Creating a Cognitive Behavioural Therapy Chatbot

A review of tools and techniques used to support patients in the mental health space and design of a Chatbot capable of delivering basic cognitive behaviour therapy.

Author:

Daniel Barrow - C1976640

Course: MSc Computing

Supervisor: Alun Preece

School: Computer Science and Informatics

Date: 5 November 2021

Abstract

This paper aims to understand the use of cognitive behavioural therapy (CBT) with patients, how technology is being used to support in its delivery and how these solutions fundamentally work. With this understanding, this paper then aims to design and develop a minimum viable product which utilising natural language processing and text analysis techniques paired with a chatbot to help facilitate the training and therapy of a patient in the basics of CBT.

Contents

Introduction 5
Project Aims and Objectives 5
Background 6
What is CBT and its advantages 6
What are the disadvantages of CBT 7
How could technology solve these problems? 8
iCBT functionality analysis 8
Introducing iCBT Conversational Agents (or Chatbots) 9
Analysis of technology behind iCBT CA's and other mental health solutions 11
Part of Speech (POS) 13
Term Frequency Inverse Document Frequency (TF-IDF) 14
Word Embedding 15
Named-Entity Recognition (NER) 16
Linguistic Inquiry and Word Count (LIWC) 16
n-gram 17
Sentiment Analysis (SA) 19
Rules-Based or Machine Learning Approach, That Is The Question? 19
Conclusion - Defining the Problem 20
Design Approach & Final Product 20
How to achieve this design approach 20
Design Discussion and Limitations 20
Objective 1 –Design Solution Which Understands The Text Inputs Of A Patient In The Context Of CBT 21
Which Technologies Can Achieve This? 21
Defining Project Dataset 22
Design Decisions 22
TF-IDF Code Explanation and Pseudocode 22
Sentiment Analysis Code Explanation and Pseudocode 25
TF-IDF Analysis 26
Sentiment Analysis 30
Characteristic Identified 33
Identifier Rules 36
Final Solution – Pseudocode: 37
Sample output: 40
Objective 1 Conclusion 40

Objective 2 & 3: Design an engaging method of introducing CBT fundamentals to patients and offer meaningful feedback and design an engaging method of introducing CBT fundamentals to patients and guiding them through the steps of CBT. 40
Background Research Design Considerations 40
Design Use-Case 41
Chatbot Transcript Design Considerations 41
Chatbot Design Pseudocode 43
Analysis 48
Manually Categorized Transcripts 48
Systems Categorization Analysis 49
Data Analysis 51
Design Shortfalls 53
Design Improvements 53
System Categorisation Analysis – With Reconfigured Category Rules 54
Data Analysis 56
Discussion 57
Research Limitations 58
Recommendations 58
Conclusion 59
Authors Research Reflection 59
References 61
Appendix 67
Appendix 1 – Text Samples Used In Corpuses 67
Appendix 2 – Full TF-IDF Analysis For ABC Corpuses 70
Appendix 3 – List of 20 most significant words inside each corpus 74
Appendix 4 – POS Tag Meaning (Penn Treebank P.O.S. Tags. 2021) 75
Appendix 5 – Sentiment Analysis of all text samples in each corpus 76
Appendix 6 – Sliding Sentiment Analysis 82
Appendix 7 – Combined tally of all tags collected within all 4 corpuses. 111
Appendix 8 – Text Processing Class Pseudocode 111
Appendix 9 – dbManager Class Pseudocode 113

Introduction

In the United Kingdom's National Health Services (NHS) provision of mental health care is under constant strain, with an ever-growing demand and depending on location a stagnant or declining workforce (British Medical Association 2020). This has led to growing need to find new and novel ways of reducing workload pressures on staff and help priorities more high risks patients over those with more conventional and easily managed conditions such as anxiety, low mood, addiction, and depression (NHS England 2017) (NHS Health Research Authority 2017). These conditions are commonly address with the use of Cognitive Behavioural Therapy (CBT), which is designed to encourage patients to assess their thinking about a specific event, identifying problem thoughts or feelings and dispute these in some way with the long-term aim of reducing specific behaviours during triggering events (Willson et al 2006).

With the above in mind and due to the structured nature of CBT, NHS trusts have begun opting to push patients to participate in more self-guided CBT with little or no practitioner contact using web-based applications (iCBT) (NHS Health Research Authority 2017). iCBT solutions though are cost effective, require much less human resource, and easily deployed to patients, they do have one major criticism, these platforms depending on practitioner support levels during treatment can have very high dropout rates and less successful treatment outcomes (Kuyken 2016) (Holfman et al 2012).

Several new products have come to market recently with a focus on providing direct interactions with patients in the form of conversational agent (CA) driven iCBT to help negate the need to have consistent practitioner support (**Knight 2017, Prochaska et al 2021)**. Though delivery of CBT in this format on its own does not lower dropout rates in comparison to conventional iCBT, if it were paired with some practitioner support dropout rates reductions far exceed that of conventional iCBT. Additionally, this method also touts patients participating in CBT lead iCBT have longer periods of mental stability in comparison to other treatment methods **(Fitzpatrick et al 2017)**.

Chatbots are described as "a software program which interprets and responds to statements made by users in ordinary natural language. It integrates computational linguistics techniques with communication over the internet" (Laduram Vishnoi 2020). Milne-Ives et al (2020) goes on to suggest Chatbots are driven by "machine learning (a statistical method of training machine learning (a statistical means of training models with data so that they can make predictions based on a variety of features) and natural language processing (NLP; the ability to recognize and analyse verbal and written language) to interact with humans via speech or text".

Project Aims and Objectives

There is a large amount of research into the use of web-based applications and chatbots in the CBT space and their efficacies, however there is little research into how these solutions are built and how they function. This project will aim to achieve the following:

1. Develop a working understanding of how CBT is delivered to patients
2. Identify and understand technology and techniques which could be used to develop patient facing CBT product,
3. Create a minimum viable product which could be used to help support mental health patients in their pursuit of CBT and develop framework for further research to take place.

Background

The purpose of this section is to explore the implementation of Cognitive Behavioural Therapy (CBT) in both a digital and nondigital context, paying particular attention to the use of Conversation Agents (CA) with the intention of developing an understanding about how CBT is delivered in either context, its advantages, and disadvantages. In addition, this section will also review what types of technologies are used to assist in Natural Language Processing with the intention of identifying technologies which are most suited to the development of a mental health product.

What is CBT and its advantages

CBT is often d escribed as an umbrella term of different types of cognitive therapies, but all these therapies follow fundamentally the same process of "cognitive assessment, reflection about experience, assignment of tasks or homework to collect more information about the linkage between thought and action, and intentional plans to modify thoughts and behaviour" (Dobson 2013). This is further supported by, Stefan Hofman et al (2017) who describes Cognitive Therapy, a type of CBT, as a way for patients accurately appraising situations they face, by asking them to think more realistically about a situation (Hofman et al 2017). Furthermore, Hofman et al (2017) describes the way a patient interprets a situation influences an emotional or behavioural response.

Dr Willem Kuyken (2016) suggests "Over the course of 46 months, 43 per cent of those who had received CBT had improved, reporting at least a 50 per cent reduction in symptoms of depression, compared with 27 per cent who continued with their usual care alone. Importantly, the study also found that this type of 'high intensity' CBT was a cost-effective treatment from the perspective of the health service." Stefan Hofman et al (2012) carried out a comprehensive meta-study reviewing the effectiveness of CBT in different treatment conditions, his research suggested that CBT depending on its application did indeed report a reasonable effectiveness rate, particularly for those who were experiencing depression or anxiety. "CBT dominates the international guidelines for psychosocial treatments, making it a first-line treatment for many disorders" because of the reasons indicated above (David et al. 2018).

CBT utilises several tools to support a patient, the ABC framework is one of these tools and is typically the first thing taught to patients participating in CBT (Willson et al 2006). ABC stands for Activating Events, Beliefs and Consequences (Willson et al 2006). The idea of this framework is to describe a triggering event, identify the beliefs or thoughts that one has about the triggering event and then describe the consequences of the triggering event and beliefs (Willson et al 2006). By following this method, a patient can "differentiate between your thoughts, feelings, and behaviours, and the trigger event" which will help towards create a grounded understanding of a situations which would then lead to the develop long-term coping strategies (Willson et al 2006). Here are some examples of ABC been applied outlined by (Wilson et al 2006):

ABC	Example 1	Example 2	Example 3
Activating Events	You went to a job interview today.	You accidently stumble head first into a flower arrangement at a party.	Your waiting for your daughter to return home after going to the cinema with friends
Beliefs	You keep thinking it did not go well at all.	You think everyone around you saw and laughed at you	You start to think that she accepts a lift home with a friend who drives recklessly

Consequences	You know feel anxious and now need to drink alcohol to reduce your nerves	You felt so embarrassed you left the party immediately	You are now convinced that she has been in a head on collision, and you start weeping.

Been able to identify and deconstruct a situation then allows the patient to recognise “catastrophic thinking” (Willson et al 2006). This is thinking which takes "a relatively minor negative event and imagining all sorts of disasters” (Willson et al 2006). The main purpose of this framework is to put a patient’s thoughts into perspective, by making them consider less terrifying explanations to an activating event and recognise that a specific way of thinking could be wrong (Willson et al 2006). With these in mind a patient can then consider next steps of the ABC framework, DE, D stands for dispute, a requires a patient to create questions which dispute their way of thinking, E stands for effect, which looks at reviewing a patient’s new thoughts because of disputing their initial thoughts about a situation (Willson et al 2006). The idea of this is to correct a patient thinking and move them away from these catastrophic thoughts (Willson et al 2006). For example

ABC	Example 1	Example 2	Example 3
Dispute	What made you think it went bad? – You do not know; you just feel it went bad What went well? You think the interviewer liked that you had experience in the role before.	Did anyone laugh? – No, you just felt embarrassed. Did anyone try to help you and make sure you were ok? The people at the party wanted to make sure you were ok and unhurt.	Is your daughter responsible? Yes, she is. Do you know which friends drive? Yes, in fact they are responsible.
Effect	You realise that your feelings are making you think negatively, in fact nothing bad really happened.	You realise that your feeling of embarrassment made you think people were laughing at you, in fact they only wanted to make sure you were ok.	You realise that, yes, they are young but its unlikely that her friends will be irresponsible because they are just not like that.

What are the disadvantages of CBT

Though there are significant clinical benefits to CBT, there are also some significant pitfalls. Beach et al (2019) states that there is currently a shortfall of 20,000 mental health professionals in the United Kingdom, coupled with a growing demand of 21% from 2016 to 2019 (roughly 300,000 more people), suggests there is a significant supply and demand issues to provide essential mental health treatment like CBT. This is supported by Waltman et al (2017) and Kumar et al. 2017, who suggests that limited resources is a key issue, specifically surrounding financial and human resource due to high caseloads, which makes it difficult to learn and deliver CBT to patients.

Furthermore, Waltman et al (2017) also indicates that there are possible inconsistencies in the delivery of CBT to patients by clinicians, because of cherry-picking elements of CBT to deliver

treatment, which could lead to suboptimal treatment outcomes. Young et al (1980) supports this stating that this approach could cause a patient to make "erratic progress" during treatment. However, Stirman et al (2013) suggests that all practitioners (27 clinicians) in their study made at least one modification to the the CBT framework, with 65% of alterations relating to "tailoring, integrating cognitive therapy into other approaches, loosening the structured elements of the session, and drift". Stirman et al (2013) states that these changes "improve clinician satisfaction with cognitive therapy, and clinicians believed that some modifications increased the intervention's acceptability to clients".

Considering the above, what efficacies of CBT? Pybis et al (2017) reports in the United Kingdom based on 33,243 research participants, "50.4% of patients made reliable improvement, 3.6% reliably deteriorated and 46.1% did not reliably change" during treatment. However, dropout rates for CBT can vary between 30% and 60% throughout the duration of treatment (Richards et al 2018). The questions remain, why is there such a high number of "reliably deteriorated" and dropout patients? Davis et al (2006), gives a simple answer patient, convenience, stating "the most common reason for refusal to participate was that the patient was not available at the required time" of a session. With the added issue indicated by Pybis et al (2017) suggesting outcomes of patients tend to improve when 8 or more CBT sessions have occurred, this could indicate one of the main issues surrounding the delivery of CBT.

Considering the above, the common problems found using CBT are related to the lack of human and financial resources related due to large demand and low supply of services, inconsistencies in delivery (though justified in some cases) and lack of convenience due to patients needing to physically see a practitioner at a specified time.

How could technology solve these problems?

Webb et al (2017) gives one solution, Internet Delivered Cognitive Behavioural Therapy (iCBT) which provides "web-based access to materials that teach participants the core cognitive skills and behavioural skills", that patients would commonly use during conventional face to face therapy. Several common benefits of using iCBT have been identified these are:

1. Allowing patients to access therapy from the comfort and convenience of their home at any time.
2. Allowing clinicians to be able to offer CBT therapies, easily and quickly with little human resource or financial cost.
3. Allowing patients required to pay for treatment using iCBT will allow them to do so either at a low price or in some cases free of charge depending on the operator.
4. Enabling patients to have less worry about stigma surrounding seeking mental health treatment.

(Donker et al 2015, Web et al 2017, Warmerdam et al 2010)

This suggests that this could be a valid solution to resolving some of the issues surrounding conventional CBT, by allowing clinicians to reduce the issues surrounding convenience of seeking therapy and the reduction human resource and financial costs associated with offering CBT to patients.

iCBT functionality analysis

Radomski et al (2019) analysed ten iCBT programmes used to support patients under the age of 19 and found the following features within them to be important to end users:

1. The ability to monitoring users progress and mood over time.
2. The ability to recommend or require homework or practice activities to be completed before the next teaching module.
3. The ability to personalise reviews of work-based homework before next teaching module.
4. The ability to tailor or personalise content based on users' characteristics or requirements.
5. The ability to offer a web-based therapist, virtual therapist, or a cartoon guide as support.

The points made above is partially reflected in the design of some iCBT products, **Webb et al (2017)** identifies that products like, Beating The Blues and MoodGym which enable patients to complete a number of taught modules used to develop skills pertaining to cognitive and behavioural improvements and encouraging the participation in completing homework relating to what was learned in each module and assessment of the current state of the patient.

Richards et al (2018) has also analysed the effectiveness of SilverCloud a iCBT solution utilised by the United Kingdom National Health Service. SilverCloud is a web-based/mobile app based iCBT solution, which gets patients to participate in one module once a week, the structure and content follows and evidenced based CBT approach which incorporates features pertaining to "self-monitoring, self-control desensitisation, gradual stimulus control, thought recording, behavioural activation, cognitive restructuring, relaxation training and challenging core beliefs" **(Richards et al, 2018)**.

What are the efficacies of iCBT

The question remains however what are the efficacies of iCBT in comparison to standard face to face CBT? Wagner et al 2013 suggests that outcomes during therapy for both iCBT and face to face CBT "yielded no significant differences". However, **Wagner et al (2013)** did note that after a 3 month follow up with their research sample iCBT patient's health remained stable, but in the case of face-to-face CBT the sample showed signs of significant depreciation in mental health.

Webb et al (2017) notes that though outcomes are shown to be better in iCBT, but dropout rates for iCBT are much higher than conventional face-to-face CBT, highlighting two studies suggesting between 57% and 74% of participants of would drop out if they had no support. However**, Richards et al (2012)** suggests that with therapist support, iCBT dropout rates fall drastically to 28% which is close to the lower end of conventional face to face CBT with between 30% to 60% dropout rates. **Richards et al (2012)**, also indicated that dropout rates also fall to 38% if administrative support is offered. On this basis, it would not be unreasonable to hypothesise, that some type of support was offered could help reduce dropout rates.

Based on this brief analysis, products in the iCBT space show to have good outcomes for patients in comparison to conventional face-to-face therapy, and a large proportion of medical providers tend to lean towards using these solutions because of their cost effectiveness. However, dropout rates from programmes incorporating iCBT without the assistance are exceptionally high, in the case of the United Kingdoms National Health Service, where there is currently a 20,000-employment shortfall even providing basic assistance during the duration of a patients iCBT programme may be difficult (Beech, J et al 2019). Which leads to the question of how could support to patients be offered patients during their iCBT programme? What if it were possible to automate this support?

Introducing iCBT Conversational Agents (or Chatbots)

In recent years research into the use of conversational agents more commonly known as chatbots and their use in delivering CBT has indicated there is evidence supporting the use of this technology in the delivery of therapy **(Suganuma et al 2018)**. However, many sources indicate that further

research into the efficacy of this as a solution is needed **(Vaidyam et al 2019, Suganuma et al 2018, Gratzer et al 2019)**.

iCBT Conversational Agents (CA) offer to guide patients through the steps of CBT in relation to their current situation, giving them an opportunity to talk about the feelings and thoughts and then getting them to participate in tasks using a mix of different tools to alter their perspective about that situation **(Knight 2017, Prochaska et al 2021)**.

Though an immerging market, there are currently two products in the iCBT CA, these are WoaBot and Wysa. WoeBot draws from a CBT framework, asks the user daily "how they are doing?", then using a combination of natural language processing and artificial intelligence Woabot is able to understand and appropriately respond to a patient's text inputs and gives guidance/tools the patient can use to manage their thoughts and feelings **(WoeBot NA)(Knight 2017) (Prochaska et al 2021).** Wysa operates in a similar way utilising artificial intelligence…..

Feature Analysis Of CA's

There is limited comprehensive research relating to feature analysis of iCBT CA products with sizable sample sizes. However, this review did identify research by Demirci (2018) who analysed the features Woabot using a small sample of 16 people was able to highlight several pitfalls to this system which could be useful to consider for future improvements Woabot and other iCBT CA products, these were:

1. The system does not explain how it works very well, leading to unrealistic user expectations and user disengagement.
2. The system gives inadequate explanations about the advantages of CBT, leading users to feel it underestimates its effectiveness.
3. The system does develop a repour with users, consequently little empathy could be developed towards the CA.
4. The system focuses only on the problem and not distracting the participant, making it difficult to focus on therapy.
5. The system asks the participant to constantly fill in surveys and carry out monotonous tasks which makes the user less inclined to converse with the CA.
6. The systems responses are repetitive and predictable which causes lower user engagement overtime.

Efficacies of iCBT Chatbots

Kathleen Fitzpatrick et al (2017) reviewed the effectiveness of the use WoeBot in young adults, their findings suggested that participants "experienced a significant reduction in depression" over a two-week period. In addition, they also suggested that this method of delivering CBT could be the solution to the dropout rates issues highlighted in iCBT products, stating "Woebot was associated with a high level of engagement with most individuals using the bot nearly every day and was generally viewed more favourably than the information-only comparison". This is further supported by **Prochaska et al (2021)**, who suggests that "brief interventions", like a daily message to users asking, "how they are doing?", can help minimise potential dropouts and improve patient outcomes.

Reviewing iCBT CA dropout rates further, Richards et al (2018), gives more general detail, suggesting that unsported iCBT CA programmes, still retain a high drop out rate of between 57 and 74%, however, if it was supported with a therapist, this rate drops to 28%, which is considerably lower than conventional CBT treatments (between 30% and 60%).

Inkster et al (2017) had similar results in the review of Wysa effectiveness in patients, suggesting that more engaged users of Wysa had a higher-than-average improvement in mood. In addition, "67.7% of user-provided feedback responses found the app experience helpful and encouraging" **(Inkster et al 2017).**

More generally, Wagner et al (2013), suggests that though patient outcomes are no different to conventional web-based iCBT, patients using iCBT CA's are more likely to remain mentally stable for longer.

Because of the above, it stands to reason that CA delivered CBT could in fact be the next step in iCBT due to its ability to remedy the issues surrounding patient dropout rates (if assisted by a professional) faced by conventional iCBT. Coupled with the benefit of reducing long term depreciation of mental health after treatment and its ability to offer the same scalability and low resources demands of conventional iCBT.

Analysis of technology behind iCBT CA's and other mental health solutions

Chatbots in the mental health space is not new concept, in fact in 1964 using Natural Language Processing (NLP) ELIZA was developed to act as a Rogerian psychotherapist (Vaidyam et al. 2019). Though ELIZA was not able to understand conversations, it was still able to elicit an emotional attachment by its users (Vaidyam et al. 2019). Though technology has developed since then, Natural Language Processing is still widely present in the creation of mental health chatbots as indicated above. Bendig et al (2019) in part supports this, suggesting that the technical implementation of chatbots in mental health setting at a basic level incorporate NLP converting the text into something a computer can understands and then using rules-based sequences which take a user through a predefined narrative based on the chatbots understanding of the text. Bendig et al (2019) then suggests that at a more advanced level chatbots can then use linguistic analysis such as sentiment analysis and 'knowledge bases or models of machine learning to construct their dialogues and AI to generate possible answers and to enhance the conversational proficiency of the bot'.

Natural Language Processing (NLP) is described as a method to allow computers to seamlessly understand human language, this is achieved by extracting meaningful information from human languages (Zhang 2021). Based on the above, it is clear that the use of NLP works as the foundation of most chatbots in the mental health space, but the question remains how is this used in mental health and how could this be helpful for the creation of a chatbot in this field?

Commonly Natural Language Processing has been used in the mental health space to help create systems which can assess patients' mental states. For example, Corcoran et al (2018) has used natural language processing to predict psychosis of patients based on text inputs. This was achieved by utilising Pythons Natural Language Tool Kit (NLTK) to pre-process text by removal of punctuation, tokenizing words (creating a list of words) from the inputted text and then finally lemmatized or inflect the words (breaking words down into their simplest forms). Then based on this, Corcoran et al (2018) then attempted to apply meaning to these words used the Touchstone Applied Science Associates (TASA) corpus which gave each word a semantic vector (A list of similar meaning words, not a necessarily a synonym but something following a similar theme for example cat/dog). Finally, Corcoran et al (2018) then used NLTK's POS-Tag function, Part of Speech tagging to assign a grammatical function definition to the words in relation to its use in a sentence. The text was then fed through a machine learning algorithm which classifies the speech and based on its characteristics would then indicate if a patient had psychosis.

Similarly, Biradar et al (2019) looks to detect depression in social media posts again using NLTK to pre-process the text, removing punctuation, tokenizing and lemmatizing the words. The output of which was then converted into a bag of words (a unique list of words that occurred 800 to 20 times in a post). A Zero Training Vector is used to record the number occurrence a word appears that is within the bag of words list. This information is then fed through to a deep neural network to identify whether a post indicates depression.

Also, Syarif et al (2019) research aims to detect mental disorders in social media and their approach bares some similarities to the two above pieces of research but also some significant differences. Firstly, they used TextBlob a Python Package to make text inputs lowercase and to resolve any typos. They then remove any intelligible words and punctuation and using NLTK and spaCy Python Packages tokenize and lemmatize the inputted text. This information is then passed through a custom lexicon which uses WordNet (to assign meaning and synonyms to each word) and SenticNet an opinion mining and sentiment analysis tool which was then used to detect emotion keywords and sentiment in text.

However, these examples aside, there is little research into how natural language processing or other technologies could be used in the context of delivering patient facing solutions to assist with patient recovery.

Based on the above, NLTK and spaCy have been identified as the primary methods of text processing and consequently merits further analysis. This could be in-part due to the need to use Python as a method to interface with these tools as Glaz et al (2021) suggests that Python was the primary language used by researchers in the creation of 222 mental health related projects.

NLTK and spaCy offers a wide range of language processing libraries and programmes as indicated in ***table 1 and 2***.

Table 1 – NLTK Features (Bird et al 2019)	
Language processing task	**Functionality**
Accessing corpora	standardized interfaces to corpora and lexicons
String processing	tokenizers, sentence tokenizers, stemmers
Collocation discovery	t-test, chi-squared, point-wise mutual information
Part-of-speech tagging	n-gram, backoff, Brill, HMM, TnT
Machine learning	decision tree, maximum entropy, naive Bayes, EM, k-means
Chunking	regular expression, n-gram, named-entity
Parsing	chart, feature-based, unification, probabilistic, dependency
Semantic interpretation	lambda calculus, first-order logic, model checking
Evaluation metrics	precision, recall, agreement coefficients
Probability and estimation	frequency distributions, smoothed probability distributions
Applications	graphical concordancer, parsers, WordNet browser, chatbots
Linguistic fieldwork	manipulate data in SIL Toolbox format

Table 2 – spaCy Features (spaCy 2015)	
NAME	**DESCRIPTION**
Tokenization	Segmenting text into words, punctuations marks etc.
Part-of-speech (POS) **Tagging**	Assigning word types to tokens, like verb or noun.
Dependency Parsing	Assigning syntactic dependency labels, describing the relations between individual tokens, like subject or object.

Lemmatization	Assigning the base forms of words. For example, the lemma of "was" is "be", and the lemma of "rats" is "rat".
Sentence Boundary Detection (SBD)	Finding and segmenting individual sentences.
Named Entity Recognition (NER)	Labelling named "real-world" objects, like persons, companies or locations.
Entity Linking (EL)	Disambiguating textual entities to unique identifiers in a knowledge base.
Similarity	Comparing words, text spans and documents and how similar they are to each other.
Text Classification	Assigning categories or labels to a whole document, or parts of a document.
Rule-based Matching	Finding sequences of tokens based on their texts and linguistic annotations, similar to regular expressions.
Training	Updating and improving a statistical model's predictions.
Serialization	Saving objects to files or byte strings.

Nasser et al (No Date) describes NLTK python library as a tool used for analysing human language data where, spaCy is more advanced and is typically used for commercial applications and typically faster most other NLP libraries. Nasser et al (No Date) goes on to suggest that for the purpose of tokenization, NLTK does in fact perform better than most libraries however, for the part of speech POS) tagging spaCy has been shown to perform significantly better.

Before taking this any further, some context on the steps of NLP is important, NLP can be broken down into two clear stages, the first is text processing and cleaning and the second is providing meaning to the text for computers to use (Manfye Goh 2020, Glaz et al 2021). The first step of text processing/cleaning in NLP is Tokenization, this is the process of breaking down blocks of text into a list of component parts, either by word or by sentence. This information is then Normalized a process which removes all irrelevant information from a dataset, this is achieved by removing stop-words (these are words that hold no meaning and punctuation) and then either lemmatizing or stemming the words in the list (Manfye Goh 2020). Lemmatizing reduces the words inside this list into their simplest forms, for example geese to goose, am and are to be, car, cars, car's and cars' to car (Stanford Edu 2021). Stemming on the other hand removes any suffixes from a word, for example "flying", the suffix would be "ing", the return value would be "fly" (Stanford Edu 2021, Srivastava 2019).

Once tokenizing and normalizing the words has taken place, the second part of NLP comes into play applying meaning to the words, there is several ways to achieve this. Glaz et al (2021) systematic review of research in the mental health space relating to NLP has describes several methods the research provided meaning to text these are, Part of speech (POS), Clinical Text Analysis and Knowledge Extraction System (cTAKES), Term Frequency Inverse Document Frequency (TF-IDF), Word Embedding, Named-Entity Recognition, Linguistic Inquiry and Word Count (LIWC), n-grams and Sentiment Analysis.

Part of Speech (POS)

POS is a method of grammatical tagging a tokenized words with information relating meaning inside the sentence, for example using NLTK, ["I", "am", "really", "sad", "today"] would return, [("I", "PRP"), ("am", "VB"), ("really", "RB"), ("sad", "JJ"), ("today", "NN")]. Ratana et al. 2019 suggests that POS tagging "is useful in terms of generating sentence classification given that the diagnostic semantics reflects common conditions in mental health based on classification taxonomies; it is the symptomatology of a condition that is used by clinicians to determine an appropriate treatment

plan". Meaning that been able to identify grammatical syntax would identify markers which could be used to classify and treat mental health conditions.

Clinical Text Analysis and Knowledge Extraction System (cTAKES)

cTAKES is a method of electronic medical record clinical free-text into a structured dataset, to make identifying critical pieces of information a much quicker process (Savova et al. 2010, Apache cTAKES 2021, Informatics 2019). For example if a "No evidence of schizophrenia" was submitted, the following would be outputted:

Disorder

- Text: Schizophrenia
- Associated Code: XYZ (This is a unique medical code for the condition)
- Certainty: 1 (how certain this categorization is correct)
- Context: Current (is this information current?)
- Relatedness to patient: true (is this related to a patient?)
- Status: negated (The patient does not have schizophrenia)

(Informatics 2019)

Term Frequency Inverse Document Frequency (TF-IDF)

TF-IDF is a form of vectorization (assigning numerical meaning to words) (Chen 2020). This method is broken down into two parts, the first TF, count the occurrences of a term in a block of text across a corpus (Moody 2017, Scott 2019). IDF measures the informativeness of the term by checking the number of times the term has been used across a corpus and then dividing it by the number of corpuses assessed. The results are then multiplied together to find the TD-IDF. This provides a guide to the terms less significance, the higher the number the more significant it is (Moody 2017, Scott 2019). To provide more detail, following Rohit Madan (2019) and Scott (2019) explanation of TF-IDF the following example will explain TD-IDF, if we have 3 journal entries:

- Journal 1: "I am sad today"
- Journal 2: "I am going to be happy today"
- Journal 3: "I am going to the zoo; it will be a great day"

Assuming that all the journals have been tokenized, normalized, and lemmatized and focusing only on Journal 1, the first step would be listing all the words in document one and then list repeated words within the entire corpus, making sure to count each occurrence of these words:

Words	Count
I	3
am	3
sad	1
today	2
going	2

The TF is then calculated using the following equation:

TF = n of repetitions of a word in a journal / x of words in a journal.

Words/Journal	Journal 1	Journal 2	Journal 3
I	0.25	0.16	0.083

am	0.25	0.16	0.083
sad	0.25	0	0
today	0.25	0.16	0
going	0	0.16	0.083

The IDF is then calculated using the following equation:

IDF = Log(number of documents) / (number of documents containing the word))

Please note if the number of documents equals the number of documents containing the work the IDF value equals 0.

Words/Journal	IDF	Journal 1	Journal 2	Journal 3
I	0	0.25	0.16	0.083
am	0	0.25	0.16	0.083
sad	1.1	0.25	0	0
today	0.41	0.25	0.16	0
going	0.41	0	0.16	0.083

Finally, it all needs to be wrapped together using the following equation,

*TF-IDF = TF * IDF*

Please note that if the IDF is 0 the TF-IDF value equals 0.

Words/Journal	Journal 1	Journal 2	Journal 3
I	0	0	0
am	0	0	0
sad	0.28	0	0
today	0.10	0.07	0
going	0	0.07	0.03

As you can see based on this table the words sad and today are more significant words in Journal 1.

Word Embedding

Before discussing Word Embedding it is important to understand what a Vector is in the context of NLP. A Vector is a collection of dimensions which help provide meaning to a word the more dimensions that are available the more information a computer has to understand that word (Codecademy 2020). In the case of Word Embedding, this method takes two words and their vectors with the intention of outlining their similarity by defining their spatial distance from one another (Codecademy 2020). This can be achieved using one or all the following three methods the Manhattan Distance (Also known as Cityblock Distance), Euclidean distance and Cosine Distance (Codecademy 2020). To explain this part, see the following example:

- Word1 = [1,2,3]
- Word2 = [2,4,6]

Manhattan Distance measures the distance defined as a sum difference between both vectors, the closer to 0 the sum is the more similar the vector is, its equation is:

Manhattan Distance(Word1, Word2) = [1-2]+[2-4]+[3-6] = 1 + 2 + 3 = 6

(Codecademy 2020)

Euclidean Distance measures the straight-line distance between the two-word vectors, the closer to 0 the sum is the more similar the vector is, its equation is:

$$\textit{Euclidean Distance(Word1, Word2)} = \sqrt{[1-2]^2+[2-4]^2+[3-6]^2} = \sqrt{2 + 4 + 6} = \sqrt{12} = 3.46$$

(Codecademy 2020)

Cosine Distance measures the angle between the two vectors rather than the line distance, if the value is zero it has no angle thus no similarity, if it is closer to one there is a similarity, the equation is:

Product = 1*2 + 2*4 + 3*6 = 28

Word 1 length = √ (1)^2 + (2)^2 + (3)^2 = 14

Word 2 length = √ (2)^2 + (4)^2 + (6)^2 = 56

Cosine Distance(Word1, Word2) = Product / (Word Length 1 * Word Length 2) = 0.036

(Codecademy 2020, GeekforGeek 2020)

Named-Entity Recognition (NER)

NER is a method of identifying entities within text such as “names of a person, organisation, location, expression of time, quantities, monetary values, percentages, etc” (Li 2018). These entities are defined by using a classifier (a list of words, sentences or phrases with tags applied to) (Li 2018 A). Using Python Programming Tutorials (2018 C) which incorporates NLTK and integrates Stanford Classifier it was possible to create an example of this in practice:

Input Text = Hi my name is Daniel, I am a student at Cardiff University.

Output = [('Hi', 'O'), ('my', 'O'), ('name', 'O'), ('is', 'O'), ('Daniel', 'PERSON'), (',', 'O'), ('I', 'O'), ('am', 'O'), ('a', 'O'), ('student', 'O'), ('at', 'O'), ('Cardiff', 'ORGANIZATION'), ('University', 'ORGANIZATION'), (',', 'O'), ('in', 'O'), ('October', 'O'), ('6th', 'O'), ('I', 'O'), ('will', 'O'), ('have', 'O'), ('my', 'O'), ('deadline', 'O'), ('for', 'O'), ('my', 'O'), ('final', 'O'), ('project', 'O'), (',', 'O'), ('and', 'O'), ('I', 'O'), ('am', 'O'), ('really', 'O'), ('stressed', 'O'), ('.', 'O')]

This classifier has been able to identify a PERSON “Daniel” and an 'ORGANIZATION' Cardiff University. NOTE: O indicates a undefined entity.

Li 2018 A, indicates it is also possible to create classifiers which could identify more custom themes like mental health conditions. This is further supported by Wang et al (2009) who experimented with Named-Entity Recognition to identify 10 types of clinical named entities within medical notes.

Linguistic Inquiry and Word Count (LIWC)

LIWC is a java-based programme and “can reveal the language use associated with a psychological disorder by classifying the input words into linguistic categories (eg, articles, numbers, and pronouns), psychological categories (eg, cognitive, affective, and social), or personal concern categories (eg, death, home, and school)” (Xu et al 2016). Using the LIWC (NA) tool it was possible to assess the following text from Young et al (1994) book, “Reinventing Your Life*”,* the following results are shown.

"I was seven years old when this happened. My farther had a heart attack at work it really hurts me to admit this, but I have only vague memories of him left. And pictures, of course. He was a big and

warm. He hugged me a lot. After he was gone, stand by the window and wait for him to come home. I guess I just couldn't accept that it had happened and it makes me cry thinking about it.

I'll never forget the feeling I had waiting for him by the window. When my husband is away, half the time I'm terrified and half the time I am crying. I feel so alone. When he gets home I am so angry at him for what he put me through. That's the irony of the whole thing. When he finally gets home, I am so mad that I don't even want to see him."

Traditional LIWC Dimension	Your Data
I-Words (I, ME, MY)	11.3
Social Words	8.2
Positive Emotions	1.9
Negative Emotions	5.0
Cognitive Processes	9.4
Summary Variables	**Your Data**
Analytic	12.4
Clout	26.5
Authenticity	56.7
Emotional Tone	1.4

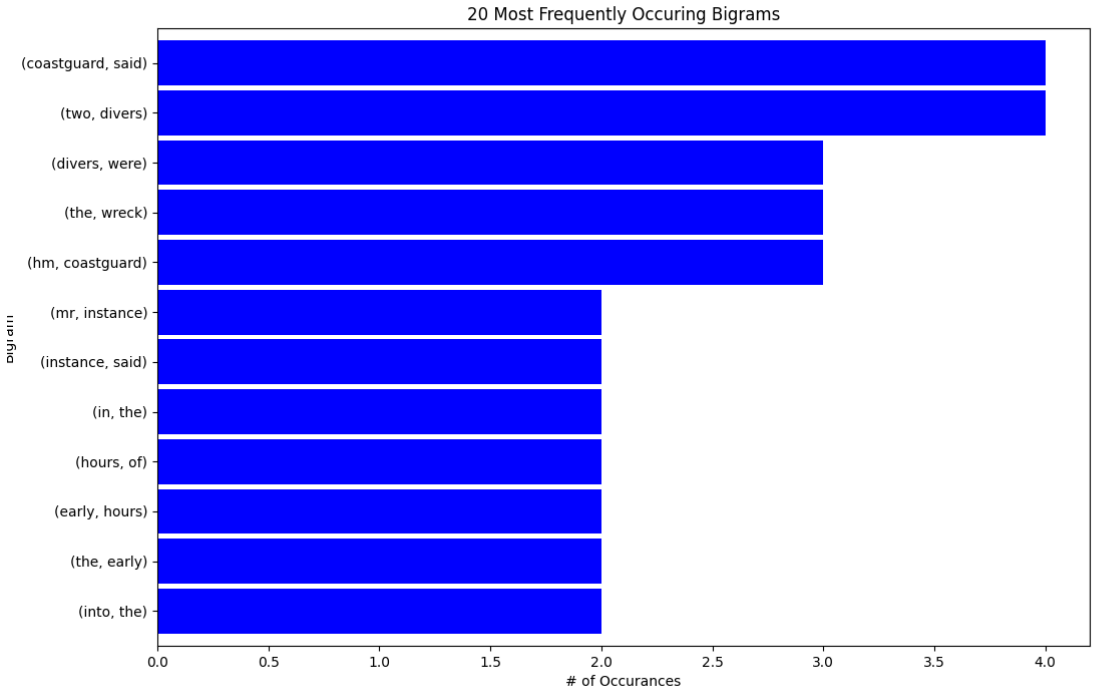
1. The data above provides several dimensions to assess the input text, these are:
2. I-Words words that gives enthuses the author in the first person,
3. Social words which are words that refer to other people,
4. Positive emotion words are those that convey positive emotion
5. Negative emotions are those words which are negative in nature
6. Cognitive processes are those words that reflect how much people are actively thinking about their writing topic (i.e. words like, think, thought, because, wonder, knowledge).
7. Analytics "captures the degree to which people use words that suggest formal, logical, and hierarchical thinking patterns"
8. "Clout refers to the relative social status, confidence, or leadership that people display through their writing or talking"
9. Authenticity When people reveal themselves in an authentic or honest way, they are more personal, humble, and vulnerable.
10. Emotional Tone is a combination of positive and negative emotion scores to assess the overall tone, numbers below 50 suggest a more negative tone.

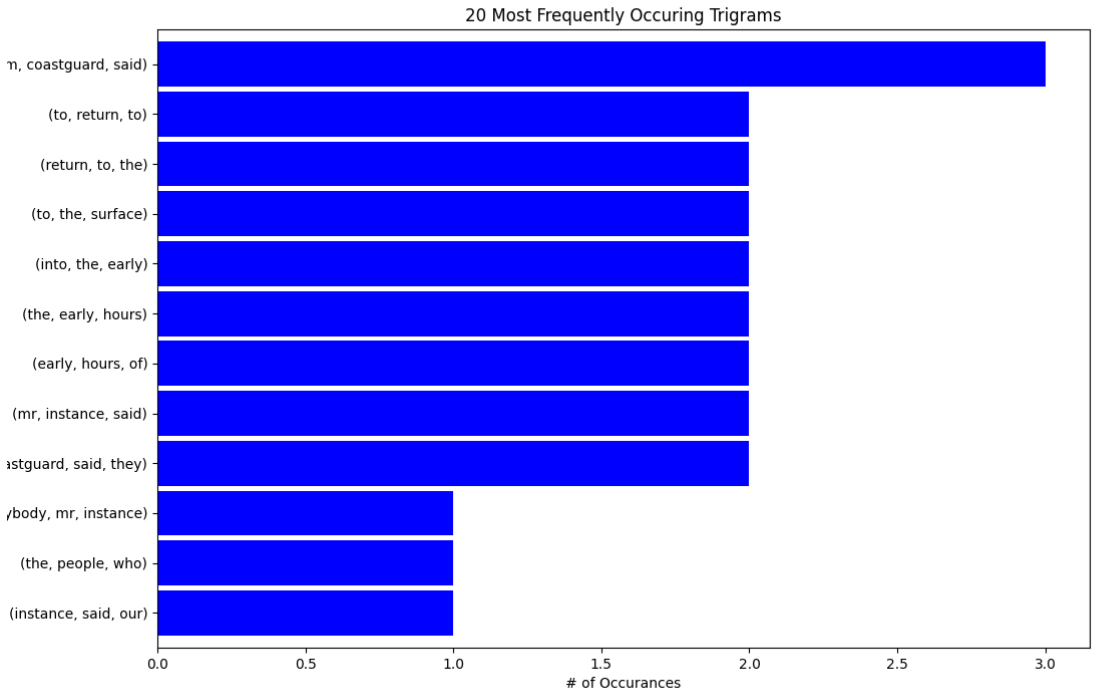
LIWC (NA) A, LIWC Results. 2021.

n-gram

Simply put, an n-gram is a way of allowing a machine to understand the meaning of a word in its context. The n in n-gram describes the number of words assessed at one time, by looking at words that are situated around a target word it is possible to create a more accurate understanding of the word's context. For example, "The staff were not friendly, terrible really", if the target word "friendly" was assessed on its own, it would be seen as positive, however, if the words "not" and "terrible" were assessed friendly in the context would be perceived as negative (Ednalyn C. De Dios 2020, Data Science Dojo 2019).

Providing a more real example, when running a news article through an n-gram the following results were identified:





The above shows, a bi-gram (2 words) and tri-gram (3 word) analysis, both have identified the phrase “coastguard said” most repeated phrase, in addition, it also indicates a context to the coastguard, that they have said something. Another example could be “early hours”, early provides the context that hours is relating to a time of day, the early morning, rather than a block of time (Ednalyn et al 2020).

Sentiment Analysis (SA)

Sentiment Analysis "can be defined as the task of detecting, extracting and classifying opinions on something" (Saad et al 2017). Typically, sentiment analysis determines a score for text based on the following headings positive, negative, and neutral. For example, using NLTK's Sentiment Analysis tool to assess the article mentioned in the N-Gram chapter the overall sentiment score is negative: 0.173, neutral: 0.796 and positive: 0.031 (Munir 2019). It is commonly expected that news articles will be fairly balanced in sentiment, consequently it is more likely to see a result which favours a neutral sentiment, however, due to the nature of the article, its has some negative sentiment embedded within it.

Rules-Based or Machine Learning Approach, That Is The Question?

The above tools provide several solutions which offer varying ways in which to interpret data it is fed, but which solution would be most appropriate to achieve the aims of this project. Before this can be answered it a decision needs to be made as to how these methods should be applied and the context in what needs to be achieved?

As a reminded, the objective of this project is to produce a system which can help a patient successfully learn and use CBT to help improve their mental health. To achieve this a system will need to be able to understand what a user is saying, this can be achieved in one of three ways, using rules approach, machine learning or a combination of both to identify structural elements and categorize text inputs into the appropriate locations (Analytics Vidhya 2020).

A rule-based approach classifies text using handcrafted linguistic rules, these rules are then used to identify relevant sections of text which fit that given category (Analytics Vidhya 2020, Dorash 2017). Whereas, a machine-based classifier, uses an "algorithms that learn to "understand" language without being explicitly programmed" (Dorash 2017). In the first instance it would analyse pre-labelled datasets to train itself how to categorize new data and then it will continue to learn from newly inputted datasets to increase its accuracy (Monkey Learn 2014). As described a hybrid system combines the two, to help provide improved results (Monkey Learn 2014, Dorash 2017).

Each approach has its benefits, a rule-based text categorisation has been described as the most flexible option as it is easy to build/adjust and debug. It is "good at capturing a specific language phenomenon" and does not require a large training corpus to develop, compared to a machine learning approach (Maryna Dorash 2017). However, this approach does have several flaws, it requires a good knowledge of linguistics, all the rules will require manual coding and debugging which can be time-consuming and finally system design can become quite complex (Maryna Dorash 2017).

In comparison, a machine-learning approach is generally more scalable due to its ability to consume new data and modify itself which negates the need to manually adjust code to account of new variables in text samples (Maryna Dorash 2017). However, one the main issue with machine-learning is the need for a large corpus of annotated data for it to learn and without it, its ability to understand and classify text would be hindered (Maryna Dorash 2017).

Conclusion - Defining the Problem

Though CBT is a widely established and effective method of treatment in the mental health space, issues surrounding human resourcing and growing demand in the United Kingdom clearly suggests a need for well-formed technological solution to help ease these pressures. Though there are historical technologies such as iCBT and new emerging solution to this problem such as CBT chatbots, there is little research into how to best apply the technologies and techniques in Natural Language Processing to help deliver mental health treatment products. This could be particularly problematic for the development of new systems because there is no clear right or wrong way to approach the design and construction of such solutions, which could impact on the time of development and the successful construction of new products.

Consequently, this project will aim to address this by create a basic framework to introduce some of the tools outlined above to demonstrate their feasibility and practicalities in this space. This framework will focus on creating a solution which can help guide patients through the fundamentals of CBT. This can then be used to help identifying suitable approaches for future projects and limit the possibility of carrying out design mistakes identified in this project.

Design Approach & Final Product

As a result of clearly defining the problem and developing an understanding of the technology available it is possible to create a more formal design approach. Firstly, it is important to define the objectives that any new features will need to create a clear solution to the problem describe above, these are:

1. Objective 1: To design a system which understands the text inputs of a patient in the context of CBT
2. Objective 2: To design a system which can offer meaningful support and feedback to a patient.
3. Objective 3: To design an engaging method of introducing CBT fundamentals to patients and guiding them through the steps of CBT.

How to achieve this design approach

To achieve the above we must look back to the technological solutions described in the background section, though iCBT has good efficacy rates, a chatbot has shown to have better engagement due to its reduced drop out levels and increased positive outcomes for patients. Consequently, the project will look to pursue a design approach which utilises a chatbot at its core. However, this leads to a question, how will this be implemented?

Design Discussion and Limitations

Though a machine learning driven approach would most likely provide the best outcome for this project, a rule-based approach has been chosen for two reasons, the first a limited amount of time to learn and implement such a solution, the second and more important point, the availability of a large enough dataset to train a statistical model. This is because the data been used in this project is considered highly sensitive consequently, it is generally quite difficult to find and would be difficult to gaining ethical approval for use. This means, it will only be possible to collect data from secondary sources such as textbooks and journals which will make building a large corpus of data quite difficult.

Considering the first aim, the system needs to be able to understand the text inputs from a patient, though several technologies and approaches have been identified that could achieve this as part of the background research for this project, it would be quite difficult to create a solution without

giving a wider consideration to the context in which a patient may be engaging with this system. The background research also discusses the ABC & DE Frameworks and how it is used to introduce patients to CBT. As the ABC (activating events, beliefs, and consequences) framework is the first component of CBT taught and due to time constraints, this project will only consider the delivery of a system which can train and guide a patient in how to use the ABC framework. To achieve this, it is proposed that the chatbot will firstly look to introduce the core concepts using examples and then ask the patient to try applying the framework, while this is happening the chatbot should be verifying what is inputted is correct and if not give examples of each part of the ABC framework. Finally, after the training is complete the chatbot will provide the patient with a journal to use and will continue to coach the patient as needed to ensure that they are accurately using the ABC framework.

In addition to this, it would also be helpful to define other design decisions and limitations, as discussed, time restrictions are a key factor in this project, consequently any programming will utilise Python. This is because Python is language which the author of this project is most proficient in and consequently will be able to develop a working prototype quickly in this language.

Objective 1 –Design Solution Which Understands The Text Inputs Of A Patient In The Context Of CBT

Before any serious work can commence on the chatbot, this project needs to define how a system could interpret the text inputs of a patient and how it intends to describe these interpretations. To achieve this firstly, this project needs to identify characteristics which could be used to categorise text as an activating event, belief or consequence.

Which Technologies Can Achieve This?

With the above in mind to achieve the second aim, this project needs to be able to determine what linguistic characteristics there are in each category type. This section will consider the technologies and techniques discussed as part of the background research and identify what would be most suitable. To achieve this some suitability criterions have been created, these are, the technology or technique must be implementable using Python and provide additional meaning and or help identify unique or important characteristics about a piece of text.

Before proceeding any further, it would be helpful to automatically disregard some of the technology and techniques that do not meet this criterion, cTAKES would be the first, this is because it is only specifically used to convert free text into medical notes and has no practical applications in this project. The second would be LIWC, though it assesses emotional context of a block of text, which would be useful in the context of this project, due to it been a java-based and not a Python-based programme it would not be possible use it in this project. Additionally, Named-Entity Recognition, though this could be useful to develop chatbot logic later on in this project, this technology cannot be used to define characteristics but rather helps detect them once parameters have been set. Finally, Word Embedding, can also be disregarded, as it does not provide an additional meaning to words or sentences but rather looks to assess their similarities only. This leaves us with Part of Speech, TF-IDF, n-grams and Sentiment Analysis as methods to assessing the datasets.

To identify structural context in a sentence there is only really one method to achieve this and that is using NLTK's (a Python library) Part of Speech tagging. This will make it possible to identify the nuances in a sentence, by specifically classifying the word based on its position and context in a sentence. This should make it possible to identify structural themes within each corpus that could be used to help classify text.

Additionally, though not providing structural context, Sentiment Analysis also offers a way of providing another variable to help characterise words, phrases and or sentences inside of a body of text. This information could in the context of this project provide valuable markers which could be used to help categorise said words, phrases and or sentences as an activating event, belief, or consequence. Furthermore, due to the nature of this project, it stands to reason that emotion would play an integral part in the text gathered in each dataset in varying degrees, if it were possible to isolate any themes using sentiment, this could be a powerful method of highlighting parts of text as a particular category.

Finally, to identifying important words or themes of words can be achieved in a few ways, the first could be n-grams, this method would assess a body of text and find word or phrase repetition in that text. TF-IDF on the other hand, is like this to degree in relation to identifying repetition, but rather solely relying on repetition it looks at significant of a word in the context to other bodies of text. This method maybe more helpful as it will filter out words which may not be relevant. However, in both cases, neither of these methods offer context, so will need to be accompanied by other methods of text analysis like part of speech to provide context.

Defining Project Dataset

As previously mentioned, the design approach will only be able to utilise secondary data, this will source from two sources, The first data source is CBT for Dummies written by Branch et al (1998), the second is Reinventing Your Life by Klosko et al (1994). Both textbooks provide real sample therapy transcripts, these transcripts will be broken down into 4 corpuses will be developed, it will contain 12 samples texts for each part of the ABC framework and random sample of 6 Reuters news articles which will be used as a baseline for any data analysis (see appendix 1 for breakdown of all corpuses).

Design Decisions

As a result of the above, the corpuses discussed will be analysed in two ways, the first, a TF-IDF analysis of every sample inside each corpus will be carried out to identify the significance of each word inside each sample. The significance would be established by comparing it against the remaining three corpuses, for example if sample one of the Action Event corpora were analysed, it would be compared against all the samples in the belief, consequence, and random news article corpuses (see appendix 2 for results).

The second would be a sentiment analysis of all the samples in each corpus, this would be done using NLTK's sentiment analysis tool in two ways the first would be to assess the sentiment of each sample (see appendix 5 for results) and the second would be to create a sliding sentiment analysis of each sentence which assesses a range of 4 words at a time to provide a more focused picture of the sample's component parts (see appendix 6 for results).

Both methods of analysis will contain part of speech tagging, this will help correlate structural components of the samples assessed to identify correlations in data which could be used as identifiers for each part of the ABC framework.

TF-IDF Code Explanation and Pseudocode

The following is an exemplification of the TF-IDF code used to assess the significance of the words in the a given corpus, this is achieved using the following steps:

1. The constructor will create several global variables, which will be used throughout the class, to store the target sentence, corpus, untokenized corpus with the target sentence added.

2. The constructor will then call on the removeStopWords method, which will remove all stop words, tokenize and convert all words to lowercase for the target sentence/all corpus entries.
3. The constructor will then call on the TermCount method, in this method will, first record every occurrence of a word, and repetitions will be tallied for both the target sentence and also all entries inside the corpus. This method will then return all words collected with a tally inside a dictionary.
4. The constructor will then call on the TermFrequency method taking TC as a parameter, this method loops through each untokenizedCorpus entry, if a word in the TC dictionary exists in this list, it will calculate the term frequency, by counting the occurrences and then dividing it by the total length of the entry sentence. If the value is greater than 0 it is recorded in a new dictionary called TF, under the word and then the entry sentence, else it is marked as 0. The TF dictionary is then returned.
5. The constructor will then call on the inverseDocumentFrequency method taking TC and TF as a parameter and creates a new dictionary called idf and copies the values of tf into it. It will then loop through each word in IDF, checking if the word is equals to the length of the global variable corpus or not, if it is the IDF of that word is marked as 0 in the idf dictionary, else it equals math.log(length of corpus / tc[word] number of occurrence). It then returns the dictionary idf.
6. Finally the constructor returns the value of the method calculate which takes the parameter of idf. This method creates a new dictionary called output. It then loops through the parameter idf creating the value of word which is then added as a new key in the output dictionary. It then loops again through the global variable untokenizedCorpus creating the variable of sentence. . If the parameter idf[word][“idf”] has a value equals or less than zero, an empty string is added to output[word][sentence]. Else idf[word][sentence]["TF"] *idf[word]["IDF"] is added to output[word][sentence]. Output is then returned which contains the words analysed, the sentences which contain that word and the TF-IDF value.

```
Class tfidf:

    Def __init__( self, sentence, corpus):

        Self.sentence = sentence

        Self.corpus = []

        Self.untokenizedcorpus = self.copus

        Self.untokenizedcorpus.insert(0, self.sentence)

        Self.removeStopWords(self.sentenece, self.corpus)

        TC = self.termCount()

        TC = self.termfrequency

        IDF = self.inverseDocumentFrequency(TF, TC)

        Return self.calculate(IDF)


    Def removeStopWords(self ,sentence, corpus)

        stopWords = list(NLTK stopwords)

        grammarRemoval = regex.compile(rules for removing all punctuation & symbols)
```

self.sentence = list comprehension, loop through return value NLTK Word Tokenize using sentence as a parameter appending to list if it is not in stopWords.

Self.sentence = list comprehension, loop through self.sentence, if word does not contain items in grammarRemoval.

Self.sentence = list comprehension, loop through sentence, making each word lowercase.

For each entry in corpus:

Temp = repeat all list comprehension above for each corpus entry.

Self.corpus.append(temp)

Def termCount(self):

Tc = {}

aTerm = {}

for word in self.sentence:

if word is in tc:

tc[word] = tc[word]+1

else:

tc[word] = 1

for entry in self.corpus:

for word in entry:

if word is in tc:

tc[word] = term[word]+1

else:

tc[word] = 1

return tc

Def termFrequency(self, tc)

Tf = {}

For word in tc:

Tf[word] = {}

For sentence in self.untokenizedcorpus:

Tf[word][sentence] = {}

If word in sentence:

Calculation = occurrences of word in sentence / length of sentence

If calculation less than 0:

Tf[word][sentence]["TF"] = calculation

Else:

Tf[word][sentence]["TF"] = 0

Else:

```
                    Tf[word][sentence][“TF”] = 0

        Return tf
Def inverseDocumentFrequency(self ,tf, tc)
        Idf = tf
        For word in idf:
                If length of self.corpus == idf[word]
                        Idf[word][“IDF”] = 0
                Else:
                        Idf[word][“IDF] = math.log(length of corpus / tc[word] number of occurrence)
        Return idf

Def calculate(self ,idf):
        Output = {}
        For word in idf:
                Output[word] = {}
                For sentence in untokenizedCorpus:
                        If idf[word][“idf”] <= 0:
                                Output[word][sentence] =  ””
                        Else:
                                output[word][sentence] = idf[word][sentence]["TF"] *idf[word]["IDF"]
        Return output
```

Sentiment Analysis Code Explanation and Pseudocode

The following is an exemplification of the Sentiment Analysis code used to assess the sentiment of 4-word phrases for each item inside every corpus, the following explains this in more detail:

1. The constructor takes a complete list of transcripts and loops through each entry named sentence. Inside this loop the constructor will add the sentence as a key to the output dictionary with a empty dictionary as its value pair. The string “SA” will then be added as a new key under sentence, with the return value of the method sentimentAnalysis taking the sentence value converted into a list as a parameter. Then the constructor will return the value of Slidingsentimentanalsysis which takes two parameters sentence and the maximum range it should assess.
2. The sentimentAnalysis takes list containing each word of a sentence, this is then passed into NLTK Sentiment Analysis Tool and then returns the scores.
3. The sentimentAnalysis takes sentence and then tokenizes and lemmatizes it. Using this it them applies pos tags to the output dictionary. Then using a while loop, the takes the

lemmatized sentence and takes a index range, starting from the beginning of the sentence at a range length specified using the range parameter.

```
Class sentimentAnalyser:
    Def __int__( self ,corpus):
        Output = {}
        For sentence in corpus:
            Output[sentence] = {}
            Output[sentence]["SA"] = sentimentAnalysis(sentences.split())
            Ssa = Slidingsetimentanalysis(sentence, 4)
        Return ssa
    Def sentimentAnalysis(self , text):
        SA = nltk. SentimentIntensityAnalyzer()
        Return sa.polarityScorer(text)
    Def slidingsenitmentanalysis(self ,sentence, range):
        Sentence = nltk.tokenize(sentence.lower())
        Sentence = nltk.lemmentize(sentence)
        Output[sentence]["POS"] = nltk.pos_tag(sentence)
        Count = 0
        While count + range <= len(sentence):
            Text = " ".join(sentence[count: range])
            Output[sentence][Text] = sentimentAnalysis(text)
```

TF-IDF Analysis

Using the analysis tools developed above, it was possible to create a large amount data which will be used to extrapolate characteristics that can be used to help identify sentences / phrases which fit into the ABC framework. Inside each table the word identified as significant will be shown, alongside its significance score, POS tag to identify its word type, word category which is used to describe the meaning of the word, and its sentiment values to help describe if the word is negative, positive or neutral.

TF-IDF Analysis – Activating Events

Using the data collected in appendix 2, the below table was created outlining the top 20 most significant words in the Action Event corpus. Using this table, it was possible to identify three key themes, the first, action words (accounting for 50% of words). The second, relates to words which identify a person (accounting for 30% of all words). The third, supporting words accounts for 20% of all words.

TF-IDF Analysis of Activating Events Text				Word Sentiment		
Word	**Total Significance**	**Word POS Tag**	**Word Type**	**Negative**	**Neutral**	**Positive**
husband	0.2111	NN	Identifier/Person	0.0	1.0	0.0

left	0.7223	NN	Action	0.0	1.0	0.0
wife	0.6933	NN	Identifier/Person	0.0	1.0	0.0
goes	0.5198	VBZ	Action	0.0	1.0	0.0
partner	0.4731	NN	Identifier/Person	0.0	1.0	0.0
separating	0.4731	VBG	Action	0.0	1.0	0.0
talking	0.3574	VBG	Action	0.0	1.0	0.0
called	0.264	VBN	Action	0.0	1.0	0.0
she	0.2163	PRP	Identifier/Person	0.0	1.0	0.0
my	0.2111	PRP$	Identifier/Person	0.0	1.0	0.0
keeps	0.1999	NNS	Action	0.0	1.0	0.0
worried	0.1798	JJ	Action	1.0	0.0	0.0
back	0.1798	RB	Support	0.0	1.0	0.0
dreading	0.1797	VBG	Action	1.0	0.0	0.0
away	0.1577	RB	Support	0.0	1.0	0.0
never	0.1577	RB	Support	0.0	1.0	0.0
come	0.1577	VB	Action	0.0	1.0	0.0
he	0.1577	PRP	Identifier/Person	0.0	1.0	0.0
party	0.1557	NN	Support	0.0	0.0	1.0
gone	0.1459	VBN	Action	0.0	1.0	0.0
			Total	**2**	**17**	**1**

Reviewing all the words identified as significant in this corpus, the below table was created and highlights 49.49% of these words are nouns or pronouns (in green) and 30.30% of these are verbs (in red), 14.14% of these words are adverbs or adjectives (in yellow), the other words relate to structural words which hold no meaning.

POS Tag	Count	%
NN	38	38.38%
VBG	12	12.12%
VBN	8	8.08%
JJ	7	7.07%
RB	7	7.07%
VBD	6	6.06%
NNS	4	4.04%
PRP	4	4.04%
PRP$	3	3.03%
IN	2	2.02%
VBZ	2	2.02%
CD	1	1.01%
DT	1	1.01%
MD	1	1.01%
VB	1	1.01%
WRB	1	1.01%
CC	1	1.01%

TF-IDF Analysis – Belief

In a similar fashion to the TF-IDF Activating Events analysis, the top 20 most significant words of the Belief corpus were selected. Using this table, it was possible to identify three themes, the first

relates words describing feelings or thoughts (40% of words). The second, relates to words which gives context to a thought or feeling sentence for example "I felt like X", "I thought x?" or "I wondered if?" (20% of words). The third, relates to words which identify a person (10% of words). The other words hold no meaning without context thus cannot be categorised.

TF-IDF Analysis of Beliefs Text				Word Sentiment		
Word	Total Significance	Word POS Tag	Word Type	Negative	Neutral	Positive
attracted	0.2803	VBN	Thought or Feeling	0.0	0.0	1.0
totally	0.2457	RB		0.0	1.0	0.0
nothing	0.2457	NN	Thought or Feeling	0.0	1.0	0.0
felt	0.2398	NN	Context	0.0	1.0	0.0
woman	0.2398	NN	Identifier	0.0	1.0	0.0
happening	0.2325	VBG		0.0	1.0	0.0
lousy	0.2325	NN	Thought or Feeling	1.0	0.0	0.0
wonder	0.2225	NN	Context	0.0	1.0	0.0
wants	0.2225	VBZ	Context	0.0	1.0	0.0
cope	0.2181	NN	Thought or Feeling	0.0	1.0	0.0
function	0.2181	NN	Thought or Feeling	0.0	1.0	0.0
might	0.2127	MD		0.0	1.0	0.0
happen	0.2127	VB		0.0	1.0	0.0
she	0.1969	PRP	Identifier	0.0	1.0	0.0
thought	0.1839	NN	Context	0.0	1.0	0.0
bad	0.1839	JJ	Thought or Feeling	1.0	0.0	0.0
storage	0.1839	NN		0.0	1.0	0.0
right	0.1737	NN	Thought or Feeling	0.0	1.0	0.0
treat	0.1737	NN	Thought or Feeling	0.0	0.0	1.0
without	0.1648	IN		0.0	1.0	0.0
			Total	2	16	2

Reviewing all the words identified as significant in this corpus, the below table was created and highlights 38.27% of these words are nouns or pronouns (in green) and 24.69% of these are verbs (in red), 13.86% of these words are adverbs or adjectives (in yellow), the other words relate to structural words which hold no meaning.

POS Tag	Count	%
NN	28	34.57%
VB	8	9.88%
RB	7	8.64%
IN	6	7.41%
VBG	6	7.41%
JJ	5	6.17%
VBN	5	6.17%
MD	4	4.94%
CD	3	3.70%
WRB	3	3.70%
PRP	2	2.47%
DT	1	1.23%
NNS	1	1.23%

RBR	1	1.23%
VBZ	1	1.23%

TF-IDF Analysis – Consequence Events

Using the data collected in appendix 2, the below table was created outlining the top 20 most significant words in the Consequence corpus. Using this table, it was possible to identify three key themes, the first, words related to actions, account for 15% of words. The second, words relating to feelings or thoughts, account for 37.5% of all words. The third, words relating to identifiers/a person, account for 20% of all words.

TF-IDF Analysis of Consequence Text				Word Sentiment		
Word	**Total Significance**	**Word POS Tag**	**Word Type**	**Negative**	**Neutral**	**Positive**
driving	0.578	VBG	Action	0.0	1.0	0.0
extremely	0.4325	RB		0.0	1.0	0.0
unsafe	0.4319	JJ	Feeling or Thought	0.0	1.0	0.0
scared	0.3846	VBN	Feeling or Thought	1.0	0.0	0.0
feel	0.3328	NN	Feeling or Thought	0.0	1.0	0.0
guess	0.3145	NN	Feeling or Thought	0.0	1.0	0.0
believe	0.3145	VB	Feeling or Thought	0.0	1.0	0.0
could	0.2727	MD		0.0	1.0	0.0
die	0.2695	NN	Feeling or Thought	1.0	0.0	0.0
made	0.2528	VBN	Action	0.0	1.0	0.0
around	0.2528	IN		0.0	1.0	0.0
he	0.2458	PRP		0.0	1.0	0.0
mad	0.2444	NN	Feeling or Thought	1.0	0.0	0.0
telling	0.2444	VBG	Action	0.0	1.0	0.0
without	0.2375	IN		0.0	1.0	0.0
always	0.2162	RB		0.0	1.0	0.0
verge	0.2162	NN		0.0	1.0	0.0
			Total	**3**	**14**	**0**

A notable observation is that the significance scores in consequences appear to be much higher than that the other parts of the ABC framework, this most likely is because a lot of these words match those shown in the random news article corpus seen in appendix 3.

Reviewing all the words identified as significant in this corpus, the below table was created and highlights 41.67% of these words are nouns or pronouns (in green) and 27.78% of these are verbs (in red), 13.89% of these words are adverbs or adjectives (in yellow), the other words relate to structural words which hold no meaning.

POS Tag	Count	%
NN	20	27.78%
RB	7	9.72%
IN	6	8.33%
VBG	6	8.33%
VBN	6	8.33%

NNS	4	5.56%
PRP	4	5.56%
JJ	3	4.17%
MD	3	4.17%
VB	3	4.17%
VBD	3	4.17%
PRP$	2	2.78%
VBZ	2	2.78%
CC	1	1.39%
CD	1	1.39%

TF-IDF Analysis Conclusion

Considering the above analysis, it is apparent there are characteristics which can be used to identify themes within each part of the ABC framework. Activating events mainly uses verbs to describe the act of doing something. This is then followed by the use nouns and pronouns which are used to identify a individual in either the first or third person. Beliefs focus on the use of nouns, verbs and adverts to describe a thought or feeling. This is followed by the used of nouns and verbs which describe the action of thinking or feeling and has some use of pronouns or nouns to identify a person. Finally, Consequences mainly utilises, nouns, verbs and adjectives to describe a feeling or thought, this is supported by the use of verbs are also used to describe an action. There does not appear to be sufficient differences in sentiment to identify any characteristics at this stage which would help categories the ABC corpus using sentiment.

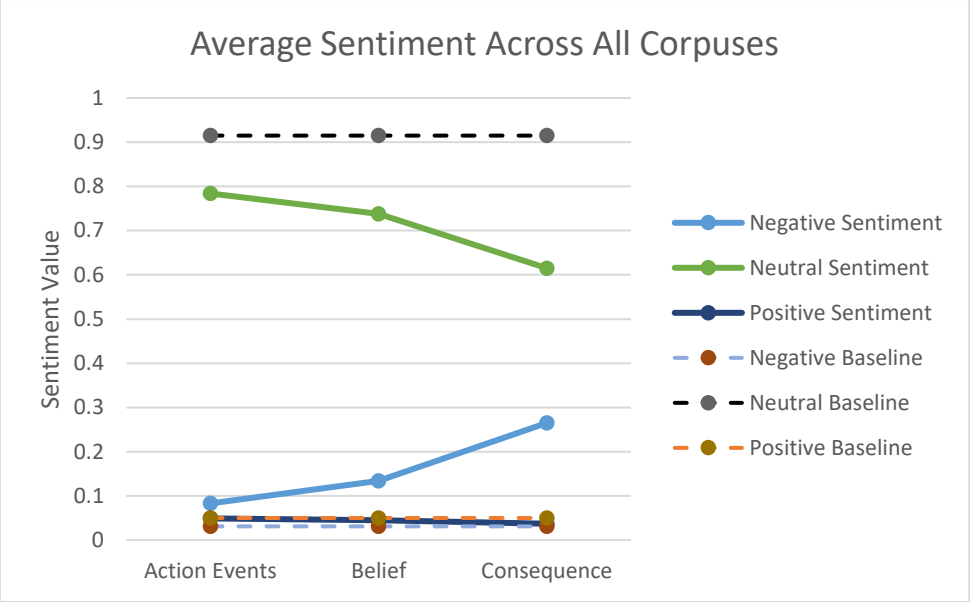
Sentiment Analysis

The sentiment of the corpuses has been analysed in two ways, the first is a overall analysis, looking at each corpus at a high level, the second, reviews each of the ABC corpuses using a sliding scale to identify phrases of interest and to provide more focused analysis.

Sentiment Analysis – Overview

The following table shows the average sentiment in each corpus, Activating Events tend to hold higher levels of neutral in sentiment, with little to no positive or negative indicators. However, as the analysis progresses further through the ABC framework negative sentiment increases, neutral sentiment decreases, and positive sentiment decreases also but at a slower rate.

Note: Random neutral news articles have been included to show that the NLTK sentiment analysis tool is balanced.



Corpus Name	Negative Sentiment	Neutral Sentiment	Positive Sentiment
Action Events	0.083	0.784	0.049
Belief	0.134	0.738	0.045
Consequence	0.265	0.615	0.037
Random News Articles Baseline	0.031	0.915	0.050

Sentiment Analysis – Sliding Analysis

A sliding sentiment analysis was produced which reviewed 4 word extracts (see appendix 6), in this data it was possible to see peaks in negative sentiment in all corpuses assed. Consequently, the table below was produced which filters out all but negative sentiment extracts and only shows the peaks in negative sentiment within each of the sample inside each corpus.

Extracted Action Event Phrases	POS Tags	Negative Sentiment	Neutral Sentiment	Positive Sentiment
I am worried he	['NN', 'VBP', 'VBN', 'PRP']	0.524	0.476	0
My wife keeps doubting	['PRP$', 'NN', 'VB', 'NN']	0.444	0.556	0
me, she's especially suspicious	['PRP', ',', 'PRP', 'VBZ', 'RB', 'JJ']	0.482	0.518	0
I have trouble trusting	['NNS', 'VBP', 'NN', 'NN']	0.422	0.156	0.422
bulb was dead. I	['JJ', 'NN', 'JJ', '.', 'NN']	0.683	0.317	0
dead. I was standing	['JJ', '.', 'JJ', 'VBP', 'VBG']	0.683	0.317	0
a sudden I broke	['DT', 'JJ', 'NN', 'VBD']	0.737	0.263	0
I was petrified. It	['NN', 'VBP', 'VBN', '.', 'PRP']	0.636	0.364	0
why I was scared.	['WRB', 'JJ', 'NN', 'VBD', '.']	0.592	0.408	0
a lot of trouble.	['DT', 'NN', 'IN', 'NN', '.']	0.574	0.426	0
would rot in hell.	['MD', 'VB', 'IN', 'NN', '.']	0.605	0.395	0
I have been dreading	['NNS', 'VBP', 'VBN', 'VBG']	0.63	0.37	0

Extracted Belief Phrases	POS Tags	Negative Sentiment	Neutral Sentiment	Positive Sentiment
think he may die	['NN', 'PRP', 'MD', 'VB']	0.565	0.435	0
she used to leave	['PRP', 'VBD', 'TO', 'VB']	0.286	0.714	0
I was totally alone.	['NN', 'VBP', 'RB', 'RB', '.']	0.534	0.466	0
I was to tired	['NN', 'VBP', 'TO', 'VB']	0.592	0.408	0
bad feelings would catch	['JJ', 'NN', 'MD', 'VB']	0.538	0.462	0
now! She has no	['RB', '.', 'PRP', 'VBD', 'DT']	0.454	0.546	0
She has no right	['PRP', 'VBD', 'DT', 'NN']	0.423	0.577	0
be a big trick,	['VB', 'DT', 'JJ', 'NN', ',']	0.375	0.625	0
I never really loved	['NN', 'RB', 'RB', 'VBD']	0.627	0.373	0
I thought something bad	['NNS', 'VBD', 'NN', 'JJ']	0.636	0.364	0
such a lousy person.	['JJ', 'DT', 'JJ', 'NN', '.']	0.636	0.364	0
I was worrying I	['NN', 'VBP', 'VBG', 'NN']	0.706	0.294	0
act like a jerk.	['NN', 'IN', 'DT', 'NN', '.']	0.407	0.169	0.424

Extracted Consequence Phrases	POS Tags	Negative Sentiment	Neutral Sentiment	Positive Sentiment
this terrifies me, I	['DT', 'VBZ', 'PRP', ',', 'VB']	0.643	0.357	0
I cry a lot	['NN', 'VBP', 'DT', 'NN']	0.756	0.244	0
I am so angry	['NN', 'VBP', 'RB', 'JJ']	0.642	0.358	0
I am worried that	['NN', 'VBP', 'VBN', 'IN']	0.524	0.476	0
am worried that Francine	['VBP', 'JJ', 'IN', 'NN']	0.423	0.577	0
makes me so lonely.	['VB', 'PRP', 'RB', 'RB', '.']	0.519	0.481	0
I felt totally crushed.	['NN', 'VBD', 'RB', 'VBN', '.']	0.607	0.393	0
a rage I was	['DT', 'NN', 'NN', 'VBP']	0.783	0.217	0
I would die without	['NN', 'MD', 'VB', 'IN']	0.661	0.339	0
I am really mad	['NN', 'VBP', 'RB', 'JJ']	0.636	0.364	0
started shaking her. I	['VBD', 'VBG', 'PRP$', '.', 'NN']	0.459	0.541	0
I really scared her!	['NNS', 'RB', 'VBD', 'PRP', '.']	0.635	0.365	0
I was extremely scared	['JJ', 'VBP', 'RB', 'JJ']	0.615	0.385	0
extremely scared to leave.	['RB', 'VBN', 'TO', 'VB', '.']	0.7	0.3	0
the verge of tears.	['DT', 'NN', 'IN', 'NN', '.']	0.388	0.612	0

Using these tables, it is possible to see a structural theme has been identified, most of these samples contain the following structure, a pronoun (like I, he, she, me) and or noun which then is followed by a series of verbs which have different themes across the ABC workflow.

In an Action Event, these verbs typically start with a non-3rd person singular present (VBP or VBZ) with words like “was”, “been” or “have”, making the sentence context present tense. It then uses a series of other verbs which are either, as a past participle (VBN) with words like worried, “petrified” or “been”, or as a gerund/present participle (VBG) with words like “standing” or “dreading” to describe an action or feeling.

Consequences are similar, non-3rd person singular present (VBP) with words like “was”, “been” or “have”, making the sentence context present tense. It then uses a combination of verbs, nouns, adjectives, or adverts to describe feelings.

Beliefs use verbs in the past tense (VBD) with words like “used”, “has” or “thought” or a modal (MD) with words like “may” or “would”. It then uses a series of other verbs like Action Events to describe a thought or feeling.

Its worth noting that some of the samples identified above do not completely conform to the structural themes that have been identified for each part of the ABC framework. Consequently, if a system was developed which filtered out these phrases using the identified structure, it may disregard some of the phrases.

Characteristic Identified

Using the information highlighted above it has been possible to develop a working understanding of the language used and structural characteristics within each part of the ABC framework. The below, will help more clearly define these components and will help develop a series of functional requirements which can be used to design a system to identify ABC characteristics.

Characteristics Identified – Action Events

When visually observing the sentences, without the use of data analysis it appears each sentence start with a personal identifier for example I, me, he, she or a name followed closely by a word which describes an action.

Reviewing the analysis above it is possible to gather a list of characteristics which could be used to identify an Action Event inside a body of text. The first is that there is a theme of language used in this corpus, which is focused on an individual or person/’s for example I, he, me, she, husband, wife etc and words that suggest action for example left, separate, gone, talking, called etc. The second is that these sentences have a highest level of neutral sentiment and the lowest negative sentiment out of the ABC corpuses. The third, is its structure commonly follows the following format a pronoun (PRP$/PRP) and or noun (NN) to identifying an individual or person/’s, a verb (VBP or VBZ) maybe used to outline the sentence in present tense and then descriptive words typically verbs (VBN or VBG), but can also include nouns (NN), adjectives (JJ) or adverbs (RB).

Pairing the data collected from the sliding sentiment analysis and TF-IDF analysis together and then filtering it to remove all but the neutral sentiment (due to action events been shown as mainly neutral). It was possible to extrapolate many extracts that fit the structural theme indicated above which would further evidence that these observed characteristics are correct. Here are some examples of the extrapolated extracts:

ID	TF-IDF Keywords	Extracts	POS TAGS	Negative Sentiment	Neutral Sentiment	Positive Sentiment
1	My, Husband, Gone	My husband has gone	['PRP$', 'NN', 'NN', 'VBN']	0	1	0
2	My, Wife, Goes	My wife goes out.	['PRP$', 'NN', 'VB', 'RP', '.']	0	1	0
3	My	My dad used to	['PRP$', 'NN', 'VBN', 'TO']	0	1	0

4	My, partner, talking	My partner is talking	['PRP$', 'NN', 'VBZ', 'VBG']	0	1	0
5	My, Husband	When my husband is	['WRB', 'PRP$', 'NN', 'VBZ']	0	1	0

In every example above you can see that identifying word (in red) is present, commonly followed by a descriptive word or action (in blue) and in between these occasionally you will find verbs which set the tense of the extract (in purple).

NOTE: Not all extracts fit these criteria, but a large proportion do.

Characteristics Identified – Beliefs

When visually observing the sentences, without the use of data analysis it appears each sentence has one or more of the following:

1. A sentence may with a synonym to "felt" for example:
    a. "I felt that Greg was attracted to this other woman"
    b. "I felt like someone who was running and running, and when I was too tired to run anymore, the bad feelings would catch up with me."
2. A sentence may start with a synonym to "think" for example:
    a. "I just thought it was happening because I was such a lousy person."
    b. "I thought something bad might happen to me if I left the storage cupboard."
    c. "I wonder if she's the one who wants to be with other men."
3. A sentence may have a negative sentiment for the entire sentence between 0.1 and 0.3

Using the data collected, it is possible to identify characteristics which could be used to highlight beliefs in a body of text. The first identifiers are the use of words which prompt a noun or verb which describes a feeling or thought for example felt, thought, think, feel, or wonder etc. This would then be led by a verbs or nouns which describes the feeling or thought for example, lousy, cope, bad, nothing, attracted etc. In addition, words which identify a person are also present, for example "woman" or "she".

The second identifier is the sentiment level, the belief corpus, has a notable increase in negative sentiment and decrease in neutral sentiment placing it firmly in between Action and Consequence in sentiment.

The third is its structural components, as mentioned above a belief sentence would typically start with a following format a pronoun (PRP$/PRP) and or noun (NN) used to identify an individual, for example I, she, wife etc. This maybe then followed by a verbs or modals to describe a past tense (VBD or MD), then descriptive words typically verbs (VBN or VBG), but can also include nouns (NN), adjectives (JJ) or adverbs (RB) are used to describe the feeling or thought.

In a similar fashion to Action Events, the pairing the results of sentiment analysis and TF-IDF was carried out, filtering out all negative and positive sentiment. A similar structure as identified above was found in a large amount of the samples.

ID	TF-IDF Keywords	Extracts	POS TAGS	Negative Sentiment	Neutral Sentiment	Positive Sentiment
1	nothing	I would have nothing	['NN', 'MD', 'VB', 'NN']	0	1	0

2	cope	I just couldn't cope	['NN', 'RB', 'MD', 'RB', 'VB']	0	1	0
3	felt	I felt that Greg	['NN', 'VBD', 'IN', 'NN']	0	1	0
4	she	She should have called	['PRP', 'MD', 'VB', 'VBN']	0	1	0
5	thought	I just thought it	['NN', 'RB', 'VBD', 'PRP']	0	1	0

In every example above you can see that a word identifying a person (in red) is present, commonly followed by a word responsible for setting the tense of the sentence (in blue) and then after this typically a word used to describe a thought or a feeling (in purple), the above data consequently supports the assessment made by visual observations of the corpus.

NOTE: Not all extracts fit these criteria, but a large proportion do.

Characteristics Identified – Consequences

When visually observing the sentences, without the use of data analysis it appears each sentence focuses on how the author of the sentence feels. This means that there appears to be a much high level of negative sentiment that any other corpus. There does not appear to be any discernible structural elements which could be used to identify a sentence which falls under the consequence category.

Using the above analysis, Consequences can be seen to be different than Action Events and Beliefs, the first important identifier is that the samples in this corpus tend to have a much higher negative sentiment score than the other corpuses, this is most likely due to the use of strong language used to describe feelings as a consequence of an Action Event and the Beliefs surrounding it for example "I am so angry" or "I am worried that", the words angry and worried are highly negative sentiment words hence a score of 0.642 and 0.524 for these samples. Structurally, the sentence structure tends to include sole focus on the author, with a noun or pronoun used (like I or me) followed closely by a feeling or action in the form of a verb (like worried, felt, angry, shaking, die). There does appear to be a higher density of adjectives in the dataset than in the other corpuses. However, there is a deviation to this structure in some sample instances.

In a similar fashion to Action Events, the pairing the results of sentiment analysis and TF-IDF was carried out, filtering out all negative and positive sentiment. A similar structure as identified above was found in a large amount of the samples.

ID	TF-IDF Keywords	Extracts	POS TAGS	Negative Sentiment	Neutral Sentiment	Positive Sentiment
1	Driving	She is driving me crazy.	['PRP', 'VBZ', 'VBG', 'PRP', 'JJ', '.']	0.375	0.625	0
2	Extremely, Scared	I was extremely scared	['JJ', 'VBP', 'RB', 'JJ']	0.615	0.385	0
3	Believe, Die	believe I would die	['NN', 'NN', 'MD', 'VB']	0.661	0.339	0
4	scared	I really scared her!	['NNS', 'RB', 'VBD', 'PRP', '.']	0.635	0.365	0

As a result of this analysis, it is possible to see that most extracts show a noun or pronoun (in yellow) followed by a verb (in green) in some cases a feeling is also ethicized using an adjective (in purple).

Identifier Rules

The above analysis will now form the foundation of the systems design by specifying the rules in which components of Action Event, Belief or Consequence can be found inside a sentence.

Action Events

Based on the above, if all three of the following rules equals true it is an Action Event:

1. Does a sentence contain the following phrase structure?

   0 or 1 pronouns or nouns. 0 or more verb, noun, adjective, adverb, determiner, preposition/subordinating conjunction or to. 1 noun, pronoun, verb, adjective, adverb or particle give up.

   Regex example:
   <NN|NNP|PRP.|PRP>+<VB.|NN|RB|JJ|TO|DT|IN><VB|NN|NNP|VB.|RB|RP|PRP|JJ|NNS>*

2. Does the phrase structure identified in point 1 have a negative sentiment below 0.083 and a neutral sentiment above 0.784?
3. Does the phrase structure identified in point 1 contain an action word and a person identifier?

   Definition
   Action Words – are words which describe an activity e.g. shout, talk, hum, walk, run, study, teach, eat etc.
   Person identifiers – are words which indicate a person e.g. he, she, me, I, husband, wife, daughter etc.

Belief

Based on the above, if all three of the following rules equals true it is an Beliefs:

1. Does a sentence contain the following phrase structure?

   1 or more pronouns or nouns. 1 or more verb, modal, or adverb.

   *Regex example: <NN|NNP|PRP.|PRP>+<VB.|MD|RB|VB>**

2. Does the phrase structure identified in point 1 have a negative sentiment below 0.134 and a neutral sentiment above 0.738?
3. Does the phrase structure identified in point 1 contain an belief word and a person identifier?

   Definition
   Belief Words – are words which indicate a thought or feeling e.g. think, thought, feel, felt, wonder etc.
   Person identifiers – are words which indicate a person e.g. he, she, me, I, husband, wife, daughter etc.

Consequence

Based on the above, if all three of the following rules equals true it is an Consequence:

1. Does a sentence contain the following phrase structure?

   0 or 1 wh-adverb. 1 or more noun or pronoun. 0 or more verb, noun, adjective, adverb

   Regex example:
   <WRB>?<NN|PRP>+<VB.|NN|RB|JJ|TO|DT|IN|MD|PRP|VB><NN.|JJ|RB|VB.|VB>*

2. Does the phrase structure identified in point 1 have a negative sentiment below 0.265 and a neutral sentiment above 0.615?
3. Does the phrase structure identified in point 1 contain an consequence word?

   Definition
   Consequence Words – are words that express a feeling e.g. angry, sad, happy, anxiety, joy etc.

Final Solution – Pseudocode:

Based on these rules the following Pseudocode was created to better visualise the creation of a class which would implement the rules indicated above and is based on python principles. This class was broken down into several methods. These are:

1. Class constructor – responsible for:
   a. Creating a word list (Identifier, Action, Belief and Consequence words) that calls on the wordDict method
   b. Creating a dictionary which contains the proposed sentiment levels of each part of the ABC framework
   c. Creating a tokenized list of sentences extracted from a text input
   d. Putting the tokenized list through a for loop that will analyses and give a ABC score (explained in more detail below) the sentences individually by calling the A,B or C method.
   e. Returning a completed outputDict

2. WordDict method – This method takes a list of words and loops through each of them twice, the first is used to create a list of synonyms which is the same word type (i.e a noun or verb) as the source word in the list and does not already exist in the supplied list of words. It then takes both lists and creates a third which are the pluralization of all words in the supplied list of words and synonyms list.
3. A, B or C method – These methods are responsible for all the rules indicated above, firstly the sentence provided is chunked using the chunked method, two parameters the supplied sentence and the regex rule (indicated above rules section), which then outputs all the chunked items from the sentence as a list. Taking this list, it will then take each item or identified phrase and check word by word, for words which are contained in the Action/Identifier/Consequence and Belief word list indicated in the constructor (Note a list will only be used if it correlates with the rules indicated above). If a word in these lists is present it will then check the sentiment of the entire phrase using the Sentiment Score method passing the phrase and the type (A, B or C, which indicates the rules it should follow), this will either return true if that phrase meets the sentiment rules or false if not. If true, it will then append the phrase into a new list, which will then be added to OutputDict.

4. Chunked Method – Takes two parameters a sentence and a regex rule, it will then use NLTK's RegexpParser to great a list of chunked phrases from the sentence.
5. Sentiment Scorer method – takes two parameters, a phrase and the type (A, B or C, which indicates the rules it should follow), using this information the phrase is checked whether it falls below the negative sentiment and is higher than the neutral sentiment indicated in the Self.sentimentRules.

1. *Break text block up into sentences:*

```
Class ADBDetect:

        Def __init__( self , text input):

                Tokenize_sentence = nltk.sent_tokenize(text input)
                Self.OutputDict = {}

                #Words for each part of the ABC framework
                Self.ActionWords = wordDict([Insert all action words in list])
                Self.IdentifierWords = wordDict([Insert all identifier words in list])
                Self.BeleifWords = wordDict([Insert all belief words in list])
                Self.ConsiquenceWords = wordDict([Insert all consequence words in list])

                Self.sentimentRules = {

                        A | B | C: {
                        Negative: …..
                        Neutral: …..
                        Positive: ….
                        }
                }

        For sentence in Tokenize_sentence:

                Self.sentence = sentence

                OutputDict[self.sentence] = {}
                OutputDict[self.sentence][Score] = {

                        aScore = a(sentence)
                        bScore = b(sentence)
                        cScore = c(sentence)

                }

        return OutputDict
```

2. *Add synonyms, plurals of list of compiled words:*

```
Def wordDict(self , list of words):

        WordList = []
        PluralList = []

        # Find synonyms of words in list of words
        for word in list of words:

                for synonyms in wordNet.synsets(word):

                        for l in synonyms.lemmas():

                                if l.name() is the same word type as word:
```

```
                    if l.name() not in list of words:

                        WordList.append(l.name)

        #Finds plural words in list of words + word list
        For words in list of words + WordList:

            Word = pattern.en.pluralize(words)
            PluralList.append(word)

    Return list of words + Word List + plural list
```

3. *Run chunked list through rules.*

#The below function will be roughly the same for each part of the ABC framework with minor differences. As a result only one function has been written to visualise this process, however, there would be in fact three separate functions for each part of the ABC framework.

```
    Def a b or c(self , sentence)

        Phrases = Chunked(sentence, chunk rule (regex example here))
        OutputList = []

        For phrase in phrases:

            VariablesPresent = False

            For word in phrase:

                If word in self.Action/Identifier/Consequence/Belief Words list:

                    VariablesPresent = True


            If VariablesPresent == True:


                If SentimentScorer (phrase, type) == True:

                    OutputList.append(phrase)

        Self.OutputDict[self.sentence]["A"|"B"|"C"] = OutputList


    Return len(OutputList)
```

4. *Chunk input sentence:*

```
    Def chunked(self , sentence, chunk rule)

        ChunkRule = "chunk rule"
        ChunkParser = RegexpParser(chunkedRule)
        Parsed = ChunkParser.parse(pos_tag(sentence))

        OutputList = []

        For item in parsed:

            #chunked items are converted into a tuple (good reference point to identify this)
            If item[0] is tuple:

                OutputList.append(item[0])
```

Return OutputList

5. *Check sentiment score:*

Def SentimentScorer(self , phrase, type):

if negative sentiment score <= self.sentimentRules[type][“Negative”] and neutral sentiment score >= self.sentimentRules[type][“Neutral”]:

return True
else:
return False

Sample output:

Input sentence: I was gardening the other day with my partner. He began to boss me around like he always does. It made me feel inadequate which really upset me.

```
{“ I was gardening the other day with my partner.” :
        {
         aScore: 1,
         bScore: 0,
         cScore: 0}
        },
  “He began to boss me around like he always does.“:
        {
         aScore: 1,
         bScore: 2,
         cScore: 3}
        },
  “It made me feel really inadequate which really upset me”:
        {
         aScore: 1,
         bScore: 2,
         cScore: 3}
        },
}
```

Objective 1 Conclusion

With this feature in place, it is now possible for a chatbot to determine whether the text it receives fulfils a specific part of the ABC framework, so now if a chatbot prompts a patient to provide examples it can firstly assess if a patient is using the framework correctly by looking for a sentence which fits whatever part of the framework it is trying to assess and also identify sentences in the patient text inputs to uses as an example as part of training.

Objective 2 & 3: Design an engaging method of introducing CBT fundamentals to patients and offer meaningful feedback and design an engaging method of introducing CBT fundamentals to patients and guiding them through the steps of CBT.

Now the “brains” have been designed, this project will now focus on achieving the other two design objectives, as reminder, the purpose of this tool is train and assist a user in approaching CBT by using the ABC framework.

Background Research Design Considerations

Thinking back to the background research, some was dedicated what patients either liked or disliked about the iCBT or Conversational Agent Solutions (See Pages 5 & 7). Though it might not be possible to address all the points outline as they may not be relevant in the context of this project, the following designs will try to consider following:

1. Patients progress will be recorded over time
2. The chatbot will explain clearly what it will do and how it will achieve this, to create realistic expectations the system.
3. The system will try to develop a repour with the user.
4. The system will try to create tailor personalised content based on a user's interactions.

Design Use-Case

With the above in mind, the project will be achieved this by creating a chatbot which clearly introduce itself and its functions, describe some CBT fundamental to the user and explain how it intends to help the user work through their problems by using the ABC framework. Upon receiving this training, the chatbot will then prompt the user to think of a recent situation which was a triggering event and ask them to describe it.

Using this description, the chatbot will then ask the user to break this situation down using the ABC principles. Each input would be checked for correctness assessing that each step has the correct information in:

1. Activating event should describe a triggering event,
2. Belief should describe how the user thought about that event,
3. Consequences should describe what emotions, behaviours and physical sensations are occurred because of that situation.

If the written format does not appear to be correct, the chatbot will then provide a recap of the information for next time using the examples constructed using the initial description provided or general examples should it not be able to understand the description.

Upon completing this training, chatbot will provide the patient with a basic journal system which they can use to continue recording their feelings using the ABC framework and using the chatbots error detection. It will then provide the patient the ability to review past entries, to see how they have performed over time.

Chatbot Transcript Design Considerations

Due to the nature of this chatbot, considerations relating to how to best approach introducing CBT and the ABC framework as well as asking a user to discuss sensitive topics will be important. Consequently, the chatbot will need to create a safe environment for the user and ensure the chatbot is effective in delivering its training. Considering this, the following design considerations will be made:

1. The chatbot will need to clearly define its function and its limitations as a help tool and not a substitute to therapy. Here are a few proposed transcript examples of this occurring:
    a. *"I just want to let you know everything we discuss is completely confidential and only you will be able to access our discussion"* – This text sets a tone of trust in place by saying that all that is discussed is completely confidential.
    b. *"I also need to stress, that I am only a tool and no substitute for professional support"* – This enthuses that the chatbot should not be used as a sole mental health support.
2. The chatbot will aim to build a rapport by using positive and fun language in places, to help make it seem more personable. Here are a few proposed transcript examples of this occurring:
    a. *"Let me introduce your ABC's! I don't mean your alphabet! :D"* – This is an attempt at light humour, to make the training slightly more engaging and fun.

b. *"Not quite right, have another go, you will get it!"* or *"Nice job! That's exactly right!"* – Providing positive encouragement to get a user to have confidence in achieving a positive outcome.

3. The chatbot will directly address the user in the transcript and ask the user questions to make a discussion more personal. Here are a few proposed transcript examples of this occurring:
   a. *"Now your turn (INSERT USERS NAME), think of a situation which made you feel back, in this situation what do you think an activating event is?"*

In addition to the above, a consideration also needs to be made relating to how to best deliver training related to the ABC framework. Though it is easy to implement, initially understanding it can be quite difficult due to its abstract nature. Consequently, the below approach should maximise the opportunity for a user to understand the principles been taught.

1. Introducing the ABC framework in principle, using the proposed transcript:

   "Understanding your emotional difficulties is one of the most important aspects of CBT, this can be achieved by breaking down a specific problem.

   I can hear you thinking how, can I do that?

   Let me introduce your ABC's! I don't mean your alphabet!

   Let me break this down for you!

   A - "Activating Event" - This is triggering events, situations in your past, present or future which make you feel or act a certain way.

   B - "Beliefs" - This is your thoughts or a belief you hold about that situation

   C - "Consequences" - This is where you think about your feelings, behaviours and sensations because of the above."

2. Exemplifying each part of the ABC framework, using the proposed transcript:

   " Now I realise that this might seem a little abstract so let me give you an example:

   A - Yesterday I went out without an umbrella, and it started to rain and now I am soaking wet

   B - I felt so stupid because, I knew it was going to rain

   C - I have to meet my friends completely wet and I am really now really anxious"

3. Asking the user to think of a situation and then apply it for each part of the ABC framework, checking each input for accuracy, for example:

   "Now your turn (INSERT USERS NAME), think of a situation which made you feel back, in this situation what do you think an activating event is?"

The input is checked it is indeed an activating event or not, if not the system will ask the user to try again. If failed for a second time the system will then provide another example.

The above approach will be adopted to give as much opportunity to educate the user in the use of each part of the framework and to try and develop their understanding on how to apply it in a practical manor.

Chatbot Design Pseudocode

As a consequence of the above the following pseudocode has been developed to visualise the basic functionality of the chatbot. This has been broken up into 4 distinctive classes:

1. The transcript class is responsible for the script and discussion rule – See below
2. The textProcessing class offers basic NLTK NLP integration to provide the chatbot with rudimentary understanding any text inputs – See appendix 8
3. The dbManager class manages a JSON database which stores user position within the chatbots lessons and journal entry for later viewing. – See appendix 9
4. The ABC Detector, this will combine the class described above to give the chatbot the ability to identify.

Class transcript:

```
# The constructor accesses the JSON database using the DdManager Class which provides a dictionary instance of the JSON file. It also provides a global instance of the username, the stage position (this is the positioning inside each transcript stage) and creating a ChatActive Bool which indicates to the system that the conversation should end or not. Then finally the StageManager method is instantiated (this controls the chat flows).

Def __init__(self)

        Db = dbManager()
        Db.selectDB()
        Self.Database = db.getDB()
        Self.UserName = database[user]
        Self.StagePosition = database[position]
        Self.Chatactive = 0
        Self.Journal = database[journal]

        Self.StageManager()

#This method controls all the chatbots chat flows, it checks the database for the users current chat stage then places the user into that chat flow.

Def StageManager(self):

        If self.databse[chatstage] == introduction

                Self.introduction()

        If self.database[chatstage] == lessonone

                Self.lessonone()

        If self.database[chatStage] == Whatever the chat position is

                Self. position method responsible for this section's transcript
```

```
#menu which is show a user upon completion of the CBT training, this will be see every time the user
relogs into the programme
If self.database[chatStage] == "none"

        Self.chat("chat", "Let me know how I can help you!")
        Self.chat("chat", "1. Write in your journal")
        Self.chat("chat", "2. See your journal")
        Self.chat("chat", "3. Relearn CBT")
        Question = Self.chat("input", "Choose option number")

        If question == 1
                Self.writeJournal()
        If question == 2
                Self.seejournal()
        If question == 3
                Self.lessonOne()


StageMethod() #This is a recursive method as we want to give the user the ability to see the final menu
at the end of each process. This can be quit by saying exit in any text input.
```

#This method controls the transcript related to introducing the chatbot and ascertaining the users name. This is achieved using a series of print and input statements which are controlled by the chat method, rather than using a vanilla print/input statement, this method applies some logic to allow the chatbot to understand whether the user would like to end the discussion, ask questions, and delay the text input into the terminal to improve readability. A series of questions are asked including what the users name is and whether a user would like to continue onto the first lesson. This first question would be processed using the textProcess class getName method, this allows the chatbot to identify the name of the user. The second uses the yesOrNo method from the textProcess class, this means that it understands If a yes or no answer is provided. Once this method ends, the chatbot saves the current position to the JSON database and will continue the conversation if the user wishes to continue.

```
Def introduction(self):

        chat(print, "whatever messages wanted to show", delay)
        Question = chat(input, "what should I call you?", delay)
        tp = textProcessing(question)
        Self. UserName = tp.getName()

        Self.db.singleDBUpdate("username", self. UserName)

        chat(chat, "Hi {self.UserName} pleasure to meet you!" delay)

        chat(chat, "whatever messages wanted to show", delay)

        Question = (input, "Do you want to start your first lesson?, delay)

        Tp = textProcessing(question)

        Answer = Tp.yesorno()

        self.db.singleDBUpdate("ChatStage", "lessonOne")

        If answer == True:

                Self.stageposition == 1

        Else:
                Self.exit()
```

 # Lesson one uses a very similar format to the introduction transcript, with one major difference, the use of DetectABC() this method calls the DetectABC class returning a dictionary with all the analysis present. This will be

then used to identify if what the user is inputting as part of the lesson is correct and will give alternatives if it is wrong.

```
Def lessonOne(self):

    If self.chatActive == 0:

        Self.welcome()
        Question = self.chat(input, "ready to learn CBT?")
        Tp = textPrcoessing(question, false)

        Answer = tp.yesOrNo()

        If answer == False:

            Self.exit()

        If answer == True:

            Self.StagePosition == 1

    If self.StagePosition == 1:

        #The chatbot will ask the user of about a time they felt something was wrong. This
        information will be passed directly into the detectABC class and the results of this will added
        to the abc1 variable as a dictionary.

        Abc1 = self.detectABC(Chat(input, "Tell me about a time were you felt bad and why))

        #Transcript for lesson starts here begins to talk about the ABC's using Chat() methods.

        Chat(print, "XYZ")

        #The chatbot will then ask the user to identify the Activating Events, Beliefs and Consequences
        in their original statement. This input would be fed directly into the CBTa statement and
        checked to see if there is an Activating event detected in the input. If it isn't, the chatbot will
        use the example of

        CBTa = self.detectABC(Chat(input, "think about the example you told me before, in this
        example can you think of what the Activating event would be?"))

        If CBTa contains a Activating Event

            Self.chat(print, "nice job!")

        Else:

            Repeat above CBTa instantiation.

            If CBTa contains a Activating Event:

                Self.chat(print, "nice job")

            Else:

                Temp = identified Activating Event inside of ABC1
                Self.chat(print, "Not quite, here is what I think it is {temp}")

        Repeat above for Belief and Consequence.

        self.db.newJournalEntry(CBTa, CBTb, CBTc) #adds the users answers to the journal
```

```
                self.exit()

Def writeJournal(self)

        #Repeats lessonOne but does not have any of the training transcripts. It will ask the user what is
        bothering them and then ask them to break the scenario up into an A,B and C category providing the
        same error checking as indicated in lessonOne.

…….

#The following allows the user to see there past journal entries, this is achieved by creating a list of entries and
assigning them with a id, the user can then select this id and the function will then index the list of journal entries
and isolate the A,B and C statement in that entry.
Def seeJournal(self)

        Entries = []
        ID = 0
        For entry in self.journal

                Entries.append(entry)
                Print(key, ":", entry)

        Question – chat("input", "Please choose entry ID number")

        CBTa = self.journal[entries[question]] ["Activating Event"]
        CBTb = self.journal[entries[question]] ["Belief"]
        CBTc = self.journal[entries[question]]["Consequence"]

        self.chat("print", f"Activating Event: {CBTa}")
      self.chat("print", f"Belief: {CBTb}")
      self.chat("print", f"Consequence: {CBTc}")

         question = self.chat("input", "Would you like to view another?\n")

          tp = textProcessing(question, False)

           answer = tp.yesOrNo()

            if answer == True:

        self.seeJournal()

#The following methods are used to support the construction of the transcript:

#Welcome produces a randomized welcome statement.

Def welcome(self)

        WelcomeStatements = list of welcome statements

        Chat("print,welcomeStatements[random.randint(o,len(welcomeStatement-1))])

#Exit produces a randomized exit statement and quits all processes.
Def exit()

        ExitStatement = list of welcome statements

        Chat("print,welcomeStatements[random.randint(o,len(welcomeStatement-1))])

#Provides custom printing/input statements to allow for delayed printing for more fluid chat like conversations
and to further logic to input statement
Def chat(type, text, delay)
```

```
        If type == print:

                Delay process using delay parameter
                Print(text)

        If type = Input:

                Delay process using delay parameter
                Input(text)

#Produces the detectABC analysis value.
Def detectABC(self , sentence)

        Abc = detectABC(sentence)
        Return abc
```

Analysis

This section will assess the effectiveness of the ABC detection functionality describe in the previous section by reviewing the system's ability to identify sentences deemed to be either an Activating Event, Belief or Consequence. This will be achieved by assessing the system based on the following success criteria:

1. Does the system identify the same sentences as a manually categorised body of text?
2. Is the system able to identify sentences which fit the ABC criteria, but were not outlined in the manually categorised body of text?

The above will be achieved utilising a dataset consisting of 9 new patient transcripts taken from Reinventing Your Life by Janet S. Klosko and Jeffrey Young.

Manually Categorized Transcripts

The following table indicates a manual categorisation of each of these transcripts, specifying at a sentence level whether it should fit an action, belief and or consequence classification based on the rules indicated in the design chapter.

ID	Transcript	A Score	Action	B Score	Belief	C Score	Consequence
1	**I haven't had a date in seven months. I haven't really met anybody who wants to date me. I really hate going to a lot of the place where I can meet people. I'm really shy. I can't talk to people. I don't expect them to like me.**	2	I haven't had a date in seven months. I haven't really met anybody who wants to date me.	3	I'm really shy. I can't talk to people. I don't expect them to like me.	1	I really hate going to a lot of the place where I can meet people.
2	**I am afraid I am going to end up entirely alone. I don't really fit in with the people at work, and my personal life is getting sparser and sparser. I just don't feel like I belong anywhere. I'm always on the outside, look in.**	1	I don't really fit in with the people at work, and my personal life is getting sparser and sparser.	1	I just don't feel like I belong anywhere.	1	I am afraid I am going to end up entirely alone.
3	**I was invited to a party last Saturday, night, and I dreaded it all week. What is it with me? Other people look forward to a party, but all week I couldn't get it off my mind. I couldn't relax. I was always on the verge of tears. I keep thinking I would get there and be so nervous, and not know what to say. I would act like a jerk. And everyone would seem better than me, better looking or smarter or more successful, and I'd have nothing to offer. And you know, that's exactly what happened. The party was like a nightmare. I couldn't wait to leave, and when I got home, I cried and cried. I know when I walk in that I'm going to be anxious. It's embarrassing to be so anxious. I'm uncomfortable, and I make other people feel uncomfortable. As soon as I get in, I know I'm going to screw up somehow. I'd say the wrong thing or do the wrong thing. I just want to crawl into a hole. it's so frustrating, because as soon as I know the person, I can carry on a normal conversations. But when I meet a stranger, I can't do it. I freeze up.**	1	I was invited to a party last Saturday, night, and I dreaded it all week.	3	I keep thinking I would get there and be so nervous, and not know what to say. I know when I walk in that I'm going to be anxious. I'm uncomfortable, and I make other people feel uncomfortable.	4	I couldn't relax. I was always on the verge of tears. The party was like a nightmare. I couldn't wait to leave, and when I got home, I cried and cried. I just want to crawl into a hole.
4	**It's like I'm alone even when I'm in a crowd. In fact, I feel most alone when I'm in a crowd. When I am with people, I would just go off into my own little world.**	1	it's like I'm alone even when I'm in a crowd	1	When I am with people, I would just go off into my own little world. I feel most alone when I'm in a crowd	1	

5	**Anthony is mad at me. He wants me to meet him at the restaurant tomorrow night. He doesn't want to have to come home first and get me. But I just can't do it. I can't get on the train. I think my panic will get so bad that i'd collapse on the street? What if something happens to me and there is no one to take care of me?**	3	He wants me to meet him at the restaurant tomorrow night. He doesn't want to have to come home first and get me.	2	I think my panic will get so bad that i'd collapse on the street? What if something happens to me and there is no one to take care of me? I just can't do it.	1	I can't get on the train
6	**I keep thinking, "What if I'm making a mistake, what if she's not the one? What if there's someone better? How do I know if she's the right one, or if I'm just settling for less? "We get along, but there just aren't fireworks. And what if I can't support her? She wants to have a lot of kids. How do I know that I can support a wife and kids? I can barely support myself. Sometimes I think I should just break up with her now, and get it over with.**	2	We get along, but there just aren't fireworks She wants to have a lot of kids	3	I keep thinking, "What if I'm making a mistake, what if she's not the one? What if there's someone better? How do I know if she's the right one, or if I'm just settling for less? How do I know that I can support a wife and kids? I can barely support myself	1	Sometimes I think I should just break up with her now, and get it over with.
7	**I'm just not the kind of person who likes to be taken care of. I just don't like it. I found myself getting really depressed and upset about it. I guess I realized that I shouldn't get so upset about it. And now I'm back in my apartment, and I'm having a hard time letting my friends do the things I need to have done for me. And the funny thing is, I know I would do the same for them in a flash, without even thinking about it."Why can't I accept help from people?**	1	I'm back in my apartment, and I'm having a hard time letting my friends do the things I need to have done for me.	1	I'm just not the kind of person who likes to be taken care of. I just don't like it.	1	I found myself getting really depressed and upset about it.
8	**When I first met Anthony at school, I used to tell him that I wished we could stay in school forever, that it never had to end. He couldn't wait to get out, for school to end, but I wanted it to go on and on. I guess it was that I felt so safe there. I knew what to expect.**	1	When I first met Anthony at school, I used to tell him that I wished we could stay in school forever, that it never had to end	2	I guess it was that I felt so safe there. I knew what to expect.	1	When I first met Anthony at school, I used to tell him that I wished we could stay in school forever, that it never had to end
9	**Usually how it goes is that all of a sudden, out of the blue, this feeling comes over me. It's like a feeling of doom. I get dizzy and light-headed and feel like I can't breathe. My heart starts pounding. And everything feels unreal. It feels like I'm having a nervous breakdown. Like I'm going to go crazy on the spot. For a long time I would run to the emergency room every time it happened. I thought I was having a heart attack or brain aneurysm or something. Those were the worst ones. I really thought I was dying. Sometimes I get this funny pressure in my head and I get scared that it's an aneurysm. But I've pretty much learned that I'm going to be all right, that I'm not going to die. I mean, it's happened so many times. It has to be a really bad one now for me to believe I'm dying. Mostly I worry now about losing my mind. I'll start screaming or raving or hearing voices, and it will never stop.**	1	Sometimes I get this funny pressure in my head and I get scared that it's an aneurysm	2	I thought I was having a heart attack or brain aneurysm or something. I really thought I was dying	4	But I've pretty much learned that I'm going to be all right, that I'm not going to die. It has to be a really bad one now for me to believe I'm dying. I'll start screaming or raving or hearing voices, and it will never stop. Sometimes I get this funny pressure in my head and I get scared that it's an aneurysm

Systems Categorization Analysis

The following table indicates an automatic categorisation of each of these transcripts, specifying at a sentence level whether it should fit an action, belief and or consequence classification using the system design and rules specified in the design chapter. In each sentence extract a yellow highlight

has been used to outline the phrase which the system has used to fit within the rules specified in each category. In addition, a score has been offered for each transcript based on the number of occurrences an action, belief or consequence phrase has occurred.

ID	Transcript	A Score	A Sentences	B Score	B Sentences	C Score	C Sentences
1	**I haven't had a date in seven months. I haven't really met anybody who wants to date me. I really hate going to a lot of the place where I can meet people. I'm really shy. I can't talk to people. I don't expect them to like me.**	0		1	I don't expect them to like me.	0	
2	**I am afraid I am going to end up entirely alone. I don't really fit in with the people at work, and my personal life is getting sparser and sparser. I just don't feel like I belong anywhere. I'm always on the outside, look in.**	0		1	I just don't feel like I belong anywhere.	0	
3	**I was invited to a party last Saturday, night, and I dreaded it all week. What is it with me? Other people look forward to a party, but all week I couldn't get it off my mind. I couldn't relax. I was always on the verge of tears. I keep thinking I would get there and be so nervous, and not know what to say. I would act like a jerk. And everyone would seem better than me, better looking or smarter or more successful, and I'd have nothing to offer. And you know, that's exactly what happened. The party was like a nightmare. I couldn't wait to leave, and when I got home, I cried and cried. I know when I walk in that I'm going to be anxious. It's embarrassing to be so anxious. I'm uncomfortable, and I make other people feel uncomfortable. As soon as I get in, I know I'm going to screw up somehow. I'd say the wrong thing or do the wrong thing. I just want to crawl into a hole. it's so frustrating, because as soon as I know the person, I can carry on a normal conversations. But when I meet a stranger, I can't do it. I freeze up.**	3	Other people look forward to a party, but all week I couldn't get it off my mind. I know when I walk in that I'm going to be anxious. But when I meet a stranger, I can't do it.	3	Other people look forward to a party, but all week I couldn't get it off my mind. I keep thinking I would get there and be so nervous, and not know what to say. I couldn't wait to leave, and when I got home, I cried and cried.	1	I was always on the verge of tears.
4	**It's like I'm alone even when I'm in a crowd. In fact, I feel most alone when I'm in a crowd. When I am with people, I would just go off into my own little world.**	0		2	In fact, I feel most alone when I'm in a crowd. When I am with people, I would just go off in to my own little world.	0	
5	**Anthony is mad at me. He wants me to meet him at the restaurant tomorrow night. He doesn't want to have to come home first and get me. But I just can't do it. I can't get on the train. I think my panic will get so bad that i'd collapse on the street? What if something happens to me and there is no one to take care of me?**	0		2	But I just can't do it. I think my panic will get so bad that i'd collapse on the street?	0	

6	**I keep thinking, "What if I'm making a mistake, what if she's not the one? What if there's someone better? How do I know if she's the right one, or if I'm just settling for less? "We get along, but there just aren't fireworks. And what if I can't support her? She wants to have a lot of kids. How do I know that I can support a wife and kids? I can barely support myself. Sometimes I think I should just break up with her now, and get it over with.**	0		4	I keep thinking, "What if I'm making a mistake, what if she's not the one? How do I know if she's the right one, or if I'm just settling for less? Sometimes I think I should just break up with her now,and get it over with.	1	Sometimes I think I should just break up with her now, and get it over with.
7	**I'm just not the kind of person who likes to be taken care of. I just don't like it. I found myself getting really depressed and upset about it. I guess I realized that I shouldn't get so upset about it. And now I'm back in my apartment, and I'm having a hard time letting my friends do the things I need to have done for me. And the funny thing is, I know I would do the same for them in a flash, without even thinking about it."Why can't I accept help from people?**	1	And the funny thing is, I know I would do the same for them in a flash, without even thinking about it.	3	I'm just not the kind of person who likes to be taken care of. I guess I realized that I shouldn't get so upset about it.	0	
8	**When I first met Anthony at school, I used to tell him that I wished we could stay in school forever, that it never had to end. He couldn't wait to get out, for school to end, but I wanted it to go on and on. I guess it was that I felt so safe there. I knew what to expect.**	1	When I first met Anthony at school, I used to tell him that I wished we could stay in school forever, that it never had to end.	3	He couldn't wait to get out, for school to end, but I wanted it to go on and on. I guess it was that I felt so safe there.	0	
9	**Usually how it goes is that all of a sudden, out of the blue, this feeling comes over me. It's like a feeling of doom. I get dizzy and light-headed and feel like I can't breathe. My heart starts pounding. And everything feels unreal. It feels like I'm having a nervous breakdown. Like I'm going to go crazy on the spot. For a long time I would run to the emergency room every time it happened. I thought I was having a heart attack or brain aneurysm or something. Those were the worst ones. I really thought I was dying. Sometimes I get this funny pressure in my head and I get scared that it's an aneurysm. But I've pretty much learned that I'm going to be allright, that I'm not going to die. I mean, it's happened so many times. It has to be a really bad one now for me to believe I'm dying. Mostly I worry now about losing my mind. I'll start screaming or raving or hearing voices, and it will never stop.**	4	Usually how it goes is that all of a sudden, out of the blue, this feeling comes over me. My heart starts pounding. Like I'm going to go crazy on the spot. Mostly I worry now about losing my mind.	4	It feels like I'm having a nervous breakdown. I thought I was having a heart attack or brain aneurysm or something. I really thought I was dying. I mean, it's happened so many times.	0	

Data Analysis

The accuracy of this tool appears to be quite low, in comparison to the manual categorisation of the transcripts, either it does not detect any appropriate sentences, or it has selected sentences which

do not fit the categories correctly. These above problems, however, do appear to be more concentrated in the action event and consequence categories, whereas for beliefs, it appears that the system on the most part has selected appropriate sentences.

In transcript 1, the sentences “I haven't had a date in seven months” and “I haven't really met anybody who wants to date me” are both sentences that could be considered action events. In both example, the identifier “I” and the action word “date” and “met” should provide sufficient information to categorise it as an action event. However, this did not occur because “date” and “met” were not included inside the action word corpus, meaning any phrases including these words would have been disregarded. By adding this into the corpus, the system would have categorised these sentences as action events.

Similarly, the sentence “I really hate going to a lot of the place where I can meet people”, should have been presented as a consequence category. This is because, the word “hate” was not included in the consequence corpus and the second was that the phrase “I really hate going” scores negative sentiment of 0.666 and neutral sentiment of 0.334 which falls above and below the sentiment threshold stipulated as part of the design process.

However, the beliefs component of the system did work partially, producing one of the three expected results “I don't expect them to like me”. However, it did not identify, “I can't talk to people”, this is because it is conveying a thought, but there is no word to euthanize this. The word “can’t” could be used as a marker for a thought or a feeling (as seen in transcript 5, see below), but because it can also be used in other context like stating a fact, this cannot be used reliably as a marker to define a belief, which highlights a limitation of this approach overall.

In transcript 5, the sentences “He wants me to meet him at the restaurant tomorrow night” and “He doesn't want to have to come home first and get me” could be deemed as action events. In both example, the identifier “he” and the action word “want/s” should provide sufficient information to categorise it as an action event. However, this did not occur, this is because the word “want” was not included inside the action word corpus. By adding this into the corpus, the system was then able to display the first example, however, the second still would not show. This because the isolated phrase “He doesn’t want” had a negative sentiment of 0.289 and neutral sentiment of 0.711, in both cases this did not meet the sentiment threshold outlined for action events which is a negative sentiment below 0.083 and a neutral sentiment above 0.784.

Furthermore, a similar issue arises in the consequence component, the sentence, “I can't get on the train”, would be the appropriate sentence for this categorisation, but it was never identified. The operative word that could be used to identify this sentence as a consequence based on the context would be “can’t”. “Can’t” in this instance has been used to convey a feeling of anxiety or fear, however, because “can’t” is also used to convey a fact, it has not been included in the consequence word corpus. This further supports the limitation indicated above, which suggests the word “can’t” cannot be used as an indicator in any of the categories. Even though the word could be used favourably to place a phrase in a category subject to its context, the designed system is simply not intelligent enough to understand the context in which it is used. Consequently, making it impossible to use this word as an indicator to categories a phrase.

Continuing with the analysis of transcript 5, in the case of beliefs, the system worked perfectly, identified two sentences “But I just can't do it” and “I think my panic will get so bad that i'd collapse on the street?”. Based on the context of the text input, both sentences convey either feeling or a thought.

In transcript 7, the sentence “I'm back in my apartment, and I'm having a hard time letting my friends do the things I need to have done for me”, should be highlighted as an action event. However, it wasn’t and in its place the sentence “And the funny thing is, I know I would do the same for them in a flash, without even thinking about it” was presented. This was shown, because wordNet has produced an identifier word of “is” and an action word of “thing”, neither of these words fits the criteria to be in either corpus and is why the system is producing a false positive. The reason the appropriate sentence was not show, is because the system does not have the action word “having” included in the action word corpus, with this added, like before however, the phrase “I'm having a hard time letting” has a negative sentiment of 0.259 and neutral sentiment of 0.741 which falls outside of the sentiment threshold outlined by the design chapter which as previously indicated is negative sentiment below 0.083 and a neutral sentiment above 0.784.

Considering the consequence category of transcript 7, the system would have needed to identify the sentence “And the funny thing is, I know I would do the same for them in a flash, without even thinking about it”. However, It didn’t, the primary reason for this is that no words that convey a feeling were detected by chunking the sentence, this is because the phrase that holds the word conveying a feeling, “funny” did not fall within the regex rule, because “funny” is an adjective and came before the noun in sentence. The current regex rules are not looking for phrase which have adjectives before a noun. To further explain, the phrase “funny thing is” has the following structure an adjective, noun, and verb (present tense, third person singular), the current regex rules are 0 or 1 Wh-adverb, 1 or more noun or pronouns followed by 1 or more Verb, Noun, Adjective, TO, Determiner, IN, Modal, Pronoun and ends with a noun, adjective, adverb or verb, hence this was not detected. In addition, even if it did two other issues would have arisen preventing the detection of this phrase. The first, is it that “funny”, is not in the consequence word corpus. The second is that the sentiment threshold for consequence primarily focuses on negative sentiment, the phrase “And the funny thing is” has a negative sentiment of 0 and a natural sentiment of 0.58, this means that even if the two other problems were resolved, this would prevent it from been identified.

However, like before, the belief category had a high level of success in identifying the appropriate phrases and sentences.

Design Shortfalls

The following analysis suggests that there are several shortfalls with the design of this system, these are as follows:

1. There are not enough relevant words inside the action and consequence word corpus to accurately identify relevant phrases.
2. The sentiment thresholds are set are too rigidly, meaning that some phrases are been missed due to its sentiment falling outside of the threshold.
3. The regex chunking rules used in the consequence category are missing key phrases.
4. Some words produced by the wordNet for the various corpuses are not relevant to that corpus.
5. The system cannot utilise some words i.e., “can’t” as category indicators because it cannot understand the context in which it is used.

Design Improvements

Though some of the design shortfalls highlighted are fundamentally because of limitations of the design approach of the algorithm, other are of more a result of inadequate rule configurations. Consequently, this report has proposed several rule configuration alterations to better assess the performance of the algorithm used, these changes are:

1. The following words will be added to the Action Word Corpus; “fit”, "with", "having", "want", "date" and "met"
2. The following words will be added to the consequence corpus; "die", "upset", "depressed", "break", "cried", "nightmare" ,"relax", "afraid", "funny" and "hate".
3. The prototype now has several unique stop word lists, which will remove identified anomalous words from the corpuses. In the identifier stop words list “is” has added and for the action stop words list “thing” has been added.
4. A new sentiment threshold will be created, though mainly arbitrary in nature, this scale has been created based on seeing the sentiment scores indicated above and is designed to emphasize the rigidity of the old sentiment threshold. Action Event phrases, now have a maximum of 0.3 negative sentiment and a minimum of 0.7 neutral sentiment, belief phrases now have a maximum of 0.6 negative sentiment and a minimum of 0.4 neutral sentiment and consequences now have a maximum of 0.9 negative sentiment and a minimum of 0.1 neutral sentiment.
5. The consequence regex rules have been updated to include the option of 1 adjective at the beginning of the a phrase, the regex now looks like, <WRB|JJ>? <NN|PRP>+ <VB.|NN|RB|JJ|TO|DT|IN|MD|PRP|VB>* <NN.|JJ|RB|VB.|VB>

System Categorisation Analysis – With Reconfigured Category Rules

The following table indicates an automatic categorisation of each of these transcripts, specifying at a sentence level whether it should fit an action, belief and or consequence classification using the system and rules specified using the above amendments to the original design. In each sentence extract a yellow highlight has been used to outline the phrase which the system has used to fit within the rules specified in each category. In addition, a score has been offered for each transcript based on the number of occurrences an action, belief or consequence phrase has occurred.

ID	Transcript	A Score	A Sentences	B Score	B Sentences	C Score	C Sentences
1	**I haven't had a date in seven months. I haven't really met anybody who wants to date me. I really hate going to a lot of the place where I can meet people. I'm really shy. I can't talk to people. I don't expect them to like me.**	3	I haven't had a date in seven months. I haven't really met anybody who wants to date me.	1	I don't expect them to like me.	1	I really hate going to a lot of the place where I can meet people.
2	**I am afraid I am going to end up entirely alone. I don't really fit in with the people at work, and my personal life is getting sparser and sparser. I just don't feel like I belong anywhere. I'm always on the outside, look in.**	0		1	I just don't feel like I belong anywhere.	1	I am afraid I am going to end up entirely alone.

3	**I was invited to a party last Saturday, night, and I dreaded it all week. What is it with me? Other people look forward to a party, but all week I couldn't get it off my mind. I couldn't relax. I was always on the verge of tears. I keep thinking I would get there and be so nervous, and not know what to say. I would act like a jerk. And everyone would seem better than me, better looking or smarter or more successful, and I'd have nothing to offer. And you know, that's exactly what happened. The party was like a nightmare. I couldn't wait to leave, and when I got home, I cried and cried. I know when I walk in that I'm going to be anxious. It's embarrassing to be so anxious. I'm uncomfortable, and I make other people feel uncomfortable. As soon as I get in, I know I'm going to screw up somehow. I'd say the wrong thing or do the wrong thing. I just want to crawl into a hole. it's so frustrating, because as soon as I know the person, I can carry on a normal conversations. But when I meet a stranger, I can't do it. I freeze up.**	4	What is it with me? Other people look forward to a party, but all week I couldn't get it off my mind. I know when I walk in that I'm going to be anxious. But when I meet a stranger, I can't do it.	5	Other people look forward to a party, but all week I couldn't get it off my mind. I couldn't relax. I keep thinking I would get there and be so nervous, and not know what to say. I couldn't wait to leave, and when I got home, I cried and cried. I just want to crawl in to a hole.	2	I couldn't relax. I was always on the verge of tears.
4	**It's like I'm alone even when I'm in a crowd. In fact, I feel most alone when I'm in a crowd. When I am with people, I would just go off into my own little world.**	1	When I am with people, I would just go off in to my own little world.	2	In fact, I feel most alone when I'm in a crowd. When I am with people, I would just go off in to my own little world.	1	It's like I'm alone even when I'm in a crowd.
5	**Anthony is mad at me. He wants me to meet him at the restaurant tomorrow night. He doesn't want to have to come home first and get me. But I just can't do it. I can't get on the train. I think my panic will get so bad that i'd collapse on the street? What if something happens to me and there is no one to take care of me?**	2	He wants me to meet him at the restaurant tomorrow night. He doesn't want to have to come home first and get me.	3	But I just can't do it. I think my panic will get so bad that i'd collapse on the street?	1	I think my panic will get so bad that i'd collapse on the street?
6	**I keep thinking, "What if I'm making a mistake, what if she's not the one? What if there's someone better? How do I know if she's the right one, or if I'm just settling for less? "We get along, but there just aren't fireworks. And what if I can't support her? She wants to have a lot of kids. How do I know that I can support a wife and kids? I can barely support myself. Sometimes I think I should just break up with her now, and get it over with.**	1	She wants to have a lot of kids.	5	I keep thinking, "What if I'm making a mistake, what if she's not the one? How do I know if she's the right one, or if I'm just settling for less? I can barely support myself. Sometimes I think I should just break up with her now, and get it over with.	1	Sometimes I think I should just break up with her now, and get it over with.

7	**I'm just not the kind of person who likes to be taken care of. I just don't like it. I found myself getting really depressed and upset about it. I guess I realized that I shouldn't get so upset about it. And now I'm back in my apartment, and I'm having a hard time letting my friends do the things I need to have done for me. And the funny thing is, I know I would do the same for them in a flash, without even thinking about it."Why can't I accept help from people?**	2	And now I'm back in my apartment, i and I'm having a hard time letting my friends do the things I need to have done for me.	4	I'm just not the kind of person who likes to be taken care of. I just don't like it. I guess I realized that I shouldn't get so upset about it.	2	I found myself getting really depressed and upset about it. I guess I realized that I shouldn't get so upset about it. And the funny thing is, I know I would do the same for them in a flash, without even thinking about it.
8	**When I first met Anthony at school, I used to tell him that I wished we could stay in school forever, that it never had to end. He couldn't wait to get out, for school to end, but I wanted it to go on and on. I guess it was that I felt so safe there. I knew what to expect.**	2	When I first met Anthony at school, I used to tell him that I wished we could stay in school forever, that it never had to end.	3	He couldn't wait to get out, for school to end, but I wanted it to go on and on. I guess it was that I felt so safe there.	0	
9	**Usually how it goes is that all of a sudden, out of the blue, this feeling comes over me. It's like a feeling of doom. I get dizzy and light-headed and feel like I can't breathe. My heart starts pounding. And everything feels unreal. It feels like I'm having a nervous breakdown. Like I'm going to go crazy on the spot. For a long time I would run to the emergency room every time it happened. I thought I was having a heart attack or brain aneurysm or something. Those were the worst ones. I really thought I was dying. Sometimes I get this funny pressure in my head and I get scared that it's an aneurysm. But I've pretty much learned that I'm going to be allright, that I'm not going to die. I mean, it's happened so many times. It has to be a really bad one now for me to believe I'm dying. Mostly I worry now about losing my mind. I'll start screaming or raving or hearing voices, and it will never stop.**	4	Usually how it goes is that all of a sudden, out of the blue, this feeling comes over me. My heart starts pounding. Like I'm going to go crazy on the spot. Mostly I worry now about losing my mind.	5	It feels like I'm having a nervous breakdown. I thought I was having a heart attack or brain aneurysm or something. I really thought I was dying. I mean, it's happened so many times. I'll start screaming or raving or hearing voices, and it will never stop.	5	Like I'm going to go crazy on the spot. Sometimes I get this funny pressure in my head and I get scared that it's an aneurysm. But I've pretty much learned that I'm going to be all right, that I'm not going to die. I'll start screaming or raving or hearing voices, and it will never stop.

Data Analysis

For example, on transcript 9, by increasing the sentiment threshold, the sentence “I'll start screaming or raving or hearing voices, and it will never stop” has been classified as a beliefs, however, there are no words in this sentence which would correlate with the themes of the belief word corpus. The same has occurred within transcript 3, a similar issue occurred in the action event category, the sentences “What is it with me?” and “Other people look forward to a party, but all week I couldn't get it off my mind” was identified. The first was identified using the phrase “is it with” and the second was identified using the phrase “my mind”, both are utilizing words which would not fit inside the identifier and action word corpuses. The reason this error as occurred is because the wordNet functionality again has created anomalous words inside of the belief corpus.

This issue appears to be prevalent throughout the system and needs to be addressed either by adding even more words to the corpuses and stop words list or wordNet is completely disregarded and a larger action, belief and consequence word corpuses are created. However, in both cases this would require more available data to identify meaningful words and time to analyse their suitability.

Within Transcript 2, no action events were identified, ideally the system should have highlighted “I don't really fit in with the people at work, and my personal life is getting sparser and sparser” as an action event, because the phrase “I don't really fit in with the people at work” would be considered an action event. The systems regex rules did identify the phrase “I don't really fit” as a possible fit, also contained the words “I” (identifier word) and “fit” (action word), however, it was disregarded because the negative sentiment was too high, hitting 0.44 and the threshold was now 0.3.

Discussion

This section started off by suggesting a success criterion for this functionality, initially it could be argued that it did not perform well against this criterion, with it not identifying a large proportion of the manually categorised sentences were not found using the analysis tool. However, once the rule improvements were applied a drastic improvement in performance was recorded, based on this it could be argued that this system succeeded in what it set out to achieve.

Though these amendments have improved the performance of identification in this dataset, there is one concern about this approach, with every new dataset, amendments will be made to further improve accuracy, however, what if these changes effect the accuracy of results previously assessed. For example, by adding a adverb to the beginning of the regex rule for the consequence category to identify a sentence that the system missed previously could this change create unwanted anomalous results? Though this could be mitigated by checking for specific keywords in the corpuses, there is a slight chance that amendments and improvements like this could damage the performance of this system. However, based on the results offered above, the structural elements that the chunking is looking for seems to accurately identify phrases needed to categorise the sentences.

Its clear that, by adding further words to the identifier, action, belief, and consequence corpuses, the systems ability to identify relevant phrases for each category has drastically improved. However, pairing these words to wordNet is causing problems by producing anomalous results. The more logical approach would be to remove this functionality from this system and replace it with a manually compiled list of appropriate words for each corpus. However, the issue with is approach is that it would require more sample data, as well as time to extrapolate these words. Which then raise question, would it not be better to pursue a machine learning approach rather than a rule based due to the increase size in dataset?

The data above suggests that even with the new changes and the improved accuracy that has come with it, the sentiment threshold still seems to be adversely affecting the system’s ability to identify some sentence types, particularly those with a positive sentiment. This could simply be because this approach cannot by nature be a one size fits all, there is simple too many combinations of phrases, with to many combinations of different types of sentiment for each category. The only thing that this approach can hope for is to identify as many phrases as possible that fit the criteria. That said, it could be argued that the data supplied as part of the design component of this project relating to the sentiment thresholds was incorrect, because the analysis of sentiment was based on sentence level sentiment and not at an extracted phrase level. This means that by removing words considered to be irrelevant to the category like the system has, it may have changed the average sentiment levels. If this system was to be redesigned, it would be advisable that sentiment is assessed in the

data by chunking phrases from the sentence and then assessing the sentiment this way to get a more accurate set of thresholds.

Research Limitations

This project has been limited by two main factors the first, time, and the second, access to patient both from accessing sensitive data perspective and talking to them as a user. This has meant that the author has had to pursue an approach to solving the problem indicated using methods which may not have been the best fit in practice but offer a solution in principle.

For example, the use of machine learning was completely disregarded because of concerns around time needed to learn how to implement it and because there was simply not enough data to create a accurate statistical model.

Furthermore, the analysis section of this report, solely focused on the "brains" of the designed solution, the ability to analyse text inputs and categorise them into the ABC framework. Rather than the design, which also included a chatbot, transcripts and a journal system. Though this was built, there was not really way to assess this, in part because of the time it would have taken to carry out user satisfaction surveys. But the main barrier was concerns around getting ethical approval and accessing clinical patients to trial such a system because of the sensitive nature of this project which meant that this part of the project was disregarded.

In addition, on reflection, a few concerns relating to the design methodology have been highlighted, the first being that there has been a implicit bias around this topic, which as bled through into the design. This is that the believed context of the text submitted to the system will always be negative, simply put that won't always be the case a user may not always want to input a negative entry. This approach seems to be effective at categorising entries with have negative sentiment, however, it has never been tested with positive entries and due to the sentiment rules in place its highly unlikely that it will be able to identify phrases which fall within any part of the ABC framework if it were positive.

Finally, sentences manually identified as an A, B or C phrase by someone who is not an expert in the field of CBT, though holding an intimate understand of CBT, he is not a medical professional nor a specialist in this area. This means that the data could be incorrectly categorised and in turn will have adversely affected the results of this project.

Recommendations

If this project were to be repeated following a rule-based approach, it would be recommended to exclude the use of sentiment analysis as a category marker due to what was discussed above. It would be advised to focus more on the types of words used in phrases and the structure of those phrases. This means that if project were to be repeated it should develop several word corpuses which can be used to define a A, B or C phrase.

Alternatively, if time and access to raw data is not a limitation, it would be strongly recommended to investigate a Machine-Learning based approach as this would enable the system to become more accurate and able to utilise a much wider dataset.

Finally, any raw data that is used as part of any further projects should be processed in conjunction with a subject matter expert in CBT as to ensure that the transcripts are correctly categorised, to reduce the possibility of errors in

Conclusion

This project set out to develop an understanding of how CBT is delivered to patients, identify and understand the technologies and techniques with the ultimate aim of creating a minimum viable product that could be used as a framework for further research into CBT products. Consequently, this project created a chatbot which can guide users through the steps of the ABC framework and assist them in implementing it. This required the Chatbot to have the ability to understand text inputs and categorise the phrases into either an Activating Event, Belief, or Consequence which could then be used to quality check interactions with the Chatbot and provide real examples of each category. Surprisingly, the algorithm designed does hold a reasonable level of accuracy, however, as seen during the analysis, there is need for more data to create a much larger corpus of words which can be used to identify an A, B or C phrase. In addition, the methods of analysing used to categorise phrases (such as sentiment analysis) are not completely convincing in their effectiveness. Paired with the issues surrounding the use of a rule-based approach there is room to increase the scope of this project, looking into the use of machine-learning would be a good next step because it would be more capability consuming large quantities of data and identifying more subtle characteristics which have not been possible to see as part of a rule-based system.

Authors Research Reflection

When asked to reflect on this project, the first thing that comes to mind is how much I enjoyed it, been able to develop subject knowledge around natural language processing and text analysis methods was something I have never done before and took great pleasure in learning and making things work as part of this project.

However, this did come with the challenge of trying to learn something rather technical with only a year of experience in computer science and at best a mediocre mathematical knowledge. This led to difficulties understanding some of the more technical elements of the project. This was particularly problematic during the background analysis of natural language processing tools and techniques used in mental health. Though some of these methods were relatively easy to grasp like tokenization, normalization, part of speech tagging, and sentiment analysis others were not, particularly those which are more mathematically orientated. Though researching these techniques was helpful to develop a basic understanding, I learned the only way I was going to truly understand them was to learn by doing, as reflected in the report. I found several examples of each technique and attempted to replicate it, once I achieved the same results, I then applied it the context of my subject matter, mental health, this approach as a result, further cemented my understanding of each technique discussed.

Another issue I encountered was time and setting realistic expectations for this project. My initial goal was far too broad and I wanted to achieve far more than would have been reasonable within the time I had. Paired with real world issues such as starting a full-time job midway through the dissertation period my expectations for this project would have been completely unattainable. Fortunately, with guidance from my supervisor, I was able to identify a scope which would be suitable for minimum viable product (MVP). What I have learned as a result of this, is to build a scope of work, then look at what features would constitute an MVP i.e. something which would show the basics idea of the product with a view of adding more to it in future.

Further discussing this projects methodology, one other element which I found difficult was direction, at the beginning of this project I felt aimless because I had not appreciated the importance

the background research prior to making some design decisions about the final product. Though at the time my supervisor did make some suggestions about the technological approach such as the use of sentiment analysis, I found myself unable to understand how to best apply it because I didn't not understand it or how different tools may complement one another. However, once I began more technical research about natural language processing, I began to understand how certain pieces of technology could work together to build a solution which could achieve my objectives. As a result of this, I have learned it is better to develop a good an understanding of a topic and tools and then create a robust design, before jumping in and trying to build something as it will reduce the need to create multiple iterations of a product.

In addition to this, I found that I had wasted a significant amount time on trying to get ethical approval for primary research. Initially I thought this would have been needed to get better access to professional insight on sample transcripts that I was intending to make/use in the development of my project. However, as the project progressed, I had found a number of good sources of real transcripts which meant I no longer needed to make up data to carry out this project.

The above exemplifies two key errors in thinking when approaching this project, planning and pre-reading. I do not enjoy reading, mainly because of my dyslexia, as a result I like to learn through doing. However, in this project, it would have vastly decreased time wastage if I had done more initial reading around the technical aspects of the subject and tried to find data sources prior to starting work on the product. This would have then intern enabled me to gain a better understand of what direction I should take the project and would have highlighted that there were sufficient secondary data sources to complete this project without the need for further data collection. In future I will endeavour to carry out more initial reading around a topic to better understand the space I am working in and better plan out my approach before jumping into the construction of a report and or product.

References

Apache cTAKESTM - clinical Text Analysis Knowledge Extraction System. 2021. Available at: https://ctakes.apache.org/faq.html#:~:text=What%20is%20cTAKES%3F,used%20by%20various%20institutions%20internationally. [Accessed: 27 July 2021].

Beach J, Boottery, S. 2019 Key areas for action on the health and care workforce. Available at: https://www.health.org.uk/publications/reports/closing-the-gap [Accessed: 24 May 2021].

Bendig, E. et al. 2019. The Next Generation: Chatbots in Clinical Psychology and Psychotherapy to Foster Mental Health – A Scoping Review. Verhaltenstherapie , pp. 1–13. Available at: https://www.karger.com/Article/FullText/501812 [Accessed: 8 July 2021].

Bird S, Klein E, Loper E. 2019. Natural Language Processing with Python – Analyzing Text with the Natural Language Toolkit (2019) Available at: https://www.nltk.org/book/ch00.html [Accessed: 18 July 2021].

British Medical Association. 2020. Mental health workforce report. Available at: https://www.bma.org.uk/advice-and-support/nhs-delivery-and-workforce/workforce/mental-health-workforce-report [Accessed: 13 April 2021].

Biradar, A. and Totad, S.G. 2019. Detecting Depression in Social Media Posts Using Machine Learning. Communications in Computer and Information Science , pp. 716–725. Available at: https://link.springer.com/chapter/10.1007/978-981-13-9187-3_64 [Accessed: 30 October 2021].

Chen, S. 2020. Getting Started with Text Vectorization - Towards Data Science. Available at: https://towardsdatascience.com/getting-started-with-text-vectorization-2f2efbec6685 [Accessed: 27 July 2021].

Codecademy. 2020. What is word embedding. Available at: https://www.codecademy.com/paths/build-chatbots-with-python/tracks/retrieval-based-chatbots/modules/nlp-word-embeddings/lessons/word-embeddings/exercises/what-is-a-word-embedding [Accessed: 30 July 2021].

Codecademy. 2020. Word Embedding Distance. Available at: https://www.codecademy.com/paths/build-chatbots-with-python/tracks/retrieval-based-chatbots/modules/nlp-word-embeddings/lessons/word-embeddings/exercises/distance [Accessed: 30 July 2021].

Corcoran, C.M. et al. 2018. Prediction of psychosis across protocols and risk cohorts using automated language analysis. World Psychiatry 17(1), pp. 67–75. Available at: https://onlinelibrary.wiley.com/doi/full/10.1002/wps.20491 [Accessed: 10 July 2021].

Data Science Dojo 2019. N-Grams in Natural Language Processing. YouTube . Available at: https://www.youtube.com/watch?v=E_mN90TYnlg&ab_channel=DataScienceDojo [Accessed: 4 September 2021].

David, D. et al. 2018. Why Cognitive Behavioral Therapy Is the Current Gold Standard of Psychotherapy. Frontiers in Psychiatry 9. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5797481/ [Accessed: 3 October 2021].

Davis, S. et al. 2006. Identifying and targeting predictors of drop-out from group cognitive behaviour therapy. Available at:

https://www.researchgate.net/publication/233448274_Identifying_and_targeting_predictors_of_drop-out_from_group_cognitive_behaviour_therapy [Accessed: 8 June 2021].

Demirci, Merve 2018. User experience over time with conversational agents: case study of woebot on supporting subjective well-being. Available at: https://open.metu.edu.tr/handle/11511/27659 [Accessed: 10 June 2021].

Dobson, K.S. 2013. The Science of CBT: Toward a Metacognitive Model of Change? Behavior Therapy 44(2), pp. 224–227. Available at: https://www-sciencedirect-com.abc.cardiff.ac.uk/science/article/pii/S0005789411000724 [Accessed: 24 May 2021].

Donker, T. 2015. Economic evaluations of Internet interventions for mental health: A systematic review. Available at: https://www.researchgate.net/publication/280695200_Economic_evaluations_of_Internet_interventions_for_mental_health_A_systematic_review [Accessed: 24 May 2021].

Cosine Similarity - GeeksforGeeks. 2020. Available at: https://www.geeksforgeeks.org/cosine-similarity/ [Accessed: 4 September 2021].

Ednalyn C. De Dios 2020. From DataFrame to N-Grams - Towards Data Science. Available at: https://towardsdatascience.com/from-dataframe-to-n-grams-e34e29df3460 [Accessed: 4 September 2021].

Fitzpatrick, K.K. et al. 2017. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. JMIR Mental Health 4(2), p. e19. Available at: https://mental.jmir.org/2017/2/e19/ [Accessed: 13 April 2021].

Word Embeddings in NLP - GeeksforGeeks. 2020. Available at: https://www.geeksforgeeks.org/word-embeddings-in-nlp/ [Accessed: 30 October 2021].

Goh, M. 2020. Text Normalization with spaCy and NLTK - Towards Data Science. Available at: https://towardsdatascience.com/text-normalization-with-spacy-and-nltk-1302ff430119 [Accessed: 30 October 2021].

Gratzer, D. and Goldbloom, D. 2019. Open for Business: Chatbots, E-therapies, and the Future of Psychiatry. The Canadian Journal of Psychiatry 64(7), pp. 453–455. Available at: https://journals.sagepub.com/doi/10.1177/0706743719850057 [Accessed: 24 May 2021].

Hofmann, S.G. et al. 2012. The Efficacy of Cognitive Behavioral Therapy: A Review of Meta-analyses. Cognitive Therapy and Research 36(5), pp. 427–440. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3584580/ [Accessed: 3 October 2021].

Hofmann, S.G. et al. 2013. The Science of Cognitive Therapy. Behavior Therapy 44(2), pp. 199–212. Available at: https://www-sciencedirect-com.abc.cardiff.ac.uk/science/article/pii/S0005789411000591?via%3Dihub [Accessed: 24 May 2021].

Informatics 2019. Natural Language Processing and Information Extraction for Biomedicine. Available at: https://vimeo.com/24829353 [Accessed: 27 July 2021].

Inkster, B. et al. 2018. An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. JMIR mHealth and

uHealth 6(11), p. e12106. Available at: https://mhealth.jmir.org/2018/11/e12106/ [Accessed: 15 June 2021].

Knight, W. 2017. Andrew Ng Has a Chatbot That Can Help with Depression. Available at: https://www.technologyreview.com/2017/10/18/148521/andrew-ng-has-a-chatbot-that-can-help-with-depression/ [Accessed: 10 June 2021].

Kuyken, W. 2016. Study finds CBT offers long-term benefits for people with depression — Department of Psychiatry. Available at: https://www.psych.ox.ac.uk/news/study-finds-cbt-offers-long-term-benefits-for-people-with-depression [Accessed: 3 October 2021].

Le Glaz, A. et al. 2021. Machine Learning and Natural Language Processing in Mental Health: Systematic Review. Journal of Medical Internet Research 23(5), p. e15708. Available at: https://www.jmir.org/2021/5/e15708#table4 [Accessed: 18 July 2021].

Li, S. 2018. Named Entity Recognition with NLTK and SpaCy - Towards Data Science. Available at: https://towardsdatascience.com/named-entity-recognition-with-nltk-and-spacy-8c4a7d88e7da [Accessed: 30 July 2021].

Li, S. 2018 A. Named Entity Recognition and Classification with Scikit-Learn. Available at: https://towardsdatascience.com/named-entity-recognition-and-classification-with-scikit-learn-f05372f07ba2 [Accessed: 1 August 2021].

Dorash, M 2017. Machine Learning vs. Rule Based Systems in NLP - Friendly Data - Medium. Available at: https://medium.com/friendly-data/machine-learning-vs-rule-based-systems-in-nlp-5476de53c3b8 [Accessed: 12 August 2021].

Milne-Ives, M. et al. 2020. The Effectiveness of Artificial Intelligence Conversational Agents in Health Care: Systematic Review. Journal of Medical Internet Research 22(10), p. e20346. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7644372/ [Accessed: 16 October 2021].

Munir, S. 2019. Basic Sentiment Analysis using NLTK - Towards Data Science. Available at: https://towardsdatascience.com/basic-binary-sentiment-analysis-using-nltk-c94ba17ae386 [Accessed: 4 September 2021].

NHS 2017. Digital technology helping people take action on their mental health care in North Staffordshire. Available at: https://www.nhsx.nhs.uk/digitise-connect-transform/case-studies/digital-technology-helping-people-take-action-on-their-mental-health-care-in-north-staffordshire/ [Accessed: 30 October 2021].

NHS Health Research Authority. 2017. Internet-delivered cognitive behaviour therapy (iCBT). Available at: https://www.hra.nhs.uk/planning-and-improving-research/application-summaries/research-summaries/internet-delivered-cognitive-behaviour-therapy-icbt-sussex/ [Accessed: 30 October 2021].

Laduram, V. 2020. Conversational Agent: A More Assertive Form of Chatbots. Available at: https://towardsdatascience.com/conversational-agent-a-more-assertive-form-of-chatbots-de6f1c8da8dd [Accessed: 16 October 2021].

LIWC NA. Home Page Available at: https://liwc.wpengine.com/ [Accessed: 1 August 2021].

LIWC NA A. Interpreting LIWC Output. 2015. Available at: http://liwc.wpengine.com/interpreting-liwc-output/ [Accessed: 4 September 2021].

LIWC Results. 2021. Available at: http://www.utpsyc.org/TAT/LIWCTATresults.php [Accessed: 4 September 2021].

Moody, J. 2017. What is TF-IDF? Available at: https://www.97thfloor.com/blog/what-is-tf-idf [Accessed: 27 July 2021].

Monkey Learn. 2014. Guide to Text Classification with Machine Learning & NLP. Available at: https://monkeylearn.com/text-classification/ [Accessed: 12 August 2021].

Nasser, F. et al. [no date]. Choosing an NLP Library for Analyzing Software Documentation: A Systematic Literature Review and a Series of Experiments. Available at: https://cs.adelaide.edu.au/~christoph/msr17.pdf.

Penn Treebank P.O.S. Tags. 2021. Available at: https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html [Accessed: 15 August 2021].

Prochaska, J.J. et al. 2021. A Therapeutic Relational Agent for Reducing Problematic Substance Use (Woebot): Development and Usability Study. Journal of Medical Internet Research 23(3), p. e24850. Available at: https://www.jmir.org/2021/3/e24850 [Accessed: 24 May 2021].

Python Programming Tutorials. 2021. A. Available at: https://pythonprogramming.net/chunking-nltk-tutorial/?completed=/part-of-speech-tagging-nltk-tutorial/ [Accessed: 20 July 2021].

Python Programming Tutorials. 2021. B. Available at: https://pythonprogramming.net/part-of-speech-tagging-nltk-tutorial/ [Accessed: 20 July 2021].

Python Programming Tutorials. 2021. C. Available at: https://pythonprogramming.net/named-entity-recognition-stanford-ner-tagger/ [Accessed: 1 August 2021].

Pybis, J. et al. 2017. The comparative effectiveness and efficiency of cognitive behaviour therapy and generic counselling in the treatment of depression: evidence from the 2nd UK National Audit of psychological therapies. BMC Psychiatry 17(1). Available at: https://bmcpsychiatry.biomedcentral.com/articles/10.1186/s12888-017-1370-7#Sec2 [Accessed: 8 June 2021].

Ratana, R. et al. 2019. A Comprehensive Review of Computational Methods for Automatic Prediction of Schizophrenia With Insight Into Indigenous Populations. Frontiers in Psychiatry 10. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6759015/ [Accessed: 27 July 2021].

Richards, D. et al. 2018. Digital IAPT: the effectiveness & cost-effectiveness of internet-delivered interventions for depression and anxiety disorders in the Improving Access to Psychological Therapies programme: study protocol for a randomised control trial. BMC Psychiatry 18(1). Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5833053/ [Accessed: 24 May 2021].

Madan, R 2019. TF-IDF/Term Frequency Technique: Easiest explanation for Text classification in NLP using Python (Chatbot training on words). Available at: https://medium.com/analytics-vidhya/tf-idf-term-frequency-technique-easiest-explanation-for-text-classification-in-nlp-with-code-8ca3912e58c3 [Accessed: 29 July 2021].

Saad, S. Saberi, B 2017. Sentiment Analysis or Opinion Mining: A Review. Available at: https://www.researchgate.net/publication/320762707_Sentiment_Analysis_or_Opinion_Mining_A_Review [Accessed: 1 August 2021].

Savova, G.K. et al. 2010. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. Journal of the American Medical Informatics Association 17(5), pp. 507–513. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2995668/ [Accessed: 27 July 2021].

Scott, W. 2019. TF-IDF from scratch in python on real world dataset. Available at: https://towardsdatascience.com/tf-idf-for-document-ranking-from-scratch-in-python-on-real-world-dataset-796d339a4089 [Accessed: 27 July 2021].

spaCy 101: Everything you need to know · spaCy Usage Documentation. 2015. Available at: https://spacy.io/usage/spacy-101 [Accessed: 18 July 2021].

Stanford Edu. 2021. Stemming and lemmatization. Available at: https://nlp.stanford.edu/IR-book/html/htmledition/stemming-and-lemmatization-1.html [Accessed: 20 July 2021].

Syarif, I. et al. 2019. Study on Mental Disorder Detection via Social Media Mining. 2019 4th International Conference on Computing, Communications and Security (ICCCS) . Available at: https://ieeexplore.ieee.org/abstract/document/8888096 [Accessed: 10 July 2021].

Suganuma, S. et al. 2018. An Embodied Conversational Agent for Unguided Internet-Based Cognitive Behavior Therapy in Preventative Mental Health: Feasibility and Acceptability Pilot Trial. JMIR Mental Health 5(3), p. e10454. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6092592/ [Accessed: 24 May 2021].

Srivastava, T. 2019. NLP: A quick guide to Stemming - Tushar Srivastava - Medium. Available at: https://medium.com/@tusharsri/nlp-a-quick-guide-to-stemming-60f1ca5db49e#:~:text=Stemming%20is%20basically%20removing%20the%20suffix%20from%20a%20word%20and,word%20which%20is%20%E2%80%9CFly%E2%80%9D. [Accessed: 27 July 2021].

Analytics Vidhya 2020. Text Classification in Natural Language Processing. Available at: https://www.analyticsvidhya.com/blog/2020/12/understanding-text-classification-in-nlp-with-movie-review-example-example/ [Accessed: 12 August 2021]

Vaidyam, A.N. et al. 2019. Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. The Canadian Journal of Psychiatry 64(7), pp. 456–464. Available at: https://pubmed.ncbi.nlm.nih.gov/30897957/ [Accessed: 24 May 2021]

Vimarsh Karbhari 2020. Lemmatization in chatbots - Acing AI - Medium. Available at: https://medium.com/acing-ai/lemmatization-in-chatbots-4a9e7646b9d9#:~:text=Lemmatization%20is%20very%20useful%20when,be%20much%20clearer%20and%20shorter. [Accessed: 20 July 2021].

Kumar, V. et al. 2017. The Effectiveness of Internet-Based Cognitive Behavioral Therapy in Treatment of Psychiatric Disorders. Cureus . Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5659300/ [Accessed: 24 May 2021].

Young, J. Beck, A.T. 1980. Cognitive therapy scale: Rating manual. Unpublished manuscript, 36. https://centralizedtraining.com/resources/Cognitive%20Behavioral%20Treatment%20(CBT)/CTRS%20manual.pdf

Young, K, Klosko, K. 1993. Reinventing Your Life. Published by the Penguin Group.

Warmerdam, L. et al. 2010. Cost-Utility and Cost-Effectiveness of Internet-Based Treatment for Adults With Depressive Symptoms: Randomized Trial. Journal of Medical Internet Research 12(5), p. e53. Available at: https://pubmed.ncbi.nlm.nih.gov/21169166/ [Accessed: 24 May 2021].

Wagner, B. et al. 2014. Internet-based versus face-to-face cognitive-behavioral intervention for depression: A randomized controlled non-inferiority trial. Journal of Affective Disorders 152-154, pp. 113–121. Available at: https://pubmed.ncbi.nlm.nih.gov/23886401/ [Accessed: 24 May 2021].

Waltman, S. et al. 2016. In-Session Stuck Points and Pitfalls of Community Clinicians Learning CBT: Qualitative Investigation. Available at: https://www.researchgate.net/publication/302554312_In-Session_Stuck_Points_and_Pitfalls_of_Community_Clinicians_Learning_CBT_Qualitative_Investigation [Accessed: 8 June 2021].

Wang, Y. Patrick, J. 2009. Cascading Classifiers for Named Entity Recognition in Clinical Notes. Workshop Biomedical Information Extraction. Available at: https://www.researchgate.net/publication/234758575_Cascading_classifiers_for_named_entity_recognition_in_clinical_notes [Accessed: 01/08/2021].

Webb, C.A. et al. 2017. Internet-Based Cognitive-Behavioral Therapy for Depression: Current Progress and Future Directions. Harvard Review of Psychiatry 25(3), pp. 114–122. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5421393/ [Accessed: 24 May 2021]

Willson, R. Branch, R. 2006. Cognitive Behavioural Therapy For Dummies. John Wiley & Sons.

Woebot Health. 2021. FAQ. Available at: https://woebothealth.com/faq/ [Accessed: 10 June 2021].

Xu, R. and Zhang, Q. 2016. Understanding Online Health Groups for Depression: Social Network and Linguistic Perspectives. Journal of Medical Internet Research 18(3), p. e63. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4807247/ [Accessed: 1 August 2021].

Stirman, S. et al. 2013. Community Mental Health Provider Modifications to Cognitive Therapy: Implications for Sustainability. Psychiatric Services 64(10), pp. 1056–1059. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3851297/ [Accessed: 8 June 2021].

Wu, A. et al. 2021. Smartphone apps for depression and anxiety: a systematic review and meta-analysis of techniques to increase engagement. npj Digital Medicine 4(1). Available at: https://www.nature.com/articles/s41746-021-00386-8 [Accessed: 15 June 2021].

Zhang, X (April 2021) Natural Language Processing For Human Computer Interaction (VSI-hci2) - Call for papers - Computers & Electrical Engineering - Journal - Elsevier. 2021. Available at: https://www.journals.elsevier.com/computers-and-electrical-engineering/call-for-papers/natural-language-processing [Accessed: 10 July 2021].

Appendix

Appendix 1 – Text Samples Used In Corpuses

The below table outlines all the text that is contained in all four corpuses.

Case ID	Transcript	Type	Source
1	When my husband is away, I am worried he will never come back	Action Event	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again
1	I keep thinking he may die like my father did and I will never see him again	Belief	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again
1	While he is away this terrifies me, I cry a lot and when he gets home I am so angry he put me through that.	Consequence	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again
2	My wife goes out.	Action Event	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again
2	I keep thinking about when I was a child. My mother and how she used to binge drink, sometimes she used to leave for two or three days.	Belief	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again
2	Now I am worried that Francine will leave for days and I just wait for her to come home. It makes me so lonely.	Consequence	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again
3	Greg and I were at a dinner party, and he was talking to the woman sitting next to him, and he didn't hear what I was saying to him. I got up and left.	Action Event	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again
3	I felt that Greg was attracted to this other woman	Belief	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again
3	I felt totally crushed. When he called me the following day I flew into a rage I was so angry!	Consequence	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again
4	My partner is talking about separating	Action Event	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again
4	I just couldn't cope without him. I can't function in the world alone.	Belief	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again
4	I guess I believe I would die without him	Consequence	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again
5	Greg left me.	Action Event	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again
5	I was totally alone. There was just nothing at all.	Belief	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again

5	I could feel it around me, like an ache.	Consequence	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again
6	My wife keeps doubting me, she's especially suspicious about my business trips. She keeps thinking I am having affairs with other women.	Action Event	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again
6	I wonder if she's the one who wants to be with other men.	Belief	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again
6	She is driving me crazy.	Consequence	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again
7	My husband has gone away on a business trip and while I was gardening my neighbour came by to talk.	Action Event	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again
7	I felt like someone who was running and running, and when I was to tired to run anymore, the bad feelings would catch up with me.	Belief	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again
7	Talking my neighbour, it occurred to me that from the outside, it looked as though I was enjoying myself, like I was a person, who was enjoying my solitude. But I really wasn't. Its all a façade.	Consequence	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again
8	I called Valerie Monday, and it's Wednesday, and she still hasn't called me back	Action Event	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again
8	She should have called me back by now! She has no right to treat me this way!	Belief	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again
8	I am really mad and I am thinking of telling her off	Consequence	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again
9	I have just started a new relationship with Adirenne and she says she loves me, but I have trouble trusting her.	Action Event	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again
9	It's like I keep expecting the whole thing to be a big trick, Like she's gonna turn around and say, "Okay, it's over, I never really loved you, I've been tricking you all along."	Belief	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again
9	When she got home I was beside myself. I shouted at her from the top of my lungs "Where was she, who was she with, what was she doing." At one point I grabbed her and started shaking her. I really scared her!	Consequence	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again
10	Something funny happened the other day. I went into the storage closet I built, and when I switched on the light, the bulb was dead. I was standing there in the dark and all of a sudden I broke out into this cold sweat. I was petrified. It got me thinking, when I was a kit, I used to stand in a closet in the dark, shaking. My father was out there looking for me. It's funny, I never made the connection before. But that was the reason why I was scared.	Action Event	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again

10	I thought something bad might happen to me if I left the storage cupboard.	Belief	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again
10	I was extremely scared to leave.	Consequence	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again
11	My father would be fine one moment and he would then be shouting at me from the top of his lungs another moment. I was a clumsy kid and got into a lot of trouble. My dad used to tell me I would rot in hell.	Action Event	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again
11	I just thought it was happening because I was such a lousy person.	Belief	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again
11	This made me feel unsafe around him	Consequence	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again
12	I was invited to a party last Saturday Night, and I have been dreading it all week.	Action Event	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again
12	I was worrying I would get there and be really nervous and not know what to say. I would act like a jerk. And everyone would be better than me and I would have nothing to offer.	Belief	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again
12	Whenever I thought about it I would always be on the verge of tears.	Consequence	Reinvent Your Life - The Breakthrough Programme to End Negative Behaviour and Feel Great Again

Appendix 2 – Full TF-IDF Analysis For ABC Corpuses

The below outlines words the TF-IDF analysis believes is significant in 12 samples of the Action Event Corpus and calculating each word in this texts significance by comparing it against the other 3 corpuses. The below outlines the significance.

Please note that sample ID correlates with appendix 1 IDs to find text samples please find id number and type Action Event

Word	Tag	Sample ID **1**	Word	Tag	Sample ID **2**	Word	Tag	Sample ID 3
husband	NN	0.2111	wife	NN	0.6933	he	PRP	0.1245
my	PRP$	0.2111	goes	VBZ	0.5198	sitting	VBG	0.0813
worried	JJ	0.1798				hear	NN	0.0813
back	RB	0.1798				saying	VBG	0.0813
away	RB	0.1577				at	IN	0.0813
never	RB	0.1577				dinner	NN	0.0695
come	VB	0.1577				talking	VBG	0.0695
he	PRP	0.1577				woman	NN	0.0695
						next	JJ	0.0547
						got	VBD	0.045
Word	Tag	Sample ID 4	**Word**	**Tag**	**Sample ID 5**	Word	Tag	Sample ID 6
partner	NN	0.4731	left	NN	0.7223	keeps	NNS	0.1999
separating	VBG	0.4731				doubting	VBG	0.13
talking	VBG	0.3574				suspicious	JJ	0.13
						affairs	NNS	0.13
						wife	NN	0.1118
						especially	RB	0.1118
						business	NN	0.1118
						thinking	VBG	0.0989
						my	PRP$	0.0806
Word	Tag	Sample ID 7	**Word**	**Tag**	**Sample ID 8**	Word	Tag	Sample ID 9
gone	VBN	0.1459	called	VBN	0.264	started	VBN	0.1231
trip	NN	0.1459	back	RB	0.1598	new	JJ	0.0986
gardening	NN	0.1459	she	PRP	0.1598	relationship	NN	0.1426
while	IN	0.1459	still	RB	0.144	says	VBZ	0.1093
husband	NN	0.1256				loves	NNS	0.1426
neighbour	NN	0.1256				trouble	NN	0.1426
came	VBD	0.1256				trusting	VBG	0.1426
business	NN	0.1113				she	PRP	0.2163
away	RB	0.1001				but	CC	0.0986
my	PRP$	0.0833						

Word	Tag	Sample ID 10	Word	Tag	Sample ID 11	Word	Tag	Sample ID 12
closet	NN	0.0507	fine	NN	0.0664	dreading	VBG	0.1797
when	WRB	0.0507	shouting	VBG	0.0664	party	NN	0.1557
happened	VBD	0.033	clumsy	NN	0.0664	invited	VBN	0.1256
went	VBD	0.033	dad	NN	0.0664	last	JJ	0.1057
switched	VBN	0.033	tell	NN	0.0664	dreading	VBG	0.1797
bulb	NN	0.033	rot	NN	0.0664	party	NN	0.1557
sudden	JJ	0.033	top	NN	0.0575	invited	VBN	0.1256
broke	NN	0.033	kid	NN	0.0575	last	JJ	0.1057
cold	NN	0.033	lot	NN	0.0575			
looking	VBG	0.033	another	DT	0.0512			
connection	NN	0.033	he	PRP	0.0512			
storage	NN	0.0285	lungs	NNS	0.0463			
standing	VBG	0.0285	father	NN	0.0423			
stand	NN	0.0285	moment	NN	0.0423			
funny	NN	0.0253	got	VBD	0.0334			
dark	NN	0.0253	used	VBN	0.031			
reason	NN	0.0253	one	CD	0.029			
father	NN	0.0229	would	MD	0.0248			
never	RB	0.0229						
made	VBN	0.0192						
got	VBD	0.0177						

Please note that sample ID correlates with appendix 1 IDs to find text samples please find id number and type Beliefs

Word	Tag	Sample ID **1**	Word	Tag	Sample ID **7**	Word	Tag	Sample ID 3
may	MD	0.1617	**she**	PRP	0.1969	attracted	VBN	0.2803
die	NN	0.1378	**keep**	VB	0.0998	felt	NN	0.2398
thinking	NN	0.1208	**mother**	NN	0.0998	woman	NN	0.2398
never	RB	0.1208	**binge**	NN	0.0998			
he	PRP	0.1208	**sometimes**	RB	0.0998			
see	VB	0.1077	**three**	CD	0.0998			
like	IN	0.0969	**how**	WRB	0.0998			
father	NN	0.0969	**used**	VBN	0.0758			
			leave	VB	0.0748			
			when	WRB	0.0748			
			thinking	VBG	0.0668			
			two	CD	0.0668			
Word	Tag	Sample ID 4	**Word**	**Tag**	**Sample ID 5**	Word	Tag	Sample ID 6
cope	NN	0.2181	totally	RB	0.2457	wonder	NN	0.2225

function	NN	0.2181		nothing	NN	0.2457		wants	VBZ	0.2225
without	IN	0.1648						one	CD	0.0913
world	NN	0.1648								
Word	Tag	Sample ID 7		**Word**	**Tag**	**Sample ID 8**		Word	Tag	Sample ID 9
tired	VBN	0.1109		right	NN	0.1737		expecting	VBG	0.0861
run	VB	0.1109		treat	NN	0.1737		whole	JJ	0.0861
bad	JJ	0.1109		this	DT	0.1737		big	JJ	0.0861
feelings	NNS	0.1109		back	RB	0.1328		loved	VBN	0.0861
catch	NN	0.1109		called	VBN	0.1197		tricking	VBG	0.0861
someone	NN	0.0955						keep	VB	0.0744
felt	NN	0.0846						turn	NN	0.066
running	VBG	0.0846						thing	NN	0.0596
when	WRB	0.0846						never	RB	0.0596
like	IN	0.0633						really	RB	0.0596
would	MD	0.0198						around	IN	0.0498
								like	IN	0.0395
Word	Tag	Sample ID 10		Word	Tag	Sample ID 11		Word	Tag	Sample ID 12
might	MD	0.2127		happening	VBG	0.2325		worrying	VBG	0.0822
happen	VB	0.2127		lousy	NN	0.2325		nervous	JJ	0.0822
thought	NN	0.1839		thought	NN	0.1792		act	NN	0.0822
bad	JJ	0.1839						everyone	NN	0.0822
storage	NN	0.1839						better	RBR	0.0822
something	NN	0.1635						know	VB	0.0635
left	NN	0.1635						nothing	NN	0.0635
								really	RB	0.0525
								get	VB	0.0484
								would	MD	0.04
								like	IN	0.0387

Please note that sample ID correlates with appendix 1 IDs to find text samples please find id number and type Consequence

Word	Tag	Sample ID **1**		Word	Tag	Sample ID **7**		Word	Tag	Sample ID 3
he	PRP	0.2458		worried	JJ	0.0994		following	VBG	0.1401
terrifies	NNS	0.1096		leave	VB	0.0994		flew	NN	0.1401
gets	VBZ	0.1096		wait	NN	0.0994		rage	NN	0.1401
angry	JJ	0.1096		days	NNS	0.0873		totally	RB	0.1199
when	WRB	0.1096		come	VB	0.0873		felt	NN	0.1055
and	CC	0.1096		makes	VBZ	0.0873		day	NN	0.1055
cry	NN	0.0934						he	PRP	0.1055
lot	NN	0.0934						called	VBN	0.0944
home	NN	0.0934								
put	NN	0.0934								
away	RB	0.073								

Word	Tag	Sample ID 4
guess	NN	0.3145
believe	VB	0.3145
die	NN	0.2695
without	IN	0.2375
would	MD	0.0341

Word	**Tag**	**Sample ID 5**
could	MD	0.2727
feel	NN	0.2727
around	IN	0.1958
like	IN	0.1507
it	PRP	0.1507

Word	Tag	Sample ID 6
driving	VBG	0.578

Word	Tag	Sample ID 7
enjoying	VBG	0.1202
occurred	VBD	0.0788
looked	VBD	0.0788
my	PRP$	0.0764
though	IN	0.0678
really	RB	0.0601
like	IN	0.0353

Word	**Tag**	**Sample ID 8**
mad	NN	0.2444
telling	VBG	0.2444
really	RB	0.1683
thinking	VBG	0.1683

Word	Tag	Sample ID 9
she	PRP	0.1594
beside	NN	0.0683
shouted	VBN	0.0683
grabbed	NNS	0.0683
top	NN	0.059
point	NN	0.059
started	VBN	0.059
shaking	VBG	0.059
scared	VBN	0.059
home	NN	0.0472
lungs	NNS	0.0472
really	RB	0.0431
got	VBD	0.0337
my	PRP$	0.0337
one	CD	0.0291

Word	Tag	Sample ID 10
extremely	RB	0.4325
scared	VBN	0.3846

Word	Tag	Sample ID 11
unsafe	JJ	0.4319
feel	NN	0.3328
made	VBN	0.2528
around	IN	0.2528

Word	Tag	Sample ID 12
always	RB	0.2162
verge	NN	0.2162
thought	NN	0.1669
would	MD	0.0341

Appendix 3 – List of 20 most significant words inside each corpus.

The following tables identify the top 20 most significant words in each corpus, to see the complete list of TF-IDF analysis please see appendix 2. Please see appendix 4 to see POS Tag Meanings.

TF-IDF Analysis of Action Events Text		
Word	**Total Significance**	**Word POS Tag**
husband	0.2111	NN
left	0.7223	NN
wife	0.6933	NN
goes	0.5198	VBZ
partner	0.4731	NN
separating	0.4731	VBG
talking	0.3574	VBG
called	0.264	VBN
she	0.2163	PRP
my	0.2111	PRP$
keeps	0.1999	NNS
worried	0.1798	JJ
back	0.1798	RB
dreading	0.1797	VBG
away	0.1577	RB
never	0.1577	RB
come	0.1577	VB
he	0.1577	PRP
party	0.1557	NN
gone	0.1459	VBN

TF-IDF Analysis of Beliefs Text		
Word	**Total Significance**	**Word POS Tag**
attracted	0.2803	VBN
totally	0.2457	RB
nothing	0.2457	NN
felt	0.2398	NN
woman	0.2398	NN
happening	0.2325	VBG
lousy	0.2325	NN
wonder	0.2225	NN
wants	0.2225	VBZ
cope	0.2181	NN
function	0.2181	NN
might	0.2127	MD
happen	0.2127	VB
she	0.1969	PRP
thought	0.1839	NN
bad	0.1839	JJ
storage	0.1839	NN
right	0.1737	NN
treat	0.1737	NN
without	0.1648	IN

TF-IDF Analysis of Consequence Text		
Word	**Total Significance**	**Word POS Tag**
driving	0.578	VBG
extremely	0.4325	RB
unsafe	0.4319	JJ
scared	0.3846	VBN
feel	0.3328	NN
guess	0.3145	NN
believe	0.3145	VB
could	0.2727	MD
feel	0.2727	NN
die	0.2695	NN
made	0.2528	VBN
around	0.2528	IN
he	0.2458	PRP
mad	0.2444	NN
telling	0.2444	VBG
without	0.2375	IN
always	0.2162	RB

TF-IDF Analysis of News Sample Text		
Word	**Total Significance**	**Word POS Tag**
left	0.5331	NN
wife	0.4491	NN
feel	0.3739	NN
could	0.338	MD
goes	0.3212	VBZ
without	0.2817	IN
keeps	0.2771	NNS
die	0.2589	NN
extremely	0.2588	RB
talking	0.2586	VBG
scared	0.2507	VBN
around	0.2395	IN
totally	0.2283	RB
made	0.2247	VBN
something	0.2144	NN
just	0.2043	RB
called	0.1975	VBN

goes	0.2243	[('goes', 'VBZ')]		**woman**	0.1964	[('woman', 'NN')]
thought	0.2235	[('thought', 'NN')]		**husband**	0.1933	[('husband', 'NN')]
always	0.2162	[('always', 'RB')]		**thought**	0.1927	[('thought', 'NN')]

Appendix 4 – POS Tag Meaning (Penn Treebank P.O.S. Tags. 2021)

Number	Tag	Description
1.	CC	Coordinating conjunction
2.	CD	Cardinal number
3.	DT	Determiner
4.	EX	Existential *there*
5.	FW	Foreign word
6.	IN	Preposition or subordinating conjunction
7.	JJ	Adjective
8.	JJR	Adjective, comparative
9.	JJS	Adjective, superlative
10.	LS	List item marker
11.	MD	Modal
12.	NN	Noun, singular or mass
13.	NNS	Noun, plural
14.	NNP	Proper noun, singular
15.	NNPS	Proper noun, plural
16.	PDT	Predeterminer
17.	POS	Possessive ending
18.	PRP	Personal pronoun
19.	PRP$	Possessive pronoun
20.	RB	Adverb
21.	RBR	Adverb, comparative
22.	RBS	Adverb, superlative
23.	RP	Particle
24.	SYM	Symbol
25.	TO	*to*
26.	UH	Interjection
27.	VB	Verb, base form
28.	VBD	Verb, past tense
29.	VBG	Verb, gerund or present participle
30.	VBN	Verb, past participle
31.	VBP	Verb, non-3rd person singular present
32.	VBZ	Verb, 3rd person singular present
33.	WDT	Wh-determiner
34.	WP	Wh-pronoun
35.	WP$	Possessive wh-pronoun
36.	WRB	Wh-adverb

Appendix 5 – Sentiment Analysis of all text samples in each corpus

The next set of tables shows the sentiment analysis of all the samples in each corpus at a complete sample level. The random news articles have been added to show a baseline for the sentiment analysis for the rest of the corpus.

Action Events Phrases	Part of Speech Tags	Negative Sentiment	Neutral Sentiment	Positive Sentiment
When my husband is away I am worried he will never come back	['WRB', 'PRP$', 'NN', 'VBZ', 'RB', 'JJ', 'VBP', 'VBN', 'PRP', 'MD', 'RB', 'VB', 'RB']	0.167	0.833	0
Greg and I were at a dinner party, and he was talking to the woman sitting next to him, and he didn't hear what I was saying to him. I got up and left.	['NN', 'CC', 'NN', 'VBD', 'IN', 'DT', 'NN', 'NN', ',', 'CC', 'PRP', 'VBZ', 'VBG', 'TO', 'DT', 'NN', 'VBG', 'JJ', 'TO', 'PRP', ',', 'CC', 'PRP', 'VBD', 'RB', 'VB', 'WP', 'VB', 'VBP', 'VBG', 'TO', 'PRP', '.', 'NN', 'VBD', 'RB', 'CC', 'VBD', '.']	0	0.915	0.085
My partner is talking about separating	['PRP$', 'NN', 'VBZ', 'VBG', 'IN', 'VBG']	0	1	0
Greg left me.	['NN', 'VBD', 'PRP', '.']	0	1	0
My wife keeps doubting me, she's especially suspicious about my business trips. She keeps thinking I am having affairs with other women.	['PRP$', 'NN', 'VB', 'VBG', 'PRP', ',', 'PRP', 'VBZ', 'RB', 'JJ', 'IN', 'PRP$', 'NN', 'NN', '.', 'PRP', 'VBD', 'VBG', 'NN', 'VBP', 'VBG', 'NN', 'IN', 'JJ', 'NN', '.']	0.215	0.785	0
My husband has gone away on a business trip and while I was gardening my neighbour came by to talk.	['PRP$', 'NN', 'NN', 'VBN', 'RB', 'IN', 'DT', 'NN', 'NN', 'CC', 'IN', 'JJ', 'VBP', 'VBG', 'PRP$', 'NN', 'VBD', 'IN', 'TO', 'VB', '.']	0	1	0
I called Valerie Monday, and it's Wednesday, and she still hasn't called me back	['NN', 'VBD', 'NN', 'NN', ',', 'CC', 'PRP', 'VBZ', 'JJ', ',', 'CC', 'PRP', 'RB', 'VBP', 'RB', 'VBD', 'PRP', 'RP']	0	1	0
I have just started a new relationship with Adirenne and she says she loves me, but I have trouble trusting her.	['NNS', 'VBP', 'RB', 'VBN', 'DT', 'JJ', 'NN', 'IN', 'NN', 'CC', 'PRP', 'VBP', 'PRP', 'VB', 'PRP', ',', 'CC', 'NNS', 'VBP', 'NN', 'VBG', 'PRP', '.']	0.145	0.613	0.241
Something funny happened the other day. I went into the storage closet I built, and when I switched on the light, the bulb was dead. I was standing there in the dark and all of a sudden I broke out into this cold sweat. I was petrified. It got me thinking, when I was a kid, I used to stand in a closet in the dark, shaking. My father was out there looking for me. It's funny, I never made the connection before. But that was the reason why I was scared.	['NN', 'JJ', 'VBD', 'DT', 'JJ', 'NN', '.', 'VB', 'VBD', 'IN', 'DT', 'NN', 'NN', 'NN', 'VBN', ',', 'CC', 'WRB', 'NN', 'VBD', 'IN', 'DT', 'NN', ',', 'DT', 'NN', 'NN', 'JJ', '.', 'JJ', 'VBP', 'VBG', 'RB', 'IN', 'DT', 'NN', 'CC', 'DT', 'IN', 'DT', 'JJ', 'NN', 'VBD', 'RP', 'IN', 'DT', 'JJ', 'NN', '.', 'JJ', 'VBP', 'VBN', '.', 'PRP', 'VBD', 'PRP', 'NN', ',', 'WRB', 'NN', 'VBP', 'DT', 'NN', ',', 'NN', 'VBN', 'TO', 'VB', 'IN', 'DT', 'NN', 'IN', 'DT', 'NN', ',', 'VBG', '.', 'PRP$', 'NN', 'NN', 'IN', 'RB', 'VBG', 'IN', 'PRP', '.', 'PRP', 'VBZ', 'JJ', ',', 'JJ', 'RB', 'VBD', 'DT', 'NN', 'IN', '.', 'CC', 'IN', 'VBZ', 'DT', 'NN', 'WRB', 'NN', 'NN', 'VBD', '.']	0.137	0.819	0.044
My father would be fine one moment and he would then be shouting at me from the top of his lungs another moment. I was a clumsy kid and got into a lot of trouble. My dad used to tell me I would rot in hell.	['PRP$', 'NN', 'MD', 'VB', 'JJ', 'CD', 'NN', 'CC', 'PRP', 'MD', 'RB', 'VB', 'VBG', 'IN', 'PRP', 'IN', 'DT', 'NN', 'IN', 'PRP$', 'NN', 'DT', 'NN', '.', 'VB', 'VBP', 'DT', 'JJ', 'NN', 'CC', 'VBD', 'IN', 'DT', 'NN', 'IN', 'NN', '.', 'PRP$', 'NN', 'VBN', 'TO', 'VB', 'PRP', 'VB', 'MD', 'VB', 'IN', 'NN', '.']	0.149	0.777	0.074

I was invited to a party last Saturday Night, and I have been dreading it all week.	['NN', 'VBP', 'VBN', 'TO', 'DT', 'NN', 'JJ', 'JJ', 'NN', ',', 'CC', 'NNS', 'VBP', 'VBN', 'VBG', 'PRP', 'DT', 'NN', '.']	0.188	0.663	0.149
	Total Average	**0.083**	**0.784**	**0.049**

Belief Phrases	Part of Speech Tags	Negative Sentiment	Neutral Sentiment	Positive Sentiment
I am think he may die like my father did and I will never see him again	['NN', 'VBP', 'VBP', 'PRP', 'MD', 'VB', 'IN', 'PRP$', 'NN', 'VBD', 'CC', 'VB', 'MD', 'RB', 'VB', 'PRP', 'RB']	0.201	0.67	0.129
I keep thinking about when I was a child. My mother and how she used to binge drink, sometimes she used to leave for two or three days.	['JJ', 'VBP', 'VBG', 'IN', 'WRB', 'JJ', 'VBP', 'DT', 'NN', '.', 'PRP$', 'NN', 'CC', 'WRB', 'PRP', 'VBD', 'TO', 'VB', 'NN', ',', 'RB', 'PRP', 'VBD', 'TO', 'VB', 'IN', 'CD', 'CC', 'CD', 'NN', '.']	0.048	0.952	0
I felt that Greg was attracted to this other woman	['NN', 'VBD', 'IN', 'NN', 'NN', 'VBD', 'TO', 'DT', 'JJ', 'NN']	0	1	0
I just couldn't cope without him. I can't function in the world alone.	['NN', 'RB', 'MD', 'RB', 'VB', 'IN', 'PRP', '.', 'NN', 'MD', 'RB', 'VB', 'IN', 'DT', 'NN', 'RB', '.']	0.167	0.833	0
I was totally alone. There was just nothing at all.	['NN', 'VBP', 'RB', 'RB', '.', 'EX', 'JJ', 'RB', 'NN', 'IN', 'DT', '.']	0.223	0.777	0
I wonder if she's the one who wants to be with other men.	['JJ', 'VBP', 'IN', 'PRP', 'VBZ', 'DT', 'NN', 'WP', 'VBP', 'TO', 'VB', 'IN', 'JJ', 'NNS', '.']	0	1	0
I felt like someone who was running and running, and when I was to tired to run anymore, the bad feelings would catch up with me.	['NN', 'VBD', 'IN', 'NN', 'WP', 'VBP', 'VBG', 'CC', 'NN', ',', 'CC', 'WRB', 'JJ', 'VBP', 'TO', 'VB', 'TO', 'VB', 'RB', ',', 'DT', 'JJ', 'NN', 'MD', 'VB', 'RP', 'IN', 'PRP', '.']	0.214	0.702	0.084
She should have called me back by now! She has no right to treat me this way!	['PRP', 'MD', 'VB', 'VBN', 'PRP', 'RB', 'IN', 'RB', '.', 'PRP', 'VBD', 'RB', 'JJ', 'TO', 'VB', 'PRP', 'DT', 'NN', '.']	0.107	0.732	0.16
It's like I keep expecting the whole thing to be a big trick, Like she's gonna turn around and say, "Okay, it's over, I never really loved you, I've been tricking you all along.	['PRP', 'VBZ', 'IN', 'JJ', 'VBP', 'VBG', 'DT', 'JJ', 'NN', 'TO', 'VB', 'DT', 'JJ', 'NN', ',', 'IN', 'PRP', 'VBZ', 'NN', 'TO', 'VB', 'RP', 'CC', 'VB', ',', '``', 'UH', ',', 'PRP', 'VBZ', 'IN', ',', 'JJ', 'RB', 'RB', 'VBN', 'PRP', ',', 'VB', 'VBP', 'VBN', 'VBG', 'PRP', 'DT', 'RB', '.']	0.124	0.709	0.166
I thought something bad might happen to me if I left the storage cupboard.	['NNS', 'VBD', 'NN', 'JJ', 'MD', 'VB', 'TO', 'PRP', 'IN', 'NN', 'VBD', 'DT', 'NN', 'NN', '.']	0.241	0.759	0
I just thought it was happening because I was such a lousy person.	['NN', 'RB', 'VBD', 'PRP', 'VBD', 'VBG', 'IN', 'JJ', 'VBP', 'PDT', 'DT', 'JJ', 'NN', '.']	0.28	0.72	0
I was worrying I would get there and be really nervous and not know what to say. I would act like a jerk. And everyone would be better than me and I would have nothing to offer.	['NN', 'VBP', 'VBG', 'NN', 'MD', 'VB', 'EX', 'CC', 'VB', 'RB', 'JJ', 'CC', 'RB', 'VB', 'WP', 'TO', 'VB', '.', 'NN', 'MD', 'VB', 'IN', 'DT', 'NN', '.', 'CC', 'NN', 'MD', 'VB', 'JJR', 'IN', 'PRP', 'CC', 'VB', 'MD', 'VB', 'NN', 'TO', 'VB', '.']	0.182	0.682	0.136
	Total Average	**0.134**	**0.738**	**0.045**

Consequence Phrases	Part of Speech Tags	Negative Sentiment	Neutral Sentiment	Positive Sentiment
When she got home I was beside myself. I shouted at her from the top of my lungs "Where was she, who was she with, what was she doing." At one point I grabbed her and started shaking her. I really scared her!	['WRB', 'PRP', 'VBD', 'NN', 'NN', 'VBP', 'RB', 'PRP', '.', 'VB', 'VBD', 'IN', 'PRP', 'IN', 'DT', 'NN', 'IN', 'PRP$', 'NN', '``', 'WRB', 'NN', 'PRP', ',', 'WP', 'VBD', 'PRP', 'IN', ',', 'WP', 'VBD', 'PRP', 'VBG', '.', '"', 'IN', 'CD', 'NN', 'NN', 'VBD', 'PRP$', 'CC', 'VBD', 'VBG', 'PRP$', '.', 'NN', 'RB', 'VBD', 'PRP', '.']	0.121	0.838	0.042
Whenever I thought about it I would always be on the verge of tears.	['WRB', 'NN', 'VBD', 'IN', 'PRP', 'VB', 'MD', 'RB', 'VB', 'IN', 'DT', 'NN', 'IN', 'NN', '.']	0.147	0.853	0
Talking my neighbour, it occurred to me that from the outside, it looked as though I was enjoying myself, like I was a person, who was enjoying my solitude. But I really wasn't. Its all a façade.	['VBG', 'PRP$', 'NN', ',', 'PRP', 'VBD', 'TO', 'PRP', 'IN', 'IN', 'DT', 'NN', ',', 'PRP', 'VBD', 'DT', 'IN', 'JJ', 'VBP', 'VBG', 'PRP', ',', 'IN', 'NN', 'VBP', 'DT', 'NN', ',', 'WP', 'VBD', 'VBG', 'PRP$', 'NN', '.', 'CC', 'JJ', 'RB', 'JJ', 'NNP', 'NN', '.', 'PRP$', 'PDT', 'DT', 'NN', '.']	0	0.825	0.175
Now I am worried that Francine will leave for days and I just wait for her to come home. It makes me so lonely.	['RB', 'VBZ', 'VBP', 'JJ', 'IN', 'NN', 'MD', 'VB', 'IN', 'NN', 'CC', 'NN', 'RB', 'NN', 'IN', 'PRP$', 'TO', 'VB', 'NN', '.', 'PRP', 'VB', 'PRP', 'RB', 'RB', '.']	0.259	0.741	0
She is driving me crazy.	['PRP', 'VBZ', 'VBG', 'PRP', 'JJ', '.']	0.375	0.625	0
I am really mad and I am thinking of telling her off	['NN', 'VBP', 'RB', 'JJ', 'CC', 'JJ', 'VBP', 'VBG', 'IN', 'VBG', 'PRP$', 'NN']	0.28	0.72	0
I felt totally crushed. When he called me the following day I flew into a rage I was so angry!	['NN', 'VBD', 'RB', 'VBN', '.', 'WRB', 'PRP', 'VBD', 'PRP', 'DT', 'JJ', 'NN', 'VB', 'VBD', 'IN', 'DT', 'NN', 'NN', 'VBP', 'RB', 'JJ', '.']	0.333	0.667	0
I could feel it around me, like an ache.	['NN', 'MD', 'VB', 'PRP', 'IN', 'PRP', ',', 'IN', 'DT', 'NN', '.']	0.234	0.541	0.225
I was extremely scared to leave.	['JJ', 'VBP', 'RB', 'VBN', 'TO', 'VB', '.']	0.608	0.392	0
I guess I believe I would die without him	['JJ', 'NN', 'NN', 'VBP', 'NN', 'MD', 'VB', 'IN', 'PRP']	0.438	0.562	0
While he is away this terrifies me, I cry a lot and when he gets home I am so angry he put me through that.	['IN', 'PRP', 'VBZ', 'RB', 'DT', 'NNS', 'PRP', ',', 'VBZ', 'VBP', 'DT', 'NN', 'CC', 'WRB', 'PRP', 'VB', 'NN', 'NN', 'VBP', 'RB', 'JJ', 'PRP', 'VBD', 'PRP', 'IN', 'DT', '.']	0.379	0.621	0
This made me feel unsafe around him	['DT', 'VBD', 'PRP', 'VB', 'JJ', 'IN', 'PRP']	0	1	0
	Total Average	**0.265**	**0.615**	**0.037**

Random News Articles	Negative Sentiment	Neutral Sentiment	Positive Sentiment
1	0.071	0.903	0.025
2	0.032	0.891	0.077
3	0.025	0.887	0.087
4	0.032	0.902	0.066

5	0.049	0.867	0.084
6	0.071	0.836	0.093

Appendix 6 – Sliding Sentiment Analysis

The following tables shows a sliding sentiment analysis of all corpus samples, bellow you will see a heading (text sample) and snippet (part of a text sample) broken down into several components at a range of 4 words per snippet. Please see appendix 4 to see POS Tag Meanings.

Action Event Phrases	Heading or Snippet?	Tags	Negative Sentiment	Neutral Sentiment	Positive Sentiment
When my husband is away I am worried he will never come back	Heading	['WRB', 'PRP$', 'NN', 'VBZ', 'RB', 'JJ', 'VBP', 'VBN', 'PRP', 'MD', 'RB', 'VB', 'RB']	0.167	0.833	0
When my husband is	Snippet	['WRB', 'PRP$', 'NN', 'VBZ']	0	1	0
my husband is away	Snippet	['PRP$', 'NN', 'VBZ', 'RB']	0	1	0
husband is away I	Snippet	['NN', 'VBZ', 'RB', 'JJ']	0	1	0
is away I am	Snippet	['VBZ', 'RB', 'JJ', 'VBP']	0	1	0
away I am worried	Snippet	['RB', 'JJ', 'VBP', 'JJ']	0.524	0.476	0
I am worried he	Snippet	['NN', 'VBP', 'VBN', 'PRP']	0.524	0.476	0
am worried he will	Snippet	['VBP', 'JJ', 'PRP', 'MD']	0.423	0.577	0
worried he will never	Snippet	['VBN', 'PRP', 'MD', 'RB']	0.423	0.577	0
he will never come	Snippet	['PRP', 'MD', 'RB', 'VB']	0	1	0
will never come back	Snippet	['MD', 'RB', 'VB', 'RB']	0	1	0
My wife goes out.	Snippet	['PRP$', 'NN', 'VB', 'RP', '.']	0	1	0
Greg and I were at a dinner party, and he was talking to the woman sitting next to him, and he didn't hear what I was saying to him. I got up and left.	Heading	['NN', 'CC', 'NN', 'VBD', 'IN', 'DT', 'NN', 'NN', ',', 'CC', 'PRP', 'VBZ', 'VBG', 'TO', 'DT', 'NN', 'VBG', 'JJ', 'TO', 'PRP', ',', 'CC', 'PRP', 'VBD', 'RB', 'VB', 'WP', 'VB', 'VBP', 'VBG', 'TO', 'PRP', '.', 'NN', 'VBD', 'RB', 'CC', 'VBD', '.']	0	0.915	0.085
Greg and I were	Snippet	['NN', 'CC', 'NN', 'VBD']	0	1	0
and I were at	Snippet	['CC', 'NN', 'VBD', 'IN']	0	1	0

I were at a	Snippet	['NN', 'VBD', 'IN', 'DT']	0	1	0
were at a dinner	Snippet	['VBD', 'IN', 'DT', 'NN']	0	1	0
at a dinner party,	Snippet	['IN', 'DT', 'NN', 'NN', ',']	0	0.426	0.574
a dinner party, and	Snippet	['DT', 'NN', 'NN', ',', 'CC']	0	0.426	0.574
dinner party, and he	Snippet	['NN', 'NN', ',', 'CC', 'PRP']	0	0.526	0.474
party, and he was	Snippet	['NN', ',', 'CC', 'PRP', 'VBD']	0	0.526	0.474
and he was talking	Snippet	['CC', 'PRP', 'VBD', 'VBG']	0	1	0
he was talking to	Snippet	['PRP', 'VBZ', 'VBG', 'TO']	0	1	0
was talking to the	Snippet	['NN', 'VBG', 'TO', 'DT']	0	1	0
talking to the woman	Snippet	['VBG', 'TO', 'DT', 'NN']	0	1	0
to the woman sitting	Snippet	['TO', 'DT', 'NN', 'VBG']	0	1	0
the woman sitting next	Snippet	['DT', 'NN', 'VBG', 'JJ']	0	1	0
woman sitting next to	Snippet	['NN', 'VBG', 'JJ', 'TO']	0	1	0
sitting next to him,	Snippet	['VBG', 'JJ', 'TO', 'PRP', ',']	0	1	0
next to him, and	Snippet	['JJ', 'TO', 'PRP', ',', 'CC']	0	1	0
to him, and he	Snippet	['TO', 'PRP', ',', 'CC', 'PRP']	0	1	0
him, and he didn't	Snippet	['PRP', ',', 'CC', 'PRP', 'VBD', 'RB']	0	1	0
and he didn't hear	Snippet	['CC', 'PRP', 'VBD', 'RB', 'VB']	0	1	0
he didn't hear what	Snippet	['PRP', 'VBD', 'RB', 'VB', 'WP']	0	1	0
didn't hear what I	Snippet	['VBD', 'RB', 'VB', 'WP', 'VB']	0	1	0
hear what I was	Snippet	['VB', 'WP', 'VB', 'VBP']	0	1	0
what I was saying	Snippet	['WP', 'VBZ', 'VBP', 'VBG']	0	1	0
I was saying to	Snippet	['NN', 'VBP', 'VBG', 'TO']	0	1	0
was saying to him.	Snippet	['NN', 'VBG', 'TO', 'PRP', '.']	0	1	0
saying to him. I	Snippet	['VBG', 'TO', 'PRP', '.', 'NN']	0	1	0

to him. I got	Snippet	['TO', 'PRP', '.', 'NN', 'VBD']	0	1	0
him. I got up	Snippet	['PRP', '.', 'NN', 'VBD', 'RP']	0	1	0
I got up and	Snippet	['NN', 'VBD', 'RB', 'CC']	0	1	0
got up and left.	Snippet	['VBD', 'RB', 'CC', 'VBD', '.']	0	1	0
My partner is talking about separating	Heading	['PRP$', 'NN', 'VBZ', 'VBG', 'IN', 'VBG']	0	1	0
My partner is talking	Snippet	['PRP$', 'NN', 'VBZ', 'VBG']	0	1	0
partner is talking about	Snippet	['NN', 'VBZ', 'VBG', 'IN']	0	1	0
is talking about separating	Snippet	['VBZ', 'VBG', 'IN', 'VBG']	0	1	0
Greg left me.	Heading	['NN', 'VBD', 'PRP', '.']	0	1	0
My wife keeps doubting me, she's especially suspicious about my business trips. She keeps thinking I am having affairs with other women.	Heading	['PRP$', 'NN', 'VB', 'VBG', 'PRP', ',', 'PRP', 'VBZ', 'RB', 'JJ', 'IN', 'PRP$', 'NN', 'NN', '.', 'PRP', 'VBD', 'VBG', 'NN', 'VBP', 'VBG', 'NN', 'IN', 'JJ', 'NN', '.']	0.215	0.785	0
My wife keeps doubting	Snippet	['PRP$', 'NN', 'VB', 'NN']	0.444	0.556	0
wife keeps doubting me,	Snippet	['NN', 'VB', 'VBG', 'PRP', ',']	0.444	0.556	0
keeps doubting me, she's	Snippet	['VB', 'VBG', 'PRP', ',', 'PRP', 'VBZ']	0.444	0.556	0
doubting me, she's especially	Snippet	['VBG', 'PRP', ',', 'PRP', 'VBZ', 'RB']	0.444	0.556	0
me, she's especially suspicious	Snippet	['PRP', ',', 'PRP', 'VBZ', 'RB', 'JJ']	0.482	0.518	0
she's especially suspicious about	Snippet	['PRP', 'VBZ', 'RB', 'JJ', 'IN']	0.482	0.518	0
especially suspicious about my	Snippet	['RB', 'JJ', 'IN', 'PRP$']	0.482	0.518	0
suspicious about my business	Snippet	['JJ', 'IN', 'PRP$', 'NN']	0.455	0.545	0
about my business trips.	Snippet	['IN', 'PRP$', 'NN', 'NN', '.']	0	1	0
my business trips. She	Snippet	['PRP$', 'NN', 'NN', '.', 'PRP']	0	1	0
business trips. She keeps	Snippet	['NN', 'NN', '.', 'PRP', 'VBD']	0	1	0
trips. She keeps thinking	Snippet	['NN', '.', 'PRP', 'VBD', 'VBG']	0	1	0

She keeps thinking I	Snippet	['PRP', 'VBD', 'NN', 'NN']	0	1	0
keeps thinking I am	Snippet	['VB', 'NN', 'NN', 'VBP']	0	1	0
thinking I am having	Snippet	['VBG', 'NN', 'VBP', 'VBG']	0	1	0
I am having affairs	Snippet	['NN', 'VBP', 'VBG', 'NN']	0	1	0
am having affairs with	Snippet	['VBP', 'VBG', 'NN', 'IN']	0	1	0
having affairs with other	Snippet	['VBG', 'NN', 'IN', 'JJ']	0	1	0
affairs with other women.	Snippet	['NN', 'IN', 'JJ', 'NN', '.']	0	1	0
My husband has gone away on a business trip and while I was gardening my neighbour came by to talk.	Heading	['PRP$', 'NN', 'NN', 'VBN', 'RB', 'IN', 'DT', 'NN', 'NN', 'CC', 'IN', 'JJ', 'VBP', 'VBG', 'PRP$', 'NN', 'VBD', 'IN', 'TO', 'VB', '.']	0	1	0
My husband has gone	Snippet	['PRP$', 'NN', 'NN', 'VBN']	0	1	0
husband has gone away	Snippet	['NN', 'NN', 'VBN', 'RB']	0	1	0
has gone away on	Snippet	['NN', 'VBN', 'RB', 'IN']	0	1	0
gone away on a	Snippet	['VBN', 'RB', 'IN', 'DT']	0	1	0
away on a business	Snippet	['RB', 'IN', 'DT', 'NN']	0	1	0
on a business trip	Snippet	['IN', 'DT', 'NN', 'NN']	0	1	0
a business trip and	Snippet	['DT', 'NN', 'NN', 'CC']	0	1	0
business trip and while	Snippet	['NN', 'NN', 'CC', 'IN']	0	1	0
trip and while I	Snippet	['NN', 'CC', 'IN', 'NN']	0	1	0
and while I was	Snippet	['CC', 'IN', 'JJ', 'VBP']	0	1	0
while I was gardening	Snippet	['IN', 'JJ', 'VBP', 'VBG']	0	1	0
I was gardening my	Snippet	['NN', 'VBP', 'VBG', 'PRP$']	0	1	0
was gardening my neighbour	Snippet	['NN', 'VBG', 'PRP$', 'NN']	0	1	0
gardening my neighbour came	Snippet	['VBG', 'PRP$', 'NN', 'VBD']	0	1	0
my neighbour came by	Snippet	['PRP$', 'NNS', 'VBD', 'IN']	0	1	0
neighbour came by to	Snippet	['NNS', 'VBD', 'IN', 'TO']	0	1	0
came by to talk.	Snippet	['VBD', 'IN', 'TO', 'VB', '.']	0	1	0

I called Valerie Monday, and it's Wednesday, and she still hasn't called me back	Heading	['NN', 'VBD', 'NN', 'NN', ',', 'CC', 'PRP', 'VBZ', 'JJ', ',', 'CC', 'PRP', 'RB', 'VBP', 'RB', 'VBD', 'PRP', 'RP']	0	1	0
I called Valerie Monday,	Snippet	['NN', 'VBD', 'NN', 'NN', ',']	0	1	0
called Valerie Monday, and	Snippet	['VBN', 'NN', 'NN', ',', 'CC']	0	1	0
Valerie Monday, and it's	Snippet	['NN', 'NN', ',', 'CC', 'PRP', 'VBZ']	0	1	0
Monday, and it's Wednesday,	Snippet	['NN', ',', 'CC', 'PRP', 'VBZ', 'JJ', ',']	0	1	0
and it's Wednesday, and	Snippet	['CC', 'PRP', 'VBZ', 'JJ', ',', 'CC']	0	1	0
it's Wednesday, and she	Snippet	['PRP', 'VBZ', 'JJ', ',', 'CC', 'PRP']	0	1	0
Wednesday, and she still	Snippet	['NN', ',', 'CC', 'PRP', 'RB']	0	1	0
and she still hasn't	Snippet	['CC', 'PRP', 'RB', 'VBP', 'RB']	0	1	0
she still hasn't called	Snippet	['PRP', 'RB', 'VBP', 'RB', 'VBD']	0	1	0
still hasn't called me	Snippet	['RB', 'VBP', 'RB', 'VBD', 'PRP']	0	1	0
hasn't called me back	Snippet	['NN', 'RB', 'VBD', 'PRP', 'RP']	0	1	0
I have just started a new relationship with Adirenne and she says she loves me, but I have trouble trusting her.	Heading	['NNS', 'VBP', 'RB', 'VBN', 'DT', 'JJ', 'NN', 'IN', 'NN', 'CC', 'PRP', 'VBP', 'PRP', 'VB', 'PRP', ',', 'CC', 'NNS', 'VBP', 'NN', 'VBG', 'PRP', '.']	0.145	0.613	0.241
I have just started	Snippet	['NNS', 'VBP', 'RB', 'VBN']	0	1	0
have just started a	Snippet	['VB', 'RB', 'VBN', 'DT']	0	1	0
just started a new	Snippet	['RB', 'VBD', 'DT', 'JJ']	0	1	0
started a new relationship	Snippet	['VBD', 'DT', 'JJ', 'NN']	0	1	0
a new relationship with	Snippet	['DT', 'JJ', 'NN', 'IN']	0	1	0
new relationship with Adirenne	Snippet	['JJ', 'NN', 'IN', 'NN']	0	1	0
relationship with Adirenne and	Snippet	['NN', 'IN', 'NN', 'CC']	0	1	0
with Adirenne and she	Snippet	['IN', 'NN', 'CC', 'PRP']	0	1	0

Adirenne and she says	Snippet	['NN', 'CC', 'PRP', 'VBP']	0	1	0
and she says she	Snippet	['CC', 'PRP', 'VBP', 'PRP']	0	1	0
she says she loves	Snippet	['PRP', 'VBP', 'PRP', 'VB']	0	0.448	0.552
says she loves me,	Snippet	['VB', 'PRP', 'VB', 'PRP', ',']	0	0.448	0.552
she loves me, but	Snippet	['PRP', 'VB', 'PRP', ',', 'CC']	0	0.561	0.439
loves me, but I	Snippet	['VB', 'PRP', ',', 'CC', 'NN']	0	0.46	0.54
me, but I have	Snippet	['PRP', ',', 'CC', 'NNS', 'VBP']	0	1	0
but I have trouble	Snippet	['CC', 'NNS', 'VBP', 'NN']	0.64	0.36	0
I have trouble trusting	Snippet	['NNS', 'VBP', 'NN', 'NN']	0.422	0.156	0.422
have trouble trusting her.	Snippet	['VBP', 'NN', 'VBG', 'PRP', '.']	0.365	0.27	0.365
Something funny happened the other day. I went into the storage closet I built, and when I switched on the light, the bulb was dead. I was standing there in the dark and all of a sudden I broke out into this cold sweat. I was petrified. It got me thinking, when I was a kid, I used to stand in a closet in the dark, shaking. My father was out there looking for me. It's funny, I never made the connection before. But that was the reason why I was scared.	Heading	['NN', 'JJ', 'VBD', 'DT', 'JJ', 'NN', '.', 'VB', 'VBD', 'IN', 'DT', 'NN', 'NN', 'NN', 'VBN', ',', 'CC', 'WRB', 'NN', 'VBD', 'IN', 'DT', 'NN', ',', 'DT', 'NN', 'NN', 'JJ', '.', 'JJ', 'VBP', 'VBG', 'RB', 'IN', 'DT', 'NN', 'CC', 'DT', 'IN', 'DT', 'JJ', 'NN', 'VBD', 'RP', 'IN', 'DT', 'JJ', 'NN', '.', 'JJ', 'VBP', 'VBN', '.', 'PRP', 'VBD', 'PRP', 'NN', ',', 'WRB', 'NN', 'VBP', 'DT', 'NN', ',', 'NN', 'VBN', 'TO', 'VB', 'IN', 'DT', 'NN', 'IN', 'DT', 'NN', ',', 'VBG', '.', 'PRP$', 'NN', 'NN', 'IN', 'RB', 'VBG', 'IN', 'PRP', '.', 'PRP', 'VBZ', 'JJ', ',', 'JJ', 'RB', 'VBD', 'DT', 'NN', 'IN', '.', 'CC', 'IN', 'VBZ', 'DT', 'NN', 'WRB', 'NN', 'NN', 'VBD', '.']	0.137	0.819	0.044

Something funny happened the	Snippet	['NN', 'JJ', 'VBD', 'DT']	0	0.508	0.492
funny happened the other	Snippet	['NN', 'VBD', 'DT', 'JJ']	0	0.508	0.492
happened the other day.	Snippet	['VBD', 'DT', 'JJ', 'NN', '.']	0	1	0
the other day. I	Snippet	['DT', 'JJ', 'NN', '.', 'NN']	0	1	0
other day. I went	Snippet	['JJ', 'NN', '.', 'VB', 'VBD']	0	1	0
day. I went into	Snippet	['NN', '.', 'VB', 'VBD', 'IN']	0	1	0
I went into the	Snippet	['JJ', 'VBD', 'IN', 'DT']	0	1	0
went into the storage	Snippet	['VBD', 'IN', 'DT', 'NN']	0	1	0
into the storage closet	Snippet	['IN', 'DT', 'NN', 'NN']	0	1	0
the storage closet I	Snippet	['DT', 'NN', 'NN', 'NN']	0	1	0
storage closet I built,	Snippet	['NN', 'NN', 'NN', 'VBD', ',']	0	1	0
closet I built, and	Snippet	['NN', 'NN', 'VBN', ',', 'CC']	0	1	0
I built, and when	Snippet	['NN', 'VBN', ',', 'CC', 'WRB']	0	1	0
built, and when I	Snippet	['VBN', ',', 'CC', 'WRB', 'NN']	0	1	0
and when I switched	Snippet	['CC', 'WRB', 'NN', 'VBD']	0	1	0
when I switched on	Snippet	['WRB', 'NN', 'VBD', 'IN']	0	1	0
I switched on the	Snippet	['NN', 'VBD', 'IN', 'DT']	0	1	0
switched on the light,	Snippet	['VBN', 'IN', 'DT', 'NN', ',']	0	1	0
on the light, the	Snippet	['IN', 'DT', 'NN', ',', 'DT']	0	1	0
the light, the bulb	Snippet	['DT', 'NN', ',', 'DT', 'NN']	0	1	0
light, the bulb was	Snippet	['NN', ',', 'DT', 'NN', 'NN']	0	1	0
the bulb was dead.	Snippet	['DT', 'NN', 'NN', 'JJ', '.']	0.589	0.411	0
bulb was dead. I	Snippet	['JJ', 'NN', 'JJ', '.', 'NN']	0.683	0.317	0
was dead. I was	Snippet	['NN', 'JJ', '.', 'NN', 'VBP']	0.683	0.317	0
dead. I was standing	Snippet	['JJ', '.', 'JJ', 'VBP', 'VBG']	0.683	0.317	0
I was standing there	Snippet	['NN', 'VBP', 'VBG', 'RB']	0	1	0

was standing there in	Snippet	['NN', 'VBG', 'RB', 'IN']	0	1	0
standing there in the	Snippet	['VBG', 'RB', 'IN', 'DT']	0	1	0
there in the dark	Snippet	['RB', 'IN', 'DT', 'NN']	0	1	0
in the dark and	Snippet	['IN', 'DT', 'NN', 'CC']	0	1	0
the dark and all	Snippet	['DT', 'NN', 'CC', 'DT']	0	1	0
dark and all of	Snippet	['NN', 'CC', 'DT', 'IN']	0	1	0
and all of a	Snippet	['CC', 'DT', 'IN', 'DT']	0	1	0
all of a sudden	Snippet	['DT', 'IN', 'DT', 'JJ']	0	1	0
of a sudden I	Snippet	['IN', 'DT', 'JJ', 'NN']	0	1	0
a sudden I broke	Snippet	['DT', 'JJ', 'NN', 'VBD']	0.737	0.263	0
sudden I broke out	Snippet	['JJ', 'NN', 'VBD', 'RP']	0.583	0.417	0
I broke out into	Snippet	['NN', 'VBD', 'RP', 'IN']	0.583	0.417	0
broke out into this	Snippet	['VBD', 'RP', 'IN', 'DT']	0.483	0.517	0
out into this cold	Snippet	['RB', 'IN', 'DT', 'JJ']	0	1	0
into this cold sweat.	Snippet	['IN', 'DT', 'JJ', 'NN', '.']	0	1	0
this cold sweat. I	Snippet	['DT', 'JJ', 'NN', '.', 'NN']	0	1	0
cold sweat. I was	Snippet	['JJ', 'NN', '.', 'NN', 'VBP']	0	1	0
sweat. I was petrified.	Snippet	['NN', '.', 'JJ', 'VBP', 'VBN', '.']	0.636	0.364	0
I was petrified. It	Snippet	['NN', 'VBP', 'VBN', '.', 'PRP']	0.636	0.364	0
was petrified. It got	Snippet	['NN', 'VBD', '.', 'PRP', 'VBD']	0.538	0.462	0
petrified. It got me	Snippet	['VBN', '.', 'PRP', 'VBD', 'PRP']	0.538	0.462	0
It got me thinking,	Snippet	['PRP', 'VBD', 'PRP', 'NN', ',']	0	1	0
got me thinking, when	Snippet	['VBD', 'PRP', 'NN', ',', 'WRB']	0	1	0
me thinking, when I	Snippet	['PRP', 'NN', ',', 'WRB', 'NN']	0	1	0
thinking, when I was	Snippet	['NN', ',', 'WRB', 'NN', 'VBP']	0	1	0
when I was a	Snippet	['WRB', 'NN', 'VBP', 'DT']	0	1	0
I was a kid,	Snippet	['NN', 'VBP', 'DT', 'NN', ',']	0	1	0
was a kid, I	Snippet	['VB', 'DT', 'NN', ',', 'NN']	0	1	0
a kid, I used	Snippet	['DT', 'NN', ',', 'NN', 'VBD']	0	1	0

kid, I used to	Snippet	['NN', ',', 'NN', 'VBN', 'TO']	0	1	0
I used to stand	Snippet	['NN', 'VBN', 'TO', 'VB']	0	1	0
used to stand in	Snippet	['VBN', 'TO', 'VB', 'IN']	0	1	0
to stand in a	Snippet	['TO', 'VB', 'IN', 'DT']	0	1	0
stand in a closet	Snippet	['NN', 'IN', 'DT', 'NN']	0	1	0
in a closet in	Snippet	['IN', 'DT', 'NN', 'IN']	0	1	0
a closet in the	Snippet	['DT', 'NN', 'IN', 'DT']	0	1	0
closet in the dark,	Snippet	['NN', 'IN', 'DT', 'NN', ',']	0	1	0
in the dark, shaking.	Snippet	['IN', 'DT', 'NN', ',', 'VBG', '.']	0.362	0.638	0
the dark, shaking. My	Snippet	['DT', 'NN', ',', 'VBG', '.', 'PRP$']	0.362	0.638	0
dark, shaking. My father	Snippet	['NN', ',', 'VBG', '.', 'PRP$', 'NN']	0.362	0.638	0
shaking. My father was	Snippet	['VBG', '.', 'PRP$', 'NN', 'NN']	0.362	0.638	0
My father was out	Snippet	['PRP$', 'NN', 'NN', 'IN']	0	1	0
father was out there	Snippet	['RB', 'VBZ', 'RP', 'RB']	0	1	0
was out there looking	Snippet	['NN', 'IN', 'RB', 'VBG']	0	1	0
out there looking for	Snippet	['IN', 'RB', 'VBG', 'IN']	0	1	0
there looking for me.	Snippet	['RB', 'VBG', 'IN', 'PRP', '.']	0	1	0
looking for me. It's	Snippet	['VBG', 'IN', 'PRP', '.', 'PRP', 'VBZ']	0	1	0
for me. It's funny,	Snippet	['IN', 'PRP', '.', 'PRP', 'VBZ', 'JJ', ',']	0	0.508	0.492
me. It's funny, I	Snippet	['PRP', '.', 'PRP', 'VBZ', 'JJ', ',', 'JJ']	0	0.408	0.592
It's funny, I never	Snippet	['PRP', 'VBZ', 'JJ', ',', 'JJ', 'RB']	0	0.408	0.592
funny, I never made	Snippet	['NN', ',', 'NN', 'RB', 'VBD']	0	0.408	0.592
I never made the	Snippet	['NN', 'RB', 'VBD', 'DT']	0	1	0
never made the connection	Snippet	['RB', 'VBD', 'DT', 'NN']	0	1	0
made the connection before.	Snippet	['VBN', 'DT', 'NN', 'IN', '.']	0	1	0
the connection before. But	Snippet	['DT', 'NN', 'IN', '.', 'CC']	0	1	0
connection before. But that	Snippet	['NN', 'IN', '.', 'CC', 'IN']	0	1	0

before. But that was	Snippet	['IN', '.', 'CC', 'IN', 'NN']	0	1	0
But that was the	Snippet	['CC', 'DT', 'VBZ', 'DT']	0	1	0
that was the reason	Snippet	['DT', 'VBD', 'DT', 'NN']	0	1	0
was the reason why	Snippet	['IN', 'DT', 'NN', 'WRB']	0	1	0
the reason why I	Snippet	['DT', 'NN', 'WRB', 'NN']	0	1	0
reason why I was	Snippet	['NN', 'WRB', 'NN', 'NN']	0	1	0
why I was scared.	Snippet	['WRB', 'JJ', 'NN', 'VBD', '.']	0.592	0.408	0
My father would be fine one moment and he would then be shouting at me from the top of his lungs another moment. I was a clumsy kid and got into a lot of trouble. My dad used to tell me I would rot in hell.	Heading	['PRP$', 'NN', 'MD', 'VB', 'JJ', 'CD', 'NN', 'CC', 'PRP', 'MD', 'RB', 'VB', 'VBG', 'IN', 'PRP', 'IN', 'DT', 'NN', 'IN', 'PRP$', 'NN', 'DT', 'NN', '.', 'VB', 'VBP', 'DT', 'JJ', 'NN', 'CC', 'VBD', 'IN', 'DT', 'NN', 'IN', 'NN', '.', 'PRP$', 'NN', 'VBN', 'TO', 'VB', 'PRP', 'VB', 'MD', 'VB', 'IN', 'NN', '.']	0.149	0.777	0.074
My father would be	Snippet	['PRP$', 'NN', 'MD', 'VB']	0	1	0
father would be fine	Snippet	['NN', 'MD', 'VB', 'JJ']	0	0.625	0.375
would be fine one	Snippet	['MD', 'VB', 'JJ', 'CD']	0	0.625	0.375
be fine one moment	Snippet	['VB', 'JJ', 'CD', 'NN']	0	0.625	0.375
fine one moment and	Snippet	['JJ', 'CD', 'NN', 'CC']	0	0.625	0.375
one moment and he	Snippet	['CD', 'NN', 'CC', 'PRP']	0	1	0
moment and he would	Snippet	['NN', 'CC', 'PRP', 'MD']	0	1	0
and he would then	Snippet	['CC', 'PRP', 'MD', 'RB']	0	1	0
he would then be	Snippet	['PRP', 'MD', 'RB', 'VB']	0	1	0
would then be shouting	Snippet	['MD', 'RB', 'VB', 'VBG']	0	1	0
then be shouting at	Snippet	['RB', 'VB', 'VBG', 'IN']	0	1	0
be shouting at me	Snippet	['VB', 'VBG', 'IN', 'PRP']	0	1	0
shouting at me from	Snippet	['VBG', 'IN', 'PRP', 'IN']	0	1	0
at me from the	Snippet	['IN', 'PRP', 'IN', 'DT']	0	1	0

me from the top	Snippet	['PRP', 'IN', 'DT', 'NN']	0	0.625	0.375
from the top of	Snippet	['IN', 'DT', 'NN', 'IN']	0	0.625	0.375
the top of his	Snippet	['DT', 'NN', 'IN', 'PRP$']	0	0.625	0.375
top of his lungs	Snippet	['NN', 'IN', 'PRP$', 'NN']	0	0.625	0.375
of his lungs another	Snippet	['IN', 'PRP$', 'NN', 'DT']	0	1	0
his lungs another moment.	Snippet	['PRP$', 'NN', 'DT', 'NN', '.']	0	1	0
lungs another moment. I	Snippet	['NN', 'DT', 'NN', '.', 'NN']	0	1	0
another moment. I was	Snippet	['DT', 'NN', '.', 'NN', 'VBP']	0	1	0
moment. I was a	Snippet	['NN', '.', 'VB', 'VBP', 'DT']	0	1	0
I was a clumsy	Snippet	['NN', 'VBP', 'DT', 'NN']	0	1	0
was a clumsy kid	Snippet	['VB', 'DT', 'JJ', 'NN']	0	1	0
a clumsy kid and	Snippet	['DT', 'JJ', 'NN', 'CC']	0	1	0
clumsy kid and got	Snippet	['JJ', 'NN', 'CC', 'VBD']	0	1	0
kid and got into	Snippet	['NN', 'CC', 'VBD', 'IN']	0	1	0
and got into a	Snippet	['CC', 'VBD', 'IN', 'DT']	0	1	0
got into a lot	Snippet	['VBD', 'IN', 'DT', 'NN']	0	1	0
into a lot of	Snippet	['IN', 'DT', 'NN', 'IN']	0	1	0
a lot of trouble.	Snippet	['DT', 'NN', 'IN', 'NN', '.']	0.574	0.426	0
lot of trouble. My	Snippet	['NN', 'IN', 'NN', '.', 'PRP$']	0.474	0.526	0
of trouble. My dad	Snippet	['IN', 'NN', '.', 'PRP$', 'NN']	0.474	0.526	0
trouble. My dad used	Snippet	['NN', '.', 'PRP$', 'NN', 'VBN']	0.474	0.526	0
My dad used to	Snippet	['PRP$', 'NN', 'VBN', 'TO']	0	1	0
dad used to tell	Snippet	['NN', 'VBN', 'TO', 'VB']	0	1	0
used to tell me	Snippet	['VBN', 'TO', 'VB', 'PRP']	0	1	0
to tell me I	Snippet	['TO', 'VB', 'PRP', 'VB']	0	1	0
tell me I would	Snippet	['VB', 'PRP', 'VB', 'MD']	0	1	0
me I would rot	Snippet	['PRP', 'NN', 'MD', 'VB']	0	1	0

I would rot in	Snippet	['NN', 'MD', 'VB', 'IN']	0	1	0
would rot in hell.	Snippet	['MD', 'VB', 'IN', 'NN', '.']	0.605	0.395	0
I was invited to a party last Saturday Night, and I have been dreading it all week.	Heading	['NN', 'VBP', 'VBN', 'TO', 'DT', 'NN', 'JJ', 'JJ', 'NN', ',', 'CC', 'NNS', 'VBP', 'VBN', 'VBG', 'PRP', 'DT', 'NN', '.']	0.188	0.663	0.149
I was invited to	Snippet	['NN', 'VBP', 'VBN', 'TO']	0	1	0
was invited to a	Snippet	['RB', 'VBN', 'TO', 'DT']	0	1	0
invited to a party	Snippet	['VBN', 'TO', 'DT', 'NN']	0	0.426	0.574
to a party last	Snippet	['TO', 'DT', 'NN', 'JJ']	0	0.426	0.574
a party last Saturday	Snippet	['DT', 'NN', 'JJ', 'NN']	0	0.426	0.574
party last Saturday Night,	Snippet	['NN', 'JJ', 'JJ', 'NN', ',']	0	0.526	0.474
last Saturday Night, and	Snippet	['JJ', 'JJ', 'NN', ',', 'CC']	0	1	0
Saturday Night, and I	Snippet	['JJ', 'NN', ',', 'CC', 'NN']	0	1	0
Night, and I have	Snippet	['NN', ',', 'CC', 'NNS', 'VBP']	0	1	0
and I have been	Snippet	['CC', 'NNS', 'VBP', 'VBN']	0	1	0
I have been dreading	Snippet	['NNS', 'VBP', 'VBN', 'VBG']	0.63	0.37	0
have been dreading it	Snippet	['VBP', 'VBN', 'VBG', 'PRP']	0.531	0.469	0
been dreading it all	Snippet	['VBN', 'VBG', 'PRP', 'DT']	0.531	0.469	0
dreading it all week.	Snippet	['VBG', 'PRP', 'DT', 'NN', '.']	0.531	0.469	0

Belief Phrases	Heading or Snippet?	Tags	Negative Sentiment	Neutral Sentiment	Positive Sentiment
I am think he may die like my father did and I will never see him again	Heading	['NN', 'VBP', 'VBP', 'PRP', 'MD', 'VB', 'IN', 'PRP$', 'NN', 'VBD', 'CC', 'VB', 'MD', 'RB', 'VB', 'PRP', 'RB']	0.201	0.67	0.129
I am think he	Snippet	['NN', 'VBP', 'NN', 'PRP']	0	1	0
am think he may	Snippet	['VBP', 'VB', 'PRP', 'MD']	0	1	0

think he may die	Snippet	['NN', 'PRP', 'MD', 'VB']	0.565	0.435	0
he may die like	Snippet	['PRP', 'MD', 'VB', 'IN']	0.464	0.238	0.298
may die like my	Snippet	['MD', 'VB', 'IN', 'PRP$']	0.464	0.238	0.298
die like my father	Snippet	['NN', 'IN', 'PRP$', 'NN']	0.464	0.238	0.298
like my father did	Snippet	['IN', 'PRP$', 'NN', 'VBD']	0	0.545	0.455
my father did and	Snippet	['PRP$', 'NN', 'VBD', 'CC']	0	1	0
father did and I	Snippet	['NN', 'VBD', 'CC', 'VB']	0	1	0
did and I will	Snippet	['VBD', 'CC', 'VB', 'MD']	0	1	0
and I will never	Snippet	['CC', 'NN', 'MD', 'RB']	0	1	0
I will never see	Snippet	['NN', 'MD', 'RB', 'VB']	0	1	0
will never see him	Snippet	['MD', 'RB', 'VB', 'PRP']	0	1	0
never see him again	Snippet	['RB', 'VBP', 'PRP', 'RB']	0	1	0
I keep thinking about when I was a child. My mother and how she used to binge drink, sometimes she used to leave for two or three days.	Heading	['JJ', 'VBP', 'VBG', 'IN', 'WRB', 'JJ', 'VBP', 'DT', 'NN', '.', 'PRP$', 'NN', 'CC', 'WRB', 'PRP', 'VBD', 'TO', 'VB', 'NN', ',', 'RB', 'PRP', 'VBD', 'TO', 'VB', 'IN', 'CD', 'CC', 'CD', 'NN', '.']	0.048	0.952	0
I keep thinking about	Snippet	['JJ', 'VBP', 'VBG', 'IN']	0	1	0
keep thinking about when	Snippet	['VB', 'VBG', 'IN', 'WRB']	0	1	0
thinking about when I	Snippet	['VBG', 'IN', 'WRB', 'NN']	0	1	0
about when I was	Snippet	['IN', 'WRB', 'NN', 'VBP']	0	1	0
when I was a	Snippet	['WRB', 'NN', 'VBP', 'DT']	0	1	0
I was a child.	Snippet	['NN', 'VBP', 'DT', 'NN', '.']	0	1	0
was a child. My	Snippet	['VB', 'DT', 'NN', '.', 'PRP$']	0	1	0
a child. My mother	Snippet	['DT', 'NN', '.', 'PRP$', 'NN']	0	1	0
child. My mother and	Snippet	['NN', '.', 'PRP$', 'NN', 'CC']	0	1	0

My mother and how	Snippet	['PRP$', 'NN', 'CC', 'WRB']	0	1	0
mother and how she	Snippet	['NN', 'CC', 'WRB', 'PRP']	0	1	0
and how she used	Snippet	['CC', 'WRB', 'PRP', 'VBD']	0	1	0
how she used to	Snippet	['WRB', 'PRP', 'VBD', 'TO']	0	1	0
she used to binge	Snippet	['PRP', 'VBD', 'TO', 'VB']	0	1	0
used to binge drink,	Snippet	['VBN', 'TO', 'VB', 'NN', ',']	0	1	0
to binge drink, sometimes	Snippet	['TO', 'VB', 'NN', ',', 'RB']	0	1	0
binge drink, sometimes she	Snippet	['NN', 'NN', ',', 'RB', 'PRP']	0	1	0
drink, sometimes she used	Snippet	['NN', ',', 'RB', 'PRP', 'VBD']	0	1	0
sometimes she used to	Snippet	['RB', 'PRP', 'VBD', 'TO']	0	1	0
she used to leave	Snippet	['PRP', 'VBD', 'TO', 'VB']	0.286	0.714	0
used to leave for	Snippet	['VBN', 'TO', 'VB', 'IN']	0.286	0.714	0
to leave for two	Snippet	['TO', 'VB', 'IN', 'CD']	0.286	0.714	0
leave for two or	Snippet	['VB', 'IN', 'CD', 'CC']	0.286	0.714	0
for two or three	Snippet	['IN', 'CD', 'CC', 'CD']	0	1	0
two or three days.	Snippet	['CD', 'CC', 'CD', 'NN', '.']	0	1	0
I felt that Greg was attracted to this other woman	Heading	['NN', 'VBD', 'IN', 'NN', 'NN', 'VBD', 'TO', 'DT', 'JJ', 'NN']	0	1	0
I felt that Greg	Snippet	['NN', 'VBD', 'IN', 'NN']	0	1	0
felt that Greg was	Snippet	['VBD', 'IN', 'NN', 'NN']	0	1	0
that Greg was attacked	Snippet	['DT', 'VBZ', 'RB', 'VBN']	0	1	0
Greg was attracted to	Snippet	['NN', 'NN', 'VBD', 'TO']	0	1	0
was attracted to this	Snippet	['RB', 'VBN', 'TO', 'DT']	0	1	0
attracted to this other	Snippet	['VBN', 'TO', 'DT', 'JJ']	0	1	0
to this other woman	Snippet	['TO', 'DT', 'JJ', 'NN']	0	1	0
I just couldn't cope without him. I can't function in the world alone.	Heading	['NN', 'RB', 'MD', 'RB', 'VB', 'IN', 'PRP', '.', 'NN', 'MD', 'RB', 'VB', 'IN', 'DT', 'NN', 'RB', '.']	0.167	0.833	0

I just couldn't cope	Snippet	['NN', 'RB', 'MD', 'RB', 'VB']	0	1	0
just couldn't cope without	Snippet	['RB', 'MD', 'RB', 'VB', 'IN']	0	1	0
couldn't cope without him.	Snippet	['MD', 'RB', 'VB', 'IN', 'PRP', '.']	0	1	0
cope without him. I	Snippet	['NN', 'IN', 'PRP', '.', 'NN']	0	1	0
without him. I can't	Snippet	['IN', 'PRP', '.', 'NN', 'MD', 'RB']	0	1	0
him. I can't function	Snippet	['PRP', '.', 'NN', 'MD', 'RB', 'VB']	0	1	0
I can't function in	Snippet	['NN', 'MD', 'RB', 'VB', 'IN']	0	1	0
can't function in the	Snippet	['MD', 'RB', 'VB', 'IN', 'DT']	0	1	0
function in the world	Snippet	['NN', 'IN', 'DT', 'NN']	0	1	0
in the world alone.	Snippet	['IN', 'DT', 'NN', 'RB', '.']	0.4	0.6	0
I was totally alone. There was just nothing at all.	Heading	['NN', 'VBP', 'RB', 'RB', '.', 'EX', 'JJ', 'RB', 'NN', 'IN', 'DT', '.']	0.223	0.777	0
I was totally alone.	Snippet	['NN', 'VBP', 'RB', 'RB', '.']	0.534	0.466	0
was totally alone. There	Snippet	['NN', 'RB', 'RB', '.', 'EX']	0.433	0.567	0
totally alone. There was	Snippet	['RB', 'RB', '.', 'EX', 'NN']	0.433	0.567	0
alone. There was just	Snippet	['RB', '.', 'EX', 'VBZ', 'RB']	0.4	0.6	0
There was just nothing	Snippet	['RB', 'VBZ', 'RB', 'NN']	0	1	0
was just nothing at	Snippet	['NN', 'RB', 'NN', 'IN']	0	1	0
just nothing at all.	Snippet	['RB', 'NN', 'IN', 'DT', '.']	0	1	0
I wonder if she's the one who wants to be with other men.	Heading	['JJ', 'VBP', 'IN', 'PRP', 'VBZ', 'DT', 'NN', 'WP', 'VBP', 'TO', 'VB', 'IN', 'JJ', 'NNS', '.']	0	1	0
I wonder if she's	Snippet	['JJ', 'VBP', 'IN', 'PRP', 'VBZ']	0	1	0
wonder if she's the	Snippet	['NN', 'IN', 'PRP', 'VBZ', 'DT']	0	1	0
if she's the one	Snippet	['IN', 'PRP', 'VBZ', 'DT', 'NN']	0	1	0
she's the one who	Snippet	['PRP', 'VBZ', 'DT', 'NN', 'WP']	0	1	0
the one who wants	Snippet	['DT', 'NN', 'WP', 'VBP']	0	1	0

one who wants to	Snippet	['CD', 'WP', 'VBP', 'TO']	0	1	0
who wants to be	Snippet	['WP', 'VBP', 'TO', 'VB']	0	1	0
wants to be with	Snippet	['NN', 'TO', 'VB', 'IN']	0	1	0
to be with other	Snippet	['TO', 'VB', 'IN', 'JJ']	0	1	0
be with other men.	Snippet	['VB', 'IN', 'JJ', 'NNS', '.']	0	1	0
I felt like someone who was running and running, and when I was to tired to run anymore, the bad feelings would catch up with me.	Heading	['NN', 'VBD', 'IN', 'NN', 'WP', 'VBP', 'VBG', 'CC', 'NN', ',', 'CC', 'WRB', 'JJ', 'VBP', 'TO', 'VB', 'TO', 'VB', 'RB', ',', 'DT', 'JJ', 'NN', 'MD', 'VB', 'RP', 'IN', 'PRP', '.']	0.214	0.702	0.084
I felt like someone	Snippet	['NN', 'VBD', 'IN', 'NN']	0	0.444	0.556
felt like someone who	Snippet	['NNS', 'IN', 'NN', 'WP']	0	0.545	0.455
like someone who was	Snippet	['IN', 'NN', 'WP', 'VBP']	0	0.545	0.455
someone who was running	Snippet	['NN', 'WP', 'VBP', 'VBG']	0	1	0
who was running and	Snippet	['WP', 'VBP', 'VBG', 'CC']	0	1	0
was running and running,	Snippet	['NN', 'VBG', 'CC', 'NN', ',']	0	1	0
running and running, and	Snippet	['VBG', 'CC', 'NN', ',', 'CC']	0	1	0
and running, and when	Snippet	['CC', 'NN', ',', 'CC', 'WRB']	0	1	0
running, and when I	Snippet	['NN', ',', 'CC', 'WRB', 'NN']	0	1	0
and when I was	Snippet	['CC', 'WRB', 'NN', 'VBP']	0	1	0
when I was to	Snippet	['WRB', 'NN', 'VBP', 'TO']	0	1	0
I was to tired	Snippet	['NN', 'VBP', 'TO', 'VB']	0.592	0.408	0
was to tired to	Snippet	['NN', 'TO', 'VBN', 'TO']	0.492	0.508	0
to tired to run	Snippet	['TO', 'VB', 'TO', 'VB']	0.492	0.508	0
tired to run anymore,	Snippet	['VBN', 'TO', 'VB', 'RB', ',']	0.492	0.508	0
to run anymore, the	Snippet	['TO', 'VB', 'RB', ',', 'DT']	0	1	0
run anymore, the bad	Snippet	['NN', 'RB', ',', 'DT', 'JJ']	0.538	0.462	0
anymore, the bad feelings	Snippet	['RB', ',', 'DT', 'JJ', 'NN']	0.538	0.462	0

the bad feelings would	Snippet	['DT', 'JJ', 'NN', 'MD']	0.538	0.462	0
bad feelings would catch	Snippet	['JJ', 'NN', 'MD', 'VB']	0.538	0.462	0
feelings would catch up	Snippet	['VBG', 'MD', 'VB', 'RP']	0	1	0
would catch up with	Snippet	['MD', 'VB', 'RP', 'IN']	0	1	0
catch up with me.	Snippet	['VB', 'RP', 'IN', 'PRP', '.']	0	1	0
She should have called me back by now! She has no right to treat me this way!	Heading	['PRP', 'MD', 'VB', 'VBN', 'PRP', 'RB', 'IN', 'RB', '.', 'PRP', 'VBD', 'RB', 'JJ', 'TO', 'VB', 'PRP', 'DT', 'NN', '.']	0.107	0.732	0.16
She should have called	Snippet	['PRP', 'MD', 'VB', 'VBN']	0	1	0
should have called me	Snippet	['MD', 'VB', 'VBN', 'PRP']	0	1	0
have called me back	Snippet	['VB', 'VBN', 'PRP', 'RB']	0	1	0
called me back by	Snippet	['VBN', 'PRP', 'RB', 'IN']	0	1	0
me back by now!	Snippet	['PRP', 'RP', 'IN', 'RB', '.']	0	1	0
back by now! She	Snippet	['RB', 'IN', 'RB', '.', 'PRP']	0	1	0
by now! She has	Snippet	['IN', 'RB', '.', 'PRP', 'VBD']	0	1	0
now! She has no	Snippet	['RB', '.', 'PRP', 'VBD', 'DT']	0.454	0.546	0
She has no right	Snippet	['PRP', 'VBD', 'DT', 'NN']	0.423	0.577	0
has no right to	Snippet	['NN', 'DT', 'NN', 'TO']	0.423	0.577	0
no right to treat	Snippet	['DT', 'NN', 'TO', 'VB']	0.319	0.29	0.391
right to treat me	Snippet	['RB', 'TO', 'VB', 'PRP']	0	0.526	0.474
to treat me this	Snippet	['TO', 'VB', 'PRP', 'DT']	0	0.526	0.474
treat me this way!	Snippet	['VB', 'PRP', 'DT', 'NN', '.']	0	0.501	0.499
It's like I keep expecting the whole thing to be a big trick, Like she's gonna turn around and say, "Okay, it's over, I never really loved you, I've been tricking you all along.	Heading	['PRP', 'VBZ', 'IN', 'JJ', 'VBP', 'VBG', 'DT', 'JJ', 'NN', 'TO', 'VB', 'DT', 'JJ', 'NN', ',', 'IN', 'PRP', 'VBZ', 'NN', 'TO', 'VB', 'RP', 'CC', 'VB', ',', '``', 'UH', ',', 'PRP', 'VBZ', 'IN', ',', 'JJ', 'RB', 'RB', 'VBN',	0.124	0.709	0.166

		'PRP', ',', 'VB', 'VBP', 'VBN', 'VBG', 'PRP', 'DT', 'RB', '.']			
It's like I keep	Snippet	['PRP', 'VBZ', 'IN', 'NN', 'VBP']	0	0.444	0.556
like I keep expecting	Snippet	['IN', 'JJ', 'VBP', 'VBG']	0	0.444	0.556
I keep expecting the	Snippet	['JJ', 'VBP', 'VBG', 'DT']	0	1	0
keep expecting the whole	Snippet	['VB', 'VBG', 'DT', 'NN']	0	1	0
expecting the whole thing	Snippet	['VBG', 'DT', 'JJ', 'NN']	0	1	0
the whole thing to	Snippet	['DT', 'JJ', 'NN', 'TO']	0	1	0
whole thing to be	Snippet	['JJ', 'NN', 'TO', 'VB']	0	1	0
thing to be a	Snippet	['NN', 'TO', 'VB', 'DT']	0	1	0
to be a big	Snippet	['TO', 'VB', 'DT', 'JJ']	0	1	0
be a big trick,	Snippet	['VB', 'DT', 'JJ', 'NN', ',']	0.375	0.625	0
a big trick, Like	Snippet	['DT', 'JJ', 'NN', ',', 'IN']	0.255	0.213	0.532
big trick, Like she's	Snippet	['JJ', 'NN', ',', 'IN', 'PRP', 'VBZ']	0.211	0.351	0.439
trick, Like she's gonna	Snippet	['NN', ',', 'IN', 'PRP', 'VBZ', 'NN', 'TO']	0.211	0.351	0.439
Like she's gonna turn	Snippet	['IN', 'PRP', 'VBZ', 'NN', 'TO', 'VB']	0	0.545	0.455
she's gonna turn around	Snippet	['PRP', 'VBZ', 'NN', 'TO', 'VB', 'RP']	0	1	0
gonna turn around and	Snippet	['NN', 'TO', 'VB', 'RP', 'CC']	0	1	0
turn around and say,	Snippet	['NN', 'RP', 'CC', 'VB', ',']	0	1	0
around and say, "Okay,	Snippet	['IN', 'CC', 'VB', ',', '``', 'NN', ',']	0	1	0
and say, "Okay, it's	Snippet	['CC', 'VB', ',', '``', 'UH', ',', 'PRP', 'VBZ']	0	1	0
say, "Okay, it's over,	Snippet	['VB', ',', '``', 'UH', ',', 'PRP', 'VBZ', 'RP', ',']	0	1	0
"Okay, it's over, I	Snippet	['``', 'NN', ',', 'PRP', 'VBZ', 'IN', ',', 'NN']	0	1	0
it's over, I never	Snippet	['PRP', 'VBZ', 'IN', ',', 'JJ', 'RB']	0	1	0
over, I never really	Snippet	['IN', ',', 'JJ', 'RB', 'RB']	0	1	0
I never really loved	Snippet	['NN', 'RB', 'RB', 'VBD']	0.627	0.373	0
never really loved you,	Snippet	['RB', 'RB', 'VBN', 'PRP', ',']	0.529	0.471	0

really loved you, I've	Snippet	['RB', 'VBN', 'PRP', ',', 'VB', 'VBP']	0	0.417	0.583
loved you, I've been	Snippet	['VBN', 'PRP', ',', 'VB', 'VBP', 'VBN']	0	0.435	0.565
you, I've been tricking	Snippet	['PRP', ',', 'VB', 'VBP', 'VBN', 'VBG']	0	0.732	0.268
I've been tricking you	Snippet	['NN', 'VBP', 'VBN', 'VBG', 'PRP']	0	0.732	0.268
been tricking you all	Snippet	['VBN', 'VBG', 'PRP', 'DT']	0	0.732	0.268
tricking you all along.	Snippet	['VBG', 'PRP', 'DT', 'RB', '.']	0	0.732	0.268
I thought something bad might happen to me if I left the storage cupboard.	Heading	['NNS', 'VBD', 'NN', 'JJ', 'MD', 'VB', 'TO', 'PRP', 'IN', 'NN', 'VBD', 'DT', 'NN', 'NN', '.']	0.241	0.759	0
I thought something bad	Snippet	['NNS', 'VBD', 'NN', 'JJ']	0.636	0.364	0
thought something bad might	Snippet	['VBN', 'NN', 'JJ', 'MD']	0.538	0.462	0
something bad might happen	Snippet	['NN', 'JJ', 'MD', 'VB']	0.538	0.462	0
bad might happen to	Snippet	['JJ', 'MD', 'VB', 'TO']	0.538	0.462	0
might happen to me	Snippet	['MD', 'VB', 'TO', 'PRP']	0	1	0
happen to me if	Snippet	['VB', 'TO', 'PRP', 'IN']	0	1	0
to me if I	Snippet	['TO', 'PRP', 'IN', 'NN']	0	1	0
me if I left	Snippet	['PRP', 'IN', 'NN', 'VBD']	0	1	0
if I left the	Snippet	['IN', 'JJ', 'VBD', 'DT']	0	1	0
I left the storage	Snippet	['NN', 'VBD', 'DT', 'NN']	0	1	0
left the storage cupboard.	Snippet	['VBD', 'DT', 'NN', 'NN', '.']	0	1	0
I just thought it was happening because I was such a lousy person.	Heading	['NN', 'RB', 'VBD', 'PRP', 'VBD', 'VBG', 'IN', 'JJ', 'VBP', 'PDT', 'DT', 'JJ', 'NN', '.']	0.28	0.72	0
I just thought it	Snippet	['NN', 'RB', 'VBD', 'PRP']	0	1	0
just thought it was	Snippet	['RB', 'VBN', 'PRP', 'VB']	0	1	0
thought it was happening	Snippet	['NN', 'PRP', 'VBZ', 'VBG']	0	1	0
it was happening because	Snippet	['PRP', 'VBD', 'VBG', 'IN']	0	1	0
was happening because I	Snippet	['NN', 'VBG', 'IN', 'NN']	0	1	0

happening because I was	Snippet	['VBG', 'IN', 'NN', 'VBP']	0	1	0
because I was such	Snippet	['IN', 'JJ', 'NNS', 'JJ']	0	1	0
I was such a	Snippet	['JJ', 'VBP', 'PDT', 'DT']	0	1	0
was such a lousy	Snippet	['NNS', 'PDT', 'DT', 'JJ']	0.636	0.364	0
such a lousy person.	Snippet	['JJ', 'DT', 'JJ', 'NN', '.']	0.636	0.364	0
I was worrying I would get there and be really nervous and not know what to say. I would act like a jerk. And everyone would be better than me and I would have nothing to offer.	Heading	['NN', 'VBP', 'VBG', 'NN', 'MD', 'VB', 'EX', 'CC', 'VB', 'RB', 'JJ', 'CC', 'RB', 'VB', 'WP', 'TO', 'VB', '.', 'NN', 'MD', 'VB', 'IN', 'DT', 'NN', '.', 'CC', 'NN', 'MD', 'VB', 'JJR', 'IN', 'PRP', 'CC', 'VB', 'MD', 'VB', 'NN', 'TO', 'VB', '.']	0.182	0.682	0.136
I was worrying I	Snippet	['NN', 'VBP', 'VBG', 'NN']	0.706	0.294	0
was worrying I would	Snippet	['IN', 'VBG', 'NN', 'MD']	0.545	0.455	0
worrying I would get	Snippet	['NN', 'NN', 'MD', 'VB']	0.545	0.455	0
I would get there	Snippet	['NN', 'MD', 'VB', 'RB']	0	1	0
would get there and	Snippet	['MD', 'VB', 'RB', 'CC']	0	1	0
get there and be	Snippet	['VB', 'EX', 'CC', 'VB']	0	1	0
there and be really	Snippet	['EX', 'CC', 'VB', 'RB']	0	1	0
and be really nervous	Snippet	['CC', 'VB', 'RB', 'JJ']	0.444	0.556	0
be really nervous and	Snippet	['VB', 'RB', 'JJ', 'CC']	0.444	0.556	0
really nervous and not	Snippet	['RB', 'JJ', 'CC', 'RB']	0.444	0.556	0
nervous and not know	Snippet	['JJ', 'CC', 'RB', 'VB']	0.412	0.588	0
and not know what	Snippet	['CC', 'RB', 'VB', 'WP']	0	1	0
not know what to	Snippet	['RB', 'VB', 'WP', 'TO']	0	1	0
know what to say.	Snippet	['VB', 'WP', 'TO', 'VB', '.']	0	1	0
what to say. I	Snippet	['WP', 'TO', 'VB', '.', 'VB']	0	1	0
to say. I would	Snippet	['TO', 'VB', '.', 'NN', 'MD']	0	1	0
say. I would act	Snippet	['NN', '.', 'NN', 'MD', 'VB']	0	1	0
I would act like	Snippet	['NN', 'MD', 'VB', 'IN']	0	0.444	0.556
would act like a	Snippet	['MD', 'VB', 'IN', 'DT']	0	0.444	0.556

act like a jerk.	Snippet	['NN', 'IN', 'DT', 'NN', '.']	0.407	0.169	0.424
like a jerk. And	Snippet	['IN', 'DT', 'NN', '.', 'CC']	0.407	0.169	0.424
a jerk. And everyone	Snippet	['DT', 'NN', '.', 'CC', 'NN']	0.545	0.455	0
jerk. And everyone would	Snippet	['NN', '.', 'CC', 'NN', 'MD']	0.444	0.556	0
And everyone would be	Snippet	['CC', 'NN', 'MD', 'VB']	0	1	0
everyone would be better	Snippet	['NN', 'MD', 'VB', 'JJR']	0	0.508	0.492
would be better than	Snippet	['MD', 'VB', 'JJR', 'IN']	0	0.508	0.492
be better than me	Snippet	['VB', 'JJR', 'IN', 'PRP']	0	0.508	0.492
better than me and	Snippet	['JJR', 'IN', 'PRP', 'CC']	0	0.508	0.492
than me and I	Snippet	['IN', 'PRP', 'CC', 'VB']	0	1	0
me and I would	Snippet	['PRP', 'CC', 'VB', 'MD']	0	1	0
and I would have	Snippet	['CC', 'NN', 'MD', 'VB']	0	1	0
I would have nothing	Snippet	['NN', 'MD', 'VB', 'NN']	0	1	0
would have nothing to	Snippet	['MD', 'VB', 'NN', 'TO']	0	1	0
have nothing to offer.	Snippet	['VB', 'NN', 'TO', 'VB', '.']	0	1	0

Consequence Phrases	Heading or Snippet?	Tags	Negative Sentiment	Neutral Sentiment	Positive Sentiment
While he is away this terrifies me, I cry a lot and when he gets home I am so angry he put me through that.	Heading	['IN', 'PRP', 'VBZ', 'RB', 'DT', 'NNS', 'PRP', ',', 'VBZ', 'VBP', 'DT', 'NN', 'CC', 'WRB', 'PRP', 'VB', 'NN', 'NN', 'VBP', 'RB', 'JJ', 'PRP', 'VBD', 'PRP', 'IN', 'DT', '.']	0.379	0.621	0
While he is away	Snippet	['IN', 'PRP', 'VBZ', 'RB']	0	1	0
he is away this	Snippet	['PRP', 'VBZ', 'RB', 'DT']	0	1	0
is away this terrifies	Snippet	['VBZ', 'RB', 'DT', 'NNS']	0.586	0.414	0
away this terrifies me,	Snippet	['RB', 'DT', 'NNS', 'PRP', ',']	0.545	0.455	0
this terrifies me, I	Snippet	['DT', 'VBZ', 'PRP', ',', 'VB']	0.643	0.357	0

terrifies me, I cry	Snippet	['NNS', 'PRP', ',', 'VBZ', 'NN']	0.87	0.13	0
me, I cry a	Snippet	['PRP', ',', 'VBZ', 'VBP', 'DT']	0.756	0.244	0
I cry a lot	Snippet	['NN', 'VBP', 'DT', 'NN']	0.756	0.244	0
cry a lot and	Snippet	['NN', 'DT', 'NN', 'CC']	0.608	0.392	0
a lot and when	Snippet	['DT', 'NN', 'CC', 'WRB']	0	1	0
lot and when he	Snippet	['NN', 'CC', 'WRB', 'PRP']	0	1	0
and when he gets	Snippet	['CC', 'WRB', 'PRP', 'VB']	0	1	0
when he gets home	Snippet	['WRB', 'PRP', 'VB', 'NN']	0	1	0
he gets home I	Snippet	['PRP', 'VB', 'NN', 'NN']	0	1	0
gets home I am	Snippet	['VB', 'NN', 'NN', 'VBP']	0	1	0
home I am so	Snippet	['NN', 'NN', 'VBP', 'RB']	0	1	0
I am so angry	Snippet	['NN', 'VBP', 'RB', 'JJ']	0.642	0.358	0
am so angry he	Snippet	['VBP', 'RB', 'JJ', 'PRP']	0.545	0.455	0
so angry he put	Snippet	['RB', 'JJ', 'PRP', 'VBD']	0.545	0.455	0
angry he put me	Snippet	['JJ', 'PRP', 'VBD', 'PRP']	0.524	0.476	0
he put me through	Snippet	['PRP', 'VBD', 'PRP', 'IN']	0	1	0
put me through that.	Snippet	['VB', 'PRP', 'IN', 'DT', '.']	0	1	0
Now I am worried that Francine will leave for days and I just wait for her to come home. It makes me so lonely.	Heading	['RB', 'VBZ', 'VBP', 'JJ', 'IN', 'NN', 'MD', 'VB', 'IN', 'NN', 'CC', 'NN', 'RB', 'NN', 'IN', 'PRP$', 'TO', 'VB', 'NN', '.', 'PRP', 'VB', 'PRP', 'RB', 'RB', '.']	0.259	0.741	0
Now I am worried	Snippet	['RB', 'VBZ', 'VBP', 'JJ']	0.524	0.476	0
I am worried that	Snippet	['NN', 'VBP', 'VBN', 'IN']	0.524	0.476	0
am worried that Francine	Snippet	['VBP', 'JJ', 'IN', 'NN']	0.423	0.577	0
worried that Francine will	Snippet	['VBD', 'DT', 'NN', 'MD']	0.423	0.577	0
that Francine will leave	Snippet	['DT', 'NN', 'MD', 'VB']	0.286	0.714	0

Francine will leave for	Snippet	['NN', 'MD', 'VB', 'IN']	0.286	0.714	0
will leave for days	Snippet	['MD', 'VB', 'IN', 'NN']	0.286	0.714	0
leave for days and	Snippet	['NN', 'IN', 'NN', 'CC']	0.286	0.714	0
for days and I	Snippet	['IN', 'NN', 'CC', 'NN']	0	1	0
days and I just	Snippet	['NN', 'CC', 'NN', 'RB']	0	1	0
and I just wait	Snippet	['CC', 'VB', 'RB', 'NN']	0	1	0
I just wait for	Snippet	['NN', 'RB', 'NN', 'IN']	0	1	0
just wait for her	Snippet	['RB', 'NN', 'IN', 'PRP$']	0	1	0
wait for her to	Snippet	['NN', 'IN', 'PRP$', 'TO']	0	1	0
for her to come	Snippet	['IN', 'PRP$', 'TO', 'VB']	0	1	0
her to come home.	Snippet	['PRP$', 'TO', 'VB', 'NN', '.']	0	1	0
to come home. It	Snippet	['TO', 'VB', 'NN', '.', 'PRP']	0	1	0
come home. It makes	Snippet	['VBN', 'NN', '.', 'PRP', 'VB']	0	1	0
home. It makes me	Snippet	['NN', '.', 'PRP', 'VB', 'PRP']	0	1	0
It makes me so	Snippet	['PRP', 'VB', 'PRP', 'IN']	0	1	0
makes me so lonely.	Snippet	['VB', 'PRP', 'RB', 'RB', '.']	0.519	0.481	0
I felt totally crushed. When he called me the following day I flew into a rage I was so angy!	Heading	['NN', 'VBD', 'RB', 'VBN', '.', 'WRB', 'PRP', 'VBD', 'PRP', 'DT', 'JJ', 'NN', 'VB', 'VBD', 'IN', 'DT', 'NN', 'NN', 'VBP', 'RB', 'JJ', '.']	0.333	0.667	0
I felt totally crushed.	Snippet	['NN', 'VBD', 'RB', 'VBN', '.']	0.607	0.393	0
felt totally crushed. When	Snippet	['VBN', 'RB', 'VBN', '.', 'WRB']	0.508	0.492	0
totally crushed. When he	Snippet	['RB', 'VBN', '.', 'WRB', 'PRP']	0.508	0.492	0
crushed. When he called	Snippet	['VBN', '.', 'WRB', 'PRP', 'VBD']	0.483	0.517	0
When he called me	Snippet	['WRB', 'PRP', 'VBD', 'PRP']	0	1	0
he called me the	Snippet	['PRP', 'VBD', 'PRP', 'DT']	0	1	0

called me the following	Snippet	['VBN', 'PRP', 'DT', 'VBG']	0	1	0
me the following day	Snippet	['PRP', 'DT', 'JJ', 'NN']	0	1	0
the following day I	Snippet	['DT', 'JJ', 'NN', 'VB']	0	1	0
following day I flew	Snippet	['VBG', 'NN', 'VB', 'VBD']	0	1	0
day I flew into	Snippet	['NN', 'VB', 'VBD', 'IN']	0	1	0
I flew into a	Snippet	['NN', 'VBD', 'IN', 'DT']	0	1	0
flew into a rage	Snippet	['RB', 'IN', 'DT', 'NN']	0.643	0.357	0
into a rage I	Snippet	['IN', 'DT', 'NN', 'NN']	0.783	0.217	0
a rage I was	Snippet	['DT', 'NN', 'NN', 'VBP']	0.783	0.217	0
rage I was so	Snippet	['NN', 'NN', 'VBP', 'RB']	0.643	0.357	0
I was so angy!	Snippet	['NN', 'VBP', 'RB', 'JJ', '.']	0	1	0
I guess I believe I would die without him	Heading	['JJ', 'NN', 'NN', 'VBP', 'NN', 'MD', 'VB', 'IN', 'PRP']	0.438	0.562	0
I guess I believe	Snippet	['JJ', 'NN', 'NN', 'VBP']	0	1	0
guess I believe I	Snippet	['NN', 'NN', 'VBP', 'NN']	0	1	0
I believe I would	Snippet	['JJ', 'VBP', 'NN', 'MD']	0	1	0
believe I would die	Snippet	['NN', 'NN', 'MD', 'VB']	0.661	0.339	0
I would die without	Snippet	['NN', 'MD', 'VB', 'IN']	0.661	0.339	0
would die without him	Snippet	['MD', 'VB', 'IN', 'PRP']	0.565	0.435	0
I could feel it around me, like an ache.	Heading	['NN', 'MD', 'VB', 'PRP', 'IN', 'PRP', ',', 'IN', 'DT', 'NN', '.']	0.234	0.541	0.225
I could feel it	Snippet	['NN', 'MD', 'VB', 'PRP']	0	1	0
could feel it around	Snippet	['MD', 'VB', 'PRP', 'RP']	0	1	0
feel it around me,	Snippet	['VB', 'PRP', 'IN', 'PRP', ',']	0	1	0
it around me, like	Snippet	['PRP', 'IN', 'PRP', ',', 'IN']	0	0.545	0.455
around me, like an	Snippet	['IN', 'PRP', ',', 'IN', 'DT']	0	0.545	0.455
me, like an ache.	Snippet	['PRP', ',', 'IN', 'DT', 'NN', '.']	0.366	0.282	0.352

She is driving me crazy.	Heading	['PRP', 'VBZ', 'VBG', 'PRP', 'JJ', '.']	0.375	0.625	0
She is driving me	Snippet	['PRP', 'VBZ', 'VBG', 'PRP']	0	1	0
is driving me crazy.	Snippet	['VBZ', 'VBG', 'PRP', 'JJ', '.']	0.444	0.556	0
Talking my neighbour, it occurred to me that from the outside, it looked as though I was enjoying myself, like I was a person, who was enjoying my solitude. But I really wasn't. Its all a façade.	Heading	['VBG', 'PRP$', 'NN', ',', 'PRP', 'VBD', 'TO', 'PRP', 'IN', 'IN', 'DT', 'NN', ',', 'PRP', 'VBD', 'DT', 'IN', 'JJ', 'VBP', 'VBG', 'PRP', ',', 'IN', 'NN', 'VBP', 'DT', 'NN', ',', 'WP', 'VBD', 'VBG', 'PRP$', 'NN', '.', 'CC', 'JJ', 'RB', 'JJ', 'NNP', 'NN', '.', 'PRP$', 'PDT', 'DT', 'NN', '.']	0	0.825	0.175
Talking my neighbour, it	Snippet	['VBG', 'PRP$', 'NN', ',', 'PRP']	0	1	0
my neighbour, it occurred	Snippet	['PRP$', 'NN', ',', 'PRP', 'VBD']	0	1	0
neighbour, it occurred to	Snippet	['RB', ',', 'PRP', 'VBD', 'TO']	0	1	0
it occurred to me	Snippet	['PRP', 'VBD', 'TO', 'PRP']	0	1	0
occurred to me that	Snippet	['VBD', 'TO', 'PRP', 'IN']	0	1	0
to me that from	Snippet	['TO', 'PRP', 'IN', 'IN']	0	1	0
me that from the	Snippet	['PRP', 'IN', 'IN', 'DT']	0	1	0
that from the outside,	Snippet	['IN', 'IN', 'DT', 'NN', ',']	0	1	0
from the outside, it	Snippet	['IN', 'DT', 'NN', ',', 'PRP']	0	1	0
the outside, it looked	Snippet	['DT', 'NN', ',', 'PRP', 'VBD']	0	1	0
outside, it looked as	Snippet	['RB', ',', 'PRP', 'VBD', 'DT']	0	1	0
it looked as though	Snippet	['PRP', 'VBD', 'DT', 'NN']	0	1	0
looked as though I	Snippet	['VBD', 'DT', 'NN', 'NN']	0	1	0
as though I was	Snippet	['DT', 'NN', 'NN', 'VBP']	0	1	0
though I was enjoying	Snippet	['IN', 'JJ', 'VBP', 'VBG']	0	0.37	0.63
I was enjoying myself,	Snippet	['NN', 'VBP', 'VBG', 'PRP', ',']	0	0.37	0.63

was enjoying myself, like	Snippet	['NN', 'VBG', 'PRP', ',', 'IN']	0	0.253	0.747
enjoying myself, like I	Snippet	['VBG', 'PRP', ',', 'IN', 'NN']	0	0.145	0.855
myself, like I was	Snippet	['PRP', ',', 'IN', 'NN', 'VBP']	0	0.444	0.556
like I was a	Snippet	['IN', 'NN', 'VBP', 'DT']	0	0.286	0.714
I was a person,	Snippet	['NN', 'VBP', 'DT', 'NN', ',']	0	1	0
was a person, who	Snippet	['VB', 'DT', 'NN', ',', 'WP']	0	1	0
a person, who was	Snippet	['DT', 'NN', ',', 'WP', 'VBD']	0	1	0
person, who was enjoying	Snippet	['NN', ',', 'WP', 'VBD', 'VBG']	0	0.469	0.531
who was enjoying my	Snippet	['WP', 'VBP', 'VBG', 'PRP$']	0	0.469	0.531
was enjoying my solitude.	Snippet	['NN', 'VBG', 'PRP$', 'NN', '.']	0	0.469	0.531
enjoying my solitude. But	Snippet	['VBG', 'PRP$', 'NN', '.', 'CC']	0	0.577	0.423
my solitude. But I	Snippet	['PRP$', 'NN', '.', 'CC', 'JJ']	0	1	0
solitude. But I really	Snippet	['NN', '.', 'CC', 'VBP', 'RB']	0	1	0
But I really wasn’t.	Snippet	['CC', 'VBP', 'RB', 'JJ', 'NNP', 'NN', '.']	0	1	0
I really wasn’t. Its	Snippet	['NN', 'RB', 'JJ', 'NNP', 'NN', '.', 'PRP$']	0	1	0
really wasn’t. Its all	Snippet	['RB', 'JJ', 'NNP', 'NN', '.', 'PRP$', 'DT']	0	1	0
wasn’t. Its all a	Snippet	['NN', 'NNP', 'NN', '.', 'PRP$', 'PDT', 'DT']	0	1	0
Its all a façade.	Snippet	['PRP$', 'PDT', 'DT', 'NN', '.']	0	1	0
I am really mad and I am thinking of telling her off	Heading	['NN', 'VBP', 'RB', 'JJ', 'CC', 'JJ', 'VBP', 'VBG', 'IN', 'VBG', 'PRP$', 'NN']	0.28	0.72	0
I am really mad	Snippet	['NN', 'VBP', 'RB', 'JJ']	0.636	0.364	0
am really mad and	Snippet	['VBP', 'RB', 'JJ', 'CC']	0.538	0.462	0
really mad and I	Snippet	['RB', 'JJ', 'CC', 'NN']	0.636	0.364	0
mad and I am	Snippet	['NN', 'CC', 'NN', 'VBP']	0.615	0.385	0
and I am thinking	Snippet	['CC', 'VB', 'VBP', 'VBG']	0	1	0

I am thinking of	Snippet	['NN', 'VBP', 'NN', 'IN']	0	1	0
am thinking of telling	Snippet	['VBP', 'VBG', 'IN', 'VBG']	0	1	0
thinking of telling her	Snippet	['NN', 'IN', 'VBG', 'PRP$']	0	1	0
of telling her off	Snippet	['IN', 'VBG', 'PRP$', 'NN']	0	1	0
When she got home I was beside myself. I shouted at her from the top of my lungs "Where was she, who was she with, what was she doing." At one point I grabbed her and started shaking her. I really scared her!	Heading	['WRB', 'PRP', 'VBD', 'NN', 'NN', 'VBP', 'RB', 'PRP', '.', 'VB', 'VBD', 'IN', 'PRP', 'IN', 'DT', 'NN', 'IN', 'PRP$', 'NN', '``', 'WRB', 'NN', 'PRP', ',', 'WP', 'VBD', 'PRP', 'IN', ',', 'WP', 'VBD', 'PRP', 'VBG', '.', '''', 'IN', 'CD', 'NN', 'NN', 'VBD', 'PRP$', 'CC', 'VBD', 'VBG', 'PRP$', '.', 'NN', 'RB', 'VBD', 'PRP', '.']	0.121	0.838	0.042
When she got home	Snippet	['WRB', 'PRP', 'VBD', 'NN']	0	1	0
she got home I	Snippet	['PRP', 'VBD', 'NN', 'NN']	0	1	0
got home I was	Snippet	['VBD', 'NN', 'NN', 'VBP']	0	1	0
home I was beside	Snippet	['NN', 'NN', 'VBD', 'NN']	0	1	0
I was beside myself.	Snippet	['NN', 'VBD', 'RB', 'PRP', '.']	0	1	0
was beside myself. I	Snippet	['NN', 'NN', 'PRP', '.', 'VB']	0	1	0
beside myself. I shouted	Snippet	['RB', 'PRP', '.', 'NN', 'VBD']	0	1	0
myself. I shouted at	Snippet	['PRP', '.', 'VB', 'VBD', 'IN']	0	1	0
I shouted at her	Snippet	['NN', 'VBD', 'IN', 'PRP$']	0	1	0
shouted at her from	Snippet	['VBN', 'IN', 'PRP', 'IN']	0	1	0
at her from the	Snippet	['IN', 'PRP', 'IN', 'DT']	0	1	0
her from the top	Snippet	['PRP$', 'IN', 'DT', 'NN']	0	0.625	0.375
from the top of	Snippet	['IN', 'DT', 'NN', 'IN']	0	0.625	0.375
the top of my	Snippet	['DT', 'NN', 'IN', 'PRP$']	0	0.625	0.375

top of my lungs	Snippet	['NN', 'IN', 'PRP$', 'NN']	0	0.625	0.375
of my lungs "Where	Snippet	['IN', 'PRP$', 'NN', '``', 'WRB']	0	1	0
my lungs "Where was	Snippet	['PRP$', 'NN', '``', 'WRB', 'NN']	0	1	0
lungs "Where was she,	Snippet	['NN', '``', 'WRB', 'NN', 'PRP', ',']	0	1	0
"Where was she, who	Snippet	['``', 'WRB', 'NN', 'PRP', ',', 'WP']	0	1	0
was she, who was	Snippet	['NN', 'PRP', ',', 'WP', 'VBD']	0	1	0
she, who was she	Snippet	['PRP', ',', 'WP', 'VBD', 'PRP']	0	1	0
who was she with,	Snippet	['WP', 'VBZ', 'PRP', 'IN', ',']	0	1	0
was she with, what	Snippet	['NN', 'PRP', 'IN', ',', 'WP']	0	1	0
she with, what was	Snippet	['PRP', 'IN', ',', 'WP', 'VBD']	0	1	0
with, what was she	Snippet	['IN', ',', 'WP', 'VBD', 'PRP']	0	1	0
what was she doing."	Snippet	['WP', 'VBZ', 'PRP', 'VBG', '.', ''''']	0	1	0
was she doing." At	Snippet	['NN', 'PRP', 'VBG', '.', '''''', 'IN']	0	1	0
she doing." At one	Snippet	['PRP', 'VBG', '.', '''''', 'IN', 'CD']	0	1	0
doing." At one point	Snippet	['VBG', '.', '''''', 'IN', 'CD', 'NN']	0	1	0
At one point I	Snippet	['IN', 'CD', 'NN', 'NN']	0	1	0
one point I grabbed	Snippet	['CD', 'NN', 'NN', 'VBD']	0	1	0
point I grabbed her	Snippet	['NN', 'NN', 'VBD', 'PRP$']	0	1	0
I grabbed her and	Snippet	['NN', 'VBD', 'PRP', 'CC']	0	1	0
grabbed her and started	Snippet	['VBD', 'PRP', 'CC', 'VBD']	0	1	0
her and started shaking	Snippet	['PRP$', 'CC', 'VBD', 'VBG']	0.362	0.638	0
and started shaking her.	Snippet	['CC', 'VBD', 'VBG', 'PRP', '.']	0.362	0.638	0
started shaking her. I	Snippet	['VBD', 'VBG', 'PRP$', '.', 'NN']	0.459	0.541	0
shaking her. I really	Snippet	['VBG', 'PRP$', '.', 'NN', 'RB']	0.459	0.541	0
her. I really scared	Snippet	['PRP$', '.', 'NN', 'RB', 'VBD']	0.615	0.385	0

I really scared her!	Snippet	['NNS', 'RB', 'VBD', 'PRP', '.']	0.635	0.365	0
I was extremely scared to leave.	Heading	['JJ', 'VBP', 'RB', 'VBN', 'TO', 'VB', '.']	0.608	0.392	0
I was extremely scared	Snippet	['JJ', 'VBP', 'RB', 'JJ']	0.615	0.385	0
was extremely scared to	Snippet	['NN', 'RB', 'VBD', 'TO']	0.516	0.484	0
extremely scared to leave.	Snippet	['RB', 'VBN', 'TO', 'VB', '.']	0.7	0.3	0
This made me feel unsafe around him	Heading	['DT', 'VBD', 'PRP', 'VB', 'JJ', 'IN', 'PRP']	0	1	0
This made me feel	Snippet	['DT', 'VBD', 'PRP', 'VB']	0	1	0
made me feel unsafe	Snippet	['VBN', 'PRP', 'VB', 'JJ']	0	1	0
me feel unsafe around	Snippet	['PRP', 'VBP', 'JJ', 'IN']	0	1	0
feel unsafe around him	Snippet	['NN', 'JJ', 'IN', 'PRP']	0	1	0
Whenever I thought about it I would always be on the verge of tears.	Heading	['WRB', 'NN', 'VBD', 'IN', 'PRP', 'VB', 'MD', 'RB', 'VB', 'IN', 'DT', 'NN', 'IN', 'NN', '.']	0.147	0.853	0
Whenever I thought about	Snippet	['WRB', 'NN', 'VBD', 'IN']	0	1	0
I thought about it	Snippet	['NN', 'VBD', 'IN', 'PRP']	0	1	0
thought about it I	Snippet	['VBN', 'IN', 'PRP', 'VB']	0	1	0
about it I would	Snippet	['IN', 'PRP', 'VBZ', 'MD']	0	1	0
it I would always	Snippet	['PRP', 'RB', 'MD', 'RB']	0	1	0
I would always be	Snippet	['NN', 'MD', 'RB', 'VB']	0	1	0
would always be on	Snippet	['MD', 'RB', 'VB', 'IN']	0	1	0
always be on the	Snippet	['RB', 'VB', 'IN', 'DT']	0	1	0
be on the verge	Snippet	['VB', 'IN', 'DT', 'NN']	0	1	0
on the verge of	Snippet	['IN', 'DT', 'NN', 'IN']	0	1	0
the verge of tears.	Snippet	['DT', 'NN', 'IN', 'NN', '.']	0.388	0.612	0

Appendix 7 – Combined tally of all tags collected within all 4 corpuses.

The table below shows all the word marked as significant by the TF-IDF analysis POS Tags (see appendix 4 for POS Tag meanings).

Tally Of POS Tags For All Words Which Showed Significance In:							
Action Events		**Beliefs & Thoughts**		**Consequences**		**News Samples**	
NN	134	NN	131	NN	126	NN	258
NNS	37	NNS	36	NNS	37	NNS	104
VBG	27	JJ	26	JJ	25	JJ	60
JJ	25	VBG	23	VBG	22	VBG	46
RB	17	RB	19	RB	19	RB	39
VBN	16	VBN	17	VBN	15	VBN	38
IN	11	VB	12	IN	11	IN	20
VB	10	IN	11	VB	11	CD	17
CD	10	CD	11	CD	10	VBD	14
VBD	8	VBD	6	VBD	8	VB	12
PRP	4	VBZ	5	VBZ	5	VBZ	7
VBZ	4	MD	4	PRP	4	DT	6
DT	4	PRP	4	DT	4	PRP	5
WRB	2	DT	4	WRB	3	JJS	4
MD	2	WRB	2	PRP$	2	RBR	3
CC	2	CC	2	CC	2	JJR	3
PRP$	1	PRP$	1	MD	2	WRB	2
JJS	1	JJS	1			CC	2
						MD	2
						PRP$	1
						WP	1
						WP$	1

Appendix 8 – Text Processing Class Pseudocode

The following class provides some logic to the chatbot which allow it to understanding and further processing of text which falls outside of the detectABC class.

```
Class textProcessing

    #Takes text input on initialization and also checks to see if the text should be tokenized and lemmatized on
    initialization.

    Def __init__(text, run = False)

        Self.inputText = text

        If run == True:

            Self.tokenize()

            Self.normalize()

    Def tokenize(self)
```

```
        tokenized list of self.InputText

Def lemmatizer(self)

        Lemmentize tokenized self.inputText

Def partOfSpeach()

        Part of speech tagged self.inputText

#using a predefined searchParameter, this function will scan a body of text and then identify phrases which fit the
regex search parameter in the self.inputText

Def chunking(self, searchParameter)

        Self.partOfSpeach()

        chunkParser = RegexParser(searchParameter)

        chunked = chunkParser.parse(self.inputText)

        filteredList = []

        for subtree in chunked:

                filteredList.append(subtree.leaves())

        return filteredList

#Identifies the name of a person using a regex chunking rule which looks for NNP words

Def getName()

        FilteredList = self.chuinking("Name: {<NNP+>}")

        stopWords = ["hi", "hey"]

        newList = []

        name = ""

        for word in filteredList

                if word not in stopWords

                        newList.append(word)


        if len(newList) == 0

                return newList[0] #provides first name it detects

        else

                self.inputText = input("Sorry no name was detected, please try again")

                return self.getName()

#Allows for the detection of yes or no to a question.
```

```
Def yesOrNo()
        Yes = False
        No = False
        While Yes == False or No == False
                Self.tokenize()
                Self.normalize()
                yes = [list of yes words/phrases]
                no = [list of no words/phrases]
                for word in self.input:
                        if word in yes:
                                return True
                        if word in no
                                return False
                self.inputText = input("sorry, I don't understand, please say yes or no")
```

Appendix 9 – dbManager Class Pseudocode

This class manages data storage to a JSON file

```
dbManager(self)
        def __init__(self)
                self.database = {}
        def setJSON(self, db = {})
                write db to json file
        #writes to a single key in a database dictionary and then pushes it to a JSON file.
        def singleDBUpdate(self, key, value)
                self.database[key] = value
                self.setJSON(self.database)
        def getJSON(self)
                read JSON file convert to self.database
                return self.database
```