 1999). Lee and Seung describe NMF as about finding two non-negative matrices (W and H). Its product approximates the non-negative matrix X. They use this factorisation for dimensionality reduction or topic extraction. Topic-modelling with LDA versus NMF is about the utilisation of a generative model (Blei et al., 2003). Blei et al.'s model type allows for sets of instances to be made explainable. For them, it is why some parts of the dataset are the same. It is applied to this research. In cases of observations (that are about words collected into posts), then the model guesses that each post is a mixture of a few topics. Each word created is about one of the post's topics. Lee and Seung (1999) argue that NMF results can look like LDA. In keeping with Blei et al. (2003), LDA is related to the study of how likely or unlikely it is that things are to happen about a model. It can express doubt about the placement of topics across the posts' texts and for the assignment of words to topics.

LDA is an unsupervised algorithmic approach. NMF uses a predetermined algorithm. It can arrive at a single representation of the corpus (Dariah, 2015, 2016).

LDA (Blei, 2012) and NMF approaches can help with dimension reduction. They both provide clustering for topic identification. The researchers above demonstrate that both LDA and NMF can arrive at a representation of a corpus related to uncovering topics. As per Blei, the basic idea is to describe a post as a mixture of different topics and make use of LDA. For him, a topic is merely a collection of words that often happen with each other.

Therefore, by applying Blei's unsupervised learning, the grouping of Facebook posts can be done such that each member of each group can have the same meaning. Blei suggests that training a topic model with unsupervised learning saves the effort needed for creating labelled data and training classifiers using such labelled data for a corpus. His guidelines in the model are useful in the discovery of topics. Topic models help identify hidden relationships in the data as topic models or as n-gram models (Steyver and Griffiths, 2004, 2007).

The use of LDA validation processes is needed. The confusion matrix, together with precision-quality, f1-score, and recall, is used to validate the model (Scikit-learn, 2011; Grimmer and Stewart, 2013; Scikit, 2015). These tools are used to measure the relevance of the model for an understanding of support. LDA results are validated with a confusion matrix from Scikit-learn (2015). It uses a logistic regression model, and the validation model includes metrics of accuracy, recall, and precision scores. They define recall as the ability of the classifier to find all the positive samples. The precision is about the capacity of the classifier – not to label as positive a sample that is negative.

The accuracy score is where this function computes subset accuracy for the set of labels. These sets are predicted for a sample that accurately matches the corresponding round of accurate labels. These ways of modelling and validating the model of corpora follows the convention from Scikit-learn (2015) machine learning for Python programming and that of Caren (2015).

To, label the LDA model 'topics' and 'TOPICS', relevant linguistic research is used to compare the trigrams and random samples of posts from within a 'topic'.

Features with high-frequency on, for example, advice-seeking, advice-giving, are further used to help categorise the posts (see Section 2.2).

The aim of those as mentioned earlier multi-quantitative and qualitative approach is checking the results from the LDA machine-learning model against fundamental language-based research processes and patterns. The CL study of language ultimately offers a way to collect suitable data and, for example, to focus on the keyness of parts of the data. The corpus is robust because of the 226,385 posts that have been collected.

The CL mixed approaches can be used to test for language patterns. These are suggested by the above linguistic researchers, to discover, for example, advice-seeking, advice-giving and stance-taking patterns. It can help to see new things in a particular corpus. The identification of notable features in the discourses is made in an automatic, quantitative manner. However, a qualitative DA analysis helps to make sense of why these quantitative analysis results matter.

DA is used to analyse the corpus (Hyland, 2005; Hood, 2004, 2007, 20101, 2011; Harvey and Koteyko, 2013) and is also used to formulate the research questions and a process for answering the research questions about the particular diabetes corpus.

The DA qualitative approach is used to analyse the LDA trigrams, 'topics further' and 'TOPICS', targets and its text. The analyses are carried out and presented in Chapter 4. In applying DA analysis, it may also allow for further validation of the quantitative findings against the qualitative findings. It may be able to consider the social relations that play a part in creating such salient features in a corpus (Fairclough, 1995; Baker, 2006).

In considering the research of Hashmi (2012), DA may allow for the broader effects of the results of the LDA model to make sense. It is regarding the context of social relations about frequent types of 'topic' or pragmatic meaning, i.e., meaning in context usage as examined in the posts. Fairclough (1995) and Baker (2006) offer established ways to use corpora and DA to analyse language for features to deal with questions on the social relations of people. DA views language as a form of social practice (Fairclough, 1995).

3.7 The approaches to analysis

The approaches are used to find a support pattern, and the thesis suggests that this is a powerfully analytic and detailed thinking tool. The steps to analyse the data are listed in Figure 3-2.

3.7.1 Latent Dirichlet Allocation LDA

A primary approach to using LDA is to relate the research question to what can be achieved by the algorithm, i.e., the computer code, regarding the data analysis. A query to analyse the data is called an operationalised research question in this thesis. Furthermore, the decision-making process of the data analysis includes employing sensible guidelines for running the LDA. It is an algorithm that uses hyperparameters, and they are discussed later in this section. The process mentioned above is employed in helping to answer the research questions. It is shown later in this chapter as well as Chapters 4 and 5 about the FDUK.

The employment of machine learning is used to analyse large-scale, significant data within a reasonable timeframe instead of performing laborious manual analysis and findings, of many items across the data over longer human days. However, the assumptions of the model must be thought through and made explicit, as the use of topic-modelling is not straightforward.

Previously established research informed the LDA analysis (e.g., Blei et al., 2003; Griffiths and Steyvers, 2004). The main idea is of assigning documents in a corpus (set of documents) to latent (hidden) topics based on a vector of words, and this idea can help to understand the LDA model. In sampling-based algorithms, Gibbs sampling is the most commonly used and is about approximating the posterior with samples. Collapsed Gibbs sampling is an estimation for LDA. The process of using collapsed Gibbs sampling is described in this section. A description is given later for how the analysis is done.

A typical evaluation of LDA models involves the calculation of perplexity (Blei et al., 2003). An understanding of the fields of human conversation, diabetes discourse and AI may allow an intelligent guess as to a suitable number of LDA topics (k). Perplexity is a statistical measure of how well a probability model predicts a sample. It can be applied to LDA, for a given value of k, or an estimate of an LDA model. Then, given the theoretical word distributions represented by the topics, it can be compared to the actual topic mixtures or distribution of words in the documents. However, the statistic is somewhat meaningless on its own. However, the benefit of this statistic comes when comparing perplexity across different models with varying values of k. The model with the lowest perplexity is generally considered to be the 'best'. Perplexity measures modelling power. It is by calculating the inverse log-likelihood of unobserved documents as a decreasing function with higher likelihood, meaning a better model. Better models tend towards lower perplexity. They can suggest fewer uncertainties about the unobserved document. LDA is a widely used topic-modelling approach and has different algorithms for carrying out the analysis.

Moreover, as the number of topics increases, the LDA model becomes better, i.e., the perplexity decreases. The domain knowledge of the researcher can be essential in thinking about a suitable number of topics, or to try different values of k when interpreting the results. Perplexity can be used as one data point in the research decision process. This thesis considers investigating the topics and posts themselves and the highest probability of words associated with each one to determine whether the structure makes sense for the particular domain. The ground truth, or having a known trigram ‘topic’ or ‘TOPIC’ structure that it can be compared to, maybe useful but is not very evident in the literature.

The issue of stopwords is as crucial as the issue of punctuation when removing these to carry out LDA analysis. The set of ‘English’ stops words can be removed. After checking the results, some words that may be still shown in the corpus due to the ‘cleaning’ of the data can be the remnants of usernames or other items like HTTP addresses. Another list of ‘words’ was my_words = set([‘au’, ‘bd’, ‘ed’, ‘re’, ‘bu’]) that had to be removed.

The big problem is that LDA parameters and corpus may generate different topics every time (e.g., Scikit-learn, 2015). The researcher, therefore, needs to stabilise topic generation. The aforementioned research scientists in topic-modelling have demonstrated that the use of parameters is essential (e.g., alpha=0.01, eta=0.01, n_iter=200, random_state=1). An issue that is under discussion both in AI and linguistics is that topic-modelling in experiments can be used to produce models that vary in performance. It can be when the random state of the algorithm is changed. The difficulty is about using, for example, random state as a hyperparameter. It is not critiqued enough as to what these hyperparameters are. If a model outperforms others with a different random state, it should be considered that this particular model has overfitted a particular random state. However, hyperparameters are shown to be variables that control some high-level aspect of an algorithm’s behaviour. It is suggested, therefore, that, as opposed to normal parameters, hyperparameters cannot be automatically learned from the training data by the algorithm itself.

For this reason, a researcher can select an appropriate value based on his/her domain knowledge. It is for the semantic meaning of the hyperparameter (if any). This thesis uses different hyperparameters to get going but similar to the established research on topic-modelling. The bias in selecting hyperparameters must be counteracted. The random state is used in many randomised algorithms in Sklearn (2015) to determine the random seed passed to the pseudo-random number generator. Therefore, it does not govern any aspect of the algorithm's behaviour.

Consequently, random state values which performed well in some validation sets may not correspond to those that would perform well in a new, unseen test set. Indeed, depending on the algorithm, it might give completely different results just by changing the ordering of training samples. This thesis uses random state values but uses the same set for all of the experiments. An alternative method is to take the average accuracy of the models over a random set of random states. The thesis tries not to optimise random states, as this can produce optimistically biased performance measures.

The alpha has default 0.1, and it is the Dirichlet parameter for distribution over topics. The eta has a default of 0.01 and is the Dirichlet parameter for distribution over words. The random state can be an integer or a RandomState and is the generator used for the initial topics. A way of achieving the same training and model results is by setting the random state parameter. It can be set to the same state each time. For, example, in creating a new LDA model with any number assigned or the default 0 or 1.

The collapsed Gibbs sampler runs for 200 iterations (a slightly conservative number in order to ensure convergence). A visual inspection of log-likelihood shows that the algorithm has converged after 200 iterations with decreasing loglikelihoods given later in this section. The Python code takes about 12 minutes per run on a laptop using a multicore i7 2.5 GHz processor and 16 GB RAM. Code and physical computing must also be considered. It can limit or enhance an aspect of doing large-scale text analysis. More computing power and better code can make the process quicker. It can make large-scale text analyses possible in the first instance.

Validating the LDA model, involved a confusion matrix, and 96.2% of the data were correctly classified.

This thesis has to consider what the topics mean. It is an assumption in the literature that, generally, topic-modelling can find human-readable structures in unstructured textual data.

Firstly, the thesis uses '*device-enabled discourse purpose categories*' analysis, i.e., the usage of LDA and removing the stopwords was carried out for 25, 50 and 100 topics. Initialising and a fit for the LDA model involved a choice of the number of topics. It is an unsupervised training model. However, after assuming a number, the log-likelihood can be checked against other topic number choices. The higher the log-likelihood can be, the better the model. The 25 topics had a log-likelihood of -1325757, and the number of documents was 225,360. The vocab_size or max features was 500. There were 202,137 words due to the 'cleaning' process and keeping the stopwords. The number of iterations per run was a conservative 200, but this meant that the LDA could converge statistically and fit a model to the corpus.

These descriptions were similar for 50, but with the lower log-likelihood of -1316169, so 50 topics are the model used in the thesis. These descriptions were similar for 100 topics but with a log-likelihood of -1322230. A total of 50 topics were run ten different times (similar to 10-fold validation in the research) and gave the same results for ten runs and with same precision matrix scores. This result is due to the hyperparameter sets and to the 200 iterations required for the model to converge.

Secondly, the thesis uses '*device-enabled discourse content categories*' analysis, i.e., the usage of LDA and removing the stopwords was carried out for 25, 50 and 100 topics. For initialise and fit the LDA model, a choice of the number of topics is made as it is an unsupervised training model. However, after assuming a number, the log-likelihood can be checked against other topic number choices. The higher the log-likelihood can be, the better the model. The 25 topics had a log-likelihood of -213803, and the number of documents was 225,360. The vocab_size or max amount features for the study was a conservative 500. There were 37,824 words due to the 'cleaning' process and removal of stopwords. The number of iterations per run was a conservative 200, but this meant that the LDA could converge statistically and fit a model to the corpus. These descriptions were similar for 100 topics but with a log-likelihood of -225196. These descriptions were also similar for 50 topics, but the 50-topic model gave a log-likelihood of -216796. The log-likelihood result is lower than the 100-topic model, and this was used in the thesis for further analysis even though it is slightly higher than the 25-topic model. The thesis assumes that people talk about many topics. In a large-scale corpus with many users, there would be many topics. A significant but moderate fifty topics were used.

Ten different runs were done for the 50 topics due to an issue of the possibility of getting a different LDA model for each run. Even though the same results of trigrams and topics were gained for the ten runs, it is worth noting that the parameters used the iterations, and the alphas and betas.

Details about the LDA approach are given, as to how the LDA was configured, leading to the topics.

The use of stopwords and no stopwords in the analysis is a different way of looking at the purpose and content of the diabetes corpus.

The LDA algorithm is run for the parameters with conservative features (500), the number of topics (50) and the number of top words (10). These yield practical results (shown in Appendix Table 7-2) from having tried for smaller amounts of topics or having tried with more massive amounts of topics. This process is for dimensionality reduction. The knowledge of the domain helps to find topics that make sense. The LDA used collapsed Gibbs sampling. It is explained in Blei et al.'s paper (2003). LDA is concerned with generative probabilistic semantics. Inference using collapsed Gibbs sampling is explained in the paper by Griffiths and Steyvers (2004).

Topics and posts are selected randomly for further comparison from the LDA results and are shown in Appendix 7-1. The LDA labelling summaries are presented in the tables in Appendix 7-1. A random sample of '*device-enabled discourse purpose categories*' or 'topics' gives 'topics' 0, 24, 13, 31 and 34 for analysis from a total of 50 'topics'. 'Topic' 0 has 2,941 posts, 'topic' 24 has 1,789, 'topic' 13 has 1,966, 'topic' 31 has 1,904, and 'topic' 34 has 2,332 posts (see Appendix Table 7-8). Added together, these give 10,932 posts in the sample from a total of 225,360 posts for LDA analysis.

The decision to study the predominant and salient features meant using conservative 500 features. The posts that were thus clustered into these 'topics' of high frequency totalled 92,254. The remaining 133,106 posts were not clustered into these topics of high frequency. The research could, therefore, focus on highly used features across the corpus, 'topics' and posts. The total random sample consists of 10,932 posts from 92,254 posts and is used for further analysis. The 10,932 posts were narrowed down further into consecutive posts or threads. These posts have consecutive post-id-numbers, so it meant they linearly and chronologically followed each other in the corpus. These posts could then be studied further for the potential posting of advice with stance-taking features. This above analysis then gives seventy-three posts and their substantial quantity of text for further analysis (see Section 4.2.1 and 4.4.). Table 7-6 in Appendix 7.1.6 displays the post-id-numbers from the entire corpus in Column 1, the post-id-numbers in Column 2, the original posts in Column 3, the LDA with stopwords in Column 8 and the LDA without stopwords in Column 7.

It is not easy to label trigrams or n-grams in connection with a 'topic'. The researcher's knowledge of the domain can help, so trigrams, 'topics' and posts can tend towards suitable labels. It is best left the way the LDA presents the trigrams. CL allows for this iterative way of analysing corpora til several ways tend to give almost comparable results (Baker et al., 2008). Similar posts also tend to fall together in the LDA model (Blei et al., 2003).

For the validation of the LDA results, a confusion matrix from Scikit-learn (2015) is used with a logistic regression model (see Appendix 7.2.4 and 7.2.6).

3.7.2 Automated entity recognition and sentiment analysis

Segura-Bedmar et al. (2015) have used MeaningCloud when exploring health in social media groups and for detecting drug effects. MeaningCloud can be used to find the nouns/entities (targets) of each post and the global sentiment and is carried out in the research (MeaningCloud, 2017). The values are assigned to the 73 posts (see Appendix 7.1.6, Table 7-6). They can be positive, neutral or negative sentiment values, targets or entities. As well as offering a topics extraction feature, MeaningCloud allows the researcher to obtain the polarity associated with the entities and the concepts in the text. A sentiment model in MeaningCloud is composed of a collection of entries (Appendix 7.3). Entries are defined by a word or multiword, i.e., a group of words that can appear together in the text. Entries consist of their definition and the sentiment behaviour associated with them. They are a simple way of studying a very complex phenomenon.

The research thus uses the aforementioned automatic annotation strategies to select nouns for further analysis. The approach above is usually referred to as aspect-level sentiment analysis (MeaningCloud, 2017). This detection process is not easy to automate. The software combines NLP techniques and uses 'TOPIC' extraction for extracting the different elements that can be present in sources of information. The TOPIC detection process is also not straightforward. It is carried out by combining sophisticated NLP techniques that can allow obtaining a morphological, syntactic and semantic analysis of a text.

It can then use them to identify different types of significant elements. These can be classified according to predefined categories: entities, people, organisations, places, phone numbers, concepts, significant keywords in the text, time expressions, money expressions, quantity expressions [beta], any other expressions, alphanumeric patterns, quotations and relations.

Therefore, the global sentiment analysis that is possible to do involves providing a text and using it to determine whether it expresses a positive, negative or neutral sentiment. In order to do this, the local polarity of the different sentences in the text needs to be identified. The relationship between them is evaluated and results in a global polarity value for the whole text, or this case, the individual posts that can consist of many sentences. Besides polarity at the sentence and global level, sentiment analysis uses advanced NLP techniques to detect the polarity associated with both the entities and the concepts within the text. It can provide a reference in the relevant sentence. It can give a list of elements. It is detected with the aggregated polarity derived from all of their appearances. It takes into account the grammatical structures in which they are contained.

LIWC is used in a range of studies (e.g., Pennebaker and Davison, 1997) for chronic illness support. LIWC (2016) uses dictionaries from classes of terms related to research on psychology. An example of its use is for terms that can be related to positive or negative emotions. Positive emotion, negative emotion, certainty and health are all LIWC categories that are relevant to the thesis. These are identified through the research annotation strategies of using LIWC as automated software capable of finding terms related to these concepts. It is carried out on the seventy-three posts, and their substantial quantity of text and the values and categories are added to the individual posts (see Appendix 7.1.6, Table 7-6).

The LDA, MeaningCloud and LIWC findings are used together, in combination with the thesis to identify and understand advice, stance-taking and support amongst the peers in the FDUK. The literature review and the data analyses suggest that the advice and stance of the post can be shown to be related to the noun/entity (target). It is also related to the '*device-enabled discourse category*' as resources, '*device-enabled discourse content categories*' as objects. It covers the entity of healthcare. It is with affective stance related to positive/Negative Emotion, and global positive/Negative sentiment. It is with an epistemic stance related to Certainty.

This means, therefore, that the research can identify the elements of advice with stance-taking. The concern is who is involved, the resources, objects, and relationships over time in the posts. These are proposed in the thesis to be elements of a pattern of support. These patterns of support are also evaluated further in this thesis.

Precision testing is used to validate the findings and the results of the LDA. Comparisons of LDA, MeaningCloud and LIWC are made in the further analysis in Chapter 4, Section 4.2.

3.7.3 Linguistics forms, advice, ‘topics’ and ‘TOPICs.’ and targets

Partial manual work is needed to document and further label the support-data (see Table 7-7). The vast majority analysis is automated. It includes LDA ‘topics’ and ‘TOPICs’ as well as their corresponding trigrams, and targets. They may be matched to other relevant state-of-the-art research on advice and support in the diabetes domain (see Chapter 2 – Literature Review, and Appendix 7.2).

It includes the selection of targets in the seventy-three posts and their substantial quantity of text. The thesis focuses mainly on domain-specific targets of diabetes, blood, and pumps from many others ranging from targets of usernames, places, time, to years and is used in further analysis.

The trigrams, topics and posts were then labelled according to the types of advice-seeking, advice-giving identified in them (see Table 7-7 and Section 2.4.2). An example of advice type from state-of-the-art research (see Table 2-1) is ‘problem disclosure’ (AA_P). In contrast, another example (see Table 2-2) is ‘description of personal experience’ (AG_E), which gives an account of how the advice-giver approached the situation that the advice-seeker had described. These manual analyses and identification must be made carefully. Also, it is worth considering that ‘problem disclosure’ can be potentially ambiguous, as it can be interpreted as a request for advice, sympathy or solidarity

In this thesis, instances of the best available research design are employed and used as guides for specific and broader analysis. Other significant studies can give explanations of what this thesis refers to as ‘topics’ and ‘TOPICs’ and targets. It is together with a knowledge of the diabetes domain (Pennebaker and Davison, 1997; Brownson and Heisler, 2009; Harvey and Koteyko, 2013, p. 185; Hunt and Koteyko, 2015; Hunt et al., 2015). For instance, the use of examples from state-of-the-art research are shown in Table 2-3. They may help to identify the LDA *content* words that people use in their conversations (e.g., the word ‘diabetes’).

The keys used to label the features in the text for each post are provided in Appendix 7.1. These keys also help with the DA analysis in the next section. They contribute to identifying where and how in posts, the ‘topics’ and ‘TOPICs’ (i.e., the participant’s plans and content) are employed by the different peers to achieve their healthcare support aims.

Automated analysis, annotation, partial manual analysis and labelling are LL essential for discerning evidence of potential patterns in a corpus. However, the aim, as detailed in the next section, is to try and explain these possible patterns.

3.8 Threats to validity

This section examines the threats to the utilisation of the aforementioned mixed approaches of data analysis.

The data used, along with other data, is freely available on the internet but is also governed by strict ethical guidelines in its usage for research. It is also part of an extensive amount of big online data. The large-scale data could provide examples of language used by people in natural conversations. The study must make sensible choices for data and uses state-of-the-art research for data selection. It also develops an adapted process for carrying out the strict ethical guidelines associated with it. These two aspects of research can help in the selection and making use of appropriate data for research. The examples of data collected from FDUK can determine a lot about the behaviour of people regarding their diabetes care.

All the processes can be more challenging and tire without first understanding the benefits and limitations of the individual approaches. In the end, the results of the use of the mixed approaches are shown to compare with each other (e.g., LIWC findings versus LDA findings). The issues are considered when using this multi-method way for conducting the data analysis. The employment of a quantitative and qualitative approach could mean that the strengths of each other may improve the strengths of each one.

The corpus is vast and rich, and a methodical approach to data analysis and the formulation of appropriate research questions are essential. In considering the different ways of doing analysis gives an approach to collect the data and begin the analysis with LDA. An approach to beginning the analysis of a dataset is an issue that is important for beginning any type of analysis, and Blei et al. (2003) offers just such an approach. The process of how to begin the analysis needs to be carefully thought. It is such an entry point, and the use of different software for analysis can begin. Various approaches from a quantitative analysis viewpoint can help with finding an entry point. LDA is such an approach. However, at various points in the analysis, there need to be various sensitive measurements of the data management process and analysis (Miles and Huberman, 1994). An approach to formulating research questions and a process for answering them does benefit from the use of DA, as it can link the right questions with the right corpora.

The use of LDA for linguistics analysis must be approached with caution. Hence, a novel way of understanding the resulting topics from LDA and by not taking them for granted is through the notions of 'topics' and 'TOPICs', as used in this thesis.

These are, respectively, called '*device-enabled discourse purpose categories*' or resources and context of support; and '*device-enabled discourse content categories*' or resources and content of support. It is also not easy to assume what the number of LDA topics could be, but log-likelihood can help with the number of topics. A researcher's knowledge of the domain can also help to determine what the number of 'topics' and 'TOPICs' should be, and that they make sense. The same goes for the choice of n-grams. In this case, the choice was set on using trigrams.

The labelling of trigrams and 'topics' and 'TOPICs' and posts within them can tend towards a suitable label. Nevertheless, only after comparisons to the existing research of the domain, so it is best left the way the LDA provides them. CL does allow for this iterative way of analysing corpora until several ways tend to give almost comparable results. The assumption is similar to the LDA method, where posts are assumed to eventually cluster together under sensible 'topics' or 'TOPICs'.

There are different procedures used for different software and analytical tools. Python, R programming, LDA algorithms, MeaningCloud and MS Excel are used in this research. For example, text-mining researchers like Blei et al. (2003) use LDA, and it proves useful for the analysis of the diabetes corpora. Comparisons of results are from different software. They are used all together in a CL approach, and in answering questions that may produce comparable results. It can help a researcher to validate result (Stubbs, 1994; Krishnamurthy, 1996; Baker et al., 2008).

This thesis can use the LIWC (2016) results and compare them to the LDA and DA results for validation if they arrive at similar results. DA qualitative results help to make sense of the quantitative findings and vice versa. In the end, quantitative and qualitative analysis can broaden and deepen the overall findings.

Software-aided partial manual analysis can complement the lack of appropriate AI and automated software analysis. It is of the large-scale and complex linguistic features, thereby creating a combined process for conducting CL. The concern is the wide range used to identify linguistic concepts like ‘advice’ manually and automated identification of concepts like stance-taking but by breaking it into its elements, then identifying each element and combining them. More work on AI and Linguistic analysis is needed for the future; the limitations of the mixed method are critiqued in the next section.

3.9 Limitations

AI and topic modelling alone can produce too broad base categorisations of text data. What are the potential sophisticated linguistic concepts that underly these topic modelling patterns? However, AI can help reduce the big data and large scale text corpus from many potential linguistic dimensions to a few high-frequency ones. In counteracting these potential platforms, analytic tools and their difficulties, the thesis will bring artificial intelligence and a necessary linguistic analysis together. Such an approach may help to resolve these challenges but primarily help to develop a greater awareness of the inherent limitations. In overcoming the challenges, such an approach may firstly find a place in corpus linguistics and secondly a useful theory of support. Searching for an ‘architecture’ of these types of discourses is analogous to the search for an architecture of sentences. Linguistic forms and language are highly sophisticated. It is done to make large-scale text data analyses possible in less time and effort. It is not, therefore, a fully developed field of endeavour as yet. The study’s methodology is limited but tackles the difficulty of employing features of APPRAISAL Theory in automated analyses. It concerns the coding of the data into broad categories of advice-seeking and advice-giving with stance-taking. It uses an identification of potential elements of stance-taking in the posts. It combines them to show that postees (persons who post on social media) produce such a sophisticated linguistic concept. The development of the theory is based on only a small number of the topic number of posts and 'topics by the very nature of using primarily AI and a Linguistic analysis to guide the process. AI offers to reduce big data which has a potential for many dimensions into a few high-frequency clusters.

Even though topic-modelling can produce clusters or groups, the identification of topics can be dependent on the knowledge that the researcher has of the domain under study. Knowledge of the field of study was used to overcome issues with the automated analysis of 'topics'. It is in using existing research that identifies the types of 'topics' and 'TOPICs' that people with diabetes could potentially utilise.

A methodological challenge that the research faced concerned the immense amount of data available for analysis and the unstructured text data or the non-consecutive postings of peers. People can post at any time and comment or reply anywhere in the social media Diabetes UK page. Posts do have a linear number order based on time, but people may not follow linear comment postings all of the time. My thesis calls 'comments, replies and post', posts.

It was necessary to determine where, in the data, the analysis should start. The LDA topic-modelling approach helped with entry points into the data. It had been used in this manner in another similar research. LDA is a way of using analyses related to the study of how likely or unlikely things are to happen, and with algorithm-computer codes to analyse the data. The ideas thought to be true of algorithms/computer codes for language-based questions need a comprehensive listing of their proper usage. Topic-modelling on its own does not give context, and DA can help to give context to patterns that have been discovered. Automated content analysis, an annotation for entity recognition and sentiment for advice and stance-taking are used to help to identify aspects of support. A mixed-method quantitative methodology helped to overcome the limits of any single method.

The qualitative DA approach can be used to analyse the randomly selected topics and posts from an SFL perspective. This thesis used both quantitative and qualitative analysis in order to analyse the data in a broader and more in-depth way. Qualitative analyses are carried out on the quantitative results. The quantitative results are gained automatically by software; and also, aided manually for the identification of advice features, and comparisons of results to the state-of-the-art research in the diabetes domain. Corpus linguistics deals with such approaches to analysis. AI and software-automated analysis are shown to have limitations for the identification of sophisticated linguistic features. DAs are shown to benefit from AI and software-automated analysis.

LIWC does word counts but uses dictionaries to help find concepts within the data and is developed through research in the psychological field (e.g., words about mood and certainty).

Qualitative analysis needs samples of posts from within identified topics. These are labelled broadly by the researcher. These posts are selected randomly, so that posts selected may deal with other essential language features that can either substantiate or not substantiate the claims.

Random numbers in MS Excel are used to select random topics and posts, and other random number selections (for example, from Python) may result in other posts for analysis.

The use of different approaches (e.g., AI LDA topic-modelling, and DA qualitative analysis, and MeaningCloud and LIWC) can help with cross-checking to guard against the issues presented above.

No single software package offers a multiple, language-based analysis of corpora, especially for finding the advice and stance-taking features. The future development of a periodic table of language-based research types with comparisons to the capabilities and algorithms of existing software types will, therefore, be instrumental. It could be used for integrated analytical capabilities, such as for stance-taking. It may also help to reduce the need for any manual qualitative analysis. It may not be sufficient as qualitative analysis could always be necessary.

3.10 Conclusion

The thesis demonstrates a novel DA method to analysing the data, and a suitable corpus and analysis for the particular research questions.

Definitions of what is analysable must be considered in any approach. The annotation process can help in this regard. It can be used for sifting out concepts relevant to the analyses, which are mainly done automatically with AI software. It can also aid manual qualitative analysis.

Ethics, data collection and cleaning of the data is also of great consequence as the research considers what to leave out and on what to focus. In this case, for example, it considers the use of stopwords or otherwise in the AI LDA topic-modelling analysis. The thesis presents descriptive statistics of the corpus under analysis.

A quantitative CL method is integrated in order to analyse the high-frequencies of linguistic forms such as advice with stance-taking, questioning, humour/sarcasm patterns in the corpus. The methods include the use of LDA, LIWC and

MeaningCloud, which can ease the identification of possible elements of advice and stance-taking patterns.

A qualitative DA approach is taken to analyse the quantitative high-frequency language features found in the corpus. The framework used is based on the SFL and appraisal theory.

A research approach that helps to operationalise the research questions and collect useful data for analysis is shown to be from a CL study of language that includes DA.

Machine learning can be employed together with CL to analyse the vast diabetes corpora quantitatively.

Critical linguistic research is further employed to identify the LDA 'topics' and their posts, as well as 'TOPICs' and advice and stance-taking patterns. These analyses help to discover any support categorisations and devices that people utilise to converse for support.

DA helps to place the 'topics', 'TOPICs', targets and advice with stance-taking and posts of the topic model in the context of people's social relationships. It is a holistic way of explaining why particular categories and devices are crucial for some of the conversations that people have for their online healthcare support.

4 DATA ANALYSIS AND FINDINGS

4.1 Introduction

LDA and DA are used together and are of primary importance to the analysis. These analyses are done on the many and varied user/peer posts as one entity and then also via a random selection of topics, TOPICS, targets and individual posts with substantial text. Primarily LDA, MeaningCloud and LIWC are used together for topic-modelling, automated content analysis and annotation, entity recognition, sentiment analysis with dictionaries, and software-aided manual DA. These processes can ultimately help to identify, for example, advice with stance-taking and many other interactional activities amongst peers. A later section of this thesis will show the combined analyses concerning targets in posts. It is done in table form. It is in order to compare the results from the different analyses. It is to see whether the trends found from each one all make sense when taken together. The thesis focuses mainly on domain-specific targets of diabetes, blood, and pumps from many others ranging from targets of usernames, places, time, to years. The LDA is tested with a precision measure, and a brief overview of the process involving the mixed methods that are used for the data analysis is given in Table 4-1. Other sections explain each of the multi-methods analysis processes. Together, these make up the overall analysis of the FDUK.

The research questions are restated, and evidence from the data is provided that will assist in answering the research questions. The LDA model helps to find linguistic patterns identified in random samples of topic/TOPIC posts. DA and SFL analyses (shown later) help to answer the research questions and formulate the research claims.

A summary of the results shows in the sections that follow. The analyses are carried out with CL that involves mixed quantitative and qualitative approaches. A section on the DA is provided to help make sense of the quantitative patterns identified in the corpus.

This chapter demonstrates how the analyses proceed and details the findings. The significant findings are summarised with different examples of the analysis and results from the data.

A summary of the analysis steps is given in Table 4-1. The analysis and results are presented throughout this chapter (Appendix 7-1). It concerns potential answers to the research questions, and they are explained in detail in the sections that follow.

Table 4-1: A summary of the data analysis approach

1. **Selecting and collecting the posts (a corpus of text)**

A suitable large corpus of 218,068 posts and about 16,137 users were selected for analysis. A process for keeping to ethical standards, informed by other prominent ethics researchers was applied in this thesis (e.g., for how the data was collected and stored as a corpus). These processes are crucial decisions in CL (see Sections 2.5.3, 3.4, 3.4.1 and 3.5).

2. **Cleaning the text in the posts**

This cleaning process involved the removal of punctuation, whitespace and stopwords (or not). These are all crucial steps in analysing any corpus (see Section 3.5.4).

3. **Automatic annotation**

The evolving automatic annotation of the data was carried out through the use of different software to select salient language features. This step is vital in analysing any corpus (see Section 3.5.5).

4. **LDA with stopwords**

LDA with stopwords was used on the large 218,068 posts and their constituent text data corpus with 500 features and 50 topics to form a trigram model (device-enabled discourse purpose categories) – 'topics' (see Sections 3.7.1 and 4.2.1).

5. **LDA without stopwords**

LDA without stopwords was used on the corpus with 500 features and 50 topics to form a trigram model (device-enabled discourse content categories) – 'TOPICs' (see Sections 3.7.1 and 4.2.1).

6. **Random samples**

Random samples of topics and posts were identified for further analysis (see the tables in Appendix 7.1, Section 3.7.1 and 4.2.1).

7. **Discourses analysis: Focus on advice, but many more: events, humour/sarcasm, questioning, emotion, hope, and charity and many other interactional activities amongst peers**

Potential advice and others, for example, events, humour/sarcasm, questioning, emotion, hope, and charity patterns are matched to posts text data. It is suggested in the literature. It is carried out with minimal aided manual DA (see Section 3.7.2 and Column 4 in Appendix 7.1. tables).

8. **LIWC analysis**

The LIWC analysis was carried out for healthcare, positive emotion, negative emotion and certainty (see Sections 4.2.4 and 4.4 and Column 5 in Appendix 7.1. tables).

9. **MeaningCloud Analysis**

The MeaningCloud analysis was carried out for the global sentiment and crucial entities/targets such as Diabetes, Blood, Pump, Medication. The thesis focuses mainly on domain-specific targets of diabetes, blood, and pumps from many others ranging from targets of usernames, places, time, to years. (see Sections 4.2.3 and 4.4 and Columns 5, 9, 10, 11, 12, 13 in Appendix 7.1. tables).

10. **Discourse analysis of potential advice with stance-taking but many more: events, humour/sarcasm, questioning, emotion, hope, and charity within consecutive and non-consecutive posts**

DA was used to understand and observe the instances of primarily advice and stance-taking and other discourses. It is to compare threads for advice with stance-taking from a DA SFL perspective (see Sections 4.3 and 4.4 and Tables 4-3 to 4-7) but as a guide for the discovery of other patterns.

4.2 Quantitative Analyses

This section 4.2 provides a summary of each quantitative analysis for all posts. It also gives a summary of the findings of the diabetes corpora.

The next section gives a summary of each target, and it gives a combined overall sampled non-consecutive and consecutive post quantitative analysis and the findings from the corpus. The combined sampled quantitative and qualitative analysis is presented in Section 4.4. for all posts in each target.

4.2.1 AI LDA: Primary findings of device-enabled discourse purpose categories and content categories

Careful consideration must be given to the LDA process and the findings. The exciting thing for my thesis is that LDA clusters similar posts under a set of top trigrams. In effect, the highest frequency top trigram maybe what all those posts are inherently about, even though the human eye would not see the posts as immediately similar. As discussed in Section 3.1, different numbers of LDA topics numbering 25, 50 and 100 were tried out. It was before a decision was made to use 50 topics, in line with the log-likelihood calculations and comparison, to the state-of-the-art research findings. The results of 'topics' and 'TOPICs' compared with what is expected of them from healthcare communication studies and other researchers' work concerning diabetes, such as the results shown in Table 2-3.

Similarly, a smaller and larger number of n-grams was tried out. A total of ten runs were performed on the 50 topics. It is in order to check the consistency of the trigrams given per topic number, and the results were seen to be similar. This thesis considered trigram frequencies for use, as it can have a combination of words that may be interpretable in context. Unigrams and bigrams would not provide enough context, but trigrams could be optimal without exhaustive computing (see Section 2.5.1). The model produces the most common combinations of trigrams. These are three words that mostly follow each other. So, it is not only looking at finding clusters of single words. Nor any other number of combinations or more massive n-grams.

Stopwords also play a crucial role in this thesis. They are deemed to be relevant to the overall understanding of the corpus. It suggests that they should not be left out of the analysis, as can often be the case.

Therefore, trigrams and the (use stopwords) may make sense. It is to more easily recognise what the corpora are for and in a similar manner trigram and the (remove stopwords) to help with identifying what the corpus is all about.

Data cleaning must be engaged with for the LDA analysis (Appendix 7-2). The data needed careful cleaning as the data consists of a mixture of words, numbers, and punctuation, and the focus is on analysing the text as a whole. Keeping the stopwords (for example how and can) and question marks (?) in for instance '*how can I find?*' were as crucial as keeping the content words (such as 'diabetes') in '*good diabetes blood glucose level is about seven*'. This thesis considers the stopwords to be just as important as the content words as they are shown to be significant in understanding diabetes discourse. For example, stopwords and the question mark '?' may be utilised in the advice with stance-taking strategies. It may help people converse with each other, such as '*how can I find out about good diabetes blood glucose levels?*' to be explained later. Punctuation is also crucial in the corpus. It is removed for the LDA analysis in my thesis. What must be considered is that many other types of analyses are possible for any Discourse.

The original posts with punctuation that are shown to cluster in the LDA are once again used for further analysis. The study of other factors, such as emoji's in social media, is also widely known for its emotional or sentiment connotations. Data cleaning is essential for the analysis of the extensive text data with current software. The research decision was mainly to focus on the words in the posts for LDA analysis, but this is rectified later when other analyses are undertaken.

In the LDA model, a single post from the FDUK can belong to many different topics. However, it can be related more to a single one than many others. The LDA algorithm groups different posts into related topics and assigns topic numbers to the posts. Posts can belong to different topic numbers with various percentages of frequencies for each one. The use of LDA helps to compare posts that were made by peers. It includes precision testing in automatically allocating topics to the different user's postings. Postings can be an original post made by a peer, a comment on that post or a reply to the comment. The analysis then groups these postings under the term 'post'. The thesis focused on using the highest LDA topic frequency of the many different posts. It grouped those posts under that topic and its respective allocated number. This identification gives a single topic number for a group of posts. For example, a random selection gave Topic 24 (with stopwords) which has 1,789 posts, and these posts may be linked more closely to its similar themes (see Tables 7-2 and 7-3 in the Appendix).

There are diverse types of support 'topics', and LDA seems to be capable of finding them. However, as mentioned in the discussion in Chapter 2, the LDA 'topic' may not give the central target, or TOPIC, of the posts. In the analysis, it gave, e.g., what mostly used trigrams exists across similar posts which are called a 'topic'. The thesis suggests that the high-frequency trigrams with stopwords found in each topic can be used to help identify the 'topic' (purpose) 'device-enabled discourse purpose categories'. It is similar for the high-frequency trigrams without stopwords found in each TOPIC can help identify the 'TOPIC' (content) 'device-enabled discourse content categories' of the corpus. LDA is customarily used without stopwords to give particular grams. The top trigram is used to give labels to LDA topics and TOPICS (see Table 2-3).

For my analyses, five topics (i.e., 0, 24, 13, 31, and 34) were randomly selected. The posts within them were then selected for consecutive linear posts, and these are examined later. Large-scale text data can be examined by bringing down the scale to basic high-frequency patterns. It then becomes sensible to use AI and topic modelling for analysis.

Therefore, for Step 4 and Step 6 shown in Table 4-1, topic numbers were randomly selected, as mentioned above, and were then further analysed concerning their posts. Blei et al. (2003), describes a post as containing many different subjects. Posts are related to corpora subjects/topics. All my thesis 50 'topics' and 50 'TOPICS' categories are shown in Appendix 7.2.3 and 7.2.5 and are discussed later in this chapter.

The posts inside the LDA topics 0, 24, 13, 31 and 34 can tend to be about some aspect of the 'topic' 'device-enabled discourse purpose categories' and the 'TOPIC' 'device-enabled discourse content categories' (see Appendix 7.1).

My broad approach to exploratory research and hence to the LDA analysis can help to engage with the data systematically. It is wherein the broad purposes of the corpus are first established via analysis of a randomly selected limited number of ‘topics’ and their posts. This strategy can meet the expectation of AI machine learning and topic modelling from the newer sub-field of computing. These are further investigated to identify linguistic features of support within the well-established field of linguistic analyses.

The randomly selected posts from LDA topics 0, 24, 13, 31 and 34 are also assigned the ‘TOPICs’ to their respective posts. So, two types of similar LDA models are developed for the discourse. This process provides a sample of posts, high-frequency trigrams, topics and TOPICS that can provide trends from wherein features of support may be identified.

‘*Device-enabled discourse purpose categories*’ – ‘topic’ patterns:

Table 7-5 shows the LDA trigrams with stopwords Model, i.e., ‘device-enabled discourse purpose categories’ for 50 topics. A random sample of 5 topics is taken, and they are shown next with their **‘***device-enabled discourse purpose categories.’*

topic 0: *do you think*

topic 13: *questionmark questionmark questionmark*

topic 24: *you so much*

topic 31: *diagnosed with type*

topic 34: *you will get*

Posts were then taken from ‘topic’ 0, 24, 13, 31, and 34, where they were part of consecutive post identification numbers (ids), so they form part of a pair or a thread of a possible conversation. They could, thus, perhaps contain advice-seeking, advice-giving that may be related to similar targets and conversations. These analyses resulted in seventy-three posts and their substantial quantity of text for further analysis. The assumption is that these posts will tend to be about peers asking for and advising with stance-taking on the similar support issues, and therefore may also contain similar stance-taking features. These posts are shown in Appendix 7.1

'*Device-enabled discourse content categories*' – TOPIC' patterns:

The first three 'TOPICS' of 50 '*device-enabled discourse content categories*' from Table 7-6 are given below. However, the posts contain many different top frequency TOPICS:

TOPIC 0: feels better soon

TOPIC 1: hope sorted soon

TOPIC 2: long acting insulin

The seventy-three posts and their substantial quantity of text were then assigned their respective '*device-enabled discourse content categories*' or top TOPIC (Appendix 7.1.) They each contain one of the following TOPICS.

TOPIC 27: feel better soon

TOPIC 43: good morning hope

TOPIC 28: questionmark good luck

TOPIC 38: high blood sugars

TOPIC 19: happy new year

TOPIC 12: message add friend

TOPIC 47: feel free add

TOPIC 28: questionmark good luck

TOPIC 4: fast acting insulin

TOPIC 11: questionmark questionmark questionmark

TOPIC 21: hope comes soon

TOPIC 40: ha ha ha

TOPIC 8: Monday Friday pm

TOPIC 31: diagnosed type years

A striking observation to emerge from the data analysis of the device-enabled discourse content categories is they suggest for example discourses about advice, humour/sarcasm, questioning, emotion, hope, and charity and many other interactional activities amongst peers. They compare to the findings in the academic literature with similarities to well-established research (see Table 2-3 and Appendix 7.1.)

4.2.2 MeaningCloud: Findings of targets and global sentiment

The seventy-three posts and their substantial quantity of text that were randomly selected from the LDA 'topics' were then individually analysed using MeaningCloud. This is to find the target elements in a post, such as the concepts, time expression and quantity. The targets found in the data include, for example, diabetes, pump, and blood. The thesis focuses mainly on domain-specific targets of diabetes, blood, and pumps from many others ranging from targets of usernames, places, time, to years.

The posts that are selected for further analysis contain targets that are mentioned in the literature studies of the domain. The ones that are used relate to blood, diabetes/diabetic, and pump. These targets are found in the posts (see Table 7-1 and Section 4.4). These targets are also seen and are suggested to be related to the LDA 'POSTs'.

MeaningCloud is also used to assign values to the seventy-three posts and their substantial quantity of text. It is for global sentiment. All the words used together in a post are related to the sentiment of the entire post and not only just one single word or sentence within the post. An example of targets and the global sentiment is provided below.

Targets and global sentiment pattern

MeaningCloud gives targets of blood amongst other targets for example place, guy, kind, blood, for the following original post: 86 anonymised as '*...just came out of hospital…developed keto-acidosis and kidney infection…my **blood** glucose level was approximately fifty…. great...happy…*' The global sentiment is also given for each post as positive, neutral or negative. For example, the aforementioned original post is given a positive global sentiment.

4.2.3 LIWC: findings of emotion, certainty and healthcare

The seventy-three posts and their substantial quantity of text mentioned above from the randomly selected LDA 'topics' were then analysed individually with LIWC (see Appendix Table 7.14). LIWC gave values of either positive, neutral or negative for healthcare, emotion and certainty, and an example is provided below.

Emotion and certainty and healthcare pattern

For example: 86, '*...just came out of hospital…developed keto-acidosis and kidney infection…my blood glucose level was approximately fifty…. great...happy…*' LIWC gives high positive emotion and zeroes certainty and high healthcare values for the original post.

4.2.4 Manual identification of advice features amongst the many targets and other potential linguistics features

The seventy-three posts and their substantial quantity of text from the LDA model were also manually examined. For example, it is for advice patterns, as per the critical literature on similar work (see Tables 2.1, 2.2 and Appendix 7.1). A broad approach is taken as social media text is not straightforward for a corresponding one-to-one match. There can be many other interactional activities amongst peers. For, example, for $AA_{OI;}$ the text may suggest a question without a question mark when a person who posts askes for advice about information. A closer inspection of all the text in the entire posts can help to make a broad fit without losing the gist of the established literature on the matter. Other targets and linguistic features are also looked into, but with a focus on Advice as a potentially crucial element for support patterns.

A post can be made up of different sentences, and different targets, discourses content and purposes. These can be tackled in a single post by the person or across other posts by many different people. For example, there are targets of novomix, Levemir, lantus, blood, school, diabetes, and fourteen years, and they cover the diversity of posts by peers. The post can contain many discourse contents. For example, TOPIC 43: good morning hope potentially about Hope and Greetings, and TOPIC 28: questionmark good luck potentially about Questioning, Good Luck. The discourse can have many purposes for example topic 24: you so much potentially focusing lots on the other person; and topic 13: questionmark questionmark questionmark potentially about Questioning

The posts also tend to be about advice (section 4.3 and Appendix 7). Different types of 'Advice' features are identified which tend to match those in the established literature. The posts may contain more than one type of advice feature because a post can contain more than one sentence; it may seek advice in one sentence and advise in another. Thus, it was not easy to assign advice types to an entire post but rather to the different parts of the post. However, the post may tend towards an overall advice pattern from its inherent advice related features. This systematic way of doing this can help with the possible bias of trying to find only what one is trying to find in a corpus, and thus that expectation may be counteracted. What is not found in a corpus may be just as important as what is found in a corpus. It is a criticism in the literature.

The advice patterns that are found are comparable to those found by other researchers (see Table 2-1 and Appendix 7.1).

There tends to be, for example (AA_P) advice-asking by problem disclosure. This can be potentially ambiguous, as it can be interpreted as a request for advice, sympathy or solidarity. **Advice-asking pattern types** 62, 189527,…i am on novomix approximately three times a day…most people taking this seems to inject approximately twice a day… i find it ok for routine days…. but when i do anything different during holidays…doing more activity…eating out…find that my blood sugars go all over the place…. am considering changing to levemir and lantus AA_P … would appreciate any comments on the pros and cons AA_R , TOPIC 43: good morning hope,   topic 24: you so much, Targets: novomix, Levemir, lantus, blood

There also tends to be, for example (AG_E) advice-giving with a description of personal experience (see Table 2-2 and Appendix 7.1). It is an account of how the person dealt with the situation that the advice-seeker had described. **Advice-giving pattern types.** 56, 178116, '…when i was diagnosed at approximately fourteen years old…to say the least puberty was not fun!..potential partners would not kiss me in school…because they thought they could catch diabetes!!!...' AG_E, TOPIC 28: questionmark good luck, topic 13: questionmark questionmark questionmark;
Targets: school, diabetes, fourteen years

4.3 Combining quantitative results

The aim is to see if the research arrives at high-frequency topics, TOPICS, and targets for consecutive and non-consecutive posts. It is in for example, advice with stance-taking around key targets amongst many diverse targets and discourses; for consecutive and non-consecutive posts. This section gives the combined analysis of the random samples of the above 'topics' and the seventy-three posts and their substantial quantity of text, consecutive posts and non-consecutive posts. The analysis uses the individual mixed methods that result from Section 4.2 and combines them. It then attempts to give an analysis of any device-enabled discourse purpose and content categories and linguistics forms such as advice with stance-taking found in the corpus and broadly and amongst many other interactional activities between peers. It is about combining the individual patterns described above. It is with, for instance, examples that can provide evidence for advice with stance-taking in any given post where peers seek to help each other. For example, the targets that are found may be ‘word/s’, which are the central part of a post. Posts are shown to contain many different targets and are of high frequency. They do relate mostly to Diabetes support as would be expected but show intricate patterns.

These targets displayed in the posts are about participants’ concerns and issues that people with diabetes deal with during their lives (see Chapter 2). These can range from dealing with concerns about diabetes, blood glucose levels, and the use of insulin pumps to the use of medication. Three of these well-known targets from the literature are selected for further analysis (see Appendix 7.1).

Examples of consecutive posts are shown to contain similar targets with a variation on sentiment and with low certainty. At least three things need to be known for a given occasion of stance-taking beyond what may be overtly present in the words and structures of the stance sentence itself: (1) Who is the stance taker? (2) What is the object of stance? (3) What stance is the stance-taker responding to most? Each question points to one component of the process of interpreting stance. These are identified in the data analysis of my thesis.

The tables in Appendix 7.1 show the seventy-three posts and their substantial quantity of text with ‘posts’, ‘POSTs’, a target, a certainty value, a global sentiment value, a healthcare value, and an emotion value. The next step involves extracting, from the 73 posts, any individual posts that have the same target. Consecutive posts that pertain to one target were further analysed, for instance, advice, and stance-taking was ‘blood’, ‘diabetes’, and ‘pump’. The stance-taking in the posts may relate to the target and consecutive posts. The use of consecutive posts is to check the flow of advice-seeking and advice-giving with stance-taking.

The most striking observations to emerge from the data analysis provide evidence (see examples provided in the next section) that peers tend to utilise many patterns of advice-seeking and advice-giving with stance taking. The stances are varied for affective stance and can mostly contain low epistemic stance. Peers can employ differing but specific (and of primary importance in my thesis) ‘*device-enabled discourse purpose categories*’ and ‘*device-enabled discourse content categories*’ in closely related posts. Advice types, stance-taking types and LDA ‘topics’ can vary between posts from similar threads. There tends to be no affective stance alignment except for an epistemic stance alignment. These types of stance are also observed in posts with different targets across the sample.

There are striking observations to emerge from the data analysis. It gives evidence that, for a similar stance object in a post, people can mostly have different types of affective or similar types of epistemic stance. They can employ some particular ‘*device-enabled discourse purpose categories*’ and ‘*device-enabled discourse content categories*’, or resources within the many different contexts of support. There are also stances on different targets or ‘TOPICs’ in a single post, as well as similar stances across different posts not necessarily related to consecutive post identification numbers. The next sections offer analyses of each target from the sample.

4.3.1 Target Blood glucose levels for consecutive and non-consecutive posts

Non-consecutive posts

The individual patterns are combined and shown with the respective post's examples. They can give evidence for advice with stance-taking in posts concerning the target word, 'blood' and many other interactional activities amongst peers. There is remarkably a high-frequency stance-taking with diverse affective stance and low epistemic stance but not necessarily in consecutive posts but rather indirectly in usage across many posts about the same target.

The evidence (see Appendix Table 7-1) from this study suggests that the stance of the users may have some alignment on 'certainty'. They can tend to take up occasionally similar emotion positions or diverging global sentiment positions. These targets are part of the components of the stances, and the participants are responding to a stance on 'blood' about blood glucose levels.

They can share particular '*device-enabled discourse content categories.*'
topic 24: you so much, and topic 34: you will get.

They also tend to share different '*device-enabled discourse content categories*'. TOPIC 27: feel better soon, TOPIC 38: high blood sugars, TOPIC 43: good morning hope and TOPIC 28: questionmark good luck.

They can use advice giving features of AG_E (potential advice-giving in the social group via problem disclosure), AG_D(imperative), AG_H and advice asking features AA_P (problem disclosure) and AA_R (Explicit solicitation of advice)

They tend to share Targets of:

Targets: place, guy, kind, blood

Targets: username, long, Humalog, blood

Targets: novomix, Levemir, lantus, blood

Targets: child, blood, sugars, insulin

Consecutive posts

Post 86 and 87:

1, 86, '...just came out of hospital…developed keto-acidosis and kidney infection…my blood glucose level was approximately fifty AG_E …. great…happy…' , High positive emotion and positive global sentiment and affective stance; and zero certainty and high tentative and zero insight and epistemic stance for high healthcare, TOPIC 27: feel better soon, topic 24: you so much, targets: place, guy, kind, blood

2, 87, ...living with type 1 diabetes….about ten years……inject twice a day with humalin i - a long lasting insulin….fun!! …blood sugars are generally pretty good AG_E …. if anybody on here wants to chat about type 1 diabetes just drop me a message AG_D …, High positive emotion and global neutral sentiment affective stance; and high certainty and high tentative and zero insight and epistemic stance for high healthcare, TOPIC 38: high blood sugars, topic 24: you so much, Targets: username, long, Humalog, blood

The consecutive posts are similar for the way support is offered in the non-consecutive posts. They are about similar targets and similar stance-taking. They are not always direct consecutive support posts to persons in the thread.

4.3.2 Target Diabetes for consecutive and non-consecutive posts

Non-consecutive posts

The individual patterns from Section 4.2 are combined and shown with several examples (Appendix Table 7-2). In this section that provides evidence for advice with stance-taking in posts concerning the target word, 'diabetes' and many other interactional activities amongst peers. There is remarkably a high-frequency stance-taking with diverse affective stance and low epistemic stance but not necessarily in consecutive posts but rather indirectly in usage across many posts about the same target.

The advice with stance object 'diabetes' samples provide different consecutive posts from the overall selection of the target word, 'diabetes'. The surprising aspect of the data is in that the stance object 'diabetes'. It suggests the stance of the users may have an alignment on 'certainty' with similar zero certainty positions. Where one is low, then the other can sometimes be high certainty. They can tend to occasionally take up similar emotional positions or diverging global sentiment positions on similar TOPICs.

They can share particular '*device-enabled discourse purpose categories.'*

topic 24: you so much, topic 31: diagnosed with type, topic 34: you will get, and topic 13: questionmark questionmark questionmark

They tend to share different '*device-enabled discourse content categories*'.

TOPIC 38: high blood sugars, TOPIC 31: diagnosed type years, TOPIC 21: hope comes soon, TOPIC 11: questionmark questionmark questionmark, TOPIC 19: happy new year, TOPIC 8: Monday Friday pm, TOPIC 12: message add friend, TOPIC 47: feel free add, TOPIC 28: questionmark good luck, and TOPIC 4: fast acting insulin

They tend to utilise many advice asking features: AA_B, AA_P, AA_{OI}, AA_A, AT_P, AA_R, and advice-giving features: AG_E, AG_D, AG_H, and AG_I

They tend to share Targets of:

Targets: username, long, Humalog, diabetes

Targets: username, diabetes, type, wood

Targets: type, Diabetes, lifestyle, medication

Targets: type, diabetes, Tuesday, 12/09/2008

Targets: type, diabetic, november-94

Targets: username, type, Diabetes, pill

Targets: type, diabetes, tablets, sugar

Targets: username, type, Diabetes, insulin

Targets: hiya, username, insulin, diabetes

Targets: username, place, type, diabetes

Targets: username, diabetes, ailment, fatty

Targets: username, type, Diabetes, seven

Targets: type, married, highs, diabetic

Targets: username, course, diabetic, baby

Targets: niece, type, diabetes, couple

Targets: child, type, diabetes, honeymoon

Targets: diary, child, diabetes, thirteen-year-old

Targets: diabetic, results, injection, check

Targets: school, diabetes, fourteen years

Targets: username, diabetes, childhood, parents

Targets: blood, injections, people, diabetes

Targets: insulin, family, type, diabetes

Consecutive posts

Post 296 and 297:

4, 296, '…hi…i was diagnosed with type 2 diabetes just last year... still finding it difficult to make small changes…in my lifestyle…getting used to the medication AA_{P}…so if anyone has any tips AA_{OI} …stories AA_{B}… send me a message AA_{R} …thank you…', High positive emotion and global negative sentiment and Affective stance; and zero certainty and high tentative and zero insight and epistemic stance; for high healthcare, TOPIC 31: diagnosed type years, topic 34: you will get, Targets: type, diabetes, lifestyle, medication

5, 297, '…i was diagnosed with Type 1 diabetes approximately november 2008 AAP…', Zero positive and negative emotion and global high negative sentiment and Affective stance; and zero certainties and zero tentative and high insight and epistemic stance; for high healthcare, TOPIC 31: diagnosed type years, topic, 34: you will get, Targets: type, diabetes, Tuesday, 12/09/2008

The consecutive posts are similar for the way support is offered in the non-consecutive posts. They are about similar targets and similar stance-taking. They are not always direct consecutive support posts to persons in the thread.

4.3.3 Target Insulin pump for consecutive and non-consecutive posts

Non-consecutive posts

The individual patterns are combined. They are shown with several examples that provide evidence for advice and stance-taking in posts concerning the target word, 'pump' and many other interactional activities amongst peers. The posts suggest the use of advice patterns, and that advice includes a stance-taking element. There is remarkably a high-frequency stance-taking with diverse affective stance and low epistemic stance but not necessarily in consecutive posts but rather indirectly in usage across many posts about the same target.

The advice and stance object 'pump' samples provide different posts from the overall selection on the target word, 'pump'. The surprising aspect of the data is in that the stance object 'pump' suggests that the stance of the users may have an alignment on 'certainty' with similar zero certainty positions. Where one is low then the other can sometimes be high. They tend to take up occasionally similar emotional positions or diverging global sentiment positions on similar TOPICs.

These are the components of the stances and are the stances that the participants are responding to most. They may not always precisely match each other's levels of the stance of uncertainty, emotion or sentiment. They do have a stance-taking that both incorporates uncertainty and sentiment towards a similar target in this case 'pump'.

They can share particular '*device-enabled discourse purpose categories.*'

topic 34: you will get, topic 31: diagnosed with type, and topic 13: questionmark questionmark questionmark

They tend to share different '*device-enabled discourse content categories*'.

TOPIC 31: diagnosed type years, TOPIC 11: questionmark questionmark questionmark, TOPIC 21: hope comes soon, TOPIC 28: questionmark good luck, TOPIC 40 ha ha ha, and TOPIC 8: Monday Friday pm

They can use features of advice- asking and advice-giving with stance-taking: AG_E, AA_P, AA_I, AA_B, AA_A, AG_I, AA_R, and AA_{OI}

They tend to share Targets of:

Targets: username, type, diabetes, pump

Targets: hiya, username, insulin, pump

Targets: username, Pumps, general practitioner, type

Targets: username, lotto, pump, child

Targets: pump, work, three year

Targets: place, diabetes, type, pump

Targets: pumps, type, school, eleven years

Targets: pump, thirteen-year

Consecutive posts

Post-1029, and 1030:

9, 1029, '…hey everyone!....i was diagnosed with type 1 diabetes at approximately age eleven…i am now approximately age sixteen…i have an accu-chek insulin pump...which honestly is amazing!!. AG_E..add me if you are diabetic…plus if you want to talk AG_D…laugh…happy…', High positive emotion and neutral global sentiment and affective stance; and high certainty and low tentative and zero insight and epistemic stance; for high healthcare, TOPIC 31: diagnosed type years, topic 34: you will get, Targets: username, type, Diabetes, insulin

10, 1030, '…hiya…i was diagnosed approximately thirty-five years ago… just before approximately my fourth birthday…my diabetic consultant wants to put me on an insulin pump… i am having an approximately seventy-three-hour glucose monitor…in approximately september…waiting to go on a carbohydrate counting course AA_A …before i can go on the pump… I am hoping to be on it by christmas AA_I … i am having loads a hypo's…when i reduce the insulin it goes high..i cannot win!!. AA_P ..hope it has the same impact on my life as yours obviously has AA_B …', High positive emotion and global positive sentiment and affective stance; and low certainty low tentative and low insight and epistemic stance for high healthcare, TOPIC 11: questionmark questionmark questionmark, topic 31: diagnosed with type, Targets: hiya, username, insulin, diabetes

The consecutive posts are similar for the way support is offered in the non-consecutive posts. They are about similar targets and similar stance-taking. They are not always direct consecutive support posts to persons in the thread.

4.4 Qualitative analysis

DA helps to give a context to the analysis of top posts frequencies and clusterings and topics of the corpus (section 4.2 and 4.3). The above sections gave the individual quantitative analysis and the results of the combined qualitative analysis. This section gives a qualitative analysis of the quantitative patterns found in the consecutive and non-consecutive posts, together with examples. It is where combined quantitative language patterns and their respective examples may provide evidence for device-enabled discourse categories. And potentially advice with stance-taking in posts. It is identified in a broad way and amongst many other interactional activities amongst peers. The consecutive posts are similar to the support offered in the non-consecutive posts. They are also not always direct consecutive support posts to persons in a thread. The support seems to be given or asked of the entire group or everyone, so the entirety of the 'Social Media Diabetes Discourse'. The 'Social Media Diabetes Discourse' acts as another peer, but as the main peer. People can talk about targets anywhere in the discourse. It is at any other postings. They may take a point of view about those targets.

Firstly, we will consider the full results of the analyses. Then we will look at the 'what' and 'how' on people's expressions in the Facebook Diabetes UK posts. The sharing of information is suggested to be more about people's attitudes, opinions, and sentiments and people can respond to any post or create their post at any time. My thesis place importance on these for an online system of support.

The underlying meanings can be revealed with discourse analysis of the sampled 73 posts and their substantial quantity of text. It can give context to the Facebook Diabetes UK post, corpus and any patterns that emerge in the data analysis. The context can be inherent in the way the peers frame their issues. It concerns their chosen high-frequency targets of support, for example, ‘blood’, ‘diabetes’, and ‘pump’. The thesis focuses mainly on domain-specific targets of diabetes, blood, and pumps from many others ranging from targets of usernames, places, time, to years. The text in the posts contains discursive properties of emotive language. There is general uncertainty and peers tend to construct posts in this manner to seek and give advice. They emphasize descriptions of their targets and do take a stance towards those targets. The peers tend to utilise of primary importance device-enabled discourse categories of purpose and content to frame the disease. It is framed as relentless for example TOPIC 8: Monday Friday pm, with long descriptions of how they are coping, yet somehow managing for example topic 24: you so much and topic 13: questionmark questionmark questionmark. It is even though their health is also in an inevitable decline for example with topic 34: you will get, and TOPIC 27: feel better soon, but these devices can be about solidarity and having power over the problematic disease. These meanings tend to be imbued in the quantitative patterns. They are essential in representing online chronic illness support in a certain way by peers.

The most striking observations to emerge from the data analysis is that the high-frequency *device-enabled discourse categories* tend to relate to linguistic patterns. Thes can be advice with stance-taking about several and diverse targets of the domain. The FDP postings have examples of high levels of self-positioning that are expressed through both affective and epistemic stances. There are high levels of uncertainty for epistemic stance. There tend to be some elements of a depersonalised stance. They can position some peers as passive recipients of the targets that they converse about in the text. They can frequently express zero epistemic stance about how certain they are about their advice for targets such as ‘diabetes’ or ‘medication’. Peers tend to have support, and they are shown to be supported via the use of trigrams such as topic 0: '*do you think*'. They can utilise a passive voice, but their epistemic stance can include topic 1: '*should be able*', which can mitigate their fluency in the support.

Their usage of positivity, for example, Post 86: '*great…happy*', is also a positive form of appreciation. However, they may not always ascribe it to themselves in an agentive manner. Others presented their language in matter of fact, epistemic stances. Peers who had similar targets, for example 'blood', expressed clear uncertainties about the target as a controllable entity in their daily lives.

The sentiment and emotion that are expressed can vary. Therefore, there are broader viewpoints available on dealing with the concerns and issues in a community of support. It means that part of the attitudes towards the targets mentioned above shows variation. It is not only one type of sentiment. Opinions which may be considered in the literature to be based on people's sentiments and emotions still differ without a need to have a complete agreement on any target. People are shown to express their own diverse personal experiences about their conditions and issues in the Diabetes UK posts rather than stating facts. This sharing of 'information' is not done with high certainty. Support discourses of for example advice with stance-taking, events, humour/sarcasm, questioning, emotion, hope, and charity is not only about sharing 'information' as a fact-based system of knowledge but about finding information and support. There is more going on in 'support', and the thesis considers it to be about topics and TOPICS with, for example, advising with stance-taking. A possible suggestion is that it allows the objective medical facts and the subjective experience to come under the scrutiny of the public gaze. Participants are seen to be agents or passive participants when dealing with illness and wellness.

An example is given by post 987: '*…hi…i was diagnosed with type 2 diabetes… approximately eighteen months ago... i am afraid i largely ignore it…just pop the pills and get on with life…i do not like to feel I have an illness… never had any symptoms… thirst…great weight loss or anything... it runs in our family in later life…i am approximately sixty years old… so just accepted that I might get it one day…*'AAP*; TOPIC 31: diagnosed type years, topic 34: you will get; Targets: username, type, Diabetes, pill.* They are capable of taking on contrasting roles with varying emotions and sentiments and for advice-seeking and giving.

The discourse also contains TOPIC 31: diagnosed type years, which is potentially about discourse content of Diagnosis, Diabetes Type and Years. The discourse also contains topic 34: you will get, which is potentially about the other person and they are getting something. Other Targets of username, type, Diabetes, pill, exist and shows the richness and diversity of posts and discourses.

A single message of positivity is not the only narrative in the discourses. Differing sentiments and emotional stances towards the same targets can act as counterbalances and negotiation, and for influencing each other. Park et al. (2011) and Dwyer (2007) define influence to be 'on others' perceptions about a person, object, or event by controlling or managing the exchange of information in social interaction'. People can influence each other to be well while facing chronic illness. The posts are constructed to be about power and solidarity relations for engaging with the illness and building a network of support. There is an attempt to build friendships with people having similar issues.

The influence is also about getting others to trust the advice. People mitigate the risk among themselves. This counterbalance in the support helps when discussing and dealing with the broader results of support from, for example, participating in events and undergoing surgery. The next sections bring the analyses mentioned above together to answer the research questions, and I will present the principal findings of the current investigation.

4.5 Answering the research questions

RQ1: What attitudes, opinions, and sentiments are people expressing about their conditions and issues in Facebook Diabetes UK posts?

The results suggest that people take a variety of attitudes, and their stances for the same targets can include a negative to neutral to positive sentiment and one of uncertainty. They can have a variety of sentiments and zero certainty about the information they are sharing. They can converse about, for instance, many diverse opinions on conditions and issues ranging from their glucose levels (blood), their diabetes, the use of an insulin pump, school, place, username, events, children, greetings, and insulin

RQ2: How do people express their attitudes, opinions, and sentiments about their conditions and issues in Facebook Diabetes UK posts?

The results indicate that they share information through advice-seeking and advice-giving with stance-taking strategies. It is amongst many other linguistic forms from humour/sarcasm to greetings and many other interactional activities amongst peers.

Listed below are the operationalised research questions and claims about the users:

OQ 1.1: What are the discourse purposes? What are the particular word trigram choices used by people, often together, to express what the discourse purposes are in their posts on social media for supporting people with chronic illness?

This question considers, for example, high-frequencies, consecutive and related post. These may contain advice with stance-taker and stance-taking resources. They have been drawn upon, but to what discourse purposes, or, as the thesis calls them: for what ‘topics’.

What emerges from the results is they use and of primary importance in my thesis of *device-enabled discourse purpose categories* of for, example:

topic 0: *do you think*

topic 13: *questionmark questionmark questionmark*

topic 24: *you so much*

topic 31: *diagnosed with type*

topic 34: *you will get*

OQ 1.2: About what is the discourse? What frequently used content word trigrams relate to the discourse contents?

This question considers, for example, the advice in a post that may then contain stance-taking and related content words for support, concerning other posts, or, as the thesis calls them: for what ‘TOPICs’.

What emerges from the results is that they use and of primary importance in my thesis of *device-enabled discourse content categories* of, for example:

TOPIC 27: feel better soon

TOPIC 43: good morning hope

TOPIC 28: questionmark good luck

TOPIC 38: high blood sugars

TOPIC 19: happy new year

TOPIC 12: message add friend

TOPIC 47: feel free add

TOPIC 28: questionmark good luck

TOPIC 4: fast acting insulin

TOPIC 11: questionmark questionmark questionmark

TOPIC 21: hope comes soon

TOPIC 40: ha ha ha

TOPIC 8: Monday Friday pm

TOPIC 31: diagnosed type years

OQ 1.3: What is its primary target? What nouns/entities are discussed in these posts?

This question considers the stance-taking and related target (noun/entity) object being discussed ranging from diabetes, advice, events, humour/sarcasm, questioning, emotion, hope, to charity, across consecutive and non-consecutive posts.

The results that emerge together with the analysis provide essential insights. It is into crucial domain-specific targets such as 'blood glucose level', 'diabetes', 'children with diabetes', 'insulin pumps' and 'medication'. They show the diversity of the corpus ranging from these targets to, for example, school, children, username, events, and greetings. These are available in consecutive and non-consecutive posts.

OQ 2.1: What is the poster's stance about the certainty of their information? What is the poster's stance about the certainty of their information? How certain is the person in these posts and do people express these directly to each other in consecutive or in non-consecutive posts?

This question focuses on and considers the epistemic stance (certainty) in the giving of information in the speech acts of, for example, advice-seeking and advice-giving.

In summary, these results show that there is mostly an alignment of zero certainties about ‘blood glucose level’, ‘diabetes’, ‘children with diabetes’, ‘insulin pumps’ and ‘medication’.

There is remarkably a high-frequency stance-taking with low epistemic stance but not necessarily in consecutive posts but rather indirectly in usage across many posts about the same target.

OQ 2.2: What is the poster’s stance concerning their feelings about the information and how do they feel in these posts and do people express these directly to each other in consecutive or in non-consecutive posts?

This question considers the affective stance (attitude) in consecutive and non-consecutive posts.

In summary, these results show that there is a varied stance. It is based on the differing sentiment towards the same objects for example ‘blood glucose level’, ‘diabetes’, ‘children with diabetes’, ‘insulin pumps’ and ‘medication’. There is remarkably a high-frequency stance-taking with diverse affective stance but not necessarily in consecutive posts but rather indirectly in usage across many more non-consecutive posts about the same target.

4.6 Validation and an Awareness of Limitations

First of all, quality measures were used for the validation of the LDA topic-clustering of posts. The precision measure results suggest a good fit for the overall FDP posts and are a better fit for the users/peers’ posts. These precision measure results can be attributed to the smaller number of organisation total posts of 4% when compared to user/peer total posts of 96% (see Table 4.2 and Appendix 7.2). Therefore, the validity of the LDA topics about the organisation and user’s topics was tested with the data. It involved selecting the right topic-modelling test statistic. It is the confusion matrix with precision, f1-score, recall and support. It gave high scores for 'topics' and 'TOPICs' and particularly for peer person posts.

Secondly, the choice of topic numbers for the LDA was decided with the use of log-likelihood calculations and the researcher's knowledge of the domain. The trigrams of the 'topics' for the LDA are shown in Appendix 7-2.

Researchers employ the LDA model for finding familiar topics (Blei, 2012), which is more commonly used compared with NMF. One of its advantages is that LDA is unsupervised, and helps to remove researcher bias.

LDA topics and posts were further randomly selected for analysis and to remove researcher bias.

Thirdly, the top trigrams, 'topics' and 'TOPICs' were checked against critical linguistic research categories and devices. It is to make sure that their interpretations were sensible (see Appendix 7-2 LDA Computational Topic Model). The results are compared with each other and with examples from critical linguistic research papers. It is also done for the results on the posts within the LDA topic categories. It is carried out for the automated content analysis and annotation for entity recognition as well as sentiment analysis and potential linguistics forms such as advice with stance-taking features.

Fourthly, the qualitative DA analysis helped with seeing the context of the patterns and the sensibility of finding any such patterns in the posts and corpus and many other interactional activities amongst peers. The CL way of using LDA and DA meant that the analyses could be cyclic and iterative so that a sensible fit could be achieved.

Table 4-2: Validation of LDA Topics with stopwords

Validation of LDA Topics				
LDA	**precision**	**recall**	**f1-score**	**support**
Diabetesuk	0.666667	0.001414	0.002821	8,489
peer	0.962382	0.999972	0.980817	216,871
avg/total	0.951242	0.962358	0.943977	225,360

Finally, the LDA, automated content analysis, the annotation for entity recognition and sentiment for linguistic forms such as advice with stance-taking linguistic research approaches, as well as the DA approaches of the thesis were all checked against each other.

The *LIWC* model showed that the posts concerned healthcare. The LIWC gives an overview of the corpus of meta-data categories coded of the organisation and user posts. The LIWC classes are checked against the use of word types in the corpus. However, they are also compared to the prevalence of such words, and hence, classes in general corpora. The combined analytical approaches remove the possible self-selection of individual posts to confirm any potential claims.

These results would seem to suggest that the process of collecting and analysing a large-scale dataset from FDP meant that this thesis could try to make the claims exact but with limitations. The AI LDA model is a probabilistic model and therefore shows posts as together under ‘topics’ and ‘TOPICS’ that may be as close a fit as possible under statistical distributions. There are many different linguistic features in a post, and linguist may find other ways of identifying and looking at posts. The overall guiding aims are to find out exactly what this type of data and corpus mostly produce in terms of potential posts. These posts become the sophisticated high-frequency patterns of support.

It is a particular type of corpus under investigation, and caution must be applied. The findings might not be generalisable to all online support corpora covering many different domains.

4.7 Summary

The results suggest that a mixed approach with primarily LDA, automated content analysis and annotation for entity recognition, and sentiment analysis. These are for potential linguistic forms such as advice with stance-taking. Essential language-based approaches and DA guide it. These can be used to analyse and help in answering the research questions. LDA can be used for topic and TOPIC identification and is validated with precision measures. An example of another element of the analysis is the use of ‘certainty’ measures from LIWC.

These initial results are suggestive of evidence of a support pattern, for example, support that may be given with advice with stance-taking patterns and many other interactional activities amongst peers. Participants can employ *device-enabled discourse purpose categories or topics* and *device-enabled discourse content categories or TOPICS.* It is where people express their attitudes, opinions, and sentiment about their conditions and issues in FDP posts. It can also be about their personal experiences and concerns about the issues of, for example, blood glucose level, diabetes, children with diabetes, insulin pumps and medication.

There are potential broader issues hinted at in the discourse. These can be about influence, acting in opposition to the sometimes-dire peer conditions and issues. Peers may strive to maintain complete social solidarity and power relationships. The influence can be about wellness while facing chronic illness, low power and the need for significant solidarity relations in building a friendship. For example, it can be for trust and mitigation of risk in advice-giving and advice-seeking. The peers' commitments to each other on social media platforms are shown to be essential. They may employ primarily voses and nents with forms such as advice with stance-taking pattern amongst other linguistic forms to achieve these aims.

5 A NOVEL THEORY OF SUPPORT IN SOCIAL MEDIA DISCOURSE

5.1 Introduction

This chapter examines the novel theory of support. It considers how its production and its limitations. It looks at the evidence which is primarily an AI LDA probabilistic model of high-frequency patterns of support from the data. This high-frequencies help to categorise posts and text under topics (voses) and TOPICS (nents) and targets and linguistic forms such as advice with stance-taking in a broad manner. They would need refinement in future research. It uses real-life examples from the data of consecutive and non-consecutive posts to describe and explain this support. It is thus a close enough fit of the posts to potential identifiable and similar features from the linguistic literature and automated analyses models. The thesis suggests that the combined different software analyses and guidance from linguistic theory such as the SFL framework can help to reach some consensus. This thesis gives an interpretation of the Facebook Diabetes discourses brought together under a proposed novel theory but highlights its limitations. The thesis uses data analyses and findings to develop novel theoretical constructs to explain the observations and findings. The novel theory may be fragmented and incomplete as it is. It will undoubtedly need revision in some of its details for future research.

5.2 Support

The current study found that there is a higher alignment of uncertainty and lower alignment of emotion and sentiment. People try to engage with each other about the diverse and many different targets. The thesis focuses mainly on domain-specific targets of diabetes, blood, and pumps from many others ranging from targets of usernames, places, time, to years. Their stance is that they have something to contribute to the care of diabetes. However, their emotions and sentiments are at varying levels, and types exist from positive, neutral to negative. Emotion and sentiment are of a variety. One unexpected finding was the extent to which people attempt to self-manage their chronic illness. Therefore, peers can have different or similar stances but can take on non-agentive mannerisms.

Previous research showed that people self-select, for instance, by speaking in turn, and they are active. Crook et al. (2010) apply this idea to conversational agents, discussed in the implications section of the concluding chapter. However, the thesis finds that ‘conversation’ in the support discourse is not only linear-consecutive but also non-linear consecutive. People can post whenever, at any time and anywhere in the corpus, perhaps about common targets. However, they do take a viewpoint about their care.

One of the main difficulties with this line of reasoning is that these findings are somewhat limited by the use of a particular discourse on how people might employ salient language patterns. However, in the Facebook Diabetes discourses, people are shown to utilise sophisticated support patterns on social media discourse. It is to help each other through various conditions and issues of chronic illness in their lives.

5.2.1 Discussion

What do people usually say to each other on Facebook? How do they say it when giving or receiving support is problematic on open Facebook groups such as Diabetes UK? The posts are diverse and similar and can contain many topics from support, events, simple answers, to advertising. People also do not advise each other in consecutive linear order. There is remarkably a high-frequency stance-taking with diverse affective stance and low epistemic stance but not necessarily in consecutive posts but rather indirectly in usage across many posts about the same target. The thesis tries to bring consecutive linear post together from random selection for the analysis of advice with stance-taking and of many more for example events, humour/sarcasm, questioning, emotion, hope, and charity and the primacy of device-enabled discourse categories of purpose and content. The automated AI and computer analyses are about a 'quick' analysis and the trimming down of large-scale text data to see what remains. It is essential always to consider the removed aspects of analysis when arriving at the smallest number of 'identical' posts to analyse, e.g., stopwords.

The minimal manual labelling and analyses of the remaining posts from the machine learning is fraught with difficulty. For example, a post could potentially offer advice and ask for advice at the same time or offer these to another poster or all postees. So, Facebook group posts are both a linear fashioned support system and a database of information that people may search for answers to at any given time. A theory from this approach offers future study. It is of my thesis discovered (and of primary importance in my thesis) device-enabled discourse categories and their analyses together with that of Advice with stance-taking amongst many other discourses and many other interactional activities amongst peers.

There is potential for a finer-grained analysis of, for instance, the potential for many more categories of 'Advice' and device-enabled discourse categories. The broad grouping together of Facebook posts under the labels from other research and discovered in my work opens the door for future research.

The results suggest a possibility of *device-enabled discourse purpose categories* (voses) and *device-enabled discourse content categories* (nents) for support. The combination of these devices, (as well as advice and stance-taking features) are about patterns of language. It is in employment for support via conversations.

People can utilise these instances of natural support/conversation. Thus an answer suggests itself to the research question about language patterns that are in use. Individuals utilise them. The advice and stance-taking patterns point to a super-category of ‘Support’. In Figure 5-1, the target studied is ‘diabetes’ and ‘blood’, and peers use the ‘topic’ and ‘TOPIC’ categories.

Figure 5-1: Support illustrative with the target words, 'diabetes' and 'blood': Evidence of advice with stance-taking combinations.

72, 216361, '…people do not see the hard work that goes into trying to manage blood sugar levels AG_I …to try to keep healthy…after approximately twenty-six years since my child was diagnosed at approximately five years old with type 1 diabetes… you would have thought we would have something else to manage the condition… instead of relentless invasive blood testing and injections AG_E …how i wish for a day that my child was free from injections…', Low positive and zero negative emotion and global positive sentiment and affective stance; and zero certainty and high tentative and high insight and epistemic stance; for high healthcare, TOPIC 28: questionmark good luck, topic 13: questionmark questionmark questionmark, Targets: blood, injections, people, diabetes

73, 216362,'…i have type 1…was diagnosed approximately four years ago…after feeling unwell for approximately six months… having hypos and not realising it!...finally getting admitted to hospital as blood glucose levels were really high…after tests…being hooked up to an insulin pump…glucose drip overnight…the consultant came round the following morning… with the obligatory students…just announced this person has type 1 diabetes and will control it with insulin injections…needless to say I was shocked, stunned and devastated…i cannot express my gratitude to the staff…the continuing care I receive from my local hospital..It has been a hard learning curve for me and my family...control was difficult at the start but has been on a course…i can now count carbohydrates…it has stopped being so scary.. at times it is frightening…i have had huge problems with work… at times they just do not understand that I have to rest through illness AG_E , … until I was diagnosed I did not know anything about it... let us get the in-formation out there!.. AG_D .all my friends and family just about understand…just not fully….' , High negative emotion and global negative sentiment and affective stance; and zero certainty and low tentative and high insight and epistemic stance; for high healthcare, TOPIC 31: diagnosed type years, topic 24: you so much; Targets: insulin, family, type, diabetes

The entity of healthcare features significantly in the corpus. The affective stance can be related to positive/negative emotions and global positive/negative sentiments. The epistemic stance can be related to 'certainty' about the information shared. Therefore, the research provides an understanding of support for advice and about the stance-taking. It concerns who is involved; the resources used, the objects targeted and their relationships with each other.

These are to the components of the stances that the participants are responding. They do not always precisely match each other's stance of uncertainty, emotion or sentiment. However, They do have stance-taking that incorporates both the uncertainty and sentiment towards a target. These patterns of support are evaluated in more detail later on in the chapter.

Furthermore, a pattern may be the best practice as most other users use it. Alternatively, the best pattern may include having a good memory of conversing online or who can process the information and reply more quickly. Then again, any single user may be the only one that is using the best pattern. However, this pattern may not be a typical 'best pattern'. The assumption is that people are naturally and continuously learning; they are continually engaging in patterns that everyone else is employing to seek and give support.

Therefore, the use of CL (with machine learning LDA, entity recognition, sentiment analysis and DA) has helped to identify and compare 'topics', 'TOPICs', advice and stance-taking in posts and across the entire corpus.

The results shown above are from an approach of using LDA with nuclear language-based research and DA. This blended approach shows to benefit from an LIWC and MeaningCloud analysis of the corpora in the ways of conducting the analysis (see Chapter 3). The above evidence contributes to developing a novel theory of SSMD. In past years, however, these individual approaches have been challenged by the work of some prominent linguists. The thesis uses the combined strengths of the mixed approach to data analysis.

The machine learning LDA model, together with the primary language-based research and DA gives observations that can be a pattern for a new theory of support. Support shows to be a super-category that consists of patterns of the primary device - enabled discourse categories. It can contain advice with stance-taking and many other interactional activities amongst peers.

A machine-learning model, together with DA and LIWC and MeaningCloud, are employed to find and back up the evidence.

This evidence shows central linguistic devices. The users employ the devices repeatedly. People look for support during their online management of a disease where blood glucose levels can fluctuate wildly, as well as for their conditions and issues. People's concerns about their lives play a role in getting them to take part in the conversation.

A DA analysis of the individual posts further demonstrates the existence of the high frequency 'topics', TOPICS, targets and linguistic forms such as advice with stance-taking devices. The context can be about healthcare. Support is a more extensive influence that is acting into the conditions and issues and broader patterns in the discourse. People take action in accomplishing or completing the challenging conversations about their different lists of things to deal with most, in their lives. They have desires to reach goals as individuals involved in, for example, their long-lasting illness support. The influence can be about wellness while facing chronic illness. It is for power and solidarity to build friendship, trust, and mitigation of risk in advice-giving as well as about organisations' and users' commitment to each other on social media platforms. The pattern of support can play a role in helping the advice and stance-taking of people with chronic illnesses and can allow them to get the support they need.

What has been discovered in the data analysis and findings? The chapter shows that, firstly, people tend to converse about subjects in a purposeful way when the conversation concerns issues affecting their lives. They want to improve the handling of these matters by way of their interactions with their online conversations about support. People also use central language-based devices to construct their support and allow them to belong to the online community.

Furthermore, from the interactions between different people, the research also shows observations that prove a broader counterbalancing influence pattern of support. It can be related to community solidarity, friendship and power relations; and there is no tendency for alignment of affective stance but more for epistemic stance. Goldsmith (2000) looks at advice-seeking and personal disclosure, and this is found in the data of the research as, possibly, advice-seeking, seeking sympathy or solidarity. This thesis can broaden a definition of support to include advice, certainty, sentiment, solidarity and stance-taking. Kouper (2010) looks at advice-giving that is done by sharing details of personal experience. It is found in my thesis data. Different people can supply different sentiments about similar targets. It can be in order to help each other, but not necessarily by having the same sentiment at any given time. The influence can be about wellness. Peers face chronic illness, having low power and needing greater solidarity to build friendships, increase trust, and deal with the mitigation of risk in advice-giving. The peers' commitment to each other on social media platforms is essential.

The findings presented in Chapter 4 are the analysis of the Diabetes UK corpus. What has observed in this corpus are examples that prove the use, in general terms, of leading purpose categories, content categories, and targets. These form part of for example linguistic forms. It is for example advice with stance-taking amongst many other discourses ranging from humour/sarcasm to charity. People incorporate these in how they go about supporting themselves and each other for long-lasting illness care.

These examples are from the analysis section of the thesis. In the examples, posts for peers are examined. The keys used to label the posts are shown in the Appendix in Table 7.1. The examples give a summary of the main features of both advice and stance-taking. The actual post with the analysis of the significant features and topic numbers are provided as well.

DA analysis of the post reveals the lack of power in the certainty of the information and great solidarity in trusting others to mitigate the risk, but a commitment to share personal experiences. In this thesis, these combinations of features can be about the influence that peers can have on each other – there is a need for self-expression. They are thought of as the motivation for advice-giving, together with offering stance-taking in the ongoing support conversations.

The posts can have different combinations of the content and purpose categories related to the target. These are considered in the thesis as support, advice, and stance-taking resources. The categories tend to be about some of the conditions or issues that people face in their lives.

These conditions or issues can be the same as that of the other individuals in a supportive community. The direct experience of the peers helps to focus the amount and types of ‘topics’, ‘TOPICs’ and target categories that people can manage to talk about with each other.

The conversations also sometimes have professional input from medical experts from Diabetes UK. The corpus used in the thesis is, therefore, about healthcare. It is a critical medical area. The support cannot, therefore, be without professional input even though the thesis focuses on the peers. Thus, there is a broader context and content that may influence the peers and their conversations. The categories are about a broader issue and concern about a disease where blood glucose levels can fluctuate wildly (NHS, 2018).

These devices can exist in different combinations within the posts. In analysing the posts with DA and appraisal analysis from SFL, they are about trust in the peers about each other to help the community support itself. Power and solidarity can also be about the power that illness has over people and hence a low uncertainty in epistemic stance. Solidarity can be about the human condition that many people share and therefore has various levels of affective stance.

The approaches also lead to limitations in the research, which are then used to inform suggestions for future research, as detailed later in the concluding chapter.

During the research, random samples of topics and posts were analysed for advice with stance-taking strategy. Each strategy had different sub patterns. For example, advice-asking included the announcement of a plan of action, ‘anyone in the same boat’, persons to support oneself, requests for opinion or information, and problem disclosure, as suggested by Goldsmith (2000).

All of the salient ‘topic’ and ‘TOPIC’ categories and advice with stance-taking devices found are shown in Appendix 7-1 and 7-2. These include those categories suggested by prominent researchers, and those added on as found in this thesis. A more detailed account of a novel theory of support is given in the following section.

5.2.2 A proposed novel theory of support

The present study raises the possibility that this thesis can develop the resulting evidence into a novel theory of support for SSMD. The findings suggest that there are particular ways of doing constructs for support. It is with, for example, advice with stance-taking. It suggests emotive support and solidarity exchanges between peers. There are many other interactional activities amongst peers. The advice can be constructed for problem disclosures and sharing of experiences. It may help mitigate the risk of advising and the sharing of influential positions of knowledge, even if they are peer opinions, sentiments and attitudes.

This evidence (as discussed in the next section) consists of advice and stance-taking patterns. People and potential machines (see future research in Chapter 6) can use intricate patterns to construct support more quickly. From the analysis of the online chronic illness discourse, it shows high-frequency patterns. These can include stance-taking expressions connected with advice, sentiment and certainty. They are with *device-enabled discourse purpose categories (voses)* and *device-enabled discourse content categories (nents)*. There is remarkably a high-frequency stance-taking with diverse affective stance and low epistemic stance but not necessarily in consecutive posts but rather indirectly in usage across many posts about the same target.

Suler's (2004) disinhibition effect and the ambient affiliations of Zappavigna (2011, 2012) help to explain why people may converse with each other in the first instance. Therefore, peers may use types of 'topic', 'TOPIC', targets and linguistic forms such as advice with stance-taking.

Stahl (2006) argues that the practices of meaning-making in acts of conversation be about interaction. It is to propose, negotiate or work to reach an agreement or to display and define what are to count as the essential features of the setting, the occasion, and the typical social behaviours. Stahl argues that the context, the meanings, facts, and opinions are not made in advance but are formed by the people working as a group in many conversations. Stahl goes further to suggest that meaning is built by (and exists as) those very same activities.

In a similar, yet compelling manner, Hiss's (2013) study explores language attitudes towards 'Sámi' through analysing letters sent to the editor about a proposal to give the language co-official status with Norwegian in the town of Tromsø. In focusing on the discursive mechanisms of evaluation and stance-taking in these letters, Hiss makes use of the appraisal framework (Martin and White, 2005). He uses it to demonstrate how letter writers invoked and referred to shared values towards the issue. It is how they expressed affect, judgement and appreciation differently in their efforts to express authority over the proposed changes.

Stance allows us to reconceptualise what a 'diabetes support attitude' is and how it becomes instantiated over time, not only by individual speakers but also across speakers. As individuals express and enact stances toward diabetes and their speakers, they discursively create their subject positions ('identities'). Their interpersonal and social relations are constructed in connection with diabetes. Individuals can repeatedly express stances toward diabetes support or challenge the stances of others. This thesis can argue that these social roles and relations help 're-entextualize' ideologies of diabetes support, and can articulate attitudes towards diabetes support and speakers.

The thesis data that I examined comes from FDP. It concerns participants' diabetes support. It is somewhat limited regarding being able to see how stances enact ‘diabetes support attitudes’ and ideologies more widely. It can be helpful to consider, however, that people express their stances towards chronic illness support and speakers. They may do this regularly in various other domains on social media support groups.

An analysis of the postings can help to look at stance more carefully. It is to consider how stance is a matter of relational design through evaluation, positioning and alignment, as proposed by Du Bois (2007, p. 168).

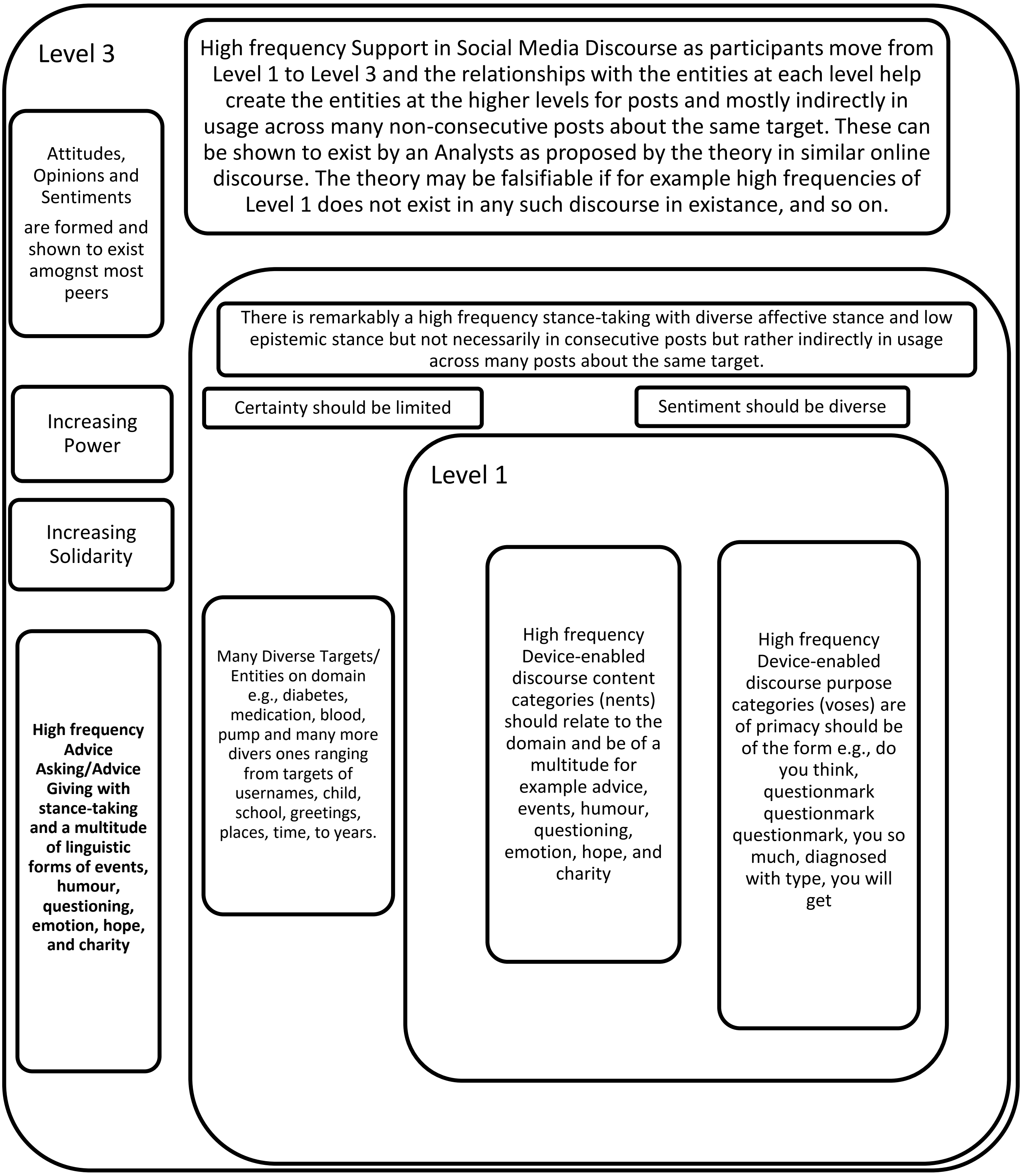
The conjecture in the thesis is that *support in social media conversations is the result of language-based high-frequency sophisticated patterns. Of primary importance are device-enabled discourse categories of purpose and content. It can include linguistic forms such as advice with stance-taking. There are many other interactional activities amongst peers. It is carried out by people in their meaningful, shared interactions. Together they influence real outcomes about people's concerns and issues. There is remarkably a high-frequency stance-taking with diverse affective stance and low epistemic stance but not necessarily in consecutive posts but rather indirectly in usage across many non-consecutive posts about the same target. People can post support at any time and not always in a linear-consecutive order. The theory may be falsified if counterexamples of such patterns in support become available.*

One of the ways to go about understanding the ‘more’ about support is to find patterns of behaviour. It can be where many people are conversing together or posting at any time in any posting order. The increasing power of ‘knowing’ by engaging with others on the platform, of seeing solidarity in taking particular actions, of decreasing risks of harmful advice and increasing trust, can all be central themes for helping users. The thesis regards these ‘topics’ as part of the broader *influencing* (Park et al., 2011; Dwyer, 2007) of patterns of support.

This thesis finds that people can use many ‘topics’, ‘TOPICs’, targets and for instance, advice with stance-taking to build and *influence* support in SSMD. There are many primary TOPICs. For example, posts can be about blood glucose levels and medication. A theory of support, for instance, can be used to help people with diabetes, to highlight their use of devices concerning advice with stance-taking, and to improve their support practice.

This theoretical model shows in table 5.1 with a consideration that attempts to clearly and fully capture the theory and explain how the components relate to each other. Diagrammatic detail is given, and the Level 1 high-frequency device enables support enabled discourse purposes and content. The representation of the theory in Table 5-1 can be utilised and applied to online support discourses by understanding the different levels. There are different levels for the components with an expanding representation in the diagram so that it is clearer how the components relate to each other in theory. The diagram also shows that the theory is falsifiable. It is if low-frequency patterns of the device-enabled discourse categories of purpose and content are not available in any similar online social media discourse. The high-frequency tendency is towards the primacy of device-enabled discourse categories of purpose and content. Other linguistics forms, for example, advice with stance-taking, events, humour/sarcasm, questioning, emotion, hope, and charity exist. The focus in my theory is on advice with stance-taking, but ultimately all the support patterns when all the elements are combined should be available. In applying this approach, advice with stance-taking helps to place the device-enabled categories into context and remarkably brings up the diverse linguistic patterns of support. The limitation of the theory is in its focus on crucial discourse aspects and targets such as diabetes, blood, pump, medication but concerning literature from the domain of diabetes and the thesis focuses mainly on domain-specific targets of diabetes, blood, and pumps from many others ranging from targets of usernames, places, time, to years. The theory may need more exceptional detail from individual future research on each component.

Table 5-1: The concise points of the theory of support for Facebook Diabetes Discourse



In general terms, this means a theory can be proposed that suggests that as shown in Table 5-1, the following concepts and relationships.

People are shown to express attitudes, opinions and sentiments about their conditions and issues in Facebook Diabetes UK posts. These tend to contain high-frequency support patterns on platforms such as the FDP, and these patterns can be used to develop a theory of support.

They may post chronologically about topics, TOPICs and targets. Alternatively, they may post in a non-linear way to cover topics, TOPICs and targets encountered across the entire online discourse.

People on social media platforms tend to support each other during chronic illnesses with a sophisticated language pattern. It includes high-frequency purpose and content language devices for the discourse constructs of support.

They make use of high frequency 'topics' or 'device-enabled discourse purpose categories'. For example: do you think, questionmark questionmark questionmark, you so much, diagnosed with type, and you will get. These can help each other to construct the discourse. They can work together on the aims of the discourse. They make use of high-frequency 'TOPICs', or 'device-enabled discourse content categories'. These combine the content of the domain with, for example, feels better soon, hope sorted soon, and long-acting insulin. These categories help each other to identify the familiar context in the corpus concerning the domain in which they need to be supported.

They primarily make use of high-frequency advice-asking and advice-giving with stance-taking and many other linguistics forms, for example, events, humour/sarcasm, questioning, emotion, hope, and charity.

There is remarkably a high-frequency stance-taking with diverse affective stance and low epistemic stance but not necessarily in consecutive posts but rather indirectly in usage across many posts about the same target. The evidence from this study suggests that the stance of the users may have an alignment on ‘certainty’. It is with similar zero certainty positions, or where there is low, then the other can sometimes be high. They can tend to take up occasionally similar emotional positions or diverging global sentiment positions on similar topics and TOPICs. These are the components of the stances. The participants can respond to, and they do not always precisely match each other’s levels of the stance of uncertainty and emotion or sentiment. However, they do have stance-taking that incorporates both uncertainty and sentiment towards a similar target.

These are in a broader context of power, solidarity, social relations, risk and trust. Hence, the uncertainty and diverse affect displayed when sharing information for support.

Each of the above points is subsequently expanded. It is to justify the supporting statistics/analyses and examples of usage from the corpus (some of these were presented in Chapter 4).

5.2.3 Social media chronic illness support discourse constructs

In this section, an explanation of support constructs and their relationships are considered.

The illustrative example of post 296 in the data analyses (Figure 5-2) does give an example of the employment of language patterns. There is remarkably a high-frequency stance-taking with diverse affective stance and low epistemic stance but not necessarily in consecutive posts but rather indirectly in usage across many posts about the same target. A summary of the illustrative example, findings leads to an explanation of their significance, (i.e., what they tell about the users). These 'topics' or 'voses' and 'TOPICs' or 'nents' can be most frequently used because they cover the opposites of illness, wellness, and vulnerability but also the motivations to deal with such illnesses. There is remarkably a high-frequency stance-taking with diverse affective stance and low epistemic stance but not necessarily in consecutive posts but rather indirectly in usage across many posts about the same target. People must cover many main 'topics' and 'TOPICs'. They may employ main advice and stance-taking devices to counteract each other's concerns and help the other gain support. People using these language patterns jointly can get support over a few or many posts.

Figure 5-2: An illustrative example of Support in Social Media Discourse

4, 296, '…hi…i was diagnosed with type 2 diabetes just last year... still finding it difficult to make small changes…in my lifestyle…getting used to the medication AA_{P}…so if anyone has any tips AA_{OI} …stories AA_{B}… send me a message AA_{R} …thank you…', High positive emotion and global negative sentiment and Affective stance; and zero certainty and high tentative and zero insight and epistemic stance; for high healthcare, TOPIC 31: diagnosed type years, topic 34: you will get, Targets: type Diabetes, lifestyle, medication 5, 297, '…i was diagnosed with Type 1 diabetes approximately november 2008 AA_{P}…', Zero positive and negative emotion and global high negative sentiment and Affective stance; and zero certainties and zero tentative and high insight and epistemic stance; for high healthcare, TOPIC 31: diagnosed type years, topic, 34: you will get, Targets: type, diabetes, Tuesday, 12/09/2008

DA analysis shows the full context of the use of 'topics', 'TOPICs', targets and devices. It can help express power and solidarity, and friendship relations of people. Their disease where blood glucose fluctuates wildly has many conditions and issues. The usage of 'topics' and 'TOPICs' categories and linguistic forms such as advice with stance-taking types tend to be about the support of people's conditions and issues. This representation is thought of as a unit of support. This social sharing of terms and problems can help in the development of support with a focus on advice with stance-taking for people to support each other online. It is thought of as, in the end, a unit of meaning. People are represented, and so are their individual needs for support.

The whole group of posts or communities of participants are made available to others to improve people's conditions and issues. The influence of different people acting in opposition to the terms and challenges of their illness exists as a more extensive pattern of dealing with people's lives, conditions and issues. There are possible opposing 'topics', 'TOPICs', targets that can balance the discourse, for instance, TOPICS on illness versus TOPICS on health.

This thesis proposes that support constructs not only act as the proper introduction of 'topics' and 'TOPICs' of support, (e.g., for discussing the long-lasting illness) nor only the use of advice and stance-taking devices for community building. Support constructs can also be a pattern where people even out, increase or decrease their influence. It can be for advice with stance-taking or in their use of many other linguistic forms. People can work to reach agreement on the meanings of certain terms in their interactions.

There is a dynamic change in the meanings of specific word-based items over a period – for example, the sense of the word '*diabetes*' in Post 296 instead of a static meaning of, for example, the pronoun 'you'. These changes are brought about using 'topic' categories, advice with stance-taking. They can point to a change in the condition of the people who take part in the support. People in different combinations can use 'topic' categories and the main advice with stance-taking devices. These frequent uses of combinations can become a pattern of sophisticated support. The use of this pattern can then help with the advice with stance-taking of people taking part in online communities. Support observations can include advice-seeking and for example, the person with diabetes as the primary person.

Advice-asking and advice-giving with stance-taking and primarily device-enabled discourse categories can produce an information-rich community. People can become the resources of the community and can utilise these support patterns to belong to a group and to participate in that group. These people can then help expand the network.

The 'topic' categories of support and language-based devices of support and influences proposed in this thesis are some of the different ways people use natural language. They use sophisticated patterns to help others and themselves. People can express what they think about in a discourse community. The community has values and rules and has people sharing direct experiences publicly. They are making sense of the healthcare social groups, which can give a sound basis for developing a theory of support constructs. This approach to 'support constructs' can lead to the optimisation of social media platforms. For example, a democratic consensus of what support is all about and for people's healthcare needs to be personalised in order to meet their goals (as discussed later in Chapter 6).

The findings of this thesis have already given a basis for developing support about the categories and language-based devices, and findings give evidence of support from a language-based viewpoint. Support is concerned with how people create and get through meanings to construct healthcare support collaboratively, or by working as a group. The evidence of support offers a grounding to analyse this social interaction. It is for an idea that is thought to be true: that people can share certain kinds of sophisticated support among themselves and others. It is to share their concerns; or when they converse and debate together in online groups. The suggestions for supporting people with chronic medical conditions and issues, and for the study of language and AI are discussed in conclusion in Chapter 6.

These patterns are about the representation of people in the 'support conversations'. They can show what an individual may need, as far as using many 'topics', advice with stance-taking for influencing each other. People have been shown to use patterns in their conversations of support. There is a need for connecting language, linguistic forms such as advice with stance-taking to support. These can bring about a fundamental change in people's lives.

Furthermore, this thesis suggests that the texts in the corpus do not act only as a means of communication. Individuals use mainly 'topic', 'TOPIC' and advice and stance-taking in a back-and-forth support-conversation. People can agree on something between themselves. They can influence a pattern of wellness in the face of worrying about their healthcare needs.

How people speak to each other is linked to keeping or creating social power relations (Foucault, 1972). This thesis shows that people can, in so doing, become a community of useful conversationalists. They are then scarce supplies of information. Ultimately, they are part of the support of other individuals who are part of this SSMD. The texts can help the group work well together and to disseminate healthcare support to a broader audience. The meanings are also about the use of the sentences constructed well from the particular high frequency 'topic', 'TOPIC', and linguistic forms such as advice with stance-taking concerning many targets.

Many people in the group may have different levels of experience in giving advice. Some can converse much more about 'topics' and 'TOPICs' than others can, and can, therefore, be vital to the group.

However, a person would want to know the full story of their diabetes. People do want to live a full life, even though their chronic illness can have a 'say' in what a person can still do or may never be able to do again (Diabetes UK, 2015).

These issues could be about the ability to use an insulin pump or about looking after their children who have diabetes. People on social media sites try to help each other with these types of issues, and also conditions like potential kidney failure.

These claims can be about the truth of the meanings of support groups and their significance. They can offer a way to question, predict, and control support items. This pattern is about solidarity and power relationships with the users. Now, this pattern can be related to, for example, advice-seeking behaviour.

Furthermore, advice-seeking and advice-giving can be a focal part of support patterns. People are reliant on others for support.

The evidence of its constructs illuminates support. These constructs are by way of people working together in a respective group. They employ language, 'topic' and 'TOPIC', advice and stance-taking in order to influence and ultimately build a shared meaning (Appendix 7-1).

5.3 Summary

The thesis shows that a new theory of support for chronic illness on social media can focus on the users themselves. The users tend to engage in healthcare descriptions, whether through the provision of information, explanation or first-hand experience.

People may naturally use particular language patterns to communicate. The high-frequency language patterns in use contain particular 'topics' or 'voses', and 'TOPICs' or 'nents'. They contain targets and linguistic forms such as advice with stance-taking strategies. For example, participants may give each other mutual support about their diabetes blood glucose levels. They may get support about their medications, and the pattern may include advice with stance-taking. There is remarkably a high-frequency stance-taking with diverse affective stance and low epistemic stance but not necessarily in consecutive posts but rather indirectly in usage across many posts about the same target.

6 CONCLUSION

6.1 Introduction

The use of DA can help with explaining the AI Topic modelling patterns discovered. Topic modelling alone is a process of a reduction of large-scale data into the manageable dimensions. The high-frequency patterns can help with building an original theory about how people behave when discussing chronic illness on Facebook. In overcoming the research challenges, such an approach may firstly find a place in corpus linguistics and secondly a useful theory of support. Searching for an 'architecture' of these types of discourses is analogous to the search for an architecture of sentences.

The study set out to explore the high-frequency features of support. It is to fill the gaps in the literature on online support patterns, discourse devices used between peers regarding their online support and linguistic forms such as advice-seeking, advice-giving with stance-taking. The chapter revisits the primary focus of the analysis of support patterns. These concern the primary importance of high frequency 'topics' (voses), 'TOPICs' (nents), targets. It included linguistic forms such as advice with stance-taking. There is remarkably a high-frequency stance-taking with diverse affective stance and low epistemic stance but not necessarily in consecutive posts but rather indirectly in usage across many posts about the same target. These have emerged in the data analysis and are used to propose a theory of support. People can use sophisticated patterns of support in social media discourse.

In general, the study found a high-frequency tendency for 'topic', 'TOPIC', targets and linguistic forms such as advice with stance-taking patterns in support of chronic illness in SSMD. Support for chronic illness on online platforms is, therefore, shown to be a multifaceted and complex social practice.

The novel theory has to be used with caution, as other future studies of different corpora regarding other online platforms may offer more insights into online support. The theory may be tested and falsified by its reliance on the existence of high-frequency elements, as demonstrated in 5.2.3.

Much of the criticism that levelled at studies of online support may relate to Hall et al.'s question about how such studies of discourse can illuminate cultural and social processes. Leonardi et al. (2012) provide a counterargument about questioning the popular narrative of technologies as tools for the service of human welfare and wellbeing. This thesis argues that in knowing how people use the technology and in facilitating that use primarily through the study of linguistic devices, there can be additional counterarguments to Leornadi's article.

6.2 Peer support

Peer support (96% of post) can consist of high-frequency language patterns tending to include primary linguistic devices: namely 'topics', 'TOPICs', targets, and linguistics forms such as advice with stance-taking. They employ as a possible 'counterbalancing influence'. For example, to report TOPICs or targets of illness. It is also to motivate the wellness of people by an alignment of uncertainties and the sharing of varying sentiments and emotions. People can share information or opinions, primarily from their experiences and personal disclosures. There is remarkably a high-frequency stance-taking with diverse affective stance and low epistemic stance but not necessarily in consecutive posts but rather indirectly in usage across many posts about the same target. The posts are complicated and have a purpose (vose) and content (nent). An example of a post (see Table 4-6), as well as its mixed methods analyses in combination, is given below:

14, 2137, '…thanks…it is very possible that the so-called diagnosed type 2 diabetes is a look-a-like ailment...often misdiagnosed by most physicians…. fatty deposits and bad cholesterol are either the culprits… or just contributing…. i have not been comfortable with this determination…so i diet…take oral meds which definitely helps AG_E … but cannot help to think…that there is more efficient medication… that would annihilate the disease totally AG_H …', High positive emotion and global negative sentiment and affective stance; and high certainty and high tentative and high insight and epistemic stance; for high healthcare, TOPIC 8: Monday Friday pm, topic 34: you will get, Targets: username, diabetes, ailment, fatty

The above example shows that posts can have a purpose and content that are latent in the discourse.

The '*device-enabled discourse purpose categories*' can be potentially about the other person and what they will get.

The *device-enabled discourse content categories*' can be potentially about Date and Time.

The posts can contain a diverse and rich variety of targets, for example, username, diabetes, ailment, and fatty.

The stance can contain concerning the target of diabetes, high positive emotion and global negative sentiment and affective stance; and high certainty and high tentative and high insight and epistemic stance; for high healthcare

Individuals use linguistic features such as advice-seeking with advice-giving with stance-taking. It can be of type AG_E (an account of how the person dealt with the situation the advice-seeker had described). It can be of type AG_H (any comment that contained explicit hedges or hedging devices, e.g., ‘I think’, ‘It seems’, or ‘Why don’t you?’).

People may go about posting, commentating and helping each other in the posts but may not necessarily be seeking only one type of alignment towards similar targets, in this case, ‘diabetes’. They can all have their lists of things to deal with and desires to reach different or similar goals in their communication of diabetes support.

The LDA models are shown with sample posts and identified language patterns (see Appendix 7.2). They show that user posts tend to categorise into ‘*device-enabled discourse purpose categories*’.

The above example shows that there can be a better linking of similar target posts versus ordering them by the time of posting on the platform. The corpus is thus usable in different ways: as a general conversational entity on many TOPICS and as a Support discourse on specific TOPICS and as a database of searchable information. Besides, it can find people with similar targets, perhaps with differing attitudes that can support each other. This process is underpinned by ‘*device-enabled discourse purpose categories*’. The corpus shows that people adopt ways to build support in diabetes discourses.

The results of the research address the research questions. They deal with the issues of how users go about conversing with each other, and about what they converse. It showed that the users adopted support patterns to build a discourse of peer support for diabetes.

6.3 Research contributions

Contribution 1: ***Theory of support in social media discourse: Sophisticated support patterns. High-frequency novel device-enabled discourse purpose (vose) and content (nent) categories. With linguistic forms such as advice with stance-taking amongst other***

linguistic forms of for example events, humour/sarcasm, questioning, emotion, hope, and charity. A high-frequency stance-taking with diverse affective stance and low epistemic stance but not necessarily in consecutive posts but rather indirectly in usage across many posts about the same target. The theory would need future refinement.

People working well together may assume that other people always say what they mean. It is a way of building trust in one other in online discourse. Trust is a necessity for support. There is remarkably a high-frequency stance-taking with diverse affective stance and low epistemic stance but not necessarily in consecutive posts but rather indirectly in usage across many posts about the same target. For example, in building a consensus in order to decrease the risk of giving the wrong support is vital to help each other deal with chronic illness. Authors generally emphasise trust (Sillence, 2010), where friendship can be regarded, for example, as influencing others to get well in the face of chronic illness. It is relevant to the ongoing conversations that make up these groups. There can be some control over what peers speak. Peers give help freely. There is a much-needed counterbalance to open collaboration with many individuals. Many people participate in order to make, for example, advice more trustworthy. A need for power and solidarity over the illness can result in particular advice and stance-taking in agreeing to correct advice and do play a crucial role in support of people. The affective stance-taking varies, and the epistemic stance tends to be zero. People do express themselves, but often show their uncertainty in the information and targets of their support. There is a diversity of discourse ranging from diabetes, children, politics, school, events, greetings to humour/sarcasm.

People are also free to express their experience of their illnesses, and they do this by using language patterns to support themselves and others. There are natural patterns that are detectable for the constructs of support. These patterns do include advice with stance-taking of people with diabetes.

The platform is also about the patients and organisations that set up the site in the first place, and it is for all users who have concerns about their issues. People can use their comments and those of other individuals in the discourse network. The platform is also there to mitigate against harmful support (e.g., to recommend visiting the doctor if their advice is inadequate).

The analysis confirms the claims of the thesis. CL (including the LDA model, MeaningCloud, LIWC, Entity Recognition, Sentiment Analysis and DA) has provided samples of posts. It helps the research to show sophisticated patterns of support for, for example, advice with stance-taking for support. The patterns show what the most utilised advice with stance-taking types in the posts are (see Chapters 4 and 5).

There are many examples of advice with stance-taking in the corpus (Appendix 7-1). Across the posts sampled, the items found cover (i.e., advice-seeking by personal disclosure and advice given by experience).

The users also converse about the broader issues of diabetes in such a way that there tends to be some broader *counterbalance of each other's sophisticated support patterns. For example, advice with stance-taking amongst many other discourses ranging from humour to charity via of primary importance 'topic' and 'TOPIC' to create shared meanings of uncertainty and one's sentiments about domain-specific targets of diabetes, blood, pump. There is remarkably a high-frequency stance-taking with diverse affective stance and low epistemic stance but not necessarily in consecutive posts but rather indirectly in usage across many posts about the same target. The thesis focuses mainly on domain-specific targets of diabetes, blood, and pumps from many others ranging from targets of usernames, places, time, to years. It is their use of such language patterns that can make communication and participation possible in the corpora.*

The found ‘topic’, ‘TOPIC’, targets and linguistic forms such as advice with stance-taking all point to a super-category of ‘support’. All central discourse and advice with stance-taking types found are in Appendix 7-2. The critical literature has already suggested some of those found.

The observed patterns in the corpus and findings need an explanation. What is their significance, and what do they reveal about the users? The found ‘topics’ and ‘TOPICs’ cover the necessary opposites of illness and wellness and vulnerability but motivation to deal with such an illness. People utilise language patterns jointly, meaning that they can get support for a few posts.

This thesis is further reviewed and evaluated in this section in order to determine its success in answering the research questions posed in Chapter 1. The operationalised questions posed in Chapter 3. The aim is that the results generated by the thesis respond to the research questions. The discussion involves the analysis of threats to the trueness of the findings and contributions. There is also a brief discussion of future directions for extending the research carried out in this thesis.

This thesis has proposed a novel theory of peer SSMD research and an example of mixed approaches for the analysis of the discourses. It has also suggested that the idea of a theory of support for SSMD can be of significance, for example, in healthcare communications. It can also help to improve the online support for long-lasting illness.

The thesis interpretation, however, overlooks (with a focus on the text as one way of going about this type of study) a study of images (Hunt, 2015) or comparisons to face-to-face doctor-patient interactions (Hunt and Harvey, 2012) in understanding wider online support networks.

The language patterns, as suggested by relevant healthcare communication linguistic researchers, were also relevant and included the idea of central target and topic. For example, people converse with each other about their blood glucose levels or medications. Well-researched patterns of advice identify any similarities in the data.

The evidence provides ways to answer research questions about the healthcare domain. The FDP site did offer a place where people with chronic diabetes may mutually support each other. The corpora used in the analyses are from SSMD for diabetes, a chronic illness, and people do go about collaborating. It is notable in *how* and *what* they converse about to each other.

The corpus helped to see groups that work well together in support work. A related coding of the data in the corpora, for example, is to discover a stance-taking pattern. The affective stance and epistemic stance shows to be part of the support. The labelling is not easy to do with automated methods. However, it works by finding the objects, sentiments and certainty, and the resources used to position the peer in their posts as part of the broader corpus interactions. It made it possible to get the purpose and context and content for broader explanations of support, as expressed by many different people. This examination is of real language data and is a test of an explanation given in the claims. It can give a stable underlying structure for support, shows in the data analysis and findings in Chapter 4. It has brought the peers' concerns to the forefront. This examination made it possible to find, compare, and test the existence of language patterns that have implications for a proposal of a novel theory of support constructs (Chapter 5). The theory of support is about what people do naturally in order to help themselves in their communications and participation. It is up to the design of online platforms to ease those communication worries.

Healthcare employed in this thesis as a potential domain for starting to understand how people undertake collaborative online support work. People can utilise language from the evidence of support that includes, for example, advice patterns. This thesis is an attempt to discuss the role of support in thinking about, for instance, healthcare advice with stance-taking. The idea of the evidence of support in the discourse is about the language-based patterns that groups who work well together can use for support work constructs.

Contribution 2: ***AI and Linguistic Analysis: Research design with predominate computing and artificial intelligence (AI) and topic-modelling, context analysis, and sentiment analysis together with guidance from linguistic analysis. The approach shows the limitations of either aspect on their own.***

For this research field, the thesis suggests a process to identify and compare high frequency 'topic', 'TOPIC', targets, discourse categories and linguistic forms such as advice with stance-taking devices in the discourse, together with an example of the FDUK corpus. There is joint employment of primarily AI machine learning and with nuclear language-based research. There is a complementary linguistic analysis. It is as separate ways of finding facts, and they are used together as one unit of analysis.

It is also not easy to find possible patterns of support. The use of CL (with machine learning LDA, Entity Recognition and Sentiment Analysis and DA) adds to the existing literature. It can help to compare high-frequency 'topic' and 'TOPIC', targets and linguistic forms such as advice with stance-taking. It is across the entirety of a large online corpus together with an analysis of individual posts.

A novel way is to look for and of primary importance in my thesis '*device-enabled discourse purpose categories*' and '*device-enabled discourse content categories*'. LDA is used in this instance, as it finds latent word usage patterns across posts and creates 'topic' and 'TOPIC' categories. The thesis argues that if trigrams are used both with and without stopwords, then LDA may give the latent devices that people use when they converse with each other. The combinations of content words with stopwords to say something about what the thesis calls '*device-enabled discourse* *purpose categories*' and '*device-enabled discourse content categories*', *respectively.* These are the resources for advice-seeking and advice-giving with stance-taking patterns.

Advice patterns identified via automated AI and software analysis helps with an aided partial manual identification. It is in comparison to other vital researchers' work. Sharing stance-taking is more complex, requiring identification of the resources, the target or TOPIC of the post, the sentiment and emotion in the posts and the certainty of the information.

Next, we will reflect on the methodology and how it affects the understanding of the data. The LDA topic-modelling may reveal most of the issues discussed on Facebook, and the results may be a product of topic-modelling. The data may contain far more variety but (albeit with not high frequency) that is not captured by the topic-modelling – an endless quantity. The analysis of ever-increasing quantities of data could show increasing patterns that people are employing for support.

However, this thesis suggests that there is primary advice with stance-taking devices. Besides, there can be endless 'support conversations', but not all of them necessarily make sense. The community tends towards the usage of these identified ones found via an integrated automated and aided partial manual labelling and DA approach. Similarly, at the other extreme, there could not be zero patterns about the behaviour or psychology of people in their online interactions. People converse with each other in particular ways to support each other about diabetes, and this thesis provides evidence of this.

These mixed approaches can help overcome some of the objections. Machine learning with LDA can help focus on the detection of latent features in a large-scale corpus. CL can contribute to focusing on data that are more significant and many more recurring patterns such as advice with stance-taking of people and broader linguistic forms. LIWC can help identify healthcare, emotions and certainty. MeaningCloud can help identify the entities or targets in a post and the global sentiment. DA focuses on the relationships and these concepts from an SFL framework. The use of low, neutral or high certainty, sentiment or emotion can gauge the affective stance of the post. DA can help give context to the quantitative patterns such as it being about the mitigation of risk between people. People must trust each other for advice, which is critical for healthcare and for gaining power and solidarity.

This mixed approach can offer a powerful way to go about analysing the text for high-frequency trends such as advice with stance-taking. This process can help make credible how a theory for SSMD forms from the observed language patterns.

The next section is concerned with evaluating the research findings and contributions.

Evaluation

It is not easy to conceive of a post in the corpus with a text/sentence of, for example, '*what is the meaning of my illness*?' A person's chronic illness may be more identifiable and supported if the posts have sophisticated patterns, as suggested by the theory in the thesis. An example is with voses, nents, targets and linguistic forms of advice with stance-taking features. It is easier to imagine the post of some individuals as standing for what they mean about themselves. It is their direct experience. It is through the many conversations in the discourse about the meaning of their conversations with others. This meaning construction matters. People converse about the sense of things about their lives as chronically ill patients. It may be difficult to do otherwise from an understanding of the progression of their disease or their regular daily self-management routines.

Another argument is that by taking a CL, the language-based approach may be an agreement with Teubert (2007a, 2007b). Each back-and-forth conversation to agree on meanings can take a ‘topic’, ‘TOPIC’, targets and linguistics forms like advice with stance-taking patterns. These patterns can be time-related or discourse-time-related and show possible continuous support of a person. The corpus, however, shows that the patterns are not only time-related but are a feature in different posts spread across the corpus history of posting. The advice with stance-taking is shared and developed by individuals working as a group over time and observed in both the visible and not visible links. These are between a post and earlier or later post in the conversation and its comparisons. The same or different people can develop a variety of posts.

Every person who takes part in the discourse tends to inherit the saved support of all earlier users and people must work to reach an agreement for dealing with issues. The corpus offers a potential advice facility but can also a searchable medium for support.

A sophisticated support pattern is available to groups that work well together. The assumption in this thesis is that groups that work well together employ a language pattern. These people can act in good faith. They may be ethical and can be bound together by a rule to achieve healthcare for each other (which is equal to any other person on the platform).

Of importance is what the thesis findings should say to the literature. It adds to what is known about language patterns that were begun in support of diabetes in social media. What is different in my thesis is that the focus is on the person with the illness. There is now a novel theory to support that person. There is a novel multi-method AI and linguistics text analysis technique. It can offer many advantages for healthcare communication.

The theory and analysis method can help linguists reach an understanding of language variation within different situations, and how variation may occur in predictable ways. Future analysis of more online studies of these types of discourse could illuminate the planning of chronic illness support on social media. Such analysis can be used to diagnose areas needing attention in peers' writing. It can then be used to make more informed planning decisions about the type of platform to use. One of the goals is for these systems to be tailored in order for peers to develop their linguistic repertoires for expressing textual, ideational and interpersonal meanings by organising and structuring texts. SFL is an approach that can assist applied linguists in addressing the issues described. Researchers can use SFL to inform online support (see Chapter 2).

My thesis applies an SFL approach. It suggests that there can be an understanding of the affective stance and epistemic stance. It is in their grammatical and lexical choices that may provide the means for peers to express their intended meanings. The analysis in this thesis has focused on advising and stance-taking features. This type of analysis of texts using the SFL framework can reveal what is valid for a particular purpose within a particular context of the situation. Comprehensive SFL studies could help to make explicit the holistic interrelationships between clause level grammar, text structure and social situations. The SFL approach adds to current practices in support studies, and the linguistic features of texts are particularly crucial for peers who are dealing with a severe and chronic illness.

6.3.1 Limitations

DA is utilised on random posts selected from within the LDA 'topic' and the 'TOPIC' of the posts of the LDA models. This approach also gives further analysis and comparison of high-frequency patterns of the LDA 'topic' and 'TOPIC', targets, linguistic forms and interactional types and their posts. It is in order to gain a context of what people were conversing with each other about and how. The reduction of large-scale text data into a few high-frequency patterns is not without its problems. Linguistics guides an understanding of the corpus. However, this is only one possible way of looking at the corpus.

Firstly, the aim is, to begin with, the analysis of a rich and extensive text-based corpus. Given this situation, it is hardly surprising that the Diabetes UK corpus needed to have an entry point into the data. The idea is to be able to analyse all posts and to see about high-frequency patterns for all main 'topics', 'TOPICs', targets and of linguistic forms such as advice with stance-taking devices. LDA offers such an approach evident in the literature about how to go about doing an 'all' analysis. At the same time, it is not easy to find out what any individual posts in a corpus would be about merely from the analysis. Finding tendencies and patterns from looking at the corpus with an LDA topic model is one way of analysing the 'all' and traditional topics.

An entry point should be arbitrarily chosen in order to dispel researcher bias. The painstaking human way of doing this could be time-exhaustive and have a potential bias. To this extent, machine learning provides a way to begin the analysis. This approach meant that the entire corpus could be covered automatically, and the analyses would be able to be conducted in an unsupervised and systematic way.

Secondly, the use of automatic analysis via algorithms meant having to understand how the algorithm goes about analysing the critical data. There is a continuous checking of the parameters. Fifty topics, for example, seemed to make sense for the model. It is rather than fewer or much more. It also, in a similar manner, uses five hundred features. However, fundamental research from linguistics was used to help understand what it is that the patterns from the LDA model were revealing. Knowledge of the domain helps with this procedure. Together, AI, entity recognition, sentiment analysis and DA can be used to identify patterns. LDA and DA analysis and the use of LIWC and MeaningCloud helped increase the reliability of the findings in the thesis.

This thesis gives a source for observations that prove the support constructs, and by support 'topic', 'TOPIC', targets and of linguistic forms such as advice with stance-taking and this device identification and categorisation. Other large-scale analysis work can be carried out on support patterns in other online technologies. Platforms for people to get help with their conditions and issues, and those in other healthcare domains, such as cancer and heart disease, may share features with the diabetes platforms. There could be distinct differences. This expansion is to understand more about the types of support, categories and devices studied in the thesis. This more extensive study can help to explore each support type language pattern. It can give more evidence. It can help provide support constructs in technology in some other support platforms. It is for people to get help about other illness concerns and issues.

An analysis of the rich data involves finding key features and using computational math-based and probabilistic approaches to reduce the data, time and work within the current limits of computational costs. The development of more relatable software is necessary based on this research in order to document, for instance, the complexity of advice with stance-taking observations that prove patterns of support.

The development of the theory was done using quantitative and qualitative-based data analysis. It is because of what is and is not essential for a whole approach, a rich analysis, and an explanation of the corpus. The arguments for patterns in particular language data can be combined with the fundamental truths about language to help develop theory. It is to explain why the patterns are found and why a proposed theory may be correct. The purpose and content of the support are as important as finding any other features in support. It is essential to consider how to go about finding the data to back this up.

A probabilistic model of language must take into consideration how groups of people work together in conversations. It must consider how people develop values and rules, how they share their direct experiences publicly, and how they produce meaning from the sentences for the issues in their lives. The theory and its predictions can be further developed through the analysis of other social media discourses for support.

6.4 Implications

This section looks at what the consequences are of this new knowledge. It gives examples of how the analysis of data can inform and improve support (see Appendix 7-1 and 7-2). The purpose of the discussion is to explain how and why linguists and social media platform designers can receive insight from these findings.

Linguists

Linguists can benefit from the ideas obtained in this thesis about people experiencing issues and concerns and about language patterns and theory of support. It can also inform about how to improve support from organisations, such as Diabetes UK, and advances in technical assistance such as social media support platform usage, design and topic-modelling for pattern extraction.

Corpus linguists

Corpus linguists can benefit from the ideas about combined approaches (with AI machine learning, entity recognition, sentiment analysis and complementary DA) for finding patterns in large corpora. In carrying out the analysis, what is and is not doable in each type of analysis is demonstrated here by the combination of AI for automated text analysis together with DA on an example corpus (in this case, Diabetes UK). AI and software can be developed for a qualitative analysis that is then carried out on the quantitative results. The quantitative results are obtained automatically by the software. They are also manually conducted for the identification of sophisticated linguistic features of, for example, advice features. The software needs to automatically make comparisons of the results from the state-of-the-art research in, for example, the diabetes domain. AI and software-automated analysis are shown to have limitations for the identification of sophisticated linguistic features. DAs are shown to benefit from AI and software-automated analysis.

Artificial Intelligence AI

AI practitioners can benefit from language patterns to create support-bots (automatic conversation) to help people profit from the more extensive scales and quicker responses for support. Within the context of this research, the functional devices can be used in a generative fashion (e.g., Shang et al., 2015). It is where the support-bot can respond from scratch to a person's need for support and converse with that person automatically. CSCW and SMCW systems can benefit from conversational bots.

The 'topic' and 'TOPIC', target from the corpus, however, can be under continuous analysis by the support-bot. It would be to produce ready retrieval-based LDA models. It can give support to any person based on what other people are saying, by applying it to that person's support query. The retrieval and generative models are combined. Kerly et al. (2006) have argued for the richer use of human language for bots that are developed to help humans in many ways. The system can be made to act in a way that embraces fuller language meanings for support in each conversation (post) between people. A study of 'meaning in language' can give understanding (see Chapter 2). There are possibilities of machines that can help people. The suggestion is made to bring people closer to supporting each other by using support-bots.

The bots can automate conversations. It is because people can go about slowly developing their 'support conversations'. The support-bot can use language patterns to predict what people are talking. It can be to link them to the right dialogue in the system of many different post threads by various people (i.e., in the discourse). The support of any individual is obtained by linking all the relevant past support to that person's request. The many 'topic' and 'TOPIC' categories, main advice, and stance-taking devices in the memory of the support, and individuals that were helped historically need to be considered. The sharing of human experiences is crucial. It is by way of the meaning of the sentences. There is a sense that helps people face their concerns and issues.

As another example, there could be a feature that people could use. It could be a button to press that could process their communications about all the previous support from other people. It can give any other person support based on previous communications. It is an important consideration. A support-bot could be programmed to do this. Then the person using the platform may know the 'topic' and 'TOPIC', targets and linguistic forms like advice and stance-taking they may use to converse with others. People may be able to ask for help with the platforms ways of doing things and speed up their support.

6.5 Future research

This section sets out suggestions for future research based on findings and the review of the literature from the thesis. Also, it attempts to incorporate the results of the research as well as suggesting directions for future research based on this work.

6.5.1 Domains

The theory needs to be checked on different SMCW groups. Predictions on a different corpus that relate to healthcare conversations can be used to test the SSMD theory. It can be trialled on CSCW and SMCW systems with their many different and particular social problems that the participants are attempting to 'solve.'

6.5.2 Automated tools

Computational math-based tools can be developed for analysing the text in online healthcare communities. It can combine quantitative and qualitative analysis and can automatically prepare a corpus for an LDA, entity recognition and sentiment analysis and a further DA analysis. The algorithms, (e.g., LDA) can also be listed but need to include thoughts about their usage.

There is also a need for automated software for the anonymisation of personal online data postings when shared with others via academic publications and theses. The online ethics process of this thesis can offer some help in a more significant project in accomplishing these tasks.

6.5.3 Chatbots

For future research and based on the findings of this thesis, it is suggested that improvements can be made to platforms where people take part in social conversations. Authors generally emphasise (e.g., Shang et al., 2015) that there are areas for improvement.

The thesis examines where exactly people cover many 'topics' and 'TOPICS' in their 'conversations' and where they produce many interactions resulting in large-scale text data on these platforms. This idea from the work of the thesis (as a secondary goal and of preliminary research for AI) is that chatbots may help people to interact quicker. It is, for example, with a more rapid search for relevant advice from the large-scale data. It can also be done from any individual's data, a person who has given that type of support. However, first of all, it is necessary to identify possible 'support patterns' from people who use these platforms and in a suitable discourse and the thesis has provided these patterns.

Turing (1950) also considered these types of problems by looking at how machines can mimic humans in conversation. He considered the possibility of thinking machines in general rather than machines that, for instance, can analyse or produce specific, meaningful sentences. By defining 'thinking' in machines and in building a thinking machine, his approach has helped to produce many technologies to help people go about their daily lives.

The need to make machines think and to define 'thinking' (Dreyfus, 1972; Searle, 1980) is an open problem in AI research (Rich et al., 2009). For them, machines that can think as humans do, are about having a computer that convinces any human that it is 'thinking' by imitating the behaviour of any other human.

Hill et al. (2015) have shown that there is a difference between current chatbot-to-human interactions and human-to-human interactions related to the 'richness' of the employment of human languages.

There are problems associated with, for example, the ‘about-ness’ of topics. Automating conversations is not without its problems. Machines do not have the connections to illness or anything in ordinary human experience. They also suffer from a lack of about-ness or a link to anything in the world.

The problem of intentionality and machines that think, as noted by Searle (1980), is shown in the Chinese room example. Humans can instantiate a program that consists of a list of instructions to complete a task but not with any suitable intentionality about the task. Searle’s thought experiment involves a person who can process the input of Chinese characters given to them by way of an access point to a closed room. They can employ information to process the outputs and give back responses without knowing the Chinese language itself. In this way, they can convince others that they are talking to a Chinese speaker. When considering the problems detailed above, developing a chatbot for support is no easy task.

Blei et al.’s. (2003) familiar topic patterns can be developed into a computational linguistic model. It can be employed to try to converse in the ways that people talk to each other but using ‘topic’ and ‘TOPIC’.

It can be achieved, for instance, with bots and a type of software that can automate human-like practices, such as producing tweets on Twitter (Shang et al., 2015). It is hard, however, to program these bots to behave like humans in conversations.

ALICE (2017) gives an example of an ‘artificial linguistic internet computer entity’ that is capable of limited intelligent conversation. It has some ability to come up with a relevant response to any text input. However, it is not easy to make the entity seem human-like when conversing.

Another example is the famous Eliza chatbot (Weizenbaum, 1966) that acted as if it was a psychiatrist. It talked to people in order to give psychiatric support. Eliza used language patterns that were sets of pre-programmed responses from typical language patterns that people employ when speaking to a psychiatrist. The bot could 'speak' about limited issues that concerned patients but were not always able to provide human-like support.

The current technology of 'generative models' (Fabian and Nicolae, 2009) has problems with parsing input from people. They need to 'understand' what people mean in the sentences they employ. Fabian and Nicolae (2009) show that the reverse process of outputs that are constructed in natural language generation may be more straightforward. It may be easier for machines to employ selected data and output responses. They discuss the problems of the occurrence of 'unknown and unexpected features in the input'.

For them, generative models can be about developing automatic responses from the entirety of the corpus data (i.e., with the use of neural networks and deep learning).

Shang et al., explain about a 'neural responding machine' based on a neural network response generator for short, textual conversations. This machine employs a general encoder-decoder framework. They show how it formalises the generation of replies. The decoding process is based on the underlying model of the input text. For them, both encoding and decoding are realised with recurrent neural networks. They claim it offers more human-like conversational qualities to chatbots.

This section explains a way of automating this support-conversation in the entirety of a human-based discourse. Previous authors have argued that by discerning patterns in corpora, one can generate meaning in sentences. AI researchers (Russell and Norvig, 2009; Hill et al., 2015) have looked at meaning in NLP done by machines. Machines have been shown to 'speak' about some topics. However, they have not been entirely taught about how to go about having conversations with humans in natural ways.

Crook et al. (2009) have demonstrated that dialogue systems can benefit from using the language-based abstractions of human dialogue acts. The work presented in this thesis could be used to develop machines utilised for support conversations focused on the meaning of sentences. The use of the language patterns studied is not about how fast a person can get the support they need. The person may not access the knowledge of the entire system of connected dialogue at once. Therefore, automating the dialogue with support-bots can help with scale and the rapid processing of, for example, advice-asking with stance-taking. A machine can be created to generate text for support automatically.

This thesis proposes that people can support each other as best they can. It is with the use of specific 'topic' (vose), 'TOPIC' (nent), target and linguistic forms such as advice with stance-taking language devices. People using posts can try to make planned-out support by standing up for each other. It can be through meaningful use of language and by creating meaning about their support generated by way of the sense of their sentences. They can use '*device-enabled discourse purpose categories*' and '*device-enabled discourse content categories.*' There is remarkably a high-frequency stance-taking with diverse affective stance and low epistemic stance but not necessarily in consecutive posts but rather indirectly in usage across many posts about the same target.

For building AI functions, the 'topic' and 'TOPIC' devices may prove valuable (see Appendix 7-1 and 7-2). The function then works as well as it is programmed to be. It may match and approximate how humans continuously develop complex support behaviour by conversing with each other in specific ways. People can sometimes act similarly to this machine-like way of support. They can tend to find a good enough solution focused on the meaning of the sentences rather than searching for all possibilities in order to find the best option. Simon (1969) calls this good enough process satisficing. To be explored further is what the thesis here suggests is that machines can mimic this manner of support, e.g., for diabetes. This approach can be thought of as satisficing versus the optimisation way of thinking.

The thesis shows that people use specific 'topics' and ways of posting more than others to form groups of sentences. These groups are used more than others are, and so can be regarded as planned-out or purposeful. My thesis suggests that meaning in the broader context of human living can be related to how a universal support machine can best help support any other human or any other computer, at any time, and anywhere. The idea of meaning for AI is about approximating language patterns for language-based meanings. It can be made up of 'topic' and 'TOPIC' categories, targets and linguistic forms like advice with stance-taking devices and counterbalances that look like support given by any ordinary person.

Support-bots may rapidly utilise language patterns in order to have conversations with large numbers of individuals and engage them in support practices. Such bots as these can be the best way to generate combinations of 'topics' and 'TOPICs' and, e.g., advice with stance-taking skills in order to speak to others naturally. It can employ a retrieval of vast databases of responses, such as the FDUK. It can utilise a combination of retrieval and generative language models where the machine must work out on its 'own' what to say next (Norvig, 2011, 2016). A combination of retrieval and generative data for patterns, as found in this study, may be necessary. It is to harvest the live interaction data from all past and present people having conversations about the support. It should be able to give the best responses at any time. It should generate the best responses from stored knowledge of possible language patterns for support.

7 APPENDIX

7.1 Random Sample Consecutive Posts with the same target

7.1.1 Consecutive and non-consecutive posts with target Blood

Table 7-1: Consecutive and non-consecutive posts with target Blood

Consecutive and non-consecutive posts with target Blood							
All no	Thread number	Anonymised randomly selected consecutive posts from Facebook with advice labels	Literature: Advice features	Meaning Cloud other targets - example first 3 targets plus if blood is a target. As in the original non-anonymised posts	MeaningCloud and LIWC Stance-taking features	LDA: device-enabled discourse content categories	Randomly selected LDA: device-enabled discourse purpose categories
1	86	'...just came out of hospital…developed keto-acidosis and kidney infection…my blood glucose level was approximately fifty AG_E …. great…happy…'	AG_E(potential advice-giving in the social group via problem disclosure)	Targets: place, guy, kind, blood	High positive emotion and positive global sentiment and affective stance; and zero certainty and high tentative and zero insight and epistemic stance for high healthcare	TOPIC 27: feel better soon	topic 24: you so much
2	87	...living with type 1 diabetes….about ten years……inject twice a day with humalin i - a long lasting insulin….fun!! …blood sugars are generally pretty good AG_E …. if anybody on here wants to chat about type 1 diabetes just drop me a message AG_D …	AG_E(potential advice-giving in the social group via problem disclosure) AG_D(advice for contact)	Targets: username, long, Humalog, blood	High positive emotion and global neutral sentiment affective stance; and high certainty and high tentative and zero insight and epistemic stance for high healthcare	TOPIC 38: high blood sugars	topic 24: you so much

62	189527	…i am on novomix approximately three times a day…most people taking this seems to inject approximately twice a day… i find it ok for routine days…. but when i do anything different during holidays…doing more activity…eating out…find that my blood sugars go all over the place…. am considering changing to levemir and lantus AA_P … would appreciate any comments on the pros and cons AA_R	AA_P AA_R	Targets: novomix, Levemir, lantus, blood	High positive emotion and global negative sentiment and Affective stance; and zero certainty and high tentative and zero insight and epistemic stance for high healthcare	TOPIC 43: good morning hope	topic 24: you so much
63	189528	…my child was on novomix… from first diagnosed at the age of approximately two… i kept my child on this for approximately six months…child was then moved to novorapid and glargine…this gives a better control over your blood glucose levels…you match the insulin to the carbohydrates...can reduce or increase accordingly… so it is a lot better to manage AG_E … i personally think AG_H -happy	AG_E AG_H	Targets: child, blood, sugars, insulin	High positive emotion and global positive sentiment and Affective stance; and zero certainty and high tentative and high insight and epistemic stance; for high healthcare	TOPIC 28: questionmark good luck	topic 34: you will get

7.1.2 Consecutive and non-consecutive posts with target Diabetes

Table 7-2: Consecutive and non-consecutive posts with target Diabetes

Consecutive and non-consecutive posts with target Diabetes							
Allno	Threadnum	Anonymised Randomly selected consecutive Posts from Facebook with advice labels	Literature: Advice features	Other targets - example first 3 targets plus if diabetes is a target As in the original non-anonymised posts	MeaningCloud and LIWC Stance-taking features	LDA: device-enabled discourse content categories	Randomly selected LDA: device-enabled discourse purpose categories

	ber						
2	87	...living with type 1 diabetes…. about ten years……inject twice a day with humalin i - a long-lasting insulin…. fun!! …blood sugars are generally pretty good…. if anybody on here wants to chat about type 1 diabetes just drop me a message…	$AG_{E(\text{potential advice-giving in the social group via problem disclosure})}$ $AG_{D(\text{advice for contact})}$	Targets: username, long Humalog, diabetes	High positive emotion and global neutral sentiment affective stance; and high certainty and high tentative and zero insight and epistemic stance for high healthcare	TOPIC 38: high blood sugars	topic 24: you so much
3	88	'…hi everyone...i have had Type 1 Diabetes for approximately eighteen years…i am approximately twenty-three years…even though i only look sixteen!.... AA_P i just wondered if anyone on here has had Diabetes for a similar amount of time AA_{OI} …experienced any of the problems associated with Type 1 Diabetes…that you hear about? AA_{OI} AA_B...touch wood…the only problem i have got at the moment is having to wear glasses!...i have just started counting carbohydrates…it looks as though i am starting to get my levels under control!...the problem is i enjoy having a few beers…with my teammates…on a weekend...i play rugby league…was a social smoker until recently AA_P …i do know I am not an idiot!... if anyone can help me out with these questions…or just fancies a chat… then drop me a line AA_R …cheers…'	AA_B AA_P AA_{OI} AA_R	Targets: username, diabetes, type wood	High negative emotion and global neutral sentiment and Affective stance; and zero certainty and high tentative and high insight and epistemic stance; for high healthcare	TOPIC 31: diagnosed type years	topic 31: diagnosed with type

4	296	'…hi…i was diagnosed with type 2 diabetes just last year... still finding it difficult to make small changes…in my lifestyle…getting used to the medication AA_P…so if anyone has any tips AA_{OI} …stories AA_B… send me a message AA_R …thank you…'	AA_P AA_B AA_{OI} AA_R	Targets: type Diabetes, lifestyle, medication	High positive emotion and global negative sentiment and Affective stance; and zero certainty and high tentative and zero insight and epistemic stance; for high healthcare	TOPIC 31: diagnosed type years	topic 34: you will get
5	297	'…i was diagnosed with Type 1 diabetes approximately november 2008 AA_P…'	AA_P	Targets: type, diabetes, Tuesday, 12/09/2008	Zero positive and negative emotion and global high negative sentiment and Affective stance; and zero certainties and zero tentative and high insight and epistemic stance; for high healthcare	TOPIC 31: diagnosed type years	topic 34: you will get
6	298	'…hi i was diagnosed with type 1…approximately november 1994 AA_P…i hate having it… also got gastroparesis…because i am diabetic…'	AA_P	Targets: type, diabetic, november-94	High negative emotion and global negative sentiment and Affective stance; and zero certainties and zero tentative and high insight and epistemic stance; for high healthcare	TOPIC 31: diagnosed type years	topic 34: you will get
7	987	'…hi…i was diagnosed with type 2 diabetes… approximately eighteen months ago... i am afraid i largely ignore it…just pop the pills and get on with life…i do not like to feel I have an illness… never had any symptoms… thirst…great weight loss or anything... it runs in our family in later life...i am approximately 60 years old… so just	AA_P	Targets: username, type Diabetes, pill	Equal negative and positive emotion and negative global sentiment and Affective stance; and high certainty and high tentative and high insight and epistemic stance; for high healthcare	TOPIC 31: diagnosed type years	topic 34: you will get

		accepted that I might get it one day AA_P …'					
8	988	…partner has gotten type 2 diabetes … quite worried …. partner takes tablets…does not have sugar...partners blood level is not consistent…. my uncle died of kidney-failure……i am scared…partner has an appointment with the diabetic doctor tomorrow AA_P….'	AA_P	Targets: type, diabetes, tablets, sugar	High negative emotion and negative global sentiment and Affective stance; and zero certainty and low tentative and zero insight and epistemic stance; for high healthcare	TOPIC 21: hope comes soon	topic 31: diagnosed with type
9	1029	'…hey everyone!....i was diagnosed with type 1 diabetes at approximately age eleven…i am now approximately age sixteen…i have an accu-chek insulin pump...which honestly is amazing!! AG_P...add me if you are diabetic…plus if you want to talk AG_D …laugh…happy …'	AG_E AG_D	Targets: username, type Diabetes, insulin	High positive emotion and neutral global sentiment and Affective stance; and high certainty and low tentative and zero insight and epistemic stance; for high healthcare	TOPIC 31: diagnosed type years	topic 34: you will get
10	1030	'…hiya…i was diagnosed approximately thirty-five years ago… just before approximately my fourth birthday…my diabetic consultant wants to put me on an insulin pump… i am having an approximately seventy-three-hour glucose monitor…in approximately september…waiting to go on a carbohydrate counting course AA_A …before i can go on the pump… I am hoping to be on it by christmas AA_I … i am having loads a hypo's…when i reduce the insulin it goes high..i cannot win!!. AA_P ..hope it has the same impact on my life as yours obviously has AA_B …'	AA_P AA_I AA_B AA_A	Targets: hiya, username, insulin, diabetes	High positive emotion and global positive sentiment and Affective stance; and low certainty low tentative and low insight and epistemic stance for high healthcare	TOPIC 11 questionmark questionmark questionmark	topic 31: diagnosed with type

13	2136	'…it is supposedly type 1, not type 2, that is hereditary….but it is not clear exactly how or why some people with diabetes in the family get it… others do not...my partner has had type 1 since approximately 2000…our approximately four years old child was diagnosed in approximately january…but our approximately 5 year old child does not have diabetes…approximately 3 of my partner's relatives died of type 1 complications in the 1950s… they were approximately fifty-one years old.. i might add…which i think was pretty good going at that time…but their sibling, my father-in-law…did not have it... type 2 is lifestyle-related…but from what i understand, in order for a bad lifestyle to give you type 2, you have to have the diabetes gene... but never ended up with type 1 AG_E …. but do not take my word for it…i am not a doctor AG_H…just done tons and tons of reading about it!...hope that helps…but do read more about it on the Diabetes UK website… it is very helpful…'	AG_E AG_H	Targets: username, place, type, diabetes	Low positive and negative emotion and negative global sentiment and Affective stance; and high certainty and high tentative and high insight and epistemic stance; for high healthcare	TOPIC 19: happy new year	topic 13: questionmark questionmark questionmark
14	2137	'…thanks…it is very possible that the so-called diagnosed type 2 diabetes is a look-a-like ailment...often misdiagnosed by most physicians…. fatty deposits and bad cholesterol are either the culprits… or just contributing…. i have	AG_E AG_H	Targets: username, diabetes, ailment, fatty	High positive emotion and global negative sentiment and Affective stance; and high certainty and high tentative and high insight and epistemic	TOPIC 8 Monday Friday pm	topic 34: you will get

		not been comfortable with this determination…so i diet…take oral meds which definitely helps AG_E … but cannot help to think…that there is more efficient medication… that would annihilate the disease totally AG_H …'			stance; for high healthcare		
19	5615	'…hello…my child is approximately seven years old…was diagnosed with type 1 diabetes…approximately eleven weeks ago… AA_P'	AA_P	Targets: username, type Diabetes, seven	Zero positive and negative emotion and high global negative sentiment and Affective stance; and zero certainty and zero tentative and high insight and epistemic stance; for high healthcare	TOPIC 31: diagnosed type years	topic 34: you will get
20	5616	'…hi there… i was approximately five years old…when i was diagnosed with type 1..i am now approximately twenty-eight years old…getting married shortly…have no complications at all…live a totally normal healthy life… everything works just fine AG_E ….the trick is to strive for good control…but if you do hit highs and lows, as long as your sugars are normally reasonable…your child should do fine…there is never a guarantee AG_H …but just do not. freak out…if she hits rough spots…. every diabetic does…hope this gives you hope… i know it has helped me to read other people's stories AT_P…'	AG_E AG_H AT_P	Targets: type, married, highs, diabetic	High positive emotion and global positive sentiment and Affective stance; and high certainty and high tentative and high insight and epistemic stance for high healthcare	TOPIC 12 message add friend	topic 34: you will get

25	8593	'…me too…i have been diabetic for approximately twenty-eight years…really struggled to have any control over it…until i did the course just over a year ago…made such a difference to my life…i felt confident enough in my control to go on and have my beautiful child approximately six months ago AG_E …i agree with the username in encouraging everyone to go on the course AG_H …'	AG_E AG_H	Targets: username, course, diabetic, baby	High positive emotion and global positive sentiment and Affective stance; and high certainty and low tentative and low insight and epistemic stance for high healthcare	TOPIC 47: feel free add	topic 13: questionmark questionmark questionmark
26	8594	'…just found out my approximately eleven-year-old niece…had been taken to hospital…diagnosed with type 1 diabetes AA_P …i have been searching for information on it for a couple of hours now...i am a bit saddening for the child…but hopefully the child will be able to cope with it …if anyone knows any support groups or advice AA_R pages for children. i could forward to my sibling for them…it would be much appreciated…sad…'	AA_P AA_R	Targets: niece, type, diabetes, couple	Low positive and negative emotion and global positive sentiment and Affective stance; and zero certainty and high tentative and high insight and epistemic stance; for high healthcare	TOPIC 31 : diagnosed type years	topic 34: you will get
27	8955	'…my children has just been diagnosed with type 1 diabetes… on approximately thursday…so we are in the honeymoon period AA_P … would love to hear from anyone…who can offer help and advice AA_R …to a mother dealing with a young child…with this condition AA_B …thanks…'	AA_P AA_R AA_B	Targets: child, type, diabetes, honeymoon	High positive emotion and high global positive sentiment and Affective stance; and zero certainty and low tentative and high insight and epistemic stance; for high healthcare	TOPIC 31 : diagnosed type years	topic 34: you will get

48	130512	'…does anyone else have problems…filling in their diary? AA_{OI} ..i have an approximately thirteen-year-old child…who always forgets to fill theirs out…we end up in a mad panic…when our hospital appointment is due!..we always try to keep up with it…but i have other smaller children…who take allot of my time…in the mornings…rely on my child doing this on her own…yet it is the one part of her diabetes we have not managed to succeed with...in the last approximately seven years of her having it! AA_P ..any hints and tips? AA_R ..thank you…'	AA_P AA_{OI} AA_R	Targets: diary, child, diabetes, thirteen-year-old	High negative emotion and global negative sentiment and Affective stance; and high certainty and high tentative and zero insight and epistemic stance; for high healthcare	TOPIC 21: hope comes soon	topic 13: questionmark questionmark questionmark
49	130513	…i know this is not helpful AG_I …been diabetic for approximately twenty years…i am still hopeless at it!... i get around it by having a blood glucose monitor…it can save my results…i can do it once a day…hopefully remember what corrections i have done…the carbohydrates i have eaten for the day... i never forget an injection… check my blood approximately nine times a day...but that wretched diary! AG_E ... grr!	AG_I AG_E	Targets: diabetic, results, injection, check	High negative emotion and global neutral sentiment and Affective stance; and high certainty and low tentative and high insight and epistemic stance; for high healthcare	TOPIC 47: feel free add	topic 24: you so much
56	178116	'…when i was diagnosed at approximately fourteen years old…to say the least puberty was not fun!..potential partners would not kiss me in school…because	AG_E	Targets: school, diabetes, fourteen years	High positive emotion and global negative sentiment and Affective stance; and zero certainties and zero tentative	TOPIC 28: questionmark good luck	topic 13: questionmark questionmark questionmark

		they thought they could catch diabetes!!!...' AG_E			and high insight and epistemic stance; for high healthcare		
57	1781 17	'…i found this really interesting AG_I ….i had a happy childhood...my parents' attitude to my diabetes…nearly approximately forty one years ago...was to pretend it was not a big deal…my insecurities were dismissed…i became very independent… partially to show i was not going to be beaten…a short, painful marriage did not really help matters….after being on my own for approximately thirteen years…i met another partner…who went out of the way to learn everything they could about diabetes...i have just started on a pump…it has been so helpful to have my partners support…not sure i would have had the courage to go ahead without them... i find it amazing that they still find me attractive…even though i have a little machine wired through my stomach! AG_E ...'	AG_I AG_E	Targets: username, diabetes, childhood, parents	High positive and negative emotion and global neutral sentiment and Affective stance; and high certainty and low tentative and high insight and epistemic stance; for high healthcare	TOPIC 4: fast acting insulin	topic 24: you so much
72	216361	'…people do not see the hard work that goes into trying to manage blood sugar levels AG_I …to try to keep healthy…after approximately twenty-six years since my child was diagnosed at approximately five years old with type 1 diabetes… you would	AG_I AG_E	Targets: blood, injections, people, diabetes	Low positive and zero negative emotion and global positive sentiment and Affective stance; and zero certainty and high tentative and high insight and epistemic	TOPIC 28: questionmark good luck	topic 13: questionmark questionmark questionmark

		have thought we would have something else to manage the condition… instead of relentless invasive blood testing and injections AG_E …how i wish for a day that my child was free from injections…'			stance; for high healthcare		
73	216362	'…i have type 1…was diagnosed approximately four years ago…after feeling unwell for approximately six months… having hypos and not realising it!...finally getting admitted to hospital as blood glucose levels were really high…after tests…being hooked up to an insulin pump…glucose drip overnight…the consultant came round the following morning…with the obligatory students…just announced this person has type 1 diabetes and will control it with insulin injections…needless to say I was shocked, stunned and devastated…i cannot express my gratitude to the staff…the continuing care I receive from my local hospital..It has been a hard learning curve for me and my family...control was difficult at the start but has been on a course…i can now count carbohydrates…it has stopped being so scary.. at times it is frightening…i have had huge problems with work… at times they just do not understand that I	AG_E AG_D	Targets: insulin, family, type, diabetes	High negative emotion and global negative sentiment and Affective stance; and zero certainty and low tentative and high insight and epistemic stance; for high healthcare	TOPIC 31: diagnosed type years	topic 24: you so much

		have to rest through illness AG_E … until I was diagnosed I did not know anything about it... let us get the in-formation out there!.. AG_D .all my friends and family just about understand…just not fully….'					

7.1.3 Consecutive and non-consecutive posts with target pump

Table 7-3: Consecutive and non-consecutive posts with target pump

Consecutive and non-consecutive posts with target pump							
All no	Thread number	Anonymised Randomly selected consecutive Posts from Facebook with advice lables	Literature: Advice features	Other targets - example first 3 targets plus if pump is a target As in the original non-anonymised posts	Meaning Cloud and LIWC Stance-taking features	LDA: device-enabled discourse content categories	Randomly selected LDA: device-enabled discourse purpose categories
9	1029	'…hey everyone!....i was diagnosed with type 1 diabetes at ap-proximately age eleven…i am now approximately age sixteen…i have an accu-chek insulin pump...which honestly is amazing!!. AG_E..add me if you are diabetic…plus if you want to talk AG_D…laugh…happy…'	AG_E AG_D	Targets: username, type, diabetes, pump	High positive emotion and neutral global sentiment and Affective stance; and high certainty and low tentative and zero insight and epistemic stance; for high healthcare	TOPIC 31: diagnosed type years	topic 34: you will get

10	1030	'…hiya…i was diagnosed approximately thirty-five years ago… just before approximately my fourth birthday…my diabetic consultant wants to put me on an insulin pump… i am having an approximately seventy-three-hour glucose monitor…in approximately september…waiting to go on a carbohydrate counting course AA_A …before i can go on the pump… I am hoping to be on it by christmas AA_I … i am having loads a hypo's…when i reduce the insulin it goes high..i cannot win!!. AA_P ..hope it has the same impact on my life as yours obviously has AA_B …'	AA_P AA_I AA_B AA_A	Targets: hiya, username, insulin, pump	High positive emotion and global positive sentiment and Affective stance; and low certainty low tentative and low insight and epistemic stance for high healthcare	TOPIC 11 questionmark questionmark questionmark	topic 31: diagnosed with type
29	22690	'…type 1…about fifty years... do not use a pump AG_E …looking for information on the internet AG_I …insulin pumps in search engine…brings up a variety of information…including contacts who will have been in a similar position to your children…need to satisfy your general practitioner or diabetic consultant or health authority… the current treatment method is not giving adequate control of blood glucose levels…it is giving rise to frequent hypos AG_E … good luck…'	AG_E AG_I	Targets: username, Pumps, general practitioner, type	High positive emotion and global positive sentiment and Affective stance; and zero certainty and high tentative and high insight and epistemic stance for high healthcare	TOPIC 21: hope comes soon	topic 34: you will get
30	22691	'…good luck with getting your child a pump AG_I username…my child is type 1...child has had it	AG_E AG_I	Targets: username, lotto, pump, child	Low positive and negative emotion	TOPIC 28: questionmark good luck	topic 31: diagnosed with type

		since approximately three years old… the child is now approximately thirteen years old…we still cannot get the pump...the hospital has changed their injections from two day to four a day… we are now learning how to count carbohydrates…we keep asking about the pump… child's nurse is all for him getting it…but the doctor will not give us it AG_E …i am beginning to think we have more chance of winning the lotto AG_I …sad…'			and global positive sentiment and Affective stance; and low certainty and low tentative and high insight and epistemic stance; for high healthcare		
58	179726	'…i have an approximately three-year-old child...on a pump since approximately january AG_E …we love it!...hard work but so worth it! AG_I ..'	AG_E	Targets: pump, work, three year	High positive emotion and global negative sentiment and Affective stance; and zero certainty and high tentative and zero insight and epistemic stance for high healthcare	TOPIC 31: diagnosed type years	topic 13: questionmark questionmark questionmark
59	179727	'…hello...i have had type 1 diabetes for approximately nine years…next month I am going on an expedition…will be trekking AA_A and camping for a month…i will be injecting while I am there…as i decided it would be difficult to take my pump with me AA_P	AA_P AA_R AA_B AA_A	Targets: place, diabetes, type, pump	Low positive and negative emotion and global negative sentiment and Affective stance; and zero	TOPIC 40 ha ha ha	topic 31: diagnosed with type

		…i was just wondering if anybody on here had done anything similar AA_B …if anyone had any tips…on how i can look after my diabetes properly…while I am over there? AA_R ..thanks…'			certainty and high tentative and high insight and epistemic stance for high healthcare		
69	2062 96	'…just looking at pumps…for my very active child…type 1…approximately eleven years old… the child dances approximately four times a week…is in all the school sports… was not keen on taking pumps on and off… to dance…but the omnipod may be the answer AA_P ….more research is needed…i think AA_{OI} …'	AA_P AA_{OI}	Targets: pumps, type, school, eleven years	High positive emotion and global negative sentiment and Affective stance; and high certainty and low tentative and high insight and epistemic stance; for zero healthcare	TOPIC 8: Monday Friday pm	topic 13: questionmark questionmark questionmark
70	2062 97	'…i will share this with my approximately thirteen-year-old child….as finally taking the plunge AA_A …to try the pump…child is very interested in this one AA_P …'	AA_P AA_A	Targets: pump, thirteen-year	High positive emotion and global positive sentiment and Affective stance; and zero certainty and high tentative and zero insight and epistemic stance; for zero healthcare	TOPIC 31: diagnosed type years	topic 13: questionmark questionmark questionmark

7.2 Topic modelling with LDA

7.2.1 LDA topic Model keep stopwords topics and number of posts

Table 7-4: LDA Model keep stopwords topics and number of posts

Number of total posts	LDA topic number	Number of total posts	LDA topic number	Number of total posts	LDA topic number
2941	0	1876	16	1889	32
1750	1	1360	17	1734	33
1201	2	2452	18	2332	34
1803	3	2374	19	1604	35
1688	4	2259	20	1516	36
1778	5	1667	21	1577	37
2092	6	1392	22	2024	38
1406	7	1566	23	1962	39
2018	8	1789	24	1868	40
2150	9	1716	25	1938	41
1447	10	2327	26	2528	42
2220	11	1563	27	1821	43
1738	12	2346	28	1135	44
1966	13	1513	29	1590	45
1732	14	2145	30	1081	46
2033	15	1904	31	1513	47
				1581	48
				2349	49
133106 were part of the analysis but showed no LDA relation to other topics of the modelled posts around the predominant topics. Also due to data analysis management decision of 50 topics and 500 features when compared to more when running topic modelling algorithms					
Total 225360					

7.2.2 Random sample topics

Table 7-5: Random sample topics

Random 5 sample topics from LDA Computational Topic Model giving topics, 0, 24, 34, 13, 31					
LDA topic number	**topic 0**	**topic 24**	**topic 34**	**topic 13**	**topic 31**
Top trigram Label for topics: device-enabled discourse purpose categories	do you think	you so much	you will get	questionmark questionmark questionmark	diagnosed with type
Number of posts	2941	1789	2332	1966	1904

7.2.3 LDA Model keep stopwords: 50 topics and device-enabled discourse purpose categories

Table 7-6: LDA Model keep stopwords: 50 topics and device-enabled discourse purpose categories

LDA Model keep stopwords: 50 topics and device-enabled discourse purpose categories	
Top topic device-enabled discourse purpose categories label from top topic trigram	From LDA Model & keep stopwords, top 9 trigrams
topic 0 do you think	Topic 0: do you think, what do you, at the moment, how do you, seems to be, do you know, have been told, it could be, been told to
topic 1 be able to	topic 1: be able to, will be able, should be able, able to help, you should be, to be able, they should be, to help you, to give you
topic 2 the rest of	topic 2: the rest of, of the day, rest of the, most of the, in the morning, of the time, out of the, at the moment, up and down
topic 3 hope you get	topic 3: hope you get, try not to, you get it, get it sorted, not to worry, you get the, to get it, to worry about, people on here
topic 4 going to be	topic 4: going to be, is going to, going to hive, mind going to, was going to, am going to, to have to, are going to, not going to
topic 5 be able to	topic 5: be able to, to do it, to be on, used to be, to be able, to do this, able to get, seem to be, need to be
topic 6 you can get	topic 6: you can get, to find out, just wondering if, find out more, wondering if anyone, you can find, was wondering if, if you can, that you can
topic 7 the first time	topic 7: the first time, for the first, this is the, all the time, to try and, try and get, had diabetes for, thought it was, dont want to
topic 8 it is very	topic 8: it is very, and it is, it is not, but it is, as it is, you have to, think it is, as well as, that it is
topic 9 don have to	topic 9: don have to, you don have, the same as, one of the, have to be, have to do, to do it, it can be, you have to
topic 10 sorry to hear	topic 10: sorry to hear, to hear that, to hear about, to hear you, you know what, you and your, do you know, that you have, that you are
topic 11 seems to be	topic 11: seems to be, it seems to, it was the, the best thing, if it is, it is the, and it was, to be the, is the best
topic 12 as long as	topic 12: as long as, if you are, you will be, long as you, will be fine, you are on, you should be, you are not, it will be

topic 13 questionmark questionmark questionmark	topic 13: questionmark questionmark questionmark, does anyone else, does anyone know, are you on, is there any, does anyone have, went to bed, thought it was, hope everyone is
topic 14 to deal with	topic 14: to deal with, diabetes for years, had diabetes for, deal with it, with it and, get on with, on with it, for years now, for years and
topic 15 if you are	topic 15: if you are, if you have, blood glucose levels, you need to, your blood glucose, it may be, you are on, your blood sugar, if you do
topic 16 if you have	topic 16: if you have, get in touch, in touch with, you have any, do you have, if you ve, you like to, if you like, you ve got
topic 17 know how you	topic 17: know how you, let us know, you get on, how you feel, us know how, how you get, let me know, hope you get, get on with
topic 18 an eye on	topic 18: an eye on, keep an eye, good luck with, make sure you, to make sure, trial and terrorcores and Cals, to do with, just in case
topic 19 you have to	topic 19: you have to, have to be, do you have, have to have, have to do, will have to, and have to, have to take, if you have
topic 20 what to do	topic 20: what to do, know what to, you need to, don know Whaddon know what, to do with, need to get, not sure what, need to do
topic 21 the last few	topic 21: the last few, couple of days, for couple of, for the last, over the last, in the last, couple of weeks, get on with, to try and
topic 22 speak to your	topic 22: speak to your, to your dsn, you need to, to speak to, you can get, as you can, if you can, if you are, as much as
Topic 23 thank you for	topic 23: thank you for, you for your, thanks for all, for all the, for all your, thanks for your, for your help, let you know, thanks for the
topic 24 you so much	topic 24: you so much, thank you so, so much for, on the pump, been on the, thanks for your, thank you all, for your help, have been on
topic 25 sounds like you	topic 25: sounds like you, it sounds like, if you are, if you don, it might be, you need to, might be worth, if you can, you can get
topic 26 to go to	topic 26: to go to, go to the, to the gym, go to bed, trying to get, to get my, to the hospital, get back to, going to the
topic 27 type for years	topic 27: type for years, been type for, for years and, had type for, had it for, for years now, years now and, and have had, and have been
topic 28 my son was	topic 28: my son was, my daughter was, the age of, son was diagnosed, daughter was diagnosed, was diagnosed with, was diagnosed at, at the age, when he was
topic 29 would like to	topic 29: would like to, you would like, if you would, would love to, to talk to, it would be, thank you for, let me know, if you are

topic 30 for me this	topic 30: for me this, me this morning, happy with that, this morning and, hope everyone is, have good day, thanks for that, went to bed, not too bad
topic 31 diagnosed with type	topic 31: diagnosed with type, my year old, was diagnosed with, with type diabetes, been diagnosed with, just been diagnosed, year old son, diagnosed with diabetes, year old daughter
topic 32 all the time	topic 32: all the time, of the time, most of the, at the moment, the time and, fed up with, all of the, one of those, some of the
topic 33 feel better soon	topic 33: feel better soon, you feel better, hope you feel, hope your ok, how are you, have you tried, well done you, have you been, one of those
topic 34 you will get	topic 34: you will get, will get there, and you will, you will have, it will be, will have to, we are all, at the moment, people on here
topic 35 people with diabetes	topic 35: people with diabetes, with type diabetes, for people with, people with type, type diabetes and, have type diabetes, with diabetes and, type and type, one of the
topic 36 you want to	topic 36: you want to, if you want, want to do, want to know, want to be, just want to, if you can, let me know, don want to
topic 37 the amount of	topic 37: the amount of, you can do, amount of insulin, what you mean, know what you, can do it, you need to, how much insulin, so you can
topic 38 blood sugar levels	topic 38: blood sugar levels, keep up the, my blood sugar, the good work, up the good, your blood sugar, my blood sugars, blood sugars are, keep it up
topic 39 in the morning	topic 39: in the morning, the morning and, in the night, up in the, in the evening, during the night, during the day, go to bed, through the night
topic 40 it would be	topic 40: it would be, the same problem, would be great, that would be, have the same, had the same, would be good, the same thing, exactly the same
topic 41 have good day	topic 41: have good day, hope you all, hope you have, you all have, have great day, you have good, good luck to, thank you all, not too bad
topic 42 does anyone know	topic 42: does anyone know, trying to get, to get my, to get it, and have been, at the moment, have been told, know how to, been told to
topic 43 all over the	topic 43: all over the, over the place, my year old, all the time, at the moment, at the time, my blood sugars, year old daughter, and it was
topic 44 feel free to	topic 44: feel free to, to add me, free to add, me if you, if you want, add me if, add me as, me as friend, you want to
topic 45 used to it	topic 45: used to it, go back to, get used to, to go back, have to go, to go on, back to the, back to normal, on the pump

topic 46 the end of	topic 46: the end of, end of the, at the end, of the day, thought it was, one of those, one of the, would have been, it was the
topic 47 been diabetic for	topic 47: been diabetic for, diabetic for years, have been diabetic, for years and, ve been diabetic, and have been, have been on, the dafne course, for years now
topic 48 hope you are	topic 48: hope you are, look at the, you are all, have look at, you need to, you are feeling, if you need, you are doing, when you are
topic 49 thanks for the	topic 49: thanks for the, the same time, at the same, for the advice, it in the, for me and, in the end, at the hospital, looking forward to

7.2.4 Validation of LDA topics Model keep stopwords and device-enabled discourse purpose categories

Table 7-7:Validation of LDA topics model, keep stopwords and device-enabled discourse purpose categories

Validation of LDA topics model, keep stopwords and device-enabled discourse purpose categories				
LDA	**precision**	**recall**	**f1-score**	**support**
Diabetesuk	0.666667	0.001414	0.002821	8,489
peer	0.962382	0.999972	0.980817	216,871
avg/total	0.951242	0.962358	0.943977	225,360

7.2.5 LDA Model remove stopwords: 50 TOPICS and device-enabled discourse content categories

Table 7-8: LDA Model remove stopwords: 50 TOPICS and device-enabled discourse content categories

LDA Model remove stopwords: 50 TOPICS and device-enabled discourse content categories	
Top TOPIC device-enabled discourse content categories from top TOPIC trigram	LDA Model remove 'English' stopwords for TOPICS: top 9 trigrams
TOPIC 0 feels better soon	TOPIC 0: feels better soon, hope feels better, hope ur ok, hope better soon, feel like im, good luck hope, oh dear hope, hope better day, hope goes ok

TOPIC 1 hope sorted soon	TOPIC 1: hope sorted soon, hypo free night, peaceful hypo free, good luck hope, night night sleep, low blood sugar, diabetes care team, insulin carb ratio, long diagnosed questionmark
TOPIC 2 long acting insulin	TOPIC 2: long acting insulin, carbs cals username, short acting insulin, Cals username, questionmark sounds like, book carbs Cals, called carbs Cals, help lose weight, carb counting questionmark
TOPIC 3 type diabetes years	TOPIC 3: type diabetes years, high blood pressure, diabetes years ago, diagnosed type diabetes, diabetes years old, lost half stone, hope tomorrow better, couple years ago, today good day
TOPIC 4 fast acting insulin	TOPIC 4: fast acting insulin, questionmark gid ref, group php questionmark, php questionmark gid, gid ref ts, half hour later, ve just got, just wanted say, levemir twice day
TOPIC 5 night time hypo	TOPIC 5: night time hypo, night time hypos, just dont know, medical exemption certificate, doing right thing, type diabetic years, parents children diabetes, continuous glucose monitor, going low night
TOPIC 6 diagnosed weeks ago	TOPIC 6: diagnosed weeks ago, test times day, ve just started, old diagnosed type, weeks ago type, years old diagnosed, accu check mobile, year old diagnosed, yr old diagnosed
TOPIC 7 just make sure	TOPIC 7: just make sure, make sure eat, breakfast lunch dinner, drink plenty fluids, need make sure, make sure ok, make sure don, control blood sugars, effect blood sugars
TOPIC 8 Monday Friday pm	TOPIC 8: monday friday pm, insulin pump users, open monday friday, really sorry hear, blood glucose levels, ll make sure, careline monday friday, pump users group, sounds like need
TOPIC 9 yr old son	TOPIC 9: yr old son, year old son, diagnosed type months, type months ago, old son type, diagnosed type just, old son diagnosed, diagnosed weeks ago, daughter diagnosed type
TOPIC 10 questionmark questionmark questionmark	TOPIC 10: questionmark questionmark questionmark, questionmark questionmark lol, insulin questionmark questionmark, advice questionmark questionmark, questionmark questionmark think, morning questionmark questionmark, questionmark questionmark thanks, high questionmark questionmark, yr old daughter
TOPIC 11 questionmark questionmark questionmark	TOPIC 11: questionmark questionmark questionmark, ok questionmark questionmark, questionmark questionmark hope, time questionmark questionmark, hope ok questionmark, diabetic questionmark questionmark, questionmark questionmark thanks, insulin questionmark questionmark, questionmark questionmark dont
TOPIC 12 message add friend	TOPIC 12: message add friend, diagnosed yrs ago, need chat just, chat just message, feel free message, just message add, lots good advice, heading right direction, hats really good

TOPIC 13 year old daughter	TOPIC 13: year old daughter, old daughter diagnosed, daughter diagnosed type, don let diabetes, diagnosed type diabetes years old, insulin times day, diabetes stop doing, diagnosed years old
TOPIC 14 hope feeling better	TOPIC 14: hope feeling better, feeling better soon, feeling better questionmark, feeling better today, type diabetes htm, diabetes htm metformin, good days bad, blood sugar questionmark, diabetes type diabetes
TOPIC 15 carb count questionmark	TOPIC 15: carb count questionmark, help good luck, ive diabetic years, best thing did, good luck tomorrow, hope makes sense, pump changed life, started carb counting, carb insulin ratio
TOPIC 16 just let know	TOPIC 16: just let know, accu chek mobile, accu chek combo, hope doing ok, let know help, ll let know, help let know, use accu chek, let know need
TOPIC 17 diagnosed months ago	TOPIC 17: diagnosed months ago, daughter diagnosed months, son diagnosed months, feel like crap, great manchester run, good bad days, months ago type, hope things better, diagnosed months old
TOPIC 18 type diabetic years	TOPIC 18: type diabetic years, im type diabetic, great north run, hi im type, im type years, fasting blood test, diabetes specialist nurse, diabetic nurse said, couple months ago
TOPIC 19 happy new year	TOPIC 19: happy new year, just quick question, questionmark thanks advance, does anybody know, let know goes, newly diagnosed type, hi does know, metformin times day, test strips lancets
TOPIC 20 world diabetes day	TOPIC 20: world diabetes day, help raise awareness, insulin pump questionmark, questionmark ve got, raise awareness type, raise awareness diabetes, awareness type diabetes, getting insulin pump, daughter type diabetes
TOPIC 21 hope comes soon	TOPIC 21: hope comes soon, long acting insulin, quick acting insulin, year old type, makes feel better, gym times week, breakfast lunch dinner, multiple daily injections, feel like im
TOPIC 22 low blood sugars	TOPIC 22: low blood sugars, hope better soon, high blood sugars, sick day rules, blood sugars high, blood sugars questionmark, just don know,blood sugars low,hope sorted soon
TOPIC 23 questionmark questionmark questionmark	TOPIC 23 : questionmark questionmark questionmark, diabetes questionmark questionmark, know questionmark questionmark, questionmark questionmark thanks, high questionmark questionmark, questionmark questionmark know, night questionmark questionmark, does anybody know, questionmark questionmark lol
TOPIC 24 ve diabetic years	TOPIC 24: ve diabetic years, finger prick test, nearly years ago, test times day, slow acting carbs, high blood pressure, fast acting carbs, just wanted say, better good luck

TOPIC 25 raise money diabetes	TOPIC 25: raise money diabetes, drink plenty water, happy new year, help raise money, drink lots water, sugar free jelly, raising money diabetes, username wowbands, hope better soon
TOPIC 26 hope feel better	TOPIC 26: hope feel better, feel better soon, really don want, ve type years, hats really good, help good luck, raise money diabetes, flu jab questionmark, free add friend
TOPIC 27 feel better soon	TOPIC 27: feel better soon, hope feel better, start feel better, hope start feel, luck let know, good luck let, drink lots water, let know goes, hope goes ok
TOPIC 28 questionmark good luck	TOPIC 28: questionmark good luck, good luck hope, diagnosed year ago, lol good luck, ve diabetic years, good luck pump, carb counting questionmark, hope don mind, questionmark page id
TOPIC 29 questionmark questionmark questionmark	TOPIC 29: questionmark questionmark questionmark, help questionmark questionmark,questionmark questionmark help, questionmark questionmark know, going questionmark questionmark, yr old daughter, ive diabetic years, insulin questionmark questionmark, yr old type
TOPIC 30 diagnosed years ago	TOPIC 30: diagnosed years ago, yr old daughter, feel free ask, type diabetes type, years ago type, doing dafne course, type diagnosed years, diabetes type diabetes, year old type
TOPIC 31 type type diabetes	TOPIC 31: type type diabetes, diagnosed type years, type years ago, sounds like good, diagnosed years ago, type diabetes years, ve type diabetes, years ago just, type years old
TOPIC 32 blood sugar levels	TOPIC 32: blood sugar levels, blood sugar high, carb counting course, high blood sugar, having bad day, morning good day, feel bit better, carbs cals book, doing carb counting
TOPIC 33 blood glucose levels	TOPIC 33: blood glucose levels, high blood glucose, people type diabetes, blood glucose level, healthy balanced diet, blood glucose testing, blood glucose monitor, blood glucose control, carbs cals app
TOPIC 34 just diagnosed type	TOPIC 34: just diagnosed type, diagnosed type diabetes, recently diagnosed type, ve just diagnosed, type diabetes year, control blood sugars, hi diagnosed type, questionmark don know, carb counting just
TOPIC 35 ve type years	TOPIC 35: ve type years, carb counting course, did dafne course, slow release metformin, hope great day, Dafne course questionmark, good morning friends, course years ago, ive diabetic years
TOPIC 36 really looking forward	TOPIC 36: really looking forward, disability discrimination act, feeling bit better, thanks guys just, diabetes support forum, diagnosed years ago, ve got questions, looking forward seeing, let know think
TOPIC 37 having good day	TOPIC 37: having good day, hope having good, type weeks ago, diagnosed type weeks, just like say, did blood test, just feel like, new blog post, injecting times day

TOPIC 38 high blood sugars	TOPIC 38: high blood sugars, blood sugars high, better safe sorry, make feel better, high blood sugar, check blood sugars, blood sugar readings, blood sugar levels, does make sense
TOPIC 39 blood sugar levels	TOPIC 39: blood sugar levels, sugar levels questionmark, blood sugar level, check blood sugar, high blood sugar, blood sugar questionmark, test blood sugar, affect blood sugar, control blood sugar
TOPIC 40 ha ha ha	TOPIC 40: ha ha ha, hope gets better, gets better soon, couple weeks ago, hope ok questionmark, lol good day, hope feeling ok, good start day, good luck hun
TOPIC 41 year old son	TOPIC 41: year old son, diagnosed type diabetes, son diagnosed type, old son diagnosed, couple weeks ago, hi year old, recently diagnosed type, newly diagnosed type, said don need
TOPIC 42 questionmark questionmark questionmark	TOPIC 42: questionmark questionmark questionmark, questionmark questionmark just, today questionmark questionmark, ideas questionmark questionmark, questionmark questionmark im, questionmark questionmark good, pump questionmark questionmark, going questionmark questionmark, just don know
TOPIC 43 good morning hope	TOPIC 43: good morning hope, hope good day, morning hope today, good morning morning, morning hope good, good day today, hope good night, hope ok today, feel like crap
TOPIC 44 children type diabetes	TOPIC 44: children type diabetes, parents children type, type diabetes questionmark, dads need know, mums dads need, need know help, diagnosed month ago, help children type, know help children
TOPIC 45 sick day rules	TOPIC 45: sick day rules, adjustment normal eating, yr old daughter, dose adjustment normal, carb counting insulin, year old daughter, sounds like doing, old daughter type, type nearly years
TOPIC 46 feel free add	TOPIC 46: feel free add, people type diabetes, children young people, inject times day, young people diabetes, ve diabetes years, free add want, chat feel free, health care professionals
TOPIC 47 feel free add	TOPIC 47: feel free add, low carb diet, point right direction, feel like eating, don feel like, free add friend, know exactly feel, dont feel like, hope good day
TOPIC 48 questionmark questionmark questionmark	TOPIC 48: questionmark questionmark questionmark, questionmark questionmark lol, diabetes questionmark questionmark, type questionmark questionmark, type type questionmark, questionmark questionmark just, day questionmark questionmark, help questionmark questionmark, like questionmark questionmark
TOPIC 49 carbs cals book	TOPIC 49: carbs cals book, swine flu jab, different times day, trying lose weight, hi just wondering, slow release carbs, accu chek aviva, people don know, glad feeling better

7.2.6 Validation of LDA (remove stop words): TOPICS, device-enabled discourse content categories

Table 7-9: Validation of LDA (with removed stop words): TOPICS, device-enabled discourse content categories

Validation of LDA (with removed stop words): TOPICS, device-enabled discourse content categories				
LDA	precision	recall	f1-score	support
Diabetesuk	0.807692	0.002474	0.004932	8,489
peer	0.962420	0.999977	0.980839	216,871
avg/total	0.956592	0.962402	0.944078	225,360

7.3 MeaningCloud model

7.3.1 MeaningCloud 'global sentiment' (examples)

Table 7-10: MeaningCloud 'global sentiment' (examples no1 to no7)

	C	D	E	F	G
	Polarity	Agreement	Subjectivity	Confidence	Irony
	P	DISAGREEMENT	SUBJECTIVE	94	NONIRONIC
h Humalin I (Long r I eat Carbs.....Fun!!	NEU	DISAGREEMENT	SUBJECTIVE	91	NONIRONIC
k 16!). he problems Counting' and it ith my team mates	NEU	DISAGREEMENT	SUBJECTIVE	92	NONIRONIC
n my lifestyle and xxxxxxxxx	N	DISAGREEMENT	SUBJECTIVE	94	NONIRONIC
	N+	AGREEMENT	OBJECTIVE	100	NONIRONIC
ic	N	AGREEMENT	SUBJECTIVE	100	NONIRONIC
ıst pop the pills and	N	DISAGREEMENT	SUBJECTIVE	76	IRONIC

Language Identification | Topi ...

7.3.2 MeaningCloud targets (examples)

Table 7-11: MeaningCloud targets (examples no1 and no2)

	Topic Form	Topic Category	Rank	Type	Theme	Frequency
	Berkshire	entity	1	Top>Location>GeoPoliticalEntity>Adm2		1
	guy	concept	1	Top>Person		1
	kind	concept	2	Top>OtherEntity>Class		1
	infection	concept	3	Top>OtherEntity>Disease	Top>LifeSciences>Medicine	1
	blood	concept	4	Top	Top>LifeSciences>Medicine	1
	10 years	time expression	1			1
	10 year	quantity	1			1
een	Tony	entity	1	Top>Person>FirstName		1
nearly	Long	entity	1	Top>Person>LastName		1
th	Humalog	entity	1	Top		1
) when	kind	concept	1	Top>OtherEntity>Class		2
then I	diabetes	concept	2	Top>OtherEntity>Disease	Top>LifeSciences>Medicine	2
sulin)	insulin	concept	3	Top		2
	bed	concept	4	Top>Location>GeographicalEntity>WaterForm>Channel		1
tty	blood	concept	5	Top	Top>LifeSciences>Medicine	1
y days	sugar	concept	6	Top>Product>Food	Top>Society>Gastronomy	1
	9 years	time expression	1			1
	a day	time expression	1			1
nat	whenever	time expression	1			1
p me a	at the moment	time expression	1			1
	every days	time expression	1			1
	1 diabetes	quantity	1			1
	9 year	quantity	1			1
	1 diabetes	quantity	1			1

7.4 LIWC model

7.4.1 LIWC class examples and overall posts

Table 7-12: LIWC class examples and overall posts

Examples	OUTPUT LABEL 2015	LIWC 2015/2007 CORRELATION	liwccat	all p2pa	all oa
	Word Count	11852.99	WC	6,574,271	386,727
it, to, no, very	Function Words	54.29	function	47.40	45.81
love, nice, sweet	Positive emotion	3.75	posemo	7.92	8.81
hurt, ugly, nasty	Negative emotion	1.83	negemo	1.83	0.82
think, know	Insight	2.13	insight	2.10	2.30
maybe, perhaps	Tentativeness	2.42	tentat	3.23	3.69
always, never	Certainty	1.27	certain	1.57	1.33
clinic, flu, pill	Health/illness	0.53	health	2.02	2.43

7.4.2 LIWC categories and LIWC 2015/2007 correlation

Table 7-13: LIWC categories and LIWC 2015/2007 correlation

LIWC Category	LIWC 2015/2007 Correlation
Language Metrics, Function Words	54.29
Grammar Other, Positive emotion	3.75
Grammar Other, Negative emotion	1.83
Grammar Other, Cognitive Processes2, Insight	2.13
Grammar Other, Cognitive Processes2, Tentativeness	2.42
Grammar Other, Cognitive Processes2, Certainty	1.27
Grammar Other, Health/illness	0.53

7.4.3 LIWC: Examples of words used in posem/negemo class

Table 7-14: LIWC: Examples of words used in posem/negemo class

31 Posemo					32 Negemo				
(:	entertain*	hugs	promising	wealthy	):	dismay*	ignorant	poorest	tragic
:)	enthus*	humor*	proud	welcom*	:(	disreput*	ignore	poorly	trauma*
accept	excel	humour*	prouder	well	abandon*	diss	ignored	poorness*	trembl*
accepta*	excelled	hurra*	proudest	wellbeing	abuse*	dissatisf*	ignores	powerless*	trick
accepted	excellence	ideal*	proudly	wellness	abusi*	distraught	ignoring	prejudic*	tricked
accepting	excellent	ily*	radian*	win	ache*	distress*	immoral*	pressur*	trickier
accepts	excellently	importance	readiness	winn*	aching*	distrust*	impatien*	prick*	trickiest
active	excelling	important	ready	wins	advers*	disturb*	impersonal	problem*	tricks
actively	excels	importantly	reassur*	wisdom	afraid	domina*	impolite*	protest	tricky
admir*	excite	impress*	reinvigor*	wise	aggravat*	doom*	inadequa*	protested	trite
ador*	excited	improve*	rejoice*	wisely	aggress	dork*	incompeten*	protesting	trivial
advantag*	excitedly	improving	relax*	wiser	aggressed	doubt*	indecis*	protests	troubl*
adventur*	excitement	incentive*	relief	wisest	aggresses	dread*	ineffect*	puk*	turmoil
affection*	exciting	innocen*	reliev*	won	aggressing	dull	inferior	punish*	twitchy
agree	fab	inspir*	resolv*	wonderful	aggression*	dumb	inferiority	pushy	ugh
agreeable	fabulous	intellect*	respect	wonderfully	aggressive	dumbass*	inhibit*	queas*	uglier
agreeableness	fabulously	intelligence	respected	worship*	aggressively	dumber	insecur*	rage*	ugliest
agreeably	fabulousness	intelligent	respectful	worthwhile	aggressor*	dumbest	insincer*	raging	ugly
agreed	fair	interest	respectfully	wow*	agitat*	dummy	insult*	rancid*	unaccept*
agreeing	fairer	interested	respecting	yay*	agoniz*	dump*	interrup*	rape*	unattractive
agreement*	fairest	interesting	reward*	yum	agony	dwell*	intimidat*	raping	uncertain*
agrees	faith*	interests	rich	yummy	alarm*	egotis*	irrational*	rapist*	uncomfortabl*
alright*	fantasi*	invigor*	richer		alone	embarrass*	irrita*	rebel*	uncontrol*
amaze*	fantastic	joke*	riches		anger*	emotional	isolat*	reek*	undesir*
amazing	fantastical	joking	richest		angrier	emptier	jaded	regret*	uneas*
amazingly	fantastically	jolly	rofl*		angriest	emptiest	jealous	reject*	unfair
amor*	fantasy	joy*	romanc*		angry	emptiness	jealousies	reluctan*	unfortunate*
amus*	fav	keen*	romantic*		anguish*	empty	jealously	remorse*	unfriendly
aok	fave	kidding	safe		annoy	enemie*	jealousy	repress*	ungrateful*
appreciat*	favor	kind	safely		annoyed	enemy*	jerk	resent*	unhapp*
approv*	favoring	kindly	safer		annoying	enrag*	jerked	resign*	unimportant
assur*	favorite	kindn*	safest		annoys	envie*	jerks	restless*	unimpress*
attract	favors	kiss*	safety		antagoni*	envious	kill*	revenge*	unkind
attracted	favour*	laidback	satisf*		anxiety	envy*	lame	ridicul*	unlov*
attracting	fearless*	laugh*	save		anxious	evil	lamely	rigid	unlucky
attraction	festiv*	legit	scrumptious*		anxiously	excruciat*	lameness	rigidity	unpleasant
attracts	fiesta*	libert*	secur*		anxiousness	exhaust*	lamer	rigidly	unprotected
award*	fine	to like	sentimental*		apath*	fail*	lamest	risk*	unsafe

7.4.4 LIWC: Examples of words used in ‘certain’ class

Table 7-15: LIWC: Examples of words used in ‘certain’ class

inguistic Inquiry and Word

such as schools or universities. Please do not share this dictio

51 Insight		52 Cause	53 Discrep	54 Tentat		55 Certain
accept	motiv*	activat*	abnormal*	allot	undecided*	absolute
accepta*	notice	affect	besides	almost	undetermin*	absolutely
accepted	noticed	affected	could	alot	unknowing	accura*
accepting	notices	affecting	could've	ambigu*	unknowingly	all
accepts	noticing	affects	couldn't	any	unknown	altogether
acknowledg*	perceiv*	aggravat*	couldnt	anybod*	unlikel*	always
admit	percept*	allow*	couldve	anyhow	unresolv*	apparent
admits	perspective	attribut*	desir*	anyone*	unsettl*	assur*
admitted	persua*	based	expect*	anything	unsure*	blatant*
admitting	ponder*	basis	hope	anytime	usually	certain*
afterthought*	prefer*	bc	hoped	anywhere	vague	clear
analy*	presum*	because	hopeful	apparently	vaguely	clearly
answer*	pretend*	bosses	hopefully	appear	vagueness	commit
appreciat*	prove*	caus*	hopes	appeared	vaguer	commitment*
assum*	proving	change	hoping	appearing	vaguest	commits
attent*	quer*	changed	ideal*	appears	variab*	committed
aware*	question	changes	if	apprehens*	varies	committing
became	questioned	changing	impossible	approximat*	vary	complete
become	questioning	compel*	inadequa*	arbitrar*	virtually	completed
becomes	questions	compliance	lack	assum*	wonder	completely
becoming	rational*	compliant	lacked	barely	wondered	completes
belief*	realization*	complied	lacking	bet	wondering	confidence
believe	realize	complies	lacks	bets	wonders	confident
believed	realized	comply*	liabilit*	betting		confidently
believes	realizes	consequen*	mistak*	blur*		correct*
believing	realizing	control*	must	border*		defined
categor*	rearrang*	cos	must'nt	chance		definite
choice*	reason*	coz	mustn't	confuse		definitely
choos*	rebel*	create	mustnt	confused		definitive*
clarif*	recall*	created	need	confuses		directly
closure	recogni*	creates	need'nt	confusing		distinct*
clue	recollect*	creating	needed	confusion*		entire*
cohere*	reconcil*	creation	needing	contingen*		especially
complex	reconsider*	creations	needn't	depend		essential
complexity	reconstruct*	creative	neednt	depended		ever
complicate	reevaluat*	creativity	needs	depending		every
complicated	refer*	cuz	normally	depends		everybod*
complicates	reflect*	deceiv*	odd	disorient*		everyday
complicating	relate*	deduc*	odder	doubt*		everyone*

7.4.5 LIWC: Category frequency

Table 7-16: LIWC: Category frequency

LIWC Category Frequency (Word and Dictionary Comparison) And LIWC Word Measures LIWC 11853 Word Measures							
	OUTPUT LABEL 2015	LIWC 2015/2007 CORRELATION	LIWC Category	All Posts User Persons	All Posts Organisation Person - diabetesuk	Questionmark Posts User Persons	Questionmark Organisation Person - diabetesuk
	Word Count	11853	WC	6,574,271	386,727	1,409,443	100,365
Summary Variable 2007	Analytical Thinking	56.34	Analytic	45.53	56.78	43.26	55.24
	Clout	57.95	Clout	49.29	89.98	57.78	93.21
	Authentic	49.17	Authentic	48.18	20.84	43.43	25.09
	Emotional Tone	54.22	Tone	57.37	74.43	46.90	70.44
Language Metrics	Words per sentence1	25.07	WPS	14.83	16.83	39.53	46.90
	Words>6 letters	15.89	Sixltr	11.87	16.25	19.61	19.79
	Dictionary words	83.95	Dic	81.88	83.49	78.43	83.00
	Function Words	54.29	function	47.40	45.81	49.07	49.24
	Total pronouns	14.99	pronoun	15.04	14.02	15.73	15.32
	Personal pronouns	9.83	ppron	9.94	9.47	9.69	10.43
	1st pers singular	4.97	i	5.40	0.12	4.34	0.13
	1st pers plural	0.72	we	0.40	3.32	0.35	2.87
	2nd person	1.61	you	2.91	5.09	3.74	6.34
	3rd pers singular	1.87	shehe	0.61	0.30	0.60	0.41
	3rd pers plural	0.66	they	0.62	0.64	0.66	0.69
	Impersonal pronouns	5.17	ipron	5.09	4.55	6.04	4.88
	Articles	6.53	article	4.22	4.26	4.30	4.42
	Prepositions	12.59	prep	9.91	12.01	9.76	12.58
	Auxiliary verbs	8.82	auxverb	9.83	9.20	10.22	9.05
	Common adverbs	4.83	adverb	5.78	5.15	5.75	5.49
	Conjunctions	5.87	conj	5.23	4.85	5.63	5.04
	Negations	1.72	negate	1.91	0.68	1.22	0.45

Grammar Other	Regular verbs	15.26	verb	18.21	17.41	17.34	16.76
	Adjectives		adj	5.01	4.29	3.89	4.00
	Comparatives		compare	2.32	1.85	1.95	1.80
	Interrogatives		interrog	1.36	1.43	2.84	2.38
	Numbers	1.98	number	3.10	1.65	0.57	0.50
	Quantifiers	2.48	quant	2.20	2.89	2.13	2.53
	Affect Words	5.63	affect	9.78	9.67	4.88	6.19
	Positive emotion	3.75	posemo	7.92	8.81	3.34	5.41
	Negative emotion	1.83	negemo	1.83	0.82	1.51	0.75
	Anxiety	0.33	anx	0.31	0.16	0.29	0.18
	Anger	0.6	anger	0.29	0.05	0.22	0.04
	Sadness	0.39	sad	0.44	0.31	0.41	0.31
	Social Words	9.36	social	8.02	16.37	9.06	17.51
	Family	0.38	family	0.36	0.24	0.31	0.30
	Friends	0.23	friend	0.29	0.20	0.21	0.17
	Female referents		female	0.50	0.22	0.43	0.29
	Male referents		male	0.58	0.25	0.51	0.31
	Cognitive Processes2	14.99	cogproc	11.50	10.80	12.55	11.57
	Insight	2.13	insight	2.10	2.30	2.32	2.56
	Cause	1.41	cause	1.40	1.33	2.18	1.91
	Discrepancies	1.45	discrep	1.98	2.28	1.94	2.12
	Tentativeness	2.42	tentat	3.23	3.69	3.97	3.81
	Certainty	1.27	certain	1.57	1.33	1.19	1.33
	Differentiation3	2.48	differ	3.10	2.36	3.19	2.21
	Perpetual Processes	2.36	percept	2.30	2.29	2.09	2.51
	Seeing	0.87	see	0.63	0.89	0.55	1.03
	Hearing	0.73	hear	0.53	0.78	0.50	0.83
	Feeling	0.62	feel	0.76	0.45	0.67	0.47
	Biological Processes	1.88	bio	4.34	3.15	4.49	2.75
	Body	0.68	body	0.75	0.33	0.76	0.34
	Health/illness	0.53	health	2.02	2.43	2.35	1.98
	Sexuality	0.28	sexual	0.03	0.02	0.04	0.03
	Ingesting	0.46	ingest	1.59	0.39	1.47	0.40
	Core Drives and Needs		drives	7.13	11.02	5.95	10.90
	Affiliation		affiliation	1.62	6.22	1.46	5.83
	Achievement	1.56	achieve	1.11	1.43	0.91	1.50

	Power		power	2.00	2.03	1.90	2.25
	Reward focus		reward	2.64	2.38	1.79	2.47
	Risk/prevention focus		risk	0.47	0.28	0.46	0.23
Time Orientation4	Past focus	4.14	focuspast	3.66	2.14	3.39	2.40
	Present focus	8.1	focuspresent	12.97	13.28	12.44	12.66
	Future focus	1	focusfuture	1.82	1.96	1.22	1.35
	Relativity	13.87	relativ	12.17	10.55	11.45	11.88
	Motion	2.06	motion	1.49	1.45	1.44	1.47
	Space	6.17	space	4.99	5.28	5.15	5.68
	Time	5.79	time	5.89	3.93	4.99	4.81
Personal Concerns	Work	2.27	work	1.20	2.66	1.22	2.17
	Leisure	1.37	leisure	0.68	1.34	0.66	1.54
	Home	0.56	home	0.22	0.15	0.17	0.15
	Money	0.7	money	0.39	0.53	0.41	0.42
	Religion	0.32	relig	0.12	0.07	0.06	0.08
	Death	0.16	death	0.03	0.01	0.03	0.01
	Informal Speech		informal	4.73	3.51	2.32	0.57
	Swear words	0.17	swear	0.20	0.00	0.18	0.00
	Netspeak		netspeak	3.09	2.78	1.48	0.14
	Assent	1.11	assent	0.80	0.36	0.40	0.25
	Nonfluencies	0.3	nonflu	0.72	0.36	0.31	0.20
	Fillers	0.4	filler	0.03	0.00	0.03	0.00
	All Punctuation5	22.9	AllPunc	20.79	22.53		
	Periods	7.91	Period	7.65	6.46		
	Commas	4.81	Comma	2.55	3.00		
	Colons	0.63	Colon	0.68	0.73		
	Semicolons	0.24	SemiC	0.05	0.04		
	Question marks	0.95	QMark	1.56	1.10		
	Exclamation marks	0.91	Exclam	3.45	1.54		
	Dashes	1.38	Dash	0.95	2.66		
	Quotation marks	1.38	Quote	0.12	0.19		
	Apostrophes	2.83	Apostro	1.67	2.40		
	Parentheses (pairs)	0.25	Parenth	0.96	0.43		
	Other punctuation	1.38	OtherP	1.15	3.98		

7.5 Facebook GDPR and data policy

7.5.1 An example from Facebook GDPR

The General Data Protection Regulation (GDPR), which goes into effect on 25 May 2018, creates consistent data protection rules across Europe. It applies to all companies that process personal data about individuals in the EU, regardless of where the company is based. Processing is defined broadly and refers to anything related to personal data, including how a company handles and manages data, such as collecting, storing, using and destroying data.

While many of the principles of this regulation build on current EU data protection rules, the GDPR has a wider scope, more prescriptive standards and substantial fines. For example, it requires a higher standard of consent for using some types of data and broadens the rights that individuals have for accessing and transferring their data. Failure to comply with the GDPR can result in significant fines – up to four per cent of global annual revenue for certain breaches.

Throughout the preparation process, Facebook is committed to the following:

Transparency

Our Data Policy defines how we process people's personal data. We'll provide education on our Data Policy to people using Facebook Company Products. We'll do this through in-product notifications and consumer education campaigns to ensure that people understand how their data is being used and the choices they have.

Control

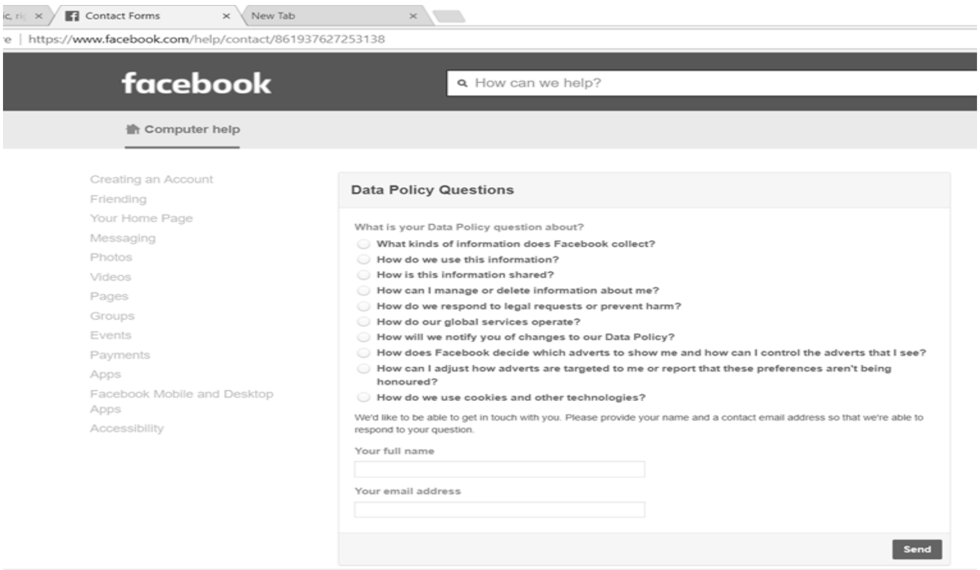
We'll continue to provide people with control over how their data is used. We've launched a new control centre to make privacy settings easier to understand and update. We also remind people as they use Facebook about how to view and edit their settings.

Accountability

We have Privacy Principles that explain how we think about privacy and data protection. We have a team of people who help ensure that we are documenting our compliance. Additionally, we meet regularly with regulators, policymakers, privacy experts and academics from around the world to keep them apprised of our practices, get feedback and continue to improve how we protect personal information.

7.5.2 An example from Facebook data policy

Table 7-17: Facebook data policy



An example of Facebook Data Policy

Date of Last Revision: September 29, 2016

We give you the power to share as part of our mission to make the world more open and connected. This policy describes what information we collect and how it is used and shared. You can find additional tools and information at Privacy Basics.

As you review our policy, keep in mind that it applies to all Facebook brands, products and services that do not have a separate privacy policy or that link to this policy, which we call the“Facebook Services” or “Services.”

I. What kinds of information do we collect?

Depending on which Services you use, we collect different kinds of information from or about you.

- **Things you do and information you provide.** We collect the content and other information you provide when you use our Services, including when you sign up for an account, create or share, and message or communicate with others. This can include information in or about the content you provide, such as the location of a photo or the date a file was created. We also collect information about how you use our Services, such as the types of content you view or engage with or the frequency and duration of your activities.
- **Things others do and information they provide.** We also collect content and information that other people provide when they use our Services, including information about you, such as when they share a photo of you, send a message to you, or upload, sync or import your contact information.

II. How do we use this information?

We are passionate about creating engaging and customized experiences for people. We use all of the information we have to help us provide and support our Services. Here's how:

Headington Campus Gipsy Lane
Oxford OX3 0BP UK

heabbott@brookes.ac.uk

- **Provide, improve and develop Services.** We are able to deliver our Services, personalize content, and make suggestions for you by using this information to understand how you use and interact with our Services and the people or things you're connected to and interested in on and off our Services.

  We also use information we have to provide shortcuts and suggestions to you. For example, we are able to suggest that your friend tag you in a picture by comparing your friend's pictures to information we've put together from your profile pictures and the other photos in which you've been tagged. If this feature is enabled for you, you can control whether we suggest that **Promote safety and security.** We use the information we have to help verify accounts and activity, and to promote safety and security on and off of our Services, such as by investigating suspicious activity or violations of our terms or policies. We work hard to protect your account using teams of engineers, automated systems, and advanced technology such as encryption and machine learning. We also offer easy-to-use security tools that add an extra layer of security to your account. For more information about promoting safety on Facebook, visit the Facebook Security Help Center.

7.6 Ethics

<u>Oxford Brookes University Ethics Committee UREC No. 120677 Ethics Approval Report</u>

UNIVERSITY RESEARCH ETHICS COMMITTEE, FACULTY OF HEALTH AND LIFE SCIENCES

Professor Rachel Harrison

Director of Studies

Department of Computing and Communication Technologies

Faculty of Technology, Design and Environment

Oxford Brookes University

Wheatley Campus

27 November 2012

Dear Professor Harrison

UREC Registration No: 120677

Methodologies for using Social Media for Collaborative Work (SMCW)

Thank you for your email of 19 November 2012 outlining your response to the points raised in my previous letter about the research study of your research student **Bazil Solomon,** and attaching the revised documents. I am pleased to inform you that, on this basis, I have given Chair's Approval for the study to begin.

The UREC approval period for this study is two years from the date of this letter, so 27 November 2014. If you need the approval to be extended please do contact me nearer the time of expiry.

In order to monitor studies approved by the University Research Ethics Committee, we will ask you to provide a (very brief) report on the conduct and conclusions of the study in a year's time. If the study is completed in less than a year, could you please contact me and I will send you the appropriate guidelines for the report.

Yours sincerely

Hazel Abbott

Chair of the University Research Ethics Committee

cc David Duce, Supervisory team

Bazil Solomon, Research Student

Pete Smith, Research Ethics Officer

Jill Organ, Graduate Office

Louise Wood, UREC Administrator

<u>UREC application extract from Ethics approval No. 120677 report</u>

LAY DESCRIPTION: Provide a brief outline of the project, including what participants will be required to do. This description must be in everyday language which is free from jargon. Please explain any technical terms or discipline-specific phrases. (No more than 350 words)

- We intend to investigate Social Media from Email, Wikis to Social Networking and integrated social media within software engineering applications, for their use in communication, cooperation i.e. collaborative work for Software Engineering. Social Media such as Facebook, Wikipedia and Twitter have features and functions that can allow large quantities of users to accomplish a wide variety of collaborative work. These range from systems building e.g. open source systems building to managing projects e.g. a large-scale engineering project to build a new type of renewable energy car with expert team member task allocation or software maintenance with question and answer forums inclusive of user input via social search or software development patterns and anti-pattern discovery, including:
- editing collaborative documents or code
- building groups and communities
- wider access to stakeholder requirements
- access to specialist skills
- searching for solutions and fixing issues
- question and answer forums for systems maintenance

- identifying software systems development patterns and anti-patterns
- Software engineering Research
- maintaining or building global systems with internationally located workforces
- managing large scale software engineering projects with good task allocations

The intention is to draw on the well-established field of CSCW and overcome some of its limitations in the process e.g. user's awareness level of other users working on a project, user motivation to keep working on the collaborative work, cultural and linguistic limits, collaborative work that may need creative solutions and large-scale coordination. So, we find technologies that have created a distinct social platform, made mass collaborations and crowdsourcing possible i.e. Social Media and those integrated into Software engineering work platforms are being put into good use for effective collaborative work. In order to better understand this and gather data for the future development of our hypotheses and systems, this project will conduct a survey of software engineers.

The online survey will be conducted with software engineers using Social Media for Collaborative Work, recruited from both within and externally to Oxford Brookes University. Existing surveys have looked at which types of Social Media is used more often, however none of these studies have examined Social Media types, tasks, motivation, emotion and value in relation to Collaborative Work . This work will also examine specifically how these factors make collaborations easier and can be used as a basis for developing suitable systems.

UREC application extract from filled in new form to update Ethics approval No. 120677

LAY DESCRIPTION:

We intend to investigate the Diabetes community through, corpus construction and analysis, surveys, texts, interviews, National Health Service (NHS) documents, Social Media from Email, Wikis, Facebook, Twitter to Social Networking and integrated social media within the Diabetes Community, for their use in communication, cooperation, coordination i.e. collaborative work for Diabetes Healthcare management via patients, Doctors, and actors involved in this process. Social Media such as Facebook, Wikipedia and Twitter have features and functions that can allow large quantities of users to accomplish a wide variety of collaborative work. Healthcare communications and collaborative work as a part of these social phenomena includes e.g. for Diabetes patient support systems with:

- constructing of collaborative comments and posts in Facebook, Twitter and dialogue in medical consultations between doctor and patient
- building of online groups and communities
- wider access to stakeholder requirements
- access to specialist skills
- searching for healthcare solutions
- question and answer forums features in patient support systems
- identifying patient support systems requirements
- maintaining or building global systems with internationally located Diabetes healthcare workers and patients
- managing large scale diabetes online activities for healthcare communications

The intention is to draw on the well-established field of Corpus linguistics and research social media collaborative work (SMCW) systems with contributions to corpus linguistics, healthcare communications and systems requirements engineering within the domain of Diabetes. Computer Supported Cooperative Work limitations can be overcome in the process e.g. user's awareness level of other users working on Diabetes activities, user and community motivation to keep working on the collaborative work activity, cultural and linguistic limits, collaborative work that may need creative solutions and large-scale coordination. So, we find technologies that have created a distinct social platform, made mass collaborations and crowdsourcing possible i.e. Social Media Groups for Diabetes Healthcare and those integrated into Diabetes work platforms are being put into good use for effective collaborative work. In order to better understand this and gather data for the development of our hypotheses and interventions, this project will conduct surveys, construct and analyse corpora, and interview actors involved in a diabetes community.

Surveys, construction and analyses of corpora, and interviews with actors involved in a diabetes community will be conducted for people involved with Diabetes healthcare, recruited from the diabetes community and from within social media. Existing surveys, construction and analyses of corpora, and interviews surveys have looked at Diabetes healthcare communication; however, none of these studies have examined social media for Diabetes healthcare collaborative work in relation to an intervention for improved collaborative work. This work will also examine specifically how these factors make collaborations easier and can be used as a basis for developing a suitable corpus linguistics technique, a healthcare diabetes collaborative work intervention and a requirement engineering technique for these types of patient support systems.

7.6.1 Oxford Brookes University Ethics Committee Updates

Table 7-18: Oxford Brookes University Ethics Committee Updates

1 UREC No. 120677 with email requesting update to research on social media data

Dear Bazil

I have now heard back from the Brookes administrator

Your UREC approval is for two years, so ran out on 27 November 2014, but the approval period can be extended. It was noted that your original study was approved to recruit participants involved in software development, so professional software engineers in industry and academia and students who develop software. The software professionals were to be recruited by social media, including Facebook, Twitter etc.

Hence, the additional work is to go to online forums for people with diabetes. If this is the the case, then the following usually applies:

Make sure no one is recruited via NHS websites / online fora

If the sites are managed, permission must be sought to post a message describing the research on the site.

For web sites that require membership, please ensure the terms and conditions of the site are adhered to.

If you have not already come across the Association of Internet Researchers, they have some useful guidelines available at:http://aoir.org/reports/ethics2.pdf

The step you should take now is to apply direct to Louise Wood, the UREC administrator, for a study change and an extension to the UREC approval.

Pete

PS - Please note I shall shortly be putting on my auto office reply as I shall be away till 13 April

2 Bazil Solomon <bazil.solomon-2011@brookes.ac.uk>

2 Apr 2015, 11:19

to Louise, Pete, Nigel, Alon, Kenneth

Dear Pete,

Thank you for the information,

"The step you should take now is to apply direct to Louise Wood, the UREC administrator, for a study change and an extension to the UREC approval".

Dear Louise: how do I go about applying for a study change and an extension to the UREC approval "the additional work is to go to online forums for people with diabetes for survey research"; and from Diabetes community emails and document data with soft systems methodology action research and data gathered from the community through unstructured interviews, observation; and from online data collected from social media sites for building and investigating corpora?

Best Regards

3 Louise Wood <louise.wood@brookes.ac.uk>

2 Apr 2015, 15:58

to Bazil, Pete, Nigel, Alon, Kenneth

Dear Bazil

In order to do this, please would you let me have one side of A4 that explains the changes to the study, and updated copies of the participant information sheet, consent form, wording for social media posting and any other information that may require updating due to the changes. Electronic copies of these are absolutely fine, I just need an adequate audit trail for the files.

Once I have received these I will liaise with the UREC Chair to request a study change and an extension of the UREC approval.

With best wishes

4 Bazil Solomon <bazil.solomon-2011@brookes.ac.uk>

2 Apr 2015, 19:25

to Louise, Pete, Nigel, Alon, Kenneth

Hi Louise,

Thanks for the information. Could you send me a link to the latest versions of the ethics forms for me to download and complete?

Best Regards

5 Louise Wood <louise.wood@brookes.ac.uk>

Attachments

9 Apr 2015, 15:35

to Bazil, Pete, Nigel, Alon, Kenneth

Hi Bazil

I don't think you need to complete a brand new E2U form for this change. If it's any use, I attach an unlocked copy of your original E2U form and you should be able to update the changes to this version.

If you do want an up to date version of the form, it is available to download at www.brookes.ac.uk/Research/Research-ethics/Ethics-review-forms

With best wishes

Louise

Attachments area

6 Bazil Solomon <bazil.solomon-2011@brookes.ac.uk>

Attachments

23 Apr 2015, 15:23

to Louise, Pete, Nigel, Alon, Kenneth

Hi Louise,

I have attached the updated document and the changes to the form that I am working on. The idea is to use data from 2011 to 2014, conduct a survey of a diabetes community, collect interview data from doctor–patient consultations, interview doctors, patients and collect data from social media. I am working on a survey questionnaire, interview questionnaire, participant information document and consent letters.

Best Regards

Attachments area

7 UREC and Information Officer GDPR report May 2018 with Points Research should comply with

1.UREC ethical statement action 1: the research needs to find and cite other studies that use publicly accessible data - they have undertaken specific procedures for preserving anonymity, e.g. name removal and reordering words

2.UREC ethical statement action 2: the research - benchmark of good practice, needs to follow this.

3.URC ethical statement action 3: the research 'having looked at this and this best practice I have done this, and it is by OBU Ethics policies. Following the advice of other researchers and on my reflection, the data has been amended because of ..., in the thesis make the change. I have learnt from the experience ... and the research needs to demonstrate the reflection on the ethics - this material with vulnerable individuals and use extra care.'

4. UREC ethical statement action 4: the researcher needs to send his extended ethical consideration section and his response to EE, in March/April time, to see amendments, and how his commentary complies with UREC policy.

As you may be aware, on the 25th of May, the EU General Data Protection Regulation and new Data Protection Act come into force. This law will influence research involving personal data.

What counts as ‘personal data’? This is data about living people from which they can be identified. As well as data containing obvious ‘identifiers’ – such as name and date of birth – this includes some genetic, biometric and online data if unique to an individual. Data that has been pseudonymised (with identifiers separated), where the dataset and identifiers are held by the same organisation, is still personal data.

Data anonymised in line with the ICO ‘Anonymisation code of practice’ is not personal data. An example of this is when identifiers are held by another organisation with an agreement that specifies no re-identification. You should be aware that the action of ‘anonymisation’ counts as processing personal data. At the time of writing, the ICO is working to update the code to reflect GDPR requirements.

If your research involves personal data, if you think the above applies to your research, please read the information here, which details how you should ensure your data processing is lawful, fair and transparent. This largely mirrors current good practice in research. Brookes also has a dedicated GDPR web page. If you have any queries about the data you hold, please contact the Information Management Team: info.sec@brookes.ac.uk.

Information Officer: There are two key areas we need to cover off with your research: 1) ‘Lawful Basis’ to process this data: Although there is currently concerns how social media data is processed, this is for genuine scientific research under Article 89 of the GDPR. You have taken necessary steps to anonymise and apply the principle of data minimisation. You have taken necessary steps to safeguard the rights and freedoms of subjects by contacting Diabetes UK and seeking their consent and advice. I am comfortable that you have a lawful basis to proceed. 2) Transparency of processing. 3) Under normal circumstances, you would have to display a privacy notice to your participants. However, as you are anonymising the quotes and to locate and provide such notice via social media would take ‘disproportionate effort’ for your scientific research (Article 14(5) (b)), you are exempt from doing so. The rest of the GDPR principles are pretty much covered in your ethics approval.

Dear student, you will be aware that the new General Data Protection Regulation (GDPR) commences on 25 May 2018 and replaces the existing UK Data Protection Act. Over the last year we have been getting the University ready to comply with the enhanced data protection requirements introduced by GDPR, both our responsibilities as a public-sector organisation, as well as the impact on individual rights for our staff, students and third-party stakeholders. A key feature of GDPR is transparency, and privacy notices are the principle way of delivering this, letting individuals know what personal information Oxford Brookes collects and why, whom we may share it with and what your rights under the legislation are. Personal data is any information that can be used to identify a single, living individual, whether it relates to private, professional or public life. It can be anything from a name, a home address, a photo, an email address, bank details, and posts on social networking websites, medical information, or a computer’s IP address. Please do read the new privacy policy. If you have any questions or would like to know more about the privacy notices or information security, please contact the Information Management team by email info.sec@brookes.ac.uk or phone 01865 484354. Useful information, IT Acceptable Use Policy, www.brookes.ac.uk/it/information-management/gdpr

7.6.2 Letter from Diabetes UK granting permission to use their data

Figure 7-1: Letter from Diabetes UK's letter granting permission to use their data, and some advice to use Townsend and Wallace's (2016) research paper.

Dear Mr Solomon

I have now heard back from our Digital Marketing Team. You are welcome to use the postings in your research on the condition you reference Diabetes UK throughout the paper.

In terms of who 'owns' the data, this is a grey area (anything a person writes belongs to them, not the Facebook page owner. However, if they allow it to be publically visible, they do not own copyright). To be safe, we would advise that you anonymise the posts by blanking out the names and any demographic details.

An example with some handy tips on an existing paper from Glasgow University which you may find helpful - http://www.gla.ac.uk/media/media_487729_en.pdf

In terms of ethics, your university should be able to help with this

Best wishes, Kamini

Kamini Shah
Acting Head of Research Funding

7.7 Glossary

1. Alpha commands how many topics a document possibly has. The lower the alpha, the lower the number of topics per documents
2. Beta commands the number of the word per document. In almost the same way, to Alpha, the lower the Beta is, the lower the number of words per topic.
3. k: the number of topics
4. Pattern: can be composed of relevant words or sequence of words; that can match any occurrence or event of the keyword or sequence of words. It is without any concern about what happened before or followed them. For example the pattern (0

you are 0) matches any event of the sequence of words "you are" having nothing to do with what happened before or followed them.

5. Postee: a person that posts on social media
6. Solidarity: reflects closeness and familiarity
7. Sparsity is a condition of containing mostly zeros

8 REFERENCES

Abdallah, Z. S., Carman. M and Haffari. G (2016). Multi-domain evaluation framework for named entity recognition tools, Computer Speech and Language 43 (2017) 34

Aikhenvald, A. (2004). Evidentiality (Oxford Linguistics), Oxford, Oxford University Press

Al Mamun, M., Ibrahim. H. M and Turin. T.C. (2015). Social Media in Communicating Health Information: An Analysis of Facebook Groups Related to Hypertension. Prev Chronic Dis 2015; 12

ALICE, Artificial Intelligence Foundation, http://www.alicebot.org, accessed 01/2017

Alpers. G. W., Winzelberg. A. J., Classen. C, Roberts. H., Dev. P and Koopman C. (2005). Evaluation of computerised text analysis in an Internet breast cancer support group. Comput Hum Behav 2005; 21: 361–76

Alter, O. (2010). Tu Vous Power Solidarity Notes, Part of the French Linguistics Notes Collection, an Upper 2.1 Package Written at Oxford University In 2010

Androutsopoulos, J. (2006). Introduction: Sociolinguistics and computer-mediated communication, Journal of Sociolinguistics 10(4): 419–438.

Association of Internet Research, (2012). Ethical Decision-Making and Internet Research. Available at: http://aoir.org/reports/ethics2.pdf

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

Baker, P. (2010). Corpora and sociolinguistics. Edinburgh: Edinburgh University Press.

Baker, P., Gabrielatos C., Khosravinik, M., Krzyzanowski, M., McEnery, T. and Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. Discourse & Society, 19(3), 273-305.

Barrera, M Jr, Glasgow. R. E., McKay. H. G., Boles. S. M and Feil. E. G. (2002). Do Internet-based support interventions change perceptions of social support? An experimental trial of approaches for supporting diabetes self-management. AMJ Community Psychol 2002; 30(5):637–54. CrossRef PubMed

Barton, D. and Lee. C. (2013). Language online: Investigating digital texts and practices. Routledge: London.

Beheshti, S. M. R., Venugopal. S., Ryu. S. H., Benatallah. B and Wang. W. (2013). Bigdata and cross-document coreference resolution: current state and future opportunities. CoRRabs/1311.3987.

BERA. (2011). Revised Ethical Guidelines for Educational Research. Southwell: British Educational Research Association. Available online at https://www.bera.ac.uk/wp-content/uploads/2014/02/BERA-Ethical-Guidelines-2011.pdf?noredirect=1, (Last accessed November 2017)

Berman, R. A., Ragnarsdóttir. H and Strömqvist. S. (2002). Discourse stance. Written Languages and Literacy, Volume 5, 2

Biber, D. (1991). Variation across speech and writing. Cambridge, UK: Cambridge University Press.

Biber, D. and Finegan, E. (1989). Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. Text 9(1): 93–124.

Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999). Longman grammar of spoken and written English. London: Longman.

Blankenship, D, C. (2010). Applied Research and Evaluation Methods in Recreation, Human Kinetics; Har/Psc Edition

Blei, D. M. (2012). Probabilistic Topic Models. Communications of the ACM, 55(4):77–84.

Blei, D. M., Ng, A.Y and Jordan, M. I. (2003). Latent Dirichlet Allocation. Journal of Machine Learning Research, 3:993–1022.

Blei, D.M and Smyth. P. (2017). Science and data science, Proceedings of the National Academy of Sciences, 2017

Boyd, D. (2010). Social Networking Sites as Networked Publics: Affordances, Dynamics and Implications. In Z. Papacharissi. A Networked Self: Identity Community and Culture on Social Network Sites. New York: Routledge

Boyd, D. and Crawford, K (2012). Critical questions for big data. Information, Communication and Society. 15 (5). 662-679.

BPA, British Psychological Association (2013). Ethics Guidelines for Internet-Mediated Research. Report available at http://www.bps.org.uk/system/files/Public%20files/inf206-guidelinesfor-internet-mediated-research.pdf

Brett, M. (2012). Topic Modeling: A Basic Introduction. http://www.fredgibbs.net/clio3workspace/blog/topic -modeling/

Brown, R and Gilman, A. (1960). “The Pronouns of Power and Solidarity.” Fishman Joshua A. (Ed.) Readings in Sociology of Language. Mouton & Co, Hague, 252-275

Brownson, C. A and Heisler. M. (2009). "The Role of Peer Support in Diabetes Care and Self-Management", The Patient: Patient-Centered Outcomes Research, Wolters Kluwer Health; Adis, vol. 2(1), pages 5-17.

Caren, N. (2015). Prof. Caren, from sociology, https://github.com/nealcaren/quant-text-fall-2014/blob/master/Topic_Modeling_Options.ipynb, Accessed 1/2015)

Chiluwa, I and Ifukor. P. (2015). War against our Children’: Stance and evaluation in #BringBackOurGirls campaign discourse on Twitter and Facebook, Sage, Discourse and Society 2015, Vol. 26(3) 267– 296,

Chomsky, N (1969), Some Empirical Assumptions in Modern Philosophy of Language, In Philosophy, Science and Method: Essays in Honour or Ernest Nagel, St. Martin’s Press

Crook, N. T., Pulman, S., Smith, C and Moore, R. (2010). Handling user interruptions in an embodied conversational agent, In Proceedings of the 9th international conference on autonomous agents and multi-agent systems (AAMAS'2010), 2010

Crook, N.T., Granell, R. And Pulman, S. (2009). Unsupervised Classification of Dialogue Acts Using a Dirichlet Process Mixture Model. In: Proceedings of SIGDIAL 2009: The 10th Annual Meeting of The Special Interest Group in Discourse and Dialogue, Pages 341–348, Queen Mary University of London, September 2009.

Crystal, D. (2001). Language and the Internet. Cambridge, U.K.: Cambridge University Press. P59 'You could make this clearer': Teachers' advice on ESL academic writing

Crystal, D. (1997). A Dictionary of linguistics and phonetics, 4th edition, Blackwell.

Dale, R. (2015). Industry Watch NLP meets the cloud, Natural Language Engineering 21 (4): 653–659, Cambridge University Press 2015.

Dariah. (2015, 2016). https://de.dariah.eu/tatom/topic_model_python.html, accessed 01/2015

Data Protection Act (2018). https://www.gov.uk/data-protection, accessed 2016, 05/2018

Decapua, A and Dunham. J. F. (1993). Strategies in the Discourse of Advice, Journal of Pragmatics 1993; 20:519–531

DeCapua, A. and Huber, L. (1995). ‘If I were you...’: Advice in American English. Multilingual 14(2): 117-132.

DeCapua, A. and Huber. L. (1993). Strategies in the discourse of advice. Journal of Pragmatics, 20:101-125

Dennis, C. (2003). Peer support within a health care context: a concept analysis. Int J Nurs Stud 2003; 40 (3): 321-32

Diabetes UK Facebook. (2015, 2016, 2017, 2018, 2019, 2020). https://www.facebook.com/diabetesuk/, accessed 01/2015 -2018

Diabetes UK Website. (2015, 2016, 2017, 2018). Http://Www.Diabetes.Org.Uk/, accessed 2015-2018

Diederich, C., Höhn. N, and Wierzbicka. A. (2012). In Locher, M. A and Limberg. H. (2012). Advice in Discourse, Pragmatics and Beyond, New Series (P&BNS)

Dreyfus, H. (1972). What Computers Can’t Do, New York: MIT Press

Du Bois, J. W. (2007). The stance triangle. In R. Englebretson (ed.) Stance-taking in discourse: Subjectivity, evaluation, interaction. Amsterdam: John Benjamins. 139–182.

Du Gay, P., Hall, S., Janes, L., Mackay, H. And Negus, K. (1997). Doing Cultural Studies: The Story of the Sony Walkman. Sage Publications (In Association with the Open University)

Dwyer, C. (2007). Digital Relationships in the Proceedings of the 40th Hawaii International Conference on System Sciences (HICSS), Hawaii, January

Eichhorn, K. C. (2008). Soliciting and providing social support over the Internet: An investigation of online eating disorder support groups. Journal of Computer-Mediated Communication, 14(1), 67-78

Elgesem. D. (2015). In Fossheim, H. and Ingierd, H. (2015). Internet Research Ethics. Available at: https://press.nordicopenaccess.no/index.php/noasp/catalog/view/3/1/9-1.

Ellis, R. J. B., Connor. U and Marshall. J. (2014). Development of patient-centric linguistically tailored psychoeducational messages to support nutrition and medication self-management in type 2 diabetes: a feasibility study, Patient Preference and Adherence 2014:8 1399–1408

Endres, D and Whitlock. L. A. (2017). The Development of a Coding Scheme to Examine Technology Use in Diabetes Self-Management, Proceedings of the Human Factors and Ergonomics Society 2017 Annual Meeting

Englebretson, R. (2007). Stance-taking in discourse: An introduction. In R. Englebretson (ed.) Stance-taking in discourse: Subjectivity, evaluation, interaction. Amsterdam: John Benjamins. 1–25.

Ess. C. (2015). In Fossheim, H. and Ingierd, H. (2015). Internet Research Ethics. Available at: https://press.nordicopenaccess.no/index.php/noasp/catalog/view/3/1/9-1

Fabian, R and Nicolae. M. A. (2009). Natural language processing implementation on Romanian ChatBot, Proceedings of the 9th WSEAS International Conference on simulation, modelling and optimization

Facebook (2015, 2016, 2017, 2018, 2019, 2020), https://www.facebook.com/FacebookUK/, (accessed 2/2015-2018)

Facebook Data Policy (2015, 2016, 2017, 2018) https://www.facebook.com/full_data_use_policy, (Last accessed November 2017, 2018)

Fairclough, N. (1989). Language and Power. London: Longman.

Fairclough, N. (1992). Discourse and Social Change. Cambridge: Polity Press.

Fairclough, N. (1995). Critical Discourse Analysis. London: Longman

Fairclough, N. (2010). Critical Discourse Analysis, 2nd Ed. Harlow: Longman.

Fan, S. B. (2010). Roles in Online Collaborative Problem Solving, 2010 IEEE Symposium on Visual Languages and Human-Centric Computing

Fernback, J (2007). Beyond the diluted community concept: A symbolic interactionist perspective on online social relations. New Media and Society, 9(1) 49-69

Ferretti, L., Wymant. C., Kendall. M., Zhao. L., Nurtay. A., Abeler-Dörner. L., Parker. M., Bonsall. D. and Fraser. C. (2020). Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing, Science 10.1126/science.abb6936 (2020)

Fossheim, H. and Ingierd, H. (2015). Internet Research Ethics. Available at: https://press.nordicopenaccess.no/index.php/noasp/catalog/view/3/1/9-1

Foucault, M. (1972). The Archaeology of Knowledge, A.M. Sheridan Smith, Trans, Pantheon Books New York

Foucault, M. (1980). Power/Knowledge: Selected Interviews and Other Writings 1972– 1977, C. Gordon, L. Marshall, J. Meplam and K. Soper (Eds). Brighton, Sussex: The Harvester Press

Fraser, B. (1990). An approach to discourse markers. Journal of Pragmatics 14, 383-95

Fraser, B. (1998). Contrastive discourse markers in English. In A. Jucker and Y. Ziv (eds), Discourse markers: Description and theory. Amsterdam/Philadelphia: Benjemins, PP301-26

Fraser, B. (1999). Pragmatic Markers? Journal of Pragmatics, 6:2.167 -190, Elsevier

Fuoli, M and Hommerberg. C. (2015). Optimising transparency, reliability and replicability: Annotation principles and inter-coder agreement in the quantification of evaluative expressions. Corpora 10(3): 315–349.

Fuoli, M. (2018). A step-wise method for annotating APPRAISAL. Functions of Language 25(2).

Gabrielatos, C., Eivind. T., Sebastian. H and Fox. S. (2010). A corpus-based sociolinguistic study of indefinite article forms in London English. Journal of English Linguistics. Pre-published 18 February 2010

Galegher, J. L. S and Kiesler. S. (1998). ‘Legitimacy, authority, and community in electronic support groups’ Written Communication, 15: 493-530

GDPR, EU. (2018). European Union, General Data Protection Regulation (GDPR) (2018), https://www.eugdpr.org/eugdpr.org.html, accessed 2018

Georgalou, M. (2014). Constructions of identity on Facebook: A discourse-centred online ethnographic study of Greek users. Doctoral dissertation. Lancaster: Lancaster University.

Georgalou, M. (2016). Place identity and social media: The Greek case of brain drain migration. Paper to be presented at the 1st International Conference on "Europe in Discourse: Identity, Diversity, Borders", Hellenic American University, Athens, 23-25 September 2016.

Goffman, E. (1981). Forms of Talk, Oxford: Basil Blackwell

Gohr, A., Hinneburg A., Schult R., Spiliopoulou M. (2009). Topic Evolution in a Stream of Documents. Proceedings of the Ninth SIAM International Conference on Data Mining. Nevada, USA.

Golder, S. A and Macy, M. W. (2011). Diurnal and seasonal mood vary with work, sleep, and day length across diverse cultures. Science, 333(6051), 1878–1881.

Goldsmith, D. (2000). Soliciting Advice: The Role of Sequential Placement in Mitigating Face Threat. Communication Monograph 2000; 67:1–19

Graham, S., Weingart S., and Milligan I. (2012). Getting Started with Topic Modeling and MALLET. The Programming Historian 2. http://programming historian.org/lessons/topic-modeling-and-mallet

Greif, I. (1988). Computer-Supported Cooperative Work, A Book of Readings, Morgan Kaufmann, 25 Dec 1988, California

Grice, P. (1975). Logic and Conversation. In Cole, P.; Morgan, J. Syntax and Semantics. 3: Speech Acts. New York: Academic Press. Pp. 41–58

Griffiths, T. L and Steyvers. M. (2004). Finding scientific topics, 5228–5235, PNAS, April 6, 2004 vol. 101 suppl. One www

Griffiths, T. L., Steyvers, M., and Tenenbaum, J. B. (2007). Topics in semantic representation. Psychological Review, 114,211-244. (Pdf) (Topic modeling toolbox) Statistical Models of Language

Grimmelman, J. (2009). Facebook and the social dynamics of privacy. Iowa Law Review, 95, 4.

Grimmer, J and Stewart. B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts, Political Analysis pp. 1–31, doi:10.1093/pan/mps028

Grudin, J. (1988). Why CSCW applications fail: problems in the design and evaluation of organizational interfaces, Proceedings of the 1988 ACM conference on Computer-supported cooperative work, P85-93, ACM, USA

Grudin, J. (1994). Computer-Supported Cooperative Work: History and Focus, IEEE, Computer, V27, N5, P19-26.

Grudin, J. (2010). CSCW: time passed, tempest, and time past, Interactions 17, 4 July 2010, ACM

Gumperz, J. (1992). Contextualization revisited. In the Contextualization of Language, P. Auer & A. di Luzio (eds), 39–54. Amsterdam/Philadelphia: John Benjamins.

Gumperz, J. J. (1982a). Discourse Strategies. Cambridge: Cambridge University Press

Gumperz, J. J. (1982b). The linguistic basis of communicative competence. In D.Tannen (ed.), Analysing Discourse: Text and Talk (Georgetown University Roundtable on Languages and Linguistics 1981). Washington, DC: George Town University Press, 323-34

Hall, C. J., Smith. P. H and Wicaksono. R. (2011). Mapping Applied Linguistics: A Guide for Students and Practitioners, Routledge, NY

Hall, D., Jurafsky D. and Manning C. D. (2008). Studying the History of Ideas Using Topic Models. Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). Pp 363-371. Hawaii, USA.

Hall, S. (1997). Representation, Cultural Representations, And Signifying Practices, Sage Publications, London

Halliday, M. (1967). In Halliday, 2005, Studies in English Language, Vol. 7 in The Collected Works. Journal of Linguistics, 3.1, 1967; 3.2, 1967; 4.2, 1968.

Halliday, M and Hasan. R. (1976). Cohesion in English. London: Longman

Halliday, M. A. K. ([1985] 1994). Introduction to functional grammar (London: Edward Arnold)

Halliday, M. A. K. (1978). Language as Social Semiotic: The Social Interpretation of Language and Meaning. London: Edward Arnold.

Halliday, M. A. K. and C. M. I. M. Matthiessen. (2004). An introduction to functional Grammar, 3rd edn (London: Hodder Arnold)

Halliday, M.A.K. (1977). Text as semantic choice in social contexts. Reprinted in full in Linguistic Studies of Text and Discourse. Volume 2 in The Collected Works of M.A.K. Halliday. Edited by J, J. Webster, London

Halliday, M.A.K. (1985). Systemic Background. In "Systemic Perspectives on Discourse, Vol. 1: Selected Theoretical Papers" from the Ninth International Systemic Workshop, James D. Benson and William S. Greaves (eds). Ablex. Reprinted in Full in Volume 3 in The Collected Works of M.A.K. Halliday, London

Halliday, M.A.K. (1992). Systemic Grammar and the Concept of a "Science of Language". In Waiguoyu (Journal of Foreign Languages), No. 2 (General Series No. 78), pp1-9. Reprinted in Full in Volume 3 in The Collected Works of M.A.K. Halliday, London

Halliday, M.A.K. (2003). Introduction: On the "architecture" of human language. In on Language and Linguistics. Volume 3 in The Collected Works of M.A.K. Halliday. Edited by Jonathan Webster, London

Halliday, M.A.K. (2004). Introduction: How Big is a Language? On the Power of Language. In the Language of Science: Volume 5 in The Collected Works of M.A.K. Edited by J.J. Webster, London.

Hamblin, C. L. (1973). Questions in Montague English. Foundations of Language 10:41-53. Reprinted in Montague Grammar, Ed. Barbara H. Partee, 247-259. New York: Academic Press.

Hanks, P. (2004). Corpus Pattern Analysis, Computational Lexicography and Lexicology, Euralex 2004 Proceedings

Hardt-Mautner, G. (1995), Only Connect, Critical Discourse Analysis and Corpus Linguistics, University of Lancaster

Harris, J. K., Mueller. N. L., Snider. D, and Haire-Joshu. D. (2013), Local Health Department Use of Twitter to Disseminate Diabetes Information, United States. Prev Chronic Dis 2013; 10

Harrison S, Barlow J. (2009). Politeness Strategies and Advice-Giving in an Online Arthritis Workshop. Journal of Politeness Research: Language, Behaviour, Culture, 2009; 5:93–111.

Harvey, K and Koteyko, N. (2013). Exploring Health and Communication, Language in Action, Routledge, London

Harvey, K. (2013). Investigating Adolescent Health Communication, a Corpus Linguistics Approach, Research in Corpus and Discourse, Bloomsbury, London

Hashmi, S. A. (2012). Said-Huntington Discourse Analyzer: A machine-learning tool for classifying and analysing discourse, Massachusetts Institute of Technology

Heritage, J. and Sefi, S. (1992). Dilemmas of Advice: aspects of the delivery and reception of advice in interactions between health visitors and first-time mothers, Chapter 12, pp. 359-417. In, Drew, P. and Heritage, J. (eds.) (1992) Talk at Work. Cambridge, Cambridge University Press. 0521376335

Herring, S. C. (2004). Computer mediated discourse analysis: An approach to researching online communities, in S. A. Barab, Kling. R. B and Gray. J. H (eds) Designing for virtual communities in the service of learning. Cambridge, UK, Higgins, C. (2012). Language attitudes as stance-taking: An interview-based study on intergenerational transmission among Native Hawaiians

Hill, J., Ford, W. R., and Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human-human online conversations and human–chatbot conversations. Computers in Human Behavior 49 (2015), 245–250.

Hinkel, E. (1994). Appropriateness of advice as L2 solidarity strategy. RELC Journal 25(2): 71–93.

Hinkel, E. (1997). Appropriateness of advice: DCT and multiple-choice data. Applied Linguistics 18(1): 1–26.

Hinkel, E. (2003a). Adverbial markers and tone in L1 and L2 students' writing. Journal of Prag Rizomilioti, V. (2006). Exploring epistemic modality in academic discourse using corpora

Hiss, F. (2013). Tromsø as a "Sámi Town"?--Language Ideologies, Attitudes, and Debates Surrounding Bilingual Language Policies, Language Policy, v12 n2 p177-196 May 2013

Hoey, M., Mahlberg, M., Stubbs, M., Teubert, W. and Sinclair, J. (2007). Text, discourse and corpora. Continuum International Publishing Group.

Hoffmann, C and Bublitz. W. (2018) (ed.), Pragmatics of Social Media, De Gruyter, Berlin, http://www.degruyter.com/view/product/458679

Hood, S. (2004). 'Appraising research: taking a stance in academic writing'. Unpublished PhD thesis. University of Technology, Sydney

Hood, S. (2007). ‘Arguing in and across disciplinary boundaries: legitimizing strategies in applied linguistic and cultural studies’ in Whittaker, R., M. O’Donnell and A. McCabe (eds) Advances in language education (London: Continuum), pp. 185–200, Google Scholar

Hood, S. (2010). Appraising research: evaluation in academic writing (London: Palgrave Macmillan)

Hood, S. (2011). ‘Writing discipline: comparing inscriptions of knowledge and knowers in academic writing’ in Christie, F. and K. Maton (eds) Disciplinarity: systemic functional and sociological perspectives (London: Continuum)

Hood, S. and J. R. Martin. (2007). ‘Invoking attitude: the play of graduation in appraising discourse’ in Webster, J., C. Matthiessen and R. Hasan (eds) Continuing discourse on language, Vol. 2 (London: Equinox), pp. 739–64

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

Hunt, D and Harvey, K. (2012). Grammar and Doctor-Patient Communication. In: C.A. Chapelle, Ed., and the Encyclopaedia of Applied Linguistics Wiley-Blackwell. 2332-2338

Hunt, D and Harvey, K. (2015). Health Communication and Corpus Linguistics: Using Corpus Tools to Analyse Eating Disorder Discourse Online. In: P. Baker and T. McEnery, Ed., Corpora and Discourse Studies Palgrave. 134-154

Hunt, D and Koteyko. N. (2015). What Was Your Blood Sugar Reading This Morning?’ Representing Diabetes Self-Management on Facebook, Discourse and Society 1–19, Sage

Hunt, D. (2015). The Many Faces of Diabetes: A Critical Multimodal Analysis of Diabetes Pages on Facebook: Language and Communication Language and Communication. 43, 72-86

Hunt, D., Koteyko, N and Gunter, B. (2015). UK Policy on Social Networking Sites and Online Health: From Informed Patient to Informed Consumer? Digital Health. 1, 2055207615592513

Hyland, K. (2005). Stance and engagement: a model of interaction in academic discourse, Discourse Studies, SAGE Publications, London, Vol 7(2): 173–192

ICO (2018), Information Commissioners Office, Anonymisation: managing data protection risk code of practice, https://ico.org.uk/for-organisations/guide-to-data-protection/anonymisation/, accessed 2018

Ihara, N. (2006). Expressions of Affect in English and Japanese Novels, Intercultural Communication Studies XV: 1 2006

Introducing the Linguistic Inquiry and Word Count (LIWC), (2016), By Dr Ryan Nichols, Philosophy, Cal State Fullerton, Orange County CA, Http://Www.Hecc.Ubc.Ca/Quantitative-Textual-Analysis/Qta-Practice/Liwcinpractice/, Accessed April 2016

Irwin, H. (1989). Health communication: the research agenda. Media Information Aust 1989; 54: 32-40.

Jakobson, R. (1971). Selected Writings vol. I, II, The Hague: Mouton.

Jefferson, G, and Lee, J. R. E. (1981). The rejection of advice: Managing the problematic convergence of a 'Troubles-Telling' and a 'Service Encounter'. Journal of Pragmatics 5: 399–422.

Jockers, M. L. (2013). Macroanalysis: Digital Methods and Literary History. University of Illinois Press.

Jones, C. (2011). Ethical issues in online research. British Educational Research Association on-line resource. Available at: https://www.bera.ac.uk/wpcontent/uploads/2014/03/Ethical-issues-in-online-research.pdf?noredirect=1

Jones, R. (2013). Health and Risk Communication: An Applied Linguistic Perspective. Abingdon: Routledge

Kaplan, A. M. and Haenlein. M. (2010). Users of the world, unite! The challenges and opportunities of social media, Business Horizons, V53, N1, P59-68 10

Karami, A and Zhou, B. (2015). Online review spam detection by new linguistic features. IConference 2015 proceedings.

Karami, A and Zhou. L. (2014a). Exploiting latent content-based features for the detection of static SMS spams. The 77th annual meeting of the Association for Information Science and Technology (ASIST).

Karami, A and Zhou. L. (2014b). Improving static SMS spam detection by using new content-based features. The 20th Americas Conference on Information Systems (AMCIS).

Karami, A., Dahl. A. A., Turner-McGrievy. G., Kharrazi and Shaw Jr. G. (2018). Characterizing diabetes, diet, exercise, and obesity comments on Twitter, International Journal of Information Management 38 (2018) 1–6

Karyotis, C., Doctor. F., Iqbal. R., James. A and Chang. V. (2017). A fuzzy computational model of emotion for cloud-based sentiment analysis, Information Sciences (2017)

Kaufman, N. (2010). Internet and information technology use in treatment of diabetes. Int J Clin Pract Suppl 2010; (166):41–6. CrossRef PubMed

Kemp, M. (2003). 'Hearts and Minds: Agency and Discourse on Distress', Anthropology and Medicine, 10; 187-205

Kerly, A., Hall. P and Bull. S. (2006). Bringing chatbots into education: Towards natural language negotiation of open learner models, Knowledge-Based Systems 20 (2007) 177–185

Kiesling, S. F. (2011). Stance in context: Affect, alignment and investment in the analysis of stance-taking. Paper presented at the i-mean 2, Context and Meaning Conference, University of the West of England, Bristol, UK, 13–15 April 2011. Retrieved from: http://www.academia.edu/1037087/Stance_in_context_Affect_alignment_and_investment_in _the_analysis_of_stance-taking (accessed 26 April 2016).

Kireyev, K., Palen. L and Anderson. K.M. (2009). Applications of Topics Models to Analysis of Disaster-Related Twitter Data, umiacs.umd.edu

Kleinberg, J. M. (2007) Challenges in mining social network data: processes, privacy and paradoxes. Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, New York, 2007.

Kokkinakis, D and Malm. M. (2013). A macroanalytic view of Swedish literature using topic modelling, Univ. of Gothenburg, Corpus Linguistics 2013

Kokkinakis, D. (2004). Reducing the Effect of Name Explosion. Proceedings of the 4th LREC Workshop: Beyond Named Entity Recognition, Semantic Labelling for NLP tasks. Pp. 1-6. Lissabon, Portugal.

Kosinski, K., Matz. S. C., Gosling. S. D., Popov. V and Stillwell. D. (2015). Facebook as a Research Tool for the Social Sciences, Opportunities, Challenges, Ethical Considerations, and Practical Guidelines. September 2015, American Psychologist 543, 2015 American Psychological Association, Vol. 70, No. 6, 543–556

Koteyko, N and Hunt, D. (2016). Performing Health Identities on Social Media: An Online Observation of Facebook Profiles: Discourse, Context & Media Discourse, Context & Media.

Koteyko, N. (2006). Corpus Linguistics and the Study of Meaning in Discourse. The Linguistics Journal, 1(2), 131-154.

Koteyko, N. (2009) ‘I Am a Very Happy, Lucky Lady, And I Am Full of Vitality!’ Analysis of Promotional Strategies on the Websites of Probiotic Yoghurt Producers. Critical Discourse Studies 6(2): 111–125.

Koteyko, N., Hunt, D and Gunter, B. (2015). Expectations in the Field of the Internet and Health: An Analysis of Claims about Social Networking Sites in Clinical Literature Sociology of Health & Illness. 37(3), 468-484

Kouper, I. (2010). The Pragmatics of Peer Advice in a LiveJournal Community. Language @Internet 2010; 7: Article 1

Krishnamurthy, R. (1996) 'Ethnic, Racial and Tribal: The Language of Racism?' in C.R. Caldas-Coulthard and M. Coulthard (eds) Texts and Practices: Readings in Critical Discourse Analysis, pp. 129–49. London: Routledge.

Labov, W. (ed.) (1972). Language in the Inner City: Studies in the Black English Vernacular. Philadelphia, USA, University Of Pennsylvania Press

Lakoff, R. (1977). What you can do with words: politeness, pragmatics and performatives. In A. Rogers, B. Wall and J.P. Murphy (eds). Proceedings of the Texas conference on performatives, Presuppositions and Implicatures. Washington, DC: Center for Applied Linguistics

Lamrichs, J. and Molder, H. (2003). Computer-mediated communication: From a cognitive to a discursive Model, New media and Society, 5:451-73

Larsson. A. O. (2015). In Fossheim, H. and Ingierd, H. (2015). Internet Research Ethics. Available at: https://press.nordicopenaccess.no/index.php/noasp/catalog/view/3/1/9-1.

Laurence, A. (2014). Antconc (Version 3.4.3) [Computer Software]. Tokyo, Japan: Waseda University. Available from Http://Www.Laurenceanthony.Net/

Lave, J. And Wenger, E. (1991). Situated Learning: Legitimate Peripheral Participation. Cambridge, UK: Cambridge University Press

Lee, D. D and Seung, H. S. (1999). Learning the Parts of Objects by Non-Negative Matrix Factorisation, Nature, Vol 401, 21 October 1999, www.Nature.Com

Lehmann, W. (1987). Language, University of Chicago Press, 15 Oct 1984

Leonardi, P. M., Nardi, P. A and Kallinikos. J. (2012). Materiality and Organizing: Social Interaction in a Technological World, University Press, Oxford

Lieberman. M. A and Goldstein. B. A. (2006). Not all negative emotions are equal: the role of emotional expression in online support groups for women with breast cancer, Psycho-Oncology 2006; 15: 160–8.

LIWC (2012). [Online]. [Accessed 2017]. Available from: http://liwc.wpengine.com/

LIWC (2016). [Online]. [Accessed 2017]. Available from: http://liwc.wpengine.com/

Lifeng, Shang, Zhengdong. Lu, Hang. Li. (2015). Neural Responding Machine for Short-Text Conversation. ACL (1) 2015: 1577-1586

Locher M.A. (2008). Relational work, politeness and identity construction. In Handbook of Applied Linguistics. Volume 2: Interpersonal Communication, G. Antos, E. Ventola & T. Weber (eds), 509–540. Berlin: Mouton de Gruyter.

Locher, M. A and Limberg. H. (2012). Advice in Discourse, Pragmatics and Beyond, New Series (P&BNS)

Locher, M. A. (2006). Advice Online. Advice-giving in an American Internet Health Column. Amsterdam/ Philadelphia: John Benjamins.

Locher, M. A. (2010). Health Internet sites: A linguistic perspective on health advice columns, Social Semiotics 20: 43–59.

Locher, M. A. (2011). Situated impoliteness: The interface between relational work and identity construction. In Situated Politeness, B. Davies, M. Haugh & A.J. Merrison (eds), 187–208, London: Continuum.

Locher, M. A. (2013). Internet advice. In the Handbook of the Pragmatics of Computer-mediated Communication, S. Herring, D. Stein & T. Virtanen (eds). Berlin: Mouton de Gruyter.

Locher, M. A. and Graham, S. L. (2010). Introduction to interpersonal pragmatics. In Handbook of Interpersonal Pragmatics, M.A. Locher & S.L. Graham (eds), 1–13. Berlin: Mouton de Gruyter.

Locher, M. A. and Watts, R. J. (2005). Politeness theory and relational work. Journal of Politeness Research 1: 9–33.

Louvign´e, S and Rubens. N. (2016). Behaviormetrika, Vol.43, No.1, 2016, 65–82, Meaning-Making Analysis and Topic Classification Ofsns Goal-Based Messages

Lüders. M. (2015). In Fossheim, H. and Ingierd, H. (2015). Internet Research Ethics. Available at: https://press.nordicopenaccess.no/index.php/noasp/catalog/view/3/1/9-1

Martin, J. R. (1992). English Text: System and Structure. John Benjamins Publishing Company: Philadelphia, Amsterdam

Martin, J. R. (2000). ‘Beyond exchange: APPRAISAL systems in English’ in Hunston, S. and G. Thompson (eds) Evaluation in Text: Authorial Stance and the Construction of Discourse (Oxford: OUP), pp. 142–75

Martin, J. R. and D. Rose. (2007). Working with discourse: meaning beyond the clause, 2nd edn (London: Continuum)

Martin, J. R. and Rose, D. (2003). Working with discourse: Meaning beyond the clause. New York: Continuum.

Martin, J. R. and White. P. R. R. (2005). The language of evaluation: appraisal in English (London: Palgrave Macmillan)

McEnery, T and Wilson. A. (2001). Corpus linguistics, 2nd edn. Edinburgh: Edinburgh University Press.

Mead, S., Hilton. D and Curtis. L. (2001). "Peer Support: A Theoretical Perspective." Psychosocial Rehabilitation Journal, 2001: 25:2, P. 134-141

MeaningCloud. (2017). https://www.meaningcloud.com/, (Last accessed September 2017)

Mickelson, K. D. (1997). ‘Seeking social support: Parents in electronic support groups, in. Kiesler (ed.) Culture of the Internet. NJ, USA: Lawrence Erlbaum Associates

Miguel, A. Benitez-Castro., Encarnación. H., Zaragoza. U and Granada. U. (2017). "Very bad. Very, very bad. Terrible altogether": A corpus-based CDA of the interviews of the Magdalene Laundries survivors, 15 Corpus Linguistics in the South, University of Cambridge

Miles, E and Winget. M. (2010). Questions Are Content: A Taxonomy of Questions in A Microblogging Environment, Asist 2010, October 22–27, USA.

Miles, M. B. And Huberman, A. M. (1994). Qualitative Data Analysis: An Expanded Sourcebook, 2nd Edition, Sage, London

Mimno, D. (2012). Computational Historiography: Data Mining in a Century of Classics Journals. J. on Computing and Cultural Heritage (JOCCH). Vol. 5:1. http://www.perseus.tufts.edu/publications/02-jocch-mimno.pdf

Mohammad, S. F., Sobhani. P, and Kiritchenko. S. (2016). Stance and Sentiment in Tweets ACM Trans. Embedd. Comput. Syst. 0, 0, Article 0 (2016), 22 pages.

Moretti, F. (2005). Graphs, maps, trees: abstract models for a literary history. R. R. Donnelley & Sons. Nelson R.K., Mimno D. and Brown T. 2012. Topic Modeling the Past. Proceedings of the DH2012, Hamburg, Germany. http://www.dh2012.unihamburg.de/

Morrow, P. R. (2012). Online advice in Japanese: Giving advice in an Internet discussion forum P255 chapter 12, In Miriam A. Locher and Holger Limberg. (2012). Advice in Discourse, Pragmatics & Beyond New Series (P&BNS)

MRC, (2018), Medical Research Council, https://www.insight.mrc.ac.uk/2018/04/16/gdpr-research-changes/, accessed 2018

Nadeau, D and Sekine. S. (2007). A survey of named entity recognition and classification. Lingvisticae Investig. 30(1), 3–26.

Nanos, A., James, A., Iqbal, R. and Hedley, Y-L. (2017). Content Summarisation of Conversation in the Context of Virtual Meetings: An enhanced TextRank, ApproachIEEE 21st International Conference on Computer Supported Cooperative Work in Design (CSCWD). IEEE, 6 p.

Narayanan, A. and Shmatikov, V. (2008) 'Robust de-anonymization of large sparse datasets (How to break anonymity of the Netflix prize dataset.)', IEEE Symposium on Security & Privacy, Oakland, CA. Available at: http://arxiv.org/pdf/cs/0610105v2.pdf

Narayanan, A. and Shmatikov, V. (2009) 'De-anonymizing social networks', IEEE Symposium on Security & Privacy, Oakland, CA. Available at: http://www.cs.utexas.edu/~shmat/shmat_oak09.pdf

NHS Diabetes. (2015, 2016, 2017, 2018). http://www.nhs.uk/Conditions/Diabetes/Pages/Diabetes.aspx, Accessed 6/2016, 2/2017, 2018

NHS Digital. (2018). https://digital.nhs.uk/, Accessed 2018

NHS Lift Psychology. (2017). https://lift.awp.nhs.uk/, accessed 1/2017

NICE. (2017). https://www.nice.org.uk/guidance/QS6/chapter/Update-information, Accessed 2/2017

Nijland, N., Seydel, E. R., Van Gemert-Pijnen J. E. W. C., Brandenburg, B., Kelders, S. M. and Will, M. (2009). Evaluation of an Internet-Based Application for Supporting Self-Care of Patients with Diabetes Mellitus Type 2, eHealth, Telemedicine, and Social Medicine, eTELEMED '09

Norvig, P. ([2011] 2016), On Chomsky and the Two Cultures of Statistical Learning. Http://Norvig.Com/Chomsky.Html, 2011. [Online; Accessed 09/2016].

Ochs, E. (1990). "Cultural Universals in the Acquisition of Language." Papers and Reports on Child Language Development, 29, pp. 1-19.

O'Reilly, T. (2009). What is Web 2.0, O'Reilly Media Inc

Oxford Brookes University Information Management. (2018). https://www.brookes.ac.uk/it/information-management/gdpr/, accessed 2018

Oxford dictionaries. (2016). Http://Www.Oxforddictionaries.Com/Definition/English/Mitigate, Accessed June 2016, 2017

Park, H. Rodgers. S and Stemmle. J. (2011). Health organizations' use of Facebook for health advertising and promotion, Journal of Interactive Advertising 12(1): 62–77.

Partington, A. (2008). The armchair and the machine: Corpus-Assisted Discourse Studies, in C. Taylor Torsello, K. Ackerley and E. Castello (eds) Corpora for University Language Teachers, Bern: Peter Lang, 189-213.

Partington, A. (2004). 'Corpora and Discourse, a Most Congruous Beast', in A. Partington, J. Morley and L. Haarman (eds) Corpora and Discourse, pp. 11–20. Bern: Peter Lang.

Partington, A. (2006). 'Metaphors, Motifs and Similes Across Discourse Types: Corpus-assisted Discourse Studies (CADS) at Work', in A. Stefanowitsch and

S. Gries (eds) Corpus-based Approaches to Metaphor and Metonymy, pp. 267–304. Berlin: Mouton de Gruyter

Pennebaker, J. W., and Davison. K. (1997). Virtual Narratives: Illness Representations in Online Support Groups, In Perceptions of Health and Illness, Harwood, Netherlands

Pennebaker, J. W., Boyd, R. L., Jordan, K., and Blackburn, K. (2015). The development and psychometric properties of LIWC2015. Austin, TX: University of Texas at Austin

Pennebaker. J. W and Francis. M. E. (1996). Cognitive, emotional, and language processes in disclosure. Cognition and Emotion 1996; 10: 601–26

Petyko, M. (2017). The use of n-gram, collocation and keyword analysis to annotate linguistically marked motives attributed to trolls in the comment threads of British political blogs, Lancaster, 15 Corpus Linguistics in the South, University of Cambridge

Pollach, I. (2012). Taming Textual Data: The Contribution of Corpus Linguistics to Computer-Aided Text Analysis, Organizational Research Methods 15(2) 263-287, Sage

Prabhu. R. (2015). In Fossheim, H. and Ingierd, H. (2015). Internet Research Ethics. Available at: https://press.nordicopenaccess.no/index.php/noasp/catalog/view/3/1/9-1

Precht, K. (2000). Patterns of stance in English. Flagstaff: Northern Arizona University.

Prince, G. (1982). Narratology: The Form and Functioning Of Narrative: Berlin, Germany, New York, USA And Amsterdam, The Netherlands: Morton

Pulman, S. (2017), theysay, http://www.theysay.io/guide/language-analytics-explained-by-prof-stephen-pulman/, (accessed June 2017)

Ramage, D., Dumais, S and Liebling, D. (2010). Characterizing Microblogs with Topic Models, Proceedings of the Fourth International AAAAI Conference on Weblogs and Social Media

Reinhart, T. (1981). Pragmatics and Linguistics: An Analysis of Sentence Topics, Philosophica 27, 1981 (1), pp. 53-94

Rich, E., Knight. K and Nair. S. B. (2009). Artificial Intelligence, 3RD Edition, Tata McGraw Hill, New Delhi

Richardson, K. P. (2003). “Health risks on the internet: Establishing credibility online.” Health, Risk and Society 5 (2): 171-184.

Richardson, K. P. (2005). Internet Discourse and Health Debates. New York: Palgrave, Macmillan.

Rizomilioti, V. (2006). Exploring epistemic modality in academic discourse using corpora. In E. Arno´ Macia`, A. Soler Cervera, & C. Rueda Ramos (Eds.), Information technology in languages for specific purposes (pp. 53-71). New York: Springer.

Rodrigues, R. G., Marques das Dores. R., Camilo-Junior. C. G and Rosa. T. C. (2016). SentiHealth-Cancer: A sentiment analysis tool to help detecting mood of patients in online social networks, International Journal of Medical Informatics 85 (2016) 80–95

Roos, I. (2003). Reacting to the diagnosis of prostate cancer: Patient learning in a community of practice, Patient education and counselling, 49: 219-24

Rose, N and Novas, C. (2004). Biological Citizenship. In: Ong A and Collier SJ (Eds) Global Assemblages. Oxford: Blackwell Publishing, Pp. 439–463.

Russell, S. J and Norvig, P. (2009). Artificial Intelligence: A Modern Approach (3rd Ed.). Upper Saddle River, New Jersey: Prentice Hall

Sarangi, S. (2000). Activity types, discourse types and interactional hybridity: The case of genetic counseling. In Discourse and Social Life, S. Sarangi and M. Coulthard (eds), 1–14. Mahwah, NJ: Lawrence Erlbaum.

Sarangi, S. and Clarke, A. (2002a). Constructing an account by contrast in counselling for childhood genetic testing. Social Science & Medicine 54: 295–308.

Sarangi, S. and Clarke, A. (2002b). Zones of expertise and the management of uncertainty in genetics risk communication. Research on Language and Social Interaction 35(2): 115–137.

Schegloff, E. A. (1984). On some questions and ambiguities in conversation. In J.M. Atkinson, and J. Hevitage (eds), Structure of social action. Cambridge: Cambridge university press 29-52

Schegloff, E. A. (1988). Presequences and indirection: applying speech act theory to ordinary conversation. Journal of Pragmatics, 12:55-62

Schiffrin, D. Tannen., D. and, Hamilton. H. E. and Schiffrin. D. (2003). The Handbook of Discourse Analysis, Blackwell Handbooks in Linguistics, 2nd

Scikit. (2015). http://scikit-learn.org/stable/auto_examples/model_selection/plot_confusion_matrix.html, accessed Jan, 2015, 2016, 2017

Scikit-Learn. (2011, 2015). Scikit-Learn: Machine Learning in Python, Pedregosa, Et Al., July 12, Pp. 2825-2830, 2011

Searle, J. R. (1969). Speech Acts. An Essay in the Philosophy of Language. Cambridge: Cambridge University Press.

Searle, J. R. (1980). Minds, Brains and Programs, Behavioural and Brain Sciences, 3 (3): 417–457

Segadal, K. (2015). In Fossheim, H. and Ingierd, H. (2015). Internet Research Ethics. Available at: https://press.nordicopenaccess.no/index.php/noasp/catalog/view/3/1/9-1

Segura-Bedmar, I., Martínez. P., Revert. R and Moreno-Schneider. J. (2015). Exploring Spanish health social media for detecting drug effects, BMC Medical Informatics and Decision Making 2015, 15 (Suppl 2): S6

Shaw, R. J and Johnson. C. M. (2011). Health information seeking and social media use on the Internet among people with diabetes. Online J Public Health Inform 2011; 3(1). PubMed

Sidnell and Stivers (2014). The Handbook of Conversation Analysis, Blackwell Publishing Ltd

Sillence, E. (2010). Seeking Out Very Like-Minded Others: Exploring Trust and Advice Issues in an Online Health Support Group. International Journal of Web Based Communities 2010; 6:376–394

Sillence, E. (2013). Giving and Receiving Peer Advice in an Online Breast Cancer Support Group, Cyberpsychology, Behaviour, and Social Networking, 16 (6). pp. 480-485. ISSN 2152-2715

Simmons, R. F. (1969). Natural Language Question- Answering Systems, Computational Linguistics, V13, N1, January 1970, Communications of the ACM, Texas

Simon, H. (1969). The Sciences of the Artificial. MIT Press, Cambridge

Sinclair, J. (2004). How to use corpora in language teaching. Amsterdam: John Benjamins.

Smith, L and Weinert C. (2000). Telecommunication support for rural women with diabetes. Diabetes Educ 2000; 26(4):645– 55. CrossRef PubMed

Solomon, B. S., Duce. D. A. and Harrison. R. (2011). Methodologies for using Social Media Collaborative Work Systems, In A. K. Chopra, F. Dalpiaz, and S. L. Lim, editors, First International Workshop on Requirements Engineering for Social Computing (RESC 2011), in conjunction with the 19th IEEE International Requirements Engineering Conference, Trento, Italy, August 29th, 2011

Solomon, B. S., Duce. D. A. and Harrison. R. (2012). Modeling Social Media Collaborative Work Systems, ICSE2012, MISE2012 Workshop, Zurich, Switzerland, June 2012

Solomon, P. (2004). "Peer support/peer provided services underlying processes, benefits, and critical ingredients", Psychiatric Rehabilitation Journal, 2004; 27(4):392-401

Spears, R and Lea. M. (1994). 'Panacea or Panopticon? The hidden power in computer-mediated communication, Communication Research, 21:427-59

Sproull, L and Kiesler. S. (1986). 'Reducing social context cues: Electronic mail in organisation communication', Management Science, 32:1492-512

Srinivasan, B. V., Natarajan, A., Sinha, R and Sinha, R. (2015). Community Reaction: From Blogs to Facebook, In Proceedings of The Second ACM IKDD Conference on Data Sciences (Cods '15). ACM, New York, NY, USA, 69-74

Stahl, G. (1999). Reflections on Webguide: Seven Issues for the Next Generation of Collaborative Knowledge Building Environments. Proceedings of Computer Supported Collaborative Learning (CSCL) (Pp. 600-610). Alto, California

Stahl, G. (2000). A Model of Collaborative Knowledge Building. In B. Fishman & S. O'Connor-Divelbiss (Eds.), Fourth International Conference on the Learning Sciences (Pp.70-77). Mahwah, New Jersey: Erlbaum

Stahl, G. (2006). Group Cognition: Computer Support for Building Collaborative Knowledge, Cambridge, MA: MIT Press

Staksrud. E. (2015). In Fossheim, H. and Ingierd, H. (2015). Internet Research Ethics. Available at: https://press.nordicopenaccess.no/index.php/noasp/catalog/view/3/1/9-1

Steele, R. (2011). A Model for Social Network-enhanced Health Communication, 2011 Ninth IEEE International Conference on Dependable, Autonomic and Secure Computing

Steen-Johansen, K and Enjolras. B. (2015). in Fossheim, H. and Ingierd, H. (2015). Internet Research Ethics. Available at: https://press.nordicopenaccess.no/index.php/noasp/catalog/view/3/1/9-1

Steyvers, M., and Griffiths, T. (2007). Probabilistic topic models. In T. Landauer, D. S. McNamara, S. Dennis, and W. Kintsch (Eds.), Handbook of Latent Semantic Analysis. Hillsdale, NJ: Erlbaum. (pdf) (topic modeling toolbox)

Stirling, E. (2014). Using Facebook as a research site and research tool. In: Sage research methods cases. Sage

Stommel, W and Koole. T. (2010). The online support group as a community. A micro-analysis of the interaction with a new member, Discourse Studies, 12: 357-78

Stubbs, M. (1983). Discourse Analysis: The Sociolinguistic Analysis of Natural Language, University of Chicago Press

Stubbs, M. (1994). Grammar, Text, and Ideology: Computer-assisted Methods in the Linguistics of Representation’, Applied Linguistics 15(2): 201–23.

Suler, J. (2004). The Online Disinhibition Effect, Cyberpsychology and Behaviour I7, P321-326

Suthers, D. (2001). Collaborative Representations: Supporting Face-To-Face and Online Knowledge-Building Discourse. Proceedings of the 34th Hawai`I International Conference on the System Sciences (Hicss-34), January 3-6, 2001, Maui, Hawai`i (CD-ROM): Institute of Electrical and Electronics Engineers, Inc. (IEEE)

Suthers, D., Dwyer, N., Vatrapu, R. And Medina, R. (2007). An Abstract Transcript Notation for Analysing Interactional Construction of Meaning in Online Learning. Proceedings of the 40th Hawai`I International Conference on the System Sciences (Hicss-40), January 3-6, 2007, Waikoloa, Hawai`i (CD-ROM): Institute of Electrical and Electronics Engineers, Inc. (IEEE)

Taavitsainen, I. and Jucker A.H. (2010). Trends and developments in historical pragmatics. In Handbook of Historical Pragmatics, I. Taavitsainen and A. H. Jucker (eds), 3–30. Berlin: De Gruyter Mouton.

Tannen, D and Trester. A.M. (2013). Discourse 2.0: Language and New Media. Washington, D.C.: Georgetown University Press. 2013.

Tannen, D. (1984). Conversational style. Norwood, Ablex

Tausczik, Y. R and Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. Journal of language and social psychology, 29(1), 24-54.

Teubert, W. (2005a). Language as An Economic Factor. In G. Barmbrook, P. Danielsson and M. Mahlberg (Eds.), Meaningful Texts: The Extraction of Semantic Information from Monolingual and Multilingual Corpora (Pp. 96-106). London; New York: Continuum.

Teubert, W. (2005b). My Version of Corpus Linguistics. International Journal of Corpus Linguistics, 10(1), 1-13.

Teubert, W. (2007a). Writing, Hermeneutics, and Corpus Linguistics. Revista Signos, 40(64), 431-453.

Teubert, W. (2007b). Language and Corpus Linguistics. In W. Teubert and A. Cermakova (Eds.), Corpus Linguistics: A Short Introduction (Pp. 1-58). London; New York: Continuum

The National Longitudinal Study of Adolescent Health. (2008). http://www.cpc.unc.edu/projects/addhealth, 2008.

Thompson, R. (2012). Looking healthy: Visualizing mental health and illness online, Visual Communication 11(4): 395–420

Torgersen, E., Gabrielatos, C., Hoffmann, S. and Fox, S. (2011). A corpus-based study of pragmatic markers in London English. Corpus Linguistics and Linguistic Theory, 7(1), 93-118.

Townsend, L and Wallace. C. (2016). Social Media Research: A Guide to Ethics, The University of Aberdeen, https://www.gla.ac.uk/media/media_487729_en.pdf, (Last accessed November 2018)

Turing, A. M. (1950). Computing Machinery and Intelligence, Mind, 49: 433-460.

Van Dijk, T. A. (1980). Macrostructures. An interdisciplinary study of global structures in discourse, interaction, and cognition. Hillsdale, NJ: Erlbaum.

Van Dijk, T. A. (1990a). Discourse analysis in the 1990s. Text 1990; 10(1/2): 133-156

Van Dijk, T. A. (1990b). The growth of discourse analysis. Discourse Soc, 1990; l (l (1): 5-16

Van Leeuwen, T. (2008). Discourse and Practice. Oxford: Oxford University Press.

Vehvilainen, S. (2010). Evaluative Advice in Educational Counseling: The Use of Disagreement in the "Stepwise Entry" to Advice, Research on Language and Social Interaction, 34:3, 371-398, DOI: 10.1207/S15327973RLSI34-3_4

Verschueren, J. (2009). Introduction: The pragmatic perspective. In Key Notions for Pragmatics [Handbook of Pragmatics Highlights 1], J. Verschueren & J. Östman (eds), 1–27. Amsterdam: John Benjamins.

Wallach, H. (2006). Topic modeling: beyond bag-of-words. Proceedings of the 23rd International Conference on Machine Learning (ICML). Pp. 977– 984. Pittsburgh, U.S. http://people.ee.duke.edu/~lcarin/icml2006.pdf

Walstrom, M. K. (2000). 'You know, who's the thinnest?' combating surveillance and creating safety in coping with eating disorders online', CyberPsychology and Behaviour, 3: 761-83

Weizenbaum, J. (1966). Eliza, a Computer Program for the Study of Natural Language Communications between Man and Machine. Comm. ACM 9, 1 (Jan. 1966), 36-45

Wenger, E. (1998). Communities of practice: learning meaning and identity, Cambridge, UK

Wilson, J. (1989). On the Boundaries of Conversation, Oxford: Pergamon Press.

Wodak, R. (2007). ‘Pragmatics and Critical Discourse Analysis: A Cross-disciplinary Inquiry’, Journal of Pragmatics and Cognition 15(1): 203–27

Yang, T. I., Torget A. J. and Mihalcea R. (2011). Topic Modeling on Historical Newspapers. Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities. Portland, OR. http://www.aclweb.org/anthology/W11-1513

Zappavigna, M. (2011). Ambient Affiliation: A Linguistic Perspective on Twitter. New Media and Society V13, P788-806.

Zappavigna, M. (2012). Discourse of Twitter and social media, how we use language to create affiliation on the web, Bloomsbury Discourse series, UK

Zappavigna, M. (2015), ‘Twitter: Social Communication in the Twitter Age’, Discourse & Communication, vol. 9, pp. 379 - 380, http://dx.doi.org/10.1177/1750481315578969c

Zappavigna, M. (2017). ‘Twitter’, in Hoffmann C; Bublitz W (ed.), Pragmatics of Social Media, De Gruyter, Berlin, pp. 201 - 224, http://www.degruyter.com/view/product/458679

Zayts, O. and Schnurr, S. (2011). Laughter as the medical providers’ resource: Negotiating informed choice in prenatal genetic counseling. Research on Language and Social Interaction, 44(1): 1–20.

Zimmer, M. (2010). ‘‘But the data is already public’’: on the ethics of research in Facebook. Ethics and Information Technology. 12:313–325

9 BIBLIOGRAPHY

Add Health. (2008). Deductive disclosure. http://www.cpc.unc.edu/projects/addhealth/data/dedisclosure, 2008.

Barnbrook, G. (2002). Defining Language: A Local Grammar of Definition Sentences. Amsterdam, Philadelphia: J. Benjamin's Publication

Berger, P. and Luckmann, T. (1966). The Social Construction of Reality: A Treatise in the Sociology of Knowledge. Penguin.

Boyd, D. M., Golder, S and Lotan, G. (2011). Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. Proceedings of the 43rd Hawaii International Conference on System Sciences. IEEE.

Brown, P and Levinson, S.C. (1987). Politeness, Some Universals in Language Usage, Cambridge University Press, UK

Brown, T. (2012). Telling New Stories about Our Texts: Next Steps for Topic Modeling in the Humanities. Proceedings of the Digital Humanities. Hamburg, Germany.

Casado, R and Younas. M. (2014). Emerging trends and technologies in big data processing, Concurrency Computation.: Practical. Experiments, 2014. DOI: 10.1002/cpe.3398

CL2017a, Hardie conference talk. (2017). Exploratory analysis of word frequencies across corpus texts, https://www.youtube.com/watch?v=ka4yDJLtSSc, accessed 2017/2018

CL2017b, Hunston conference talk. (2017). Corpus Linguistics in 2017: a personal view, https://www.youtube.com/watch?v=877NVz5GQJw, accessed 2017/2018

CLS15, (2017), Pascual Pérez-Paredes, 15 Corpus Linguistics in the South Conference- 28 October 2017, F. Education, University of Cambridge

CLS15a, (2017). Sian in Corpus Linguistics in the South Conference 2017, https://www.educ.cam.ac.uk/events/conferences/2017corpusling/, accessed 2017

CLS15b, (2017). Gabrielatos in Corpus Linguistics in the South Conference 2017, https://www.educ.cam.ac.uk/events/conferences/2017corpusling/, accessed 2017

Davis, K. A. (1995), Qualitative Theory and Methods in Applied Linguistics Research, Tesol Quarterly Vol. 29, No. 3, Autumn 1995

De Saussure, F. (1915). Course in General Linguistics, edited by Charles Bally and Albert Sechehaye in Collaboration with Albert Riedlinger, Translated, With an Introduction and Notes by Wade Baskin, McGraw-Hill Book Company, New York Toronto London

Denzin, N. K and Lincoln, Y. S. (1994). The Discipline and Practice of Qualitative Research, the Sage Handbook of Qualitative Research

Edwards, J. (2009). Language and Identity: An introduction, Cambridge University Press

Ekman, P. (1972). Universals and Cultural Differences in Facial Expressions of Emotions. In Cole, J. (Ed.), Nebraska Symposium on Motivation (Pp. 207-282). Lincoln, N.B.: University of Nebraska Press

Eysenbach, G. (2008). Medicine 2.0: Social Networking, Collaboration, Participation, Apomediation, and Openness. Journal of Medical Internet Research 10(3): E22.

Forgas, J. P. (1995), Mood and Judgment: The Affect Infusion Model (Aim). Psychological Bulletin, Vol 117(1), Jan 1995, 39-66

Forgas, J. P. (2011). The Upside of Feeling Down: The Benefits of Negative Mood for Social Cognition and Social Behaviour, Sydney Symposium of Social Psychology, 2011: Social Thinking and Interpersonal Behaviour

Forgas, J. P. (2013). Don't Worry, Be Sad! On the Cognitive, Motivational, And Interpersonal Benefits of Negative Mood, Current Directions in Psychological Science June 2013 Vol. 22 No. 3 225-232

Fox, S. (2011). Peer-To-Peer Healthcare. Pew Internet and American Life Project. Available At: Http://Www.Pewinternet.Org/Files/Old-Media/Files/Reports/2011/Pew_P2phealthcare_2011. Pdf (Accessed March 2014).

Guthrie, A.M. (1997). On the systematic deployment of okay and mmhmm in academic advising sessions. Pragmatics 7(3): 397–415.

Halliday, M. A. K. and C. M. I. M. Matthiessen. (1999). Construing experience through meaning: a language-based approach to cognition (London: Cassell. Republished in 2006, London and New York: Continuum)

Harvey, K. J., 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.

Hausser, R. (1983). On Questions. In Questions and Answers, Ed. F. Kiefer, 97-158. Dordrecht: Reidel.

Hawn, C. (2009). Take Two Aspirin and Tweet Me in the Morning: How Twitter, Facebook, and Other Social Media Are Reshaping Health Care, Health Affairs, 28/2:361-368, Http://Content.Healthaffairs.Org/Content/28/2/361.Long

Honeycutt, C and Herring, S. (2009). Beyond Microblogging: Conversation and Collaboration Via Twitter. Proceedings of the 42nd Hawaii International Conference on System Sciences. IEEE

Hunt, D and Churchill, R. (2013). Diagnosing and Managing Anorexia Nervosa in UK Primary Care: A Focus Group Study: Family Practice. 30(4), 459-465

Jensen, J. F. (1998). Interactivity: Tracing a New Concept in Media and Communication Studies. Nordicom Review 19(1): 185–204.

Karttunen, L. (1977). Syntax and Semantics of Questions, Linguistics and Philosophy, 1(1), 3–44.

Kuhn, T. (1970). The Structure of Scientific Revolutions, Chicago University Press, Chicago, Il

Lakoff, R. (1973). The logic of politeness; or, minding your p's and q's. Papers from 9th regional meeting of the Chicago linguistic society: 292-305.

Lischinsky, A. (2011). In Times of Crisis: A Corpus Approach to the Construction of the Financial Crisis Annual Reports. Critical Discourse Studies, 8(3): 153-168

Marlow, M. L. and Giles. H. (2008). ‘‘Who you tink You, talkin Propah?’’: Pidgindemarginalized. Journal of Multicultural Discourses 3: 53-68.

Marlow, M. L. and Giles. H. (2010). ‘‘We won't get ahead speaking like that!’ Expressing and managing language criticism in Hawai'i', Journal of Multilingual and Multicultural Development, 31: 3, 237 — 251

Marwick, A. E. And Boyd, D. (2011). ‘I Tweet Honestly, I Tweet Passionately’: Twitter Users, Context Collapse, And the Imagined Audience. New Media and Society 13: 114-133

Nguyen, D. Q., Billingsley, R., Du, L and Johnson, M. (2015). Improving Topic Models with Latent Feature Word Representations, Transactions of the Association for Computational Linguistics, Vol. 3, pp. 299-313.

Peirce, C. S. (1934). Collected Papers: Volume V. Pragmatism and Pragmaticism. Cambridge, MA, USA: Harvard University Press

Punch, K. F. (2004). Introduction to Social Research, Quantitative and Qualitative Approaches, 2nd Edition, UK

Reddy, M. C., Bardram, J. and Gorman, P. (2010). CSCW Research in Healthcare: Past, Present, and Future, CSCW 2010, February 6–10, 2010, Savannah, Georgia, USA

Seale, C. (2005). New Directions for Critical Internet Health Studies: Representing Cancer Experience on the Web. Sociology of Health & Illness 27(4): 515–540

Sebastiani, F. (2002). Machine Learning in Automated Text Categorization. ACM Computing Surveys, 34(1):1–47

Sinclair, J. (1991). Corpus, concordance, collocation, UK

Sinclair, J. (1995). Collins Cobuild English Dictionary. London: Harper-Collins. 2010; 6:376–394

Socher, R., Alex Perelygin, Jean Wu, Jason Chuang, Christopher Manning, Andrew Ng and Christopher Potts. (2013). Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank Conference on Empirical Methods in Natural Language Processing (EMNLP 2013)

Thornbury, S. (2010). 'What Can a Corpus Tell Us About a Discourse?' In the Routledge Handbook of Corpus Linguistics Edited by Anne O'Keeffe, Michael McCarthy, Routledge, UK

Van Dijk, T. A. (1988). News as discourse. Hillsdale, NJ: Lawrence Erlbaum, 1988.

Van Dijk, T. A. (1990). Discourse may be rule governed and highly structured, or ad hoc and context-bound, which explains variations across different contexts.

Vásquez, C. (2011). Complaints online: The case of TripAdvisor. Journal of Pragmatics 43: 1707- 1717.

Vygotsky, L. S. (1978). Mind in Society, Cambridge, Ma: Harvard University Press

White M, and Dorman S. M, (2001), Receiving Social Support Online: Implications for Health Education, 2001 Dec; 16(6):693-707.

Wittgenstein, L. (1953). Philosophical Investigations. Oxford: Basil Blackwell 1985

Wodak, R. and Meyer, M. (2011). Methods of Critical Discourse Analysis, Sage, London