Formalization and Modeling of Human Values for Recipient Sentiment Prediction

Doctor of Philosophy

OBINNA CHINEDU ONYIMADU

Informatics Research Centre

Henley Business School

September 2017

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Obinna Chinedu Onyimadu

Abstract

Sentiment analysis is viewed generally as a text classification problem involving the prediction of the semantic orientation of a text. Much of the analysis has focused on the sentiment expressed in the sentence or by the writer but not the sentiment of the recipient. For example, the sentence "*Housing costs have dropped significantly*" might be assigned a negative classification by a sentiment analysis model, however humans from different works of life might express different sentiments. A landlord will likely express a negative sentiment while a renter might express a positive sentiment. Therefore, traditional sentiment analysis methods fail to capture the human centric aspects that motivate diverse sentiments.

Additionally, attempts at predicting recipient sentiment have involved considerable human effort in the form of content analysis and empirical surveys, making the process expensive and time-consuming. Thus, the aim of this research is to develop a method of recipient sentiment analysis that is devoid of human input in the form of annotations or empirical surveys. The approach taken in this research involves applying a model of human values towards recipient sentiment prediction. The justification for this approach is based on the well-established principle that values influence human behaviour of which sentiment is a form. Therefore, if a persons' values can be modelled quantitatively, when presented with some text, in theory the sentiment of the recipient can be predicted.

This research proposes that the application of values in developing sentences is a generative process, that can be represented as a language model. A mechanism called Feature Switching (FS) that enables the determination of recipient's sentiment from the value language model is also discussed. The resulting sentiment prediction model has an accuracy in the range of 72.2%-72.5% which is in and about the range of performance of existing systems which make use of content analysis and human annotated data.

Acknowledgements

I would like to express my gratitude to my supervisor, Dr Keiichi Nakata for his excellent supervision and support for making this thesis a reality. His invaluable suggestions and critical comments were key to the completion of this thesis.

I would also like to acknowledge staff at the IRC and my friends for the many enlightening discussions we had at various points during this research.

Finally, I want to acknowledge God almighty, without whom I would have been unable to complete this thesis.

Obinna Chinedu Onyimadu

September 2017

Dedications

This thesis is dedicated to my parents, brothers and especially my wife. You all made the journey a lot easier and enjoyable. God Bless.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1. Introduction

The needs and demands brought about by 21st century content explosion has framed the research discourse around sentiment analysis (SA) considerably with most of the emphasis on the sentiment expressed in the text i.e. classifying the sentence or utterance as positive, negative or neutral, which does not necessarily translate to the sentiment of the reader or hearer of the utterance. For example, the utterance *"Google shares might drop next week"*, though quite subjective and expressive of a clearly negative sentiment cannot from a human point of view be classified as negative. Associating this example with a negative semantic orientation without considering human and contextual dimensions as traditional SA methodologies would, produces a one-size fits all depiction of sentiment that would not yield a different sentiment orientation if the subject of the sentence were for example changed from '*Google*' to '*Yahoo*' i.e. *"Yahoo shares might drop next week"*. Assuming the reader of the sentence, has some stake in Google, it is very likely that the utterance would induce negative sentiment or behaviour. Conversely, if the reader/hearer had a stake in a rival company, then from a human centric perspective, the sentiment polarity would be the converse. This behaviour is absent in traditional SA methodologies and algorithms, which will typically predict or classify the sentiment of the author and not the reader. Even SA tasks such as stance detection and emotion classification, are aimed at classifying the stance or emotion disseminated in the sentence and not the emotion or stance of a reader/hearer – recipient[1]. Therefore, there is a gap in current SA research around the prediction of a recipient's sentiment towards an utterance/sentence, because current SA approaches do not incorporate the subjective human centric aspects which determine the sentiments of individuals. This problem is called the author-reader standpoint because the sentiment expressed by the author of a sentence does not necessarily translate to that expressed by the recipient (Liu, 2012). Thus, in this thesis, a new approach for addressing the author-reader standpoint[2] is developed and implemented.

There have been very few attempts at solving the author-user standpoint problem. However, a common approach has involved the application of theories from sociology such as Affect Control Theory (Ahothali & Hoey, 2015; Heise, 1987; Mejova, 2012), appraisal theory (Bloom, 2011) and frame semantics (Bhowmick, 2009; Fillmore, 1982). Thus, the basis for this research's approach to solving this problem also stems from a social concept, human values. It proposes that the sentiment of a reader towards an utterance is a form of human behaviour, which in turn is influenced and determined by the values of the reader or hearer (Templeton et al, 2011a, 2011b). In fact, the definition of values as fundamental

[1] From this point onwards, the term 'recipient' is used to represent readers/hearers

[2] An implementation of the model is carried out on political data focusing on two timely subjects – The EU and Immigration. A major reason for focusing on the political domain was because of a previous KTP research carried out by the author of this thesis on the development of Information Retrieval and SA tools for UK Parliamentary debates and data. The research resulted in the development and implementation of a commercial product called Semantris. In developing Semantris, access was made available to debates, relevant data and contacts.

abstract coordinators of behaviour and guides for preference of one situation over another (Rokeach, 1973), substantiates the link between sentiment as a form of human behaviour and values. It is values that guides preference for one state, substance, entity, concept, idea etc. over another and sentiment in itself is an expression of preference for one state over another. Therefore, to develop a model that predicts reader or hearer sentiment, such a model needs to incorporate the values of the recipient. This is depicted in figure 1, where a sentence is passed through a sentiment model augmented with a model of human values to yield human centric sentiment predictions.

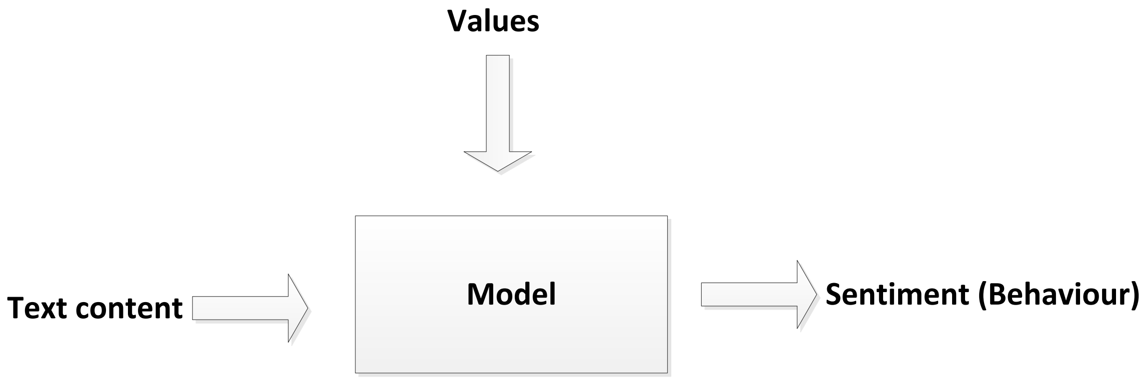


Figure 1: Illustration of Value and Sentiment Model

A major benefit in the application of values in sentiment prediction is to enable the prediction of sentiment in objective evaluative sentences. Objective evaluative sentences are sentences which do not contain any explicit sentiments, are not subjective and whose sentiments are implied[3]. Traditional SA methodologies and algorithms have succeeded in identifying these types of sentences but there has not been much success at assigning the polarity. Part of the problem lies in the one-size-fits all nature of SA, as well as the inability of current SA approaches to match the utterance to diverse sentiment holders. Consider the statements "*We will exit the EU*", *"I paid £1000 for the new office software"*. Both these statements are objective and do not carry any explicit sentiments, however both contain implied sentiments. In exiting the 'EU', a person with values consistent with remaining in the 'EU' will view it as negative and similarly, a person with open-source software values will perceive the latter statement as negative. Traditional sentiment analysis methodologies are unlikely to make these differentiations, and in fact some SA implementations are likely to assign the first sentence a negative polarity because of the presence of the linguistic unit 'exit'. For such Objective Evaluative Sentences (OES), the use of traditional sentiment analysis methods in the determination of the sentiment would result in poor or spurious predictions primarily because of the absence of human centric

[3] The task of identifying objective evaluative sentences is not part of our research scope

reference points. This research proposes that the inclusion of human values in a model can remedy this problem.

There has been several research works to show how human values influence behaviour. Schwartz (1992) and Schwartz (n.d(a)), showed that on the subject of gay marriage, people with traditional values were more likely to be in opposition or have a negative sentiment. Conversely, people with hedonistic values were most likely to view the subject positively. Typical approaches to modelling human values have focused extensively on empirical surveys and interviewing value holders. These approaches are expensive to implement, require considerable human input and normally result in an enumerated list of values making them rigid and inapplicable to other domains.

Another approach involves modeling values from text, however, most approaches for modeling values from text involve a combination of document or text content analysis and empirical surveys of human value holders (Cheng and Fleischmann, 2010; Scott, 1965; Takayama et al, 2014). Again, these approaches are expensive, time consuming, require considerable human investment in annotations and result in a list of value concepts that are not adaptable to other domains. These methodological gaps validate the need for an approach that is flexible, easily applicable to multiple domains and finally one that can be implemented without the express need of human annotations or content analysis.

The next section details the aims and objectives of this research.

1.1 Aims and Objectives

The research problem is defined as follows:

The aim of this research is to develop a method of sentiment analysis that can predict the sentiment of the recipient without the use of explicit human annotations.

The question is, given a sentence with implicit or explicit sentiments, can a model of an individual's values be applied towards the prediction of the individual's sentiment polarity? In other words, can the behaviour (sentiment) of the hearer or reader of an utterance or piece of text be predicted from a model of his/her values?

To accomplish this, the research attempts to formalize and implement a methodology for modelling human values from text and apply this model to a SA model for the prediction of recipient sentiment for objective evaluative and subjective sentences. The approach in this research is quite distinct from other approaches in that it is devoid of human input in the form of empirical surveys or content analysis in the development of the value model. Similarly, this research's sentiment prediction methodology and model will also not require human input unlike traditional models which require annotated training data.

Given these aims, the research objectives are as follows:

- To develop a representation of an individual's value orientation.
- To formalize an approach for identifying and applying values to sentiment prediction.
- To implement the value model appropriate for a recipient oriented sentiment analysis.
- To implement a recipient sentiment prediction model that is based around the value model.
- To evaluate the precision of the model against human benchmarks as well as the effect of semantic enhancements on the overall model.

Having defined the aims and objectives of this thesis, the next section describes the structure of the thesis.

1.2 Overview of Research Methodology

The approach of this research is based also on modeling abstract unobservable human values from observable spoken utterances and textual content made by value holders. This research proposes that observable utterances made by value holders are a function of the values held by the value holders and such utterances can be modeled via a generative model. Implementing such a generative model entails the identification of parameters and components which make up human values, the formalization of these parameters and the formulation of the generative model. Following the modeling of values, this research presents a unique approach for applying the generative model in predicting sentiment using two contributions. Firstly, the sentiment of a recipient is formulated as the difference between the likelihood of a recipient making an utterance and the likelihood of the same recipient in making an utterance with a contrary sentiment. Secondly, this likelihood is implemented using a unique approach called Feature Switching. The methodology adopted in carrying out this research is based on Design Science Research (DSR) primarily because the outcome of this research are artifacts. Therefore, the methodology focuses on an awareness of the research problem, followed by proposing an approach to solving the problem. The suggested approach and its implementation is founded on existing theory and research. It is finally evaluated using a dataset drawn from the political sector.

1.3 Thesis Structure

The rest of this thesis is structured as follows:

- Chapter 2, introduces sentiment analysis, briefly describing its history and applications in research and industry. It provides a general description of the approaches to recipient sentiment analysis, and discuss some of the existing literature and their associated deficiencies.
- Chapter 3 discusses the theoretical foundations of values as well as its relationship with sentiment analysis. It describes the main approaches to modeling values and

the associated issues. In addition, conceptual definitions of values drawn from a wide array of research are discussed. These definitions provide clues which are used in chapter 5 and 6 in model implementation.

- Chapter 4, discusses the methodology, which is based on design science (Hevner et al, 2004).
- Chapter 5 focuses on the design of the value model and sentiment prediction model. The extraction of components which make up human values is described in a process called values decomposition. The parameters which make up the values are expressed mathematically and transformed into a model. In addition, the intuition for the application of values to sentiment which termed Value Sentiment Model (VSM) is also described.
- Chapter 6 outlines the algorithms for sentiment prediction and a description of its implementation.
- Chapter 7 describes a complete implementation, which focuses on the two subjects the European Union and Immigration. Value holders are actual UK political parties. This chapter describes the data and provides two related model implementations. The first implementation is a generic model called 'm1' whose data preparation and feature vectors are corpus independent. The second model is a modification of 'm1' featuring additional semantically enhanced features and data preparation processes. Part of the goal is to determine if the inclusion of semantically enhanced features improves the generic model.
- Chapter 8 describes tests and an evaluation of the models. Finally, chapter 9, highlights the contributions and suggests some future work.

In conclusion, this thesis offers an approach for recipient SA that does not require human input or annotations and adopts human centric values in making its prediction. The next chapter offers a review of relevant SA literature.

2. Sentiment Analysis as a Field of Study

The chapter serves two purposes. It offers a literature review of sentiment analysis (SA) including its history and methodologies. This introductory background paves the way for the second purpose which is a review of some of the literature and methodologies applied in recipient sentiment prediction. Since research in recipient sentiment analysis is scarce, this review covers all known works across all domains. It also highlights some of the limitations in existing methodologies applied towards the recipient sentiment prediction.

2.1 Brief History of Sentiment Analysis

The evolution of the web from a purely informational system to a hub for generating user content and carrying out transactions, including delivering navigational, informational and transactional content and processes spurred the need for SA. In fact, the terms 'sentiment analysis' and 'opinion mining' did not appear until 2003 (Dave et al, 2003; Nasukawa and Yi, 2003). This is not to say that work in this domain was non-existent before the 21st century. Earlier research focused on beliefs (Carbonell, 1979; Su et al, 2008) followed subsequently by research into the interpretation and detection of viewpoints, subjectivity, affects, sentiment lexicon, metaphor interpretation (Wiebe, 1990; Wiebe, 1994; Wiebe et al, 1999; Hearst, 1992; Wiebe, 1994; Kantrowitz, 2000; Wiebe and Riloff, 2005; Choi and Claire, 2009; Hatzivassiloglou and McKeown, 1997; Jindal and Liu, 2006a; Somasundaran and Wiebe, 2009; Liu, 2012).

Commercially, it was not until the early 2000s that the demand for understanding and computationally evaluating opinions in text exploded. Social networking sites like Twitter[4], Facebook[5] and Myspace[6], retail and consumer sites like Amazon[7] and eBay[8], blogging applications like WordPress[9] and Blogger[10] availed a global audience the means to generate, create, share and evaluate content. The result of this was an increased demand for understanding online user behaviour. E-commerce and social networking sites sought techniques for monitoring, finding and distilling their user and audience perception by monitoring reviews, comments and online chatter. An unintended but inevitable consequence of this vast content explosion was the availability of information on all types of matters. Information capable of providing solutions to problems of decision-making and information seeking but embedded deep in vast amounts of text. The availability of vast amounts of data related to any subject meant that people could make more informed decisions by mining through the web for the opinion of others. Political changes have also been triggered and engineered through the force of opinionated postings in social media,

---

[4] Twitter.com

[5] Facebook.com

[6] Myspace.com

[7] Amazon.com

[8] Ebay.com

[9] Wordpress.com

[10] Blogger.com

influencing people and contributing to a shift in business and public opinion. Events like the Arab spring[11] and the election of Barack Obama[12] were influenced by the forces of social media opinions and sentiments. Liu (2012) points out that industrial and research activities around sentiment detection have in turn flourished and 'spread to almost every possible domain' and the key objective is to develop and improve ways of finding out what people are saying and how people perceive products, services, events and topical subjects.
In both research and industry, this computational study of opinions, sentiment or private states - Private states are states of an individual that are not open to objective observation or verification such as beliefs and emotion (Quirk et al, 1985) - has largely been perceived as a natural language text classification problem spawning several dimensions of sub-problems, which include:

- Opinion detection - Research suggests two main types of opinions, regular opinions and comparative opinions (Jindal and Liu, 2006a, 2006b; Zhang et al, 2011; Liu, 2012)[13]. According to Liu (2012), regular opinions express sentiment on a particular entity or aspect of the entity. For example, *"The new MacBook pro has a long battery life."* In this example, a positive sentiment is expressed on an aspect – *'battery life'* - of the entity *'The new MacBook'*. Several sub-tasks related to opinion detection include:
    - Opinion spam detection which involves the detection of spammers, bogus blogs and fake opinion (Jindal and Bing, 2007; Jindal and Bing, 2008).
    - Opinion summarization involves a quantitative or qualitative summary of opinion or sentiments on aspects or features of a product, subject or entity (Das and Chen, 2001; Hu and Liu, 2004; Liu et al, 2005).
- Stance detection[14] involves detecting if an individual's stance is for or against a subject (Ganter and Strube, 2009; Greene and Resnik, 2009; Somasundaran and Wiebe, 2009).
- Emotion detection is the task of classifying or assigning an emotional category to a sentence or utterance. Categories include, anger, disgust, fear, happiness, sadness etc. (Liu et al, 2003; Chaumartin, 2007; Tsoumakas & Katakis, 2007; Ahothali & Hoey, 2015).

Computationally, the task of SA or opinion mining involves identifying semantic orientation (Ding et al, 2008; Melville et al, 2009) - a measure of the positivity, negativity or neutrality - of a sentence or some aspect or segment of the sentence (Benamara et al,

[11] http://www.bbc.co.uk/news/world-middle-east-12813859 - Last accessed 16/09/2014
[12] http://www.journalism.org/2012/08/15/how-presidential-candidates-use-web-and-social-media/ - Last accessed on 16/09/2014
[13] Appendix 1 provides further discussion on opinion types.
[14] Stance detection should not be confused for reader/hearer sentiment because in stance detection, the utterance or sentence whose sentiment is predicted or classified is that of the writer/author/speaker. For example, given a sentence *X*, the task of stance detection is to detect the stance of the speaker, whereas in predicting the sentiment of a reader *R*, the task is to predict the sentiment of *R* towards *X*.

2011; Thomas et al, 2006; Wiebe et al 2001). Usually it might involve an initial determination of the sentence subjectivity, a task called subjectivity identification. Since most opinions are subjective, the presence of subjective clues is a very good indicator of opinion presence (Wiebe et al, 1999; Turney, 2002; Li et al, 2010; Liu, 2011; Melville et al, 2009).

A key question to consider in SA is whose sentiment orientation is being estimated. Consider the sentence: "*Google shares might drop next week*". The traditional approach taken by most SA methodologies is to determine if the sentence is subjective or objective. In this example, the sentence is identified as subjective because of the presence of subjective clues like the phrase '*might drop*'. Subsequently, additional linguistic clues are processed and applied in estimating the sentiment of the objective sentence, which in this example is negative and represents the sentiment expressed by the speaker. Consider a slightly different variation of the same sentence: "*Google shares dropped last week*". In this instance, the speaker is making an objective statement of fact, that is implicitly negative, but negative for who? For one hearer, such an utterance is negative whereas for another it is positive or even neutral. Consider also the case of an objective sentence without any implied semantic orientation like: "*The UK will exit the EU*". Exiting the EU could be positive or negative depending on who the recipient is. Assuming the hearer is guided by pro-EU values, then he/she is more likely to have a negative sentiment towards the utterance and vice-versa if the recipient has anti-EU values. Therefore, if a SA model is aware of who the recipient of an utterance is and could incorporate the values or human centric social factors of the recipient, then it should theoretically be able to predict the sentiment of the recipient to the utterance regardless of whether it is implicitly or explicitly subjective/objective. Making such sentiment prediction is the objective of this thesis and so this chapter reviews some of the literature around recipient sentiment prediction. Due to the dearth of research in this area, related work around stance detection and emotion detection are reviewed. Since the underlying approaches used in author SA and recipient SA are identical, the next section provides a brief review of the techniques and followed up with a discussion of the associated problems.

2.2 SA Methods and Problems

The first half of this section focuses on SA methods, while the second part considers the approaches and problems associated with recipient sentiment prediction.

Methods

Techniques adopted in accomplishing subjectivity identification and semantic orientation classification are diverse ranging from rule based system (Tong, 2001; Zhang and Baudin, 2011; Liu, 2012) to supervised/unsupervised machine learning models (Benamara et al, 2011; Hatzivassiloglou and McKeown, 1997; Joachims, 1999; Neviarouskaya et al, 2009; Liu, 2012; Mohtarami et al, 2013). For example, in Yessenalina et al (2010), the sentiment of a document was predicted by first extracting relevant subjective sentences from documents and then using linguistic features extracted in the sentences to infer the overall sentiment expressed by the writer of the sentence, while Chumartin (2007), applied a rule based approach for emotion classification of news headlines.

A common theme in methodologies is the identification and utilization of linguistic features and patterns in sentences such as word and part-of-speech (POS) frequency as well as subjectivity clues. For instance, Finn et al (2002), implemented an unsupervised approach in identifying subjective content in the form of the separation of reviews from other content. They developed a classifier based on the relative frequency of each POS in a document, and this approach out-performed a bag of word classifier with custom built features. Dave et al (2003) makes use of a supervised approach using a corpus of tagged reviews in separating positive reviews from negative reviews. Greene and Resnik (2009) applied lexical semantics and syntax in identifying implicit sentiments embedded in sentences.

Although heuristic rule-based systems are quite simple, they work considerably well for simple regular sentences but fail to deal with context dependent opinion words as well as very long complex sentences. In addition, because they are based on observed patterns in the sentence, they cannot be applied to all sentences. They also fail to capture the contextual intent of the sentence since the words in sentences could possess multiple meanings.

Other solutions have involved the adoption of:

- Corpora based techniques which rely on syntactic or co-occurrence patterns in a corpus for the discovery of domain specific features (Hatzivassiloglou and McKeown, 1997; Wiebe and Riloff, 2005).

- Opinion lexicons or dictionaries which are essentially a comprehensive list of opinion words used as external reference resources and they remain popular in research and industry (Turney, 2002; Hu and Liu, 2004; Kim and Hovy, 2004; Popescu and Etzioni, 2005; Qiu et al, 2009). Das and Chen (2001) implemented a classifier based on a manually crafted lexicon in determining if postings on a board correlated with stock prices. The words in the lexicon were grouped based on manually assigned polarities – positive, negative, neutral. Using the prior polarity of words in the lexicon, the overall polarity of the sentence is aggregated. Even then, as in simple rule based classifiers, they face serious limitations, crucially because of the lack of context (Polanyi and Zaenen, 2006). For instance, using an opinion lexicon, the two utterances *"This film sucks"* and *"This vacuum cleaner sucks well"* will be assigned negative polarities because the word *'sucks'* in the lexicon has a negative semantic orientation even though its contextual use is clearly different. Another deficiency in lexicon based approaches is that the lexicon does not contain all possible words and so in some domains or datasets, the lack of coverage becomes a problem.

- Knowledge Base Approaches which involve the use of domain knowledge bases like ontologies in identifying aspects or segments of a sentence that express a sentiment (Su et al, 2008; Titov and McDonald, 2008).

- Using discourse structures embedded in text in performing SA related tasks such as distinguishing implicit opinions and explicit opinions (Benamara et al, 2011). Asher et al (2008) implemented an annotation schema for a fine grained contextual opinion analysis using discourse relations. Somasundaran (2010) proposed a discourse level treatment to improve sentence based polarity classification and to recognize the overall stance.

Problems

For supervised learning models, feature identification and annotation typically involve the use of human annotators who judge, assess and annotate training data for ground truth. Sufficiently high inter-annotator agreement will subsequently reveal a collection of statistically ideal features, which can be used to develop a model. However, a commonly ignored flaw of feature collection by human annotators unique to sentiment analysis is natural human bias. Toprak et al (2010) reports that high inter-annotator disagreement in distinguishing polar facts from inherently evaluative language because of diverse user views and opinion. Since human opinions are naturally subjective the potential of inter-annotator agreement for subjective sentences is bound to be low. This will be the case in a sentence such as *"Housing costs have dropped significantly",* where annotators such as landlords and renters, with different interest will most likely not share the same sentiments and are thus likely to provide different annotations. Thus, the human element in the form of individual or group interests can skew and influence the model. This emphasizes the importance of human centric qualities like values in the determination of sentiment and emphasizes the need for this research.

Another evaluative consideration often missing from sentiment analysis involves the parties involved in the signification process and the relationship that exist between them. Consider example 1 below, which illustrates four sentences.

Example 1

Text 1: *"Osama Bin Laden is such a good guy.*

Text 2: *"George Bush is such a good guy"*

Text 3: *"We can have the dog for dinner"*

Text 4: *"We can have the roast for dinner"*

Example 2[15]

POS Tagged Text 1: *Osama/**NNP** Bin Laden/**NNP** is/**VBZ** such/**JJ** a/**DT** nice/**JJ** guy/**NN***

POS Tagged Text 2: *George/**NNP** Bush/**NNP** is/**VBZ** such/**JJ** a/**DT** nice/**JJ** guy/**NN***

POS Tagged Text 3: *We/**PRP** can/**MD** have/**VB** the/**DT** dog/**NN** for/**IN** dinner/**NN***

POS Tagged Text 4*: We/**PRP** can/**MD** have/**VB** the/**DT** roast/**NN** for/**IN** dinner/**NN***

By applying lexical and syntactic clues, texts *1* and *2* are lexically similar – where lexical similarity is a function of word order, frequency of common terms and most importantly part of speech order (Metzler and Croft, 2005; Metzler et al, 2007) – essential features used in building SA models. Surface representation of lexical similarity suggests very high similarity (applying part of speech tagging to the sentences in example 1 results in the POS tagged sentences in example 2 which shows that the sentences are similar as they share the same part of speech). Therefore, interpreting the sentiment of text 1 for instance from its syntactic constituents should yield the same semantic orientation as text 2. This logic is flawed, because from a human centric point of view, the sentiment polarity is not just contingent on the make-up of the text but also on the hearer/speaker and their relationship to the principals or expressed subject matter – *'Osama Bin Laden'* and *'George Bush'*. Clearly, personal associations will influence the assignation of polarities and this is not captured by current SA approaches.

Furthermore, going by Stamper's theory of sign formulation (Stamper, 1973; Stamper, 1992) which describes a sentence's syntax as one that observes all the formulaic and grammatical tenets of the language, texts 1 and 2 are syntactically similar. Both texts are constructed correctly; tenses and punctuation are correctly applied and located. Apart from the constituent principal nouns, they contain the same words and bear the same semantic interpretation. However, pragmatically, they are clearly different. What current sentiment analysis techniques fail to capture are the underlying influencers of sentiment such as the impact of the existing relationships between authors and recipients. Also, in example 1, texts 3 and 4 are syntactically and semantically similar. They are objective, featuring clearly delineated sentiments, yet critically humans from different cultures with diverse attitudes towards the subject will express very different sentiments. For example, certain cultures will most likely assign a negative polarity to text 3 and a positive to text 4 because of the societal and legal values prohibiting the consumption of *'the dog'* - a domestic pet - while some other cultures will appraise both signs indifferently. What this illustrates is that values and context, diverse as they are, constitute a vital influence in sentiment classification and represents a vital element absent in current sentiment analysis state of the art.

[15] POS tags are in bold font. Tags used were based on the hepple POS tagger, see https://gate.ac.uk/sale/tao/splitap7.html#x39-784000G – last accessed 10/01/2017

2.3 Review of Author/Reader Stand Point

As discussed, a significant proportion of SA has focused on the sentiment of the author, and as such the literature on recipient sentiment prediction is quite limited. Therefore, in addition to the recipient sentiment prediction models discussed in this section, this review is augmented with some of the research on detecting recipient emotion and stance. Stance detection is the task of classifying perspectives e.g. for or against something and emotion detection involves identifying the emotion expressed in a sentence e.g. anger, disgust, fear, happiness, sadness and surprise (Strapparava & Mihalcea, 2007; Tang and Chen, 2011). These two subjects are considered because they are subparts of the field of sentiment analysis. Methodologies applied in predicting recipient sentiment can be divided into two:

- Text and knowledgebase methodologies which are based on identifying patterns and features in text, utilizing an external domain knowledgebase or ontology, identifying and applying discourse patterns in sentences and utterances.

- Social theoretic approaches – Tend to incorporate social theories which model or describe abstract human behaviour in sentiment or emotion detection. Examples of social theories include Affect Control Theory (Heise, 1979, 2006, 2007), Frames (Ruppenhofer, 2013; Ruppenhofer et al, 2016), Profile of Mood States (POMS) (McNair et al, 2003; Norcross et al, 2006) and Appraisal theory (Roseman & Smith, 2001; Scherer et al, 2001).

Subsequent sections describe some of the works and their limitation with respect to both recipient sentiment prediction as well as the sentiment expressed by the author.

2.3.1 Writer/Author Emotion and Stance

The works and methods described in this section are directed at predicting the emotion or stance of the writer. The approaches for emotion detection can be divided into three groups:

- **Use of a Tagged Corpus**

  This involves the use of emotion-tagged corpus to detect the emotions of the authors. Here, authors identify and tag sentences which portray possible emotions. The tagged sentences are subsequently used to train a classifier in predicting sentiments. (Yang et al, 2007a; Yang et al 2007b; Yang et al, 2008). Mihalcea & Liu (2006) implemented a corpus based approach to classify blog posts from LiveJournal into 'happy' and 'sad' category. In applying tagged corpus, a ground truth of correct annotations is applied in training the model. This annotation process is time consuming, expensive and requires considerable human effort.

- **Use of an Affect Lexicon**

  Affect lexicons consist of several emotion categories containing relevant words that are synonymous with emotions. Subasic and Huettner (2001) implemented a fuzzy logic based system in classifying documents. It consisted of a manually constructed lexicon. Words in the lexicon were associated with affect categories specifying the intent and centrality of the word. For example, the word 'mayhem' was associated with violence. Similarly, Balahur et al (2009) implemented a model for classifying emotion by applying affect, opinion and attitude lexicon. Words are assigned ratings according to a set of emotion classes. Due to the versatile and open-ended characteristic of spoken and written language, it is impossible for manually generated list of affects to cover all possible contexts or meanings of a word thereby making this approach rigid. Finally, from a human centric perspective a word signifying one emotion, might signify a different emotion for another individual. In light of this there is a need for an approach that is flexible, loosely structured and capable of catering to diverse contexts.

- **Knowledgebases**

  Liu et al (2003), applied relationship from the Openmind Commonsense database[16] as well as a manually composed set of ground truths to assign affect categories to linguistic units. The affect categories were – happy, sad, anger, fear, disgust, surprise. The limitations of this approach are:

  - The ground truths and knowledgebase does not cover all possible circumstances.
  - Ground truths involve human involvement and annotation.

Stance detection has been applied in a variety of sectors and domains ranging from politics (Thomas et al, 2006; Somasundaram and Wiebe, 2009, 2010) to online debates on a variety of subjects (Murakami and Raymond, 2010).

In Thomas et al (2006), they investigate "whether one can determine from the transcripts of U.S. Congressional floor debates whether the speeches represent support of or opposition to proposed legislation". They choose not to classify speeches in isolation rather using discourse segments which illustrate agreement between speakers. Although their system had an accuracy of 71.28% it was based on an annotated set of training sentences. In addition, they did not consider the actual speakers or the relationship between the speakers in the design of the system.

Lin (2006) and Lin et al (2006), propose an unsupervised learning approach for detecting the perspective at the sentence and document level called Latent Perspective Model (LPM). The model is evaluated on articles related to the Israeli-Palestinian Conflict. They show that perspectives can be learned squarely from word usage and also obtained high accuracies of about 86.9%. Their application of just lexical units and clues in the

[16] http://conceptnet.io/ - Last accessed 10/04/2017

determination of stance justifies this thesis' goal of modelling values and predicting hearer sentiment from lexical clues without human annotations.

Somasundaran and Wiebe, (2009, 2010), implemented a system for recognizing the stance of speakers in an online debate by applying discourse relevant factors derived from an arguing lexicon. The motive behind the introduction of an arguing lexicon is that people having a debate express their subjective expressions by using argumentative terms that enforce their stance. Therefore, the aim of the discourse relevant factors is to differentiate statements of a stance from statements where a person is merely making a concession. Their application focused on political domain and obtained an overall accuracy of 63.93%. However, the arguing lexicon was derived from a manually annotated corpus, thus, human involvement.

A recurring theme in the approaches considered is the involvement of humans. In the next section, where recipient stance or emotion is considered, the same theme is also observed.

2.3.2 Recipient Emotion and Stance Detection

Like the approaches mentioned in the previous section, Tang and Chen (2011) performed emotion detection of writer and recipient emotion in a chat room. Their approach required the identification and annotation of emotions in both the reader and recipient content. Recipients (readers) in the chat network label sentences with optional quantifying emotions like 'Likes', 'Shares', 'Gives', 'Hates', 'Wants', 'Wishes', 'Needs', 'Will', 'Hopes', 'Asks' etc. In addition, contributions in the chat room were labelled as positive or negative. As such, sentences were labelled for their emotional content and mapped to a semantic orientation, thereby ensuring that both linguistic and human centric features were captured and harnessed in the model development. Additional human centric features like the social relations between writers and their behaviour were also captured and fed into the model. The eventual supervised model was shown to have an accuracy in the range of 80.67% and 88.37% for predicting the reader's emotion. Although this model performs quite well, it required considerable human involvement in the annotation of the content and even in the collection of user centric behaviour.

Similarly, Lin et al (2007) and Lin and Chen (2008) adopted a familiar approach in estimating the emotion of readers from a manually tagged Yahoo! Kimo news corpus. A corpus was tagged based on eight emotional classes by humans and linguistic features such as character bigrams, presence of emotional words and content metadata were extracted and applied in their model. They reported an accuracy of 76.88%. Again, in this approach it is noticeable that the introduction of human annotations and the division of reader emotion into classes makes the approach quite rigid since human emotions could belong to more than one class.

Sridhar et al (2014) implemented a stance detection approach using both linguistic and the structural arrangement of the debates in online posts as features in classifying the stance on gun control and gay marriage. Linguistic features such as the length of a speech, word counts, discourse cues and punctuation count are applied. The most unique feature applied was the incorporation of author information. However, this implementation was dependent on hand annotated stances for each sentence in the training set. That is each sentence in the training set contained a marker saying if it was pro or anti a subject. They obtain an average F1 score of 74% for the positive class.

From this review of existing work on stance detection, a group of recurrent limitations are observed: The need for human annotation for ground truth which will be fed to a model, the non-inclusion of human features besides Sridhar et al (2014) and Tang and Chen (2011), the dependence on manually constructed lexicon or knowledgebase. In the next section, socio-theoretic approaches which typically incorporate human behaviour are considered.

2.3.3 Socio-Theoretic Approaches

This section, discusses Affect Control Theory (ACT), a sociology theory that has been applied in SA and in particular recipient sentiment prediction.

ACT is a social psychological theory of human interaction (Heise, 2007). It suggests that "certain cultural norms dictate the affective meanings of words that people in a culture with a common language share". It computes this affective meaning of an event or concept - events or concepts are expressed as words - in a multi-dimensional semantic space (Robinson and Smith-Lovin, 2006; Mejova, 2012) that consists of Evaluation, Potency and Activity (EPA).

In ACT, empirical equations are derived for a wide-ranging set of situations associated with an event. The affective sentiment of cultures is derived or measured using a survey technique called semantic differential derived by Osgood et al (1957), the basis of which is not so different from the empirical surveys applied in modeling values. Basically, individuals with knowledge of a culture rate concepts on a numerical scale with opposing adjectives at each end. In fact, a database of concepts expressed as words and their average EPA ratings derived from survey participants who are knowledgeable about their culture has been collected in Heise (2010). For instance, in the example given by Ahothali and Joey (2015), the culturally shared EPA for the concept 'mother' in Ontario Canada is given as [2.74, 2.04, 0.67] which is interpreted as quite good, quite powerful and slightly active. Whereas in the same place, the concept 'daughter' has an EPA of [2.18, -0.01, 1.92], which is interpreted as quite good, less powerful and more active than mother. These values are derived from ACT equations. ACT lexicons have been compiled for several countries and cultures including USA, Canada, Germany, China and Northern Ireland (Robinson and Smith-Lovin, 2006; Mejova, 2012). Additionally, ACT lexicons have also been developed for groups within societies such as religious groups (Smith-Lovin and Douglas, 1992), state troopers (Heise, 1979) and internet users (King, 2001).

Mejova (2012), showed that sentiment orientation classifiers which make use of ACT lexicons outperforms traditional SA classifiers. Mejova showed that using three variations of ACT compared to a sentiment analysis algorithm, the accuracy of the system was between 71.9% and 80.3%. The accuracy of the positive class polarity was between 64.2% and 85.7%, while the accuracy of the negative polarity class was between 77.5% and 78.1%. However, it was indicated in the experiments of Mejova (2012) that a major flaw in the ACT approach is that the ACT lexicon is limited and so does not necessarily account for all possible words that can be used to describe a situation. However, Ahothali and Joey (2015) implemented an approach for increasing the dataset or vocabulary of ACT words. Unlike the work of Mejova (2012) which focused on the sentiment of the reader, Ahothali and Joey (2015) applied ACT in analysing reader sentiment towards factual objective content. They computed reader sentiment using ACT equations and evaluated their approach against traditional SA approaches on news headlines. This resulted in a precision of between 68% and 82%. They also showed like Mejova (2012) better performance compared to traditional SA methods.

A unique benefit of ACT is that due to the lexicons and equations obtained for each culture, using the approach in Ahothali and Joey (2015) or Mejova (2012), it is possible to predict the sentiment of a recipient in cultures for which there exists a lexicon. Nevertheless, these lexicons are limited and do not encompass all cultures or situations. More so, the lexicon is generated by empirical surveys, which involve considerable human effort and time. Therefore, while ACT clearly incorporates human centric features, there is still a gap in the research methodologies for an approach that is independent of human annotations or input and one that is not dependent on a knowledgebase of human values.

Other socio-theoretic approaches used in SA have focused on the sentiment of the writer. One such theory involves frames. Frames “capture the background knowledge that competent speakers use when producing and understanding utterances” (Ruppenhofer, 2013). The fundamental idea behind frames is that people understand the meaning of a word based on the frames they evoke and that these frames are “story fragments which serve to connect a group of words to a bundle of meanings” (Ruppenhofer et al, 2016). Ruppenhofer et al (2016) illustrates with an example, where the term avenger evokes the Revenge frame, which describes a complex series of events and the group of participants involved in the event. The knowledgebase of frames is collated from human annotations of sentences, involving the identification of possible frames expressed in the sentence and the participants. FrameNet[17] is a knowledgebase of words and their usage frames. As a resource, frames capture the contextual implication of words and participants involved in the discourse and so this makes it ideal for SA related tasks. Frames have been applied in aspects of sentiment analysis including the identification of multiple opinions, identification of opinion source and opinion target (Ruppenhofer, 2013). More so, the work of Bhomwick et al (2009) which classifies the emotion of readers from sentences is the only work that was identified in this research that applies frames in recipient emotion. The emotion of readers was categorised into four classes -disgust, fear, happiness and sadness. They showed that the inclusion of word frames as feature vectors performed better than

[17] https://framenet.icsi.berkeley.edu/fndrupal/: Last accessed 20.03.2017

the use of just words and their POS. The overall F1 score of their approach was 82.1%. Nevertheless, the use of frames highlights the gap in the research in that the inclusion of frames in the SA methodology requires the annotation or labelling of sentences into emotion classes as was the case in Bhomwick et al (2009). By having a fixed set of emotion classes, the application is already restricted to those classes and thus unable to account for variations in emotion or even sentences that portray multiple emotions.

In this section, socio-theoretic principles that have been applied to SA and recipient sentiment prediction were identified. Their limitations have also been articulated. The next section concludes the entire review.

2.4 Conclusion

Research into recipient sentiment prediction is limited and the accuracy of the methodologies range from 66% to about 88%. In addition, the approaches applied till date involve considerable human involvement either in the identification and annotation of ground truth for a learning algorithm or in the development of knowledgebases or lexicons. These approaches are quite expensive, lacking flexibility and restricted to the domains or data set for which the application was designed. More so, the use of lexicons and knowledgebases do not always cover all circumstances or new unseen words. There is thus a need for a methodology that does not require human annotations of ground truths and is flexible enough to handle new terms and contexts.

Existing literature also shows that linguistic features augmented with social features can significantly improve the performance of recipient sentiment prediction models. Since these social features are abstract and unseen, they are encoded as linguistic features of observed words and patterns as is the case in frames. This provides a justification in this research for modeling unseen abstract values as observed textual expressions. The next chapter reviews the existing literature on values.

3. Values as a Field of Study

It has been shown that the reliance on linguistic clues alone is insufficient in predicting the sentiment of a reader and that the inclusion of human centric features can potentially improve the accuracy of a sentiment prediction model. Human values are a type of social construct that can influence human behaviour for which sentiment is a type. This chapter proposes that a person's value determines his/her sentiment. In other words, the sentiment an individual may express about a subject or issue is determined by the values held by the individual. Consequently, an understanding of human values and its formalization can theoretically improve the precision and accuracy of sentiment prediction.

To this end, this chapter provides a review of existing value conceptualizations, instruments and models, including inherent deficiencies associated with addressing the research problem. It begins with an exploration into the definitions of values.

3.1 Definition of Values

There's been a distinct lack of uniformity in the definition and formalization of values, with different fields across social sciences and humanities proffering diverse definitions (Hitlin, 2003; Hitlin and Piliavin, 2004; Rohan, 2000). Values have always been perceived as an abstract concept. In fact, Perry (1926) defined it as a philosophical concept or belief associated closely with virtuous living and morality. Similarly, Williams (1979) expressed values as interests, pleasures, likes, preferences, moral obligations, desires, wants, goals, needs, aversions, attractions and many other kind of selective orientations (Perry, 1926). Rokeach (Rokeach, 1973) attempted to provide a uniform definition and conceptualization of values defining values as "abstract fundamental coordinators of behaviour". 'Abstract' representing an unquantifiable, non-physical entity and 'coordinators of behaviour' implying that for any form of behaviour for which sentiment is a type, values represent the primary causal factor. Similarly, Verplanken and Holland (2002) expressed values as "latent variables that have explanatory value for the choices people make". Schwartz (1996), Feather (1995) and Bardi and Schwartz (2003) reinforce this notion of values as causative to behaviour referring to values as principal determinants of behaviour and attitude.

Critically, four conceptual definitions emerge from literature. In the first, values are portrayed as beliefs (Rokeach, 1973). Essentially the belief that a "specific mode of conduct or end state of existence is personally or socially preferable to an opposite mode of conduct or end state of existence". Schwartz (1994, p.4) augments this notion of values as enduring beliefs stating that values "reflect the desirability of an end state or mode of conduct that transcends specific situations". Schwartz indicates that it is this state of belief that motivates action on the part of the individual or group. Such action could be making a decision, performing an act or the expression of an attitude, behaviour or sentiment. This conceptual perspective of values also highlights the significance of values as a precursor to the expression of sentiments.

The second perspective is the notion of values as concepts or principles. Kluchorn (1951) describes values as explicit or implicit concepts that are determining factors for choices. Guth and Taguiri (1965, p.7) define values as explicit or implicit conception, which act as a "guide to determining what is desirable". Hutcheon (1972) indicate that values are concepts that point out why a behaviour is acceptable or which state or behaviour is most acceptable from a set of options. Similarly, Braithwaite and Blamey (1998, p.364) define values as principles for actions "encompassing abstract goals in life and modes of conduct that an individual or a collective considers preferable across contexts and situations". Typically, such abstract concepts are mapped to real conceptual entities such as policies, rules, guides or principles. In addition, they are prescriptive because they express actions that need to be taken.

A third expression of values is as motivations and it has its basis in the satisfaction of human needs (Smith et al, 1956; Rokeach, 1960). Human needs are scarce and diverse and so values serve as a means for determining what motivations are most expedient. For instance, a person might value freedom because he/she is motivated by the need to be independent and self-sufficient. Such motivations can be expressed as the 'why' behind actions or behaviour. An example of such a value conceptualization is Schwartz Value Proposition (SVP) (Schwartz, 1992, 2012; Schwartz and Boehnke, 2004; Schwartz and Rubel-Lifschitz, 2005), which assumes that humans share universal values because motivations and needs are broadly the same across all cultures. Values as motivations contends that values are the motivating force behind any action.

Finally, values are also conceptualized as what is important to an individual or group of people. For example, Friedman, Kahn and Boring (2006, p.349) define values as "what a person or group of people consider important in life". This perspective stems from the notion that what individuals or groups of people consider as important or hold in high esteem is bound to influence or determine the actions they take and the attitudes they display. For instance, a business that considers the development of open source software as important to its operational model would not only invest in the development and use of more open source software but will likely discourage the use of proprietary applications.

With these definitions established, the next section considers some of the value formalizations.

3.2 Classification, Formalization and Application of Values

Research in the application and formalization of values has been an on-going task in the social sciences and humanities. Expectedly, these fields have evolved a significant proportion of the theory and methodologies adopted today in the classification, formalization and application of values. The task of formalizing values involves the detection of value motivations and items, their categorization into inventories or classes and finally aggregation into value orientations. Methods used in accomplishing these aims are centred principally on empirical surveys (Scott, 1965; Schwartz, 1994; Schwartz, 2004a, 2004b; Schwartz, 2012, McDonald and Gandz, 1991; Braithwaite and Scott, 1991), content analysis (Cheng and Fleischmann, 2010; Callicott et al, 2000; Ishita et al, 2010) and human/theoretical analysis (Rokeach, 1973).

Classifying values involves identifying value types and the structures and relationship between them. Typically, this process involves research efforts towards "enumerating the theoretically limited number of values that exist in the world and efforts towards categorizing those values into particular types" (Henry & Reyna, 2007, p.274). This entails detecting explicit and implicit values by analysing recorded communication in textual materials like speeches, debates, testimonies, reports and utterances. This detection is carried out by researchers and domain/subject experts and ends up with an enumerated list of value types or items. Value type enumeration of this sort always results in a wide and diverse collection of classifications because the researchers and domain experts could have different perspectives, the class of subject matters is almost infinite and the experiences of the domain experts are always quite diverse. Since the classification of a value is dependent on the perspective from which it is seen, Rescher (1969) describes six perspectives from which values can be classified. They include:

1. The subscribership to the value – This perspective classifies the values based on the individual or group that takes ownership of it. Typical value types here could include personal values, professional values and national values.

2. The object at issue – One of the features identified earlier in the review of values is the fact the values refer to entities, objects or states. Classifying values from the perspective of the object refers to classifications based on the referenced objects or entity. Examples include environmental values- where the object of value is the environment, thing or entity value where the values are about an entity or thing. Each object classification would have distinct set of value types that are associated with it. For example, values related to forestation like anthropocentric values and bio-centric values (Bengston et al, 2004) would never be mentioned in the same classification as values related to a football player or a political party.

3. The benefits at issue – In this case, values are classified based on the benefits accrued. For example, are the benefits economic, moral, intellectual, physical or religious. This classification also fits with the definition of values as means for determining the preferred state that is a state that is more beneficial to the value holder.

4. The purpose at issue - Here the perspective of the classification focuses on the purpose to be realized.

5. The relationship between subscriber and beneficiary – Classifications from this perspective focus on if the values can be classified as self-oriented or egocentric.

6. The relationship of the value to other values – Values are classified based on their relationship. For instance, Schwartz Value Theory, explicates the structure of the

dynamic relations amongst value types. For example, the pursuit of achievement values will typically conflict with the pursuit of benevolence values. In addition, pursuit of achievement and power values is usually quite compatible (Schwartz, 2012).

Understanding of perspective is typically the step that precedes the actual classification. Without an understanding of the perspective, values remain abstract and disjoint.

3.3 Classification Methodology

Classification consists of two stages – the identification of value items followed by categorization of items to inventories. Once this is accomplished, the values can be aggregated to determine value holder's value orientation.

3.3.1 Selection and Identification of Value Items

This process involves "enumerating the theoretically limited number of values that exist in the world" (Henry & Reyna, 2007, p.274). Once the perspective of the value is determined, a list of concepts that encapsulate the value and its goals are enumerated. These concepts which are normally words or expressions are called the value items. In literature, value items and concepts are used synonymously. Identifying the value items is usually subject to the goal of the researcher, the subject domain and the surrounding contexts. The item words or phrases are normally nouns, verbs or adjectives that reflect expected desires and actions. The enumerated words could be sourced intuitively, that is the researcher/s uses his/her intuition or experience to itemize a list of expected items. They could also be derived from reviewing literature and conducting surveys on domain experts. For instance, in Schwartz (1994) the goal was to identify a set of basic human values to which 56 basic human value items were identified, Scott (1965), identified 12 value items for the goal of identifying personal traits for ideal relations, Kahle et al (1988), identified 9 value items required for formalizing values to measure consumer attitudes and behaviour and finally, Crace and Brown (1996), in developing values for decision making itemized 14 value items.

Items describe abstract values and can be statistically graded e.g. a scale of 1-10 or 1-5. The score provides a quantifiable measure of the individual's application of the value. This empirical measure however has certain deficiencies. Human input required in judging and scoring is subjective. It can also be quite expensive and time-consuming depending on the number of respondents and items. Additional disadvantages to using value items include:

- The list of items could be too long and could confuse respondents.
- There is no way to determine an exhaustive list of items which capture all possible values especially because values have been shown to vary with time and because contexts evolve (Rokeach, 1973).
- The list of items is also subjective.

Finally, the items are categorized into value inventories.

3.3.2 Categorization of items to inventories

Value Inventory (VI) is a model that represents a set of value types and their constituent items. In this research, VI is defined as a list of items that provide explicit categories for the analysis of human values. The value type is the name assigned to the inventory. Identification of the value types is premised on perspective and because there can be multiple perspectives, there exists a wide array of VIs of diverse origins, purposes and contexts. For instance, Schwartz's (1992, 2012) value conceptualization focused on the notion that human values are based on motivational goals and needs that are basically universal across all cultures and peoples. Hence, the set of values identified by Schwartz are generic and applicable to social issues (Appendix A2.1 tabulates Schwartz's VI). Rokeach (1973) conceptualized values from two perspectives, where in one case, the values are perceived as a set of ultimate goals called terminal values, in the second, they are perceived as modes of behaviour. The result of this was a list of 36 value items categorized into terminal and instrumental value types (see Appendix A2.2). Bernthal (1962), proposed a hierarchy of values for management decisions based purely on rational reasoning. The inventories contained – the business firm level, economic system level, societal level and individual level. Personal Values Questionnaire (PVQ) (England, 1967) comprised of 66 value items organized into 5 categories (see appendix A2.3). Inventories are thus instruments expressing and modeling value classifications and are subsequently used in determining values.

3.4 Value Inventories

In this section, some popular inventories and their limitations are described.

3.4.1 Rokeach Value Survey (RVS)

RVS was developed in 1973 to show a theoretical connection between values and behaviour. Rokeach arrived at the inventory through an initial rational and intuitive selection of value items from reviewing literature and observing personality traits (Rokeach, 1973). The outcome was two value categories (Instrumental and Terminal values) made up jointly of 36 value items (see appendix A3.2). RVS is based primarily on intuition and while it has received significant reference, a major criticism is that because it is based on human intuition it has no statistical or empirical basis.

3.4.2 Schwartz Value Inventory (SVI)

Schwartz (1992, 2012) proposed a universal value framework called SVI. It was based on the assumption that human motivations and needs are vastly the same across all cultures. The inventory was derived from surveys conducted in 44 countries and the Rokeach Value Survey. The survey from which the inventory was derived is called the Schwartz Value Survey (SVS). SVI describes ten motivationally unique values from 3 universal requirements of human condition. These are the needs for survival as biological organisms, the need for coordinated social interaction and finally, the welfare needs of groups. SVI models the dynamic relations of congruence and conflict among value types. This relationship has been empirically proven in various research projects. For instance, in

Schwartz and Rubel-Lifschitz (2005, 2009), on the question of whether 'gay and lesbians should be free to live as they like', it was discovered that people with conformity and traditional values correlated negatively with accepting personal freedom for gay people while people with hedonistic and universalistic values correlated positively with freedom for gays. This also showed a strong correlation between conformity and tradition, hedonism and universalism values as well as the conflict between traditional values and hedonism values. Evidence of the veracity of this value structure has been established in samples from 67 countries (Schwartz, n.d(b); Schwartz, 1992).

The SVI has been widely applied in research, primarily because it cuts across several cultures and its applicability to generic social issues. For instance, SVI has been applied towards exploring the relationship between behaviour and value conflict (Schwartz, 1992, 2007; Schwartz and Bilsky, 1987). In marketing research, it has been applied to explain specific aspects of customer behaviour (Grunert and Juhl, 1995). It has also been used to explain the relationship between values and party affiliations (Schwartz, 1996; Capara et al, 2006). Despite its wide acceptance and use, the SVI has several limitations. It is limited to generic social values and it is difficult to apply to specific domains and contexts such as in the office, organization or a sports club. Secondly, Cheng et al (2010) reported significant inter coder disagreement in applying the SVI towards recognition of values in net neutrality policy debates. This is because of significant ambiguity in relating the meanings of value items to content. Furthermore, the fact that there are over 56 basic value items covered across 10 values, increases the likelihood of inter coder disagreement as respondents would view comments and their associated values differently.

3.4.3 Personal Values Questionnaire (PVQ)

PVQ's design is like the RVS in that the values were derived from literature and observation of human behaviour (England, 1967). It was designed for application in a business context, to study the values applied by business managers. The initial set of values were derived from 200 concepts and subsequently trimmed to 66 concepts by domain experts and real-life business managers. Like the RVS, the concepts were grouped into five categories (see appendix A2.3). Unlike RVS, PVQ is context and domain specific partly because it is aimed at businesses but also because of the empirical input of domain experts in trimming the initial list of value items from 200 to 66. However, 66 value items represent a sizeable list of words to consider and raises the risk of inter-coder disagreement. Another flaw of PVQ is that some of the concepts do not in themselves constitute values (Cheng and Fleischmann, ASSIST, 2010). For instance, concepts such as employees, customers and government are not actual values because they are not expressions of motivation.

3.4.4 List of Values (LOV)

LOV is based on the importance of people in value fulfilment (Kahle et al, 1998). It was first designed to measure consumer attitude with its focus on personal values that apply to people's daily lives (Cheng and Fleischmann, 2010). LOV's value types are founded on a combination of RVS, Maslow's hierarchy of needs and existing values literature. The LOV values include fun and enjoyment, warm relationships, self-fulfilment, being well respected, sense of accomplishment, security, self-respect, sense of belonging and excitement. The list of value items is limited and does not necessarily encompass all possible values expressed by humans. It also does not capture the fact that multiple value items can be expressed by a person at any time.

Appendices A2.4-A2.8, describes other value inventories. Inventories also serve to distinguish groups and identify their Value Orientation (VO). According to Kluckhorn (1951), VOs are "a set of linked propositions embracing both value and existential elements". People have more than one value and a collection of related value types form their orientation. For example, a person who identifies himself/herself, as politically conservative will hold a collection of values on issues. It is the sum of these values that form their orientation and their grouping as leftist, Marxist etc. Computing the aggregate involves the use of empirical surveys in collating individual scores on inventories.

This section has shown that values are viewed as VIs, but how are these inventories built?

3.5 Building a Value Inventory

In this section, the process through which VIs are built are discussed. Formalized value models are applied towards tasks such as determining people's values or automatically identifying values from text. Cheng and Fleischmann (2010) highlight three methods for inventory development. They are:

Rational-Theoretical Inventories – These are inventories conceptualized purely from rational human intuition or a priori inventories. Examples include RVS, LOV and the PVQ. Issues with this approach are: The values specified are not verifiable since they are not grounded in empirical analysis. Secondly, they do not apply to all possible contexts or situations: in other words, they are not a one-size fits all inventory. Thirdly, the approach is entirely subjective since there's no way of determining the number of values or items that constitute the inventory. According to Hofstede (1980, p7), "inspection of the number of instruments designed to measure human values makes it clear that the universe of all human values is not defined and that each author has made his or her own subjective selection from this unknown universe, with little consensus among authors".

Empirical Inventories – Empirical approaches involve directly assessing the value items from subjects through surveys, interviews, focus groups, content analysis on a representative sample of individuals. Participants are required to rank or manually rate derived value items drawn from human motivations according to relative importance (Braithwaite and Scott, 1991). Examples of empirically derived value inventories include Bird and Water's managerial moral standards (1987), the Personal Value Scale (PVS) (Scott, 1965).

Empirical surveys result in a definition of the values held by the individuals including behaviours and attitudes induced by the values. Associated problems include: Surveys are time consuming and can incur considerable expense especially considering if the domain is quite complex and spans several groups of interests and stakeholders. Secondly, as human values change over time, empirical surveys become quite impractical as the cost of conducting them would rise significantly. Domain knowledge is also required when carrying out empirical surveys. Researchers must be aware of the domain, its nuances and vocabulary. This requirement further amplifies the complexity and practicality of empirical surveys. In addition, since empirical methods require human participation, it is liable to several biases including: self-selection and participation bias. Added to this is the risk of participants not answering questions correctly due to insufficient reflection, self-deception, conscious or subconscious withholding of information.

Content Analysis (CA) is another form of empirical analysis that is quite common in value research (Cheng, et al, 2012). Neuendorf (2002, p.1) defines it as "the systematic, objective quantitative analysis of message characteristics". Essentially, CA involves the use of coders or human annotators in detecting values embedded in content. The coders apply their judgement in highlighting the values associated with content and consequently group them into inventory types. Once sufficient inter-coder agreement is achieved the annotated inventory forms the value model. CA techniques have been used for 50+ years in various fields including journalism, sociology, psychology and business studies (Krippendorf, 2012). As such, it has become an effective method for "tracking markets, political leanings and emerging ideas" as well as "exploring individual human minds" (Krippendorf, 2012, p.1). Some benefits of applying CA to values discovery include its unobtrusiveness (Morris, 1994), ease of use for testing hypothesis (Schmidt, 2007) and the fact that humans can recognize subtle, unconscious and implicit values in content is hugely beneficial. Limitations of CA include the following:

- Coding can be quite expensive and time consuming especially when the corpus is large Takayama et al, (2014).
- Coders are prone to human bias and coding errors (Ho and Quinn, 2008).
- Coders requires domain knowledge.
- Coders are required to agree on value types. These types are often vastly subjective resulting in inter-coder disagreement (Takayama et al, 2014; Cheng et al, 2012).

Theoretical-Empirical Inventories – These inventories are a hybrid of empirical and rational theoretical methods. A popular example is the SVI. By combining rational theoretical and empirical methods, the goal is to provide an empirical basis for the value items selected rationally.

3.6 Conclusion

Values are abstract unseen entities which influence behaviour. Approaches in modeling or applying values demand extensive human input in the form of empirical surveys and content analysis. In addition, the models of values are an enumerated list of words which shows that unseen values can be expressed as texts. This also justifies the view in this research that if values can be modelled from text then they can be realistically embedded in sentiment prediction. However, several issues exist with inventories

- The inventory of texts is often too large and when applied by humans in identifying values embedded in sentences, can result in disagreements and misunderstanding.
- Some words in the inventories do not really constitute values.
- Most inventories are completely domain specific and cannot be applied to other contexts. This is because the perspective from which they are formulated are tied to a domain.
- Inventories themselves are too structured. Human beings can make judgements based on more than one value and inventories are incapable of capturing this.
- Finally, approaches to value orientations are based purely on human judgement.

In conclusion, although value models exist, they are rigid, domain dependent and require considerable human input. However, the fact that values can be related to textual expressions presents an opportunity to explore new ways of automating the process of modeling values and incorporating them into sentiment prediction. Before addressing the research approach, the research methodology is described in the next section.

4. Research Methodology

The purpose of this research involves implementing a solution for recipient sentiment prediction, through a model of human values. In addition, part of the challenge is to adopt a methodology that is devoid of human input or annotations. This would involve creating an innovative technique for classifying or modeling values from text, creating an algorithm for SA that can harness the implemented value model and finally applying the SA algorithm towards predicting the recipient's sentiment. Therefore, since these target objectives involve the creation of innovative methodologies as well as a design implementation, the research methodology adopted is based on Design Science Research (DSR) methodology. DSR is a body of knowledge which promotes the design and implementation of artifacts in solving information systems research problems (Simon, 1996). In the next section, DSR is briefly introduced. Following this, the DSR guidelines are related to the tasks that will be carried out towards the accomplishment of the design objective.

4.1 DSR Methodology

According to Hevner et al (2004), in DSR, the knowledge and understanding of a problem domain and its solution are achieved through the building and application of artifacts. The design of innovative artifacts are premised on existing knowledge and bodies of work that have been applied, tested and proven (Markus et al. 2002; Walls et al. 1992).

In addition to providing an innovative solution to a problem, the essence of research is to generate new knowledge that can be included to the already existing body of knowledge on the research subject. DSR methodology accomplishes this task through the implementation of processes or activities drawn from existing knowledge, resulting in artifacts. The evaluation of the artifacts "provides feedback information as well as a better understanding of the problem" while also improving the quality of the artifact and the design process (Hevner et al, 2004, p.78). The outcome is an iterative 'build' and 'evaluate' loop, where the 'build' process involves the sequence of steps and activities applied in building the artifact and the 'evaluate' process entails the evaluation and assessment of the artifact. This back and forth loop between modifying the process and artifact continues until the final artifact is attained (Markus et al, 2002).

What then are artifacts? March and Smith (1995) identifies four artifact types. They include:

- Constructs - These artifacts are vocabularies and symbols. According to Söhon (1983) constructs "provide the language in which problems and solutions are defined and communicated" (Hevner et al, 2004, p.78). Constructs describe a way of representing the abstract problem so that models can be built from them.

- Models - Models are a representation of the real-world problem derived from constructs. According to Hevner et al (2004, p.78), "Models aid problem and solution understanding and frequently represent the connection between problem and solution components enabling exploration of the effects of design and decisions and changes in the real world.".

- Methods - These include algorithms and processes that provide guidance as to how to solve a problem. They could be formal mathematical algorithms or textual descriptions of best practices approaches or a hybrid combination of both.

- Instantiations – These artifacts are typically prototype systems which show how constructs, models or methods are implemented in an actual functioning system. Instantiations are important in DSR research because they "demonstrate feasibility, enabling concrete assessment of an artifacts suitability to its intended purpose" (Hevner et al, 2004, p.79).

It is important to differentiate basic system design or software development from DSR. The difference lies in the nature of the research problem and the solutions provided. According to Hevner et al, (2004, p.81), conventional development or system design entails the application of existing knowledge to organizational problems using best practice artifacts and accepted methodologies/techniques. Conversely, DSR is aimed at unsolved problems in "unique innovative ways or solved problems in more effective or efficient ways". In addition, the research contribution identified using DSR is also a key difference between routine design and DSR. Since the research problem in this thesis is unique and the approach in using values to model sentiments is also innovative and untried, accounts for the use of DSR methodology in this thesis. Hevner et al (2004, p.82) provides 7 guidelines for applying DSR. The guidelines are described as follows:

"Design-science research requires the creation of an innovative, purposeful artifact (Guideline 1) for a specified problem domain (Guideline 2). Because the artifact is purposeful, it must yield utility for the specified problem. Hence, thorough evaluation of the artifact is crucial (Guideline 3). Novelty is similarly crucial since the artifact must be innovative, solving a heretofore unsolved problem or solving a known problem in a more effective or efficient manner (Guideline 4). In this way, design-science research is differentiated from the practice of design. The artifact itself must be rigorously defined, formally represented, coherent, and internally consistent (Guideline 5). The process by which it is created, and often the artifact itself, incorporates or enables a search process whereby a problem space is constructed and a mechanism posed or enacted to find an effective solution (Guideline 6). Finally, the results of the design-science research must be communicated effectively (Guideline 7) both to a technical audience (researchers who will extend them and practitioners who will implement them) and to a managerial audience (researchers who will study them in context and practitioners who will decide if they should be implemented within their organizations)".

Table 1, which is reproduced from Hevner et al, (2004) illustrates the guidelines that must be taken by a researcher in applying DSR methodology. The next section shows how these guidelines are applied in the completion of this research.

4.2 Design Steps

The DSR guidelines are mapped to steps to be taken in accomplishing this research. Each step encapsulates a DSR guideline and refers to a chapter or chapters in this thesis. Figure

2 illustrates these steps. Each step is mapped to a DSR output and its representative chapters in this thesis. The subsequent sections discuss the steps in their order.

4.2.1 Step 1: Awareness of problem

DSR 'guideline 2' in table 1, surmises that the DSR requires the creation of an innovative, purposeful artifact for a specified problem domain. Without a knowledge of the problem domain, there can be no research, thus, identifying the problem domain is the first step in this research. The awareness of the research problem involves identifying and outlining the research gap by accessing a knowledge base of existing research, processes and artifacts in order to identify and formulate the gaps in research. The output of this stage is a documented proposal of the research problem and ways by which the eventual solution would be evaluated. In this thesis, chapters 1, 2 and 3 constitute the output of the problem awareness. In chapter 1, the research problem is briefly discussed – the gap in research is identified and the overall aim and objectives are clearly elucidated. Chapters 2 and 3 reviews the existing knowledge and provide a more detailed insight into the current limitations.

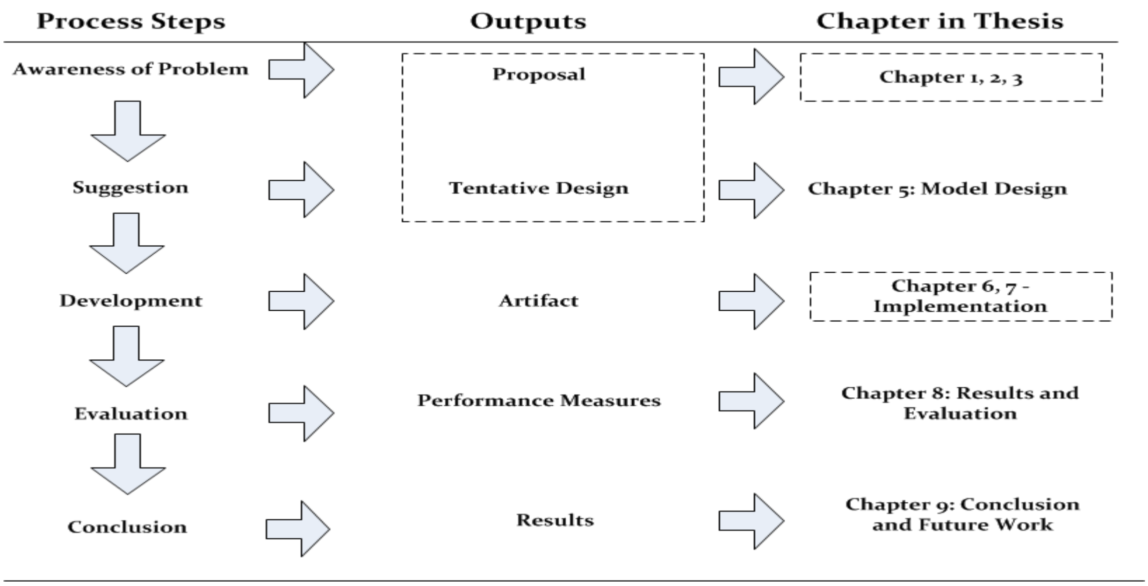


Figure 2: Design Science Research Process Model

4.2.2 Step 2: Suggestion

This step involves the creation of a tentative design or prototype, which is why it is called 'Suggestions'. It also maps to guideline 5 of the DSR guidelines in table 1. Quoting Hevner et al, (2004, p82), "... The artifact itself must be rigorously defined, formally represented, coherent, and internally consistent (Guideline 5) ...". Rigor addresses the way the research is conducted. It comprises of the mathematical formalisms, theorems, foundations and principles applied in describing or formalizing the problem. Research rigor is predicated on the selection of appropriate techniques from literature to develop a theory or artifact. According to Peirce et al. (1965, p51), "suggestions are drawn from existing knowledge/theory base for the problem area". The outcome is thus an initial set of formalisms or designs that describe the problem. Therefore, chapter 5, describes a design which focuses on the modeling of values from text via the identification of clues embedded in values definition. A formalism for representing values via a process called value decomposition is also described.

The artifacts obtained from this step include a structural representation of values, a translation of the values into a mathematical model. In addition, the design proposes a new and innovative way of expressing the sentiment of the hearer, and from this perspective a value sentiment model design that incorporates the modelled values is described.

Table 1: Design Science Research Guidelines reproduced from (Hevner et al, 2004, p. 83)

Guideline	Description
Guideline 1: Design as an Artifact	Design-science research must produce a viable artifact in the form of a construct, a model, a method, or an instantiation.
Guideline 2: Problem Relevance	The objective of design-science research is to develop technology-based solutions to important and relevant business problems.
Guideline 3: Design Evaluation	The utility, quality, and efficacy of a design artifact must be rigorously demonstrated via well-executed evaluation methods.
Guideline 4: Research Contributions	Effective design-science research must provide clear and verifiable contributions in the areas of the design artifact, design foundations, and/or design methodologies.
Guideline 5: Research Rigor	Design-science research relies upon the application of rigorous methods in both the construction and evaluation of the design artifact.
Guideline 6: Design as a Search Process	The search for an effective artifact requires utilizing available means to reach desired ends while satisfying laws in the problem environment.
Guideline 7: Communication of Research	Design-science research must be presented effectively both to technology-oriented as well as management-oriented audiences.

4.2.3 Step 3: Development

In this step, design artifacts which include processes for addressing the problem and an implementation (instantiation artifact) are developed and implemented. This step encompasses guidelines 1 and 6 of the DSR in table 1. In this thesis, this step is spread over chapters 6 and 7. Chapter 6 focuses on the description of the model implementation, including an elucidation of the value sentiment algorithm. It describes all the process applied in implementing the model ranging from data preparation to the actual model implementation. Chapter 7, focuses on how the implementation in chapter 6 is applied to the research test corpus from the political domain. It is the outcome of this implementation that are evaluated in the next step.

4.2.4 Step 4: Evaluation

This step involves the evaluation of the artifacts and falls under guideline 3 in table 1. It also entails the identification and explanation of any deviations from the expected results. Modifications to the model and their effects are also described. The output of this section includes evaluation measures and their results. In chapter 8 the testing and evaluation of the model is described. The evaluation measures used are precision, recall and F-Score (Korphage, 1997; Yang and Liu, 1999; Jurafsky and Martin, 2009). Precision is the fraction of retrieved documents that are relevant in other words, precision is a measure of how accurately the model performed. Recall on the other hand is the fraction of relevant documents retrieved. Other evaluation measures such as accuracy and misclassification rate are also computed to enable comparison with other systems whose performances are based on accuracy or misclassification rate. In addition, in this chapter, the evaluation scenarios and setups are also described.

4.2.5 Step 5: Conclusions

This step marks the finale of the research. The research results are judged and written up. Chapter 9 of this research provides a description of the research conclusions. It provides an examination of the final results against the outlined aims and objectives. This section maps to guidelines 4 and 7 in table 1. In addition to summing up the conclusions of this research, the contributions i.e. the knowledge gained, facts learned and the issues/limitations associated with the research are discussed. Suggestions on possible future research are also outlined.

4.3 Conclusion

In conclusion, DSR methodology, offers a methodology that is fit for this thesis since it provides steps for the creation and implementation of new and innovative artifacts, and this is analogous to the objective of this thesis. Applying this methodology, the next step in this document is the suggestion step and this is described in the next chapter as the model design.

5. Model Design – A model of Sentiment and Values

In previous chapters, it's been shown that values as abstract, latent entities are principal determinants of human behaviour and that the expression of a recipient's sentiment regarding a subject matter is a form of behaviour. In this chapter, a methodology for formalizing and modeling abstract unseen values is described. The description of the model stems from the notion that although values are abstract and unseen, they are implicitly observed in the behaviour of individuals and the utterances and sentences they make. According to Entman (1993, p2), a speaker conveying a message about an entity or subject will "select aspects of a perceived reality, and make it more salient in a communicating text, in such a way as to promote" a particular view point. This research suggests that there is a link between the choice of lexical units used in describing an event and the underlying values held by the speaker. Therefore, the verbal descriptions or linguistic units used in describing an event portray the underlying values held by the speaker towards the event or subject matter. This section commences with the theoretical characteristics of values as drawn from literature, followed by a description and portrayal of the common approach to value models. Through this characterization and portrayal, the parameters necessary for the development of the model are identified.

5.1 Characteristics of Values

Values refer to preference for one state over another and so communicate and emphasize priority. When one prefers a state or event over another, then such a person will prioritize that state over others and will emphasize such a value in his expressions. Priority for one state over another suggests that the nature of values is hierarchical. The choice of a state in the hierarchy is subject to the underlying context and situation. For instance, some values will have a higher priority in an office environment while in a family setting it might have no place. In an experiment held in Germany and Israel to determine the value priorities of adolescents between the ages of 9 and 18, it was discovered that value priorities vary with different life contexts e.g. the value prioritized in school is different from the value priority when amongst family members (Migration and Societal Integration Research Consortium, n.d). In the experiment, it was shown that in school, achievement values were more prominent while tradition values were more dominant in the family. As such it was concluded that, "Value priorities between contexts will differ to adapt to contextual demands" (Migration and Societal Integration Research Consortium, n.d). Several characteristics can thus be drawn:

1. Values are abstract quantities that account for the behaviour and attitude of individuals.
2. Values refer to goals, objectives, objects, entities, end states or actions – Schwartz (n.d(a)) described values as the criteria through which people use to evaluate actions, people and events.
3. Values vary in priority amongst individuals and groups.
4. Values account for a person or group's preference for an entity over another.

5. Values are contextual.
6. The application of values is a function of the contextual setting and situation. However, there exist a class of values, which transcend specific situations. Such values are called instrumental values and they are behavioural and moral codes (Rokeach, 1973).
7. People's values also evolve over time - Bengston et al (2004), identified a change in values regarding forest areas in America, noting a steady shift from anthropocentric values to more bio-centric values in the 21st century.
8. Values are hierarchical and ordered by importance relative to one another (Schwartz, 2012). According to Schwartz (2012, p4) "people's values form an ordered system of priorities that characterize them as individuals". Some people will prioritize open-source values over profit making while others might prioritize achievement over justice (Schwartz, 2012).

With these characteristics established, the next section explores the modeling of values and sentiment.

5.2 Description of Value-Sentiment Model Processes

To differentiate the model in this research from contemporary value models, it is vital to provide an overview of how values are modelled and applied to sentiment prediction. From literature, research in understanding, formalizing and applying values involves value identification/modeling and value application. In this research these processes, that is, the journey from abstract values to full sentiment prediction is illustrated as a five-stage process which for the purposes of this thesis is called Value-Sentiment Model (VSM). Figure 3 depicts this five-stage process.

Each stage of the VSM has a goal. In figure 3, the methodologies and outcome of each goal are depicted. The journey from abstract values to sentiment prediction is a process where the output of one stage is fed to the next. This accounts for the arrows shown in figure 3, that point from one stage to the stage directly below.

The processes described in order are:

Stage 1: Identifying the perspective, motivation and subject of the value

This step involves questions such as identifying the subject of the value or the states and actions addressed by the value? What are the expectations of the value holders? who are the subscribers to the value? what scenarios could exist for the value to be applied? what purpose will the values serve? As shown in figure 3, these questions are typically answered by interviewing value holders through surveys or questionnaires. The outcome of this process includes:

a. A clearer understanding of the value's purpose.

b. A clear definition of the value's subjects.

c. An understanding of who the value holder is and the relationship between the value holder and other stakeholders.

d. The identification and itemization of the scenarios where the values would be applied.

This stage is followed by the second stage, which involves the identification of value items.

STAGE 1	GOAL	Understand the goals and motivations of the value
	METHODOLOGY	Human intervention – interviewing value holders, surveys, questionnaires
	OUTCOME	List of value subjects and entities, identification of scenarios and perspectives
STAGE 2	GOAL	Identifying value items
	METHODOLOGY	Human intervention – Theoretical enumeration of value items, content analysis, surveys, interviews
	OUTCOME	Enumerated list of value items
STAGE 3	GOAL	Categorization of enumerated value items into a value inventory
	METHODOLOGY	Human intervention – Value holders use domain knowledge to create categories and assign items to categories
	OUTCOME	Value inventory made up of categorized value items
STAGE 4	GOAL	Discover value orientation
	METHODOLOGY	Human intervention – numerical and linguistic grading
	OUTCOME	Identification of the value holders leaning based on numerical score
STAGE 5	GOAL	Application of value orientation to sentiment prediction
	METHODOLOGY	Human intervention – Identifying value items from input sentence and inferring the likely orientation from the value orientation model
	OUTCOME	Sentiment polarity of the input sentence

Figure 3: Illustration of the Five-Stage VSM

Stage 2: Identifying value items

A clarification of perspectives, motivations and subjects paves the way for the identification of value items. As discussed in chapter 3, the value items are a list of concepts that encapsulate the value and its enumerated goals. The identification of goals in the first stage greatly simplifies this process. The methodology for identifying value items could be purely theoretical, where the value holder or researcher simply enumerates a list of possible concepts based on his/her knowledge of the goals and perspective (domain knowledge). Alternatively, it could also be carried out through content analysis or empirical surveys of content – documents, speeches, utterances and sentences made by the value holder.

Stage 3: Categorization of value items into inventories

The items derived in the second stage are simply a list of relevant concepts that make up the value. Typically, this list is normally long, containing repeated concepts. Therefore, the goal of this stage is to shrink the list and eliminate repetition by grouping the items into user friendly categories that are based on the perspective and goals identified in the first step. The methodology here will also involve human influence through approaches such as content analysis or empirical surveys. Humans would propose category names and map each item to a category.

Stage 4: Discover Value Orientation

The outcome of the third stage is a value model that consists of a set of value categories and their relevant items. In this stage, the objective is to identify the orientation of each value holder towards a subject matter by assigning grades and scores that are representative of their orientation under certain scenarios. These grades could be numeric or linguistic. In this stage, human intervention is required. The process would involve presenting value holders with a subject and a scenario for which they would be required to assess their orientations relative to each value item on a scale, for instance, 1 to 10 (in this instance a form of numerical grading). The total score would represent a measure of the value holder's orientation. Since the scores are numeric, it offers a unique benefit in that they can be visualized on plots and used to identify groups and clusters of people with similar values or extreme values.

In linguistic grading, the value holders would use words such as comparative adjectives to describe their preferences. This would normally be presented in the form of a questionnaire or interview. For instance, the value holder could be asked a question such as, "*In scenario X would you prefer more of subject A, less of subject A or none of A*" or "*In scenario X which is the best position A, B or C*". The adjectives posed in this sort of questions are meant to elicit the orientation and preference of the value holder.

Stage 5: Application of value orientation model to Sentiment Prediction

So far, the modeling process has progressed from abstract unobservable values to a structured measurable model of individual values under given scenarios. The objective in this stage is to apply this structured model towards predicting the sentiment of an utterance. It is a form of value application. The methodology involves humans inferring from the sentence, the subject matter, and its end state (this would be an item in the value inventory) and then comparing it against the known orientation of the value holder. The utterance is positive if the inferred preference matches the value orientation of the value holder and negative if it does not and thus, the outcome is the sentiment polarity of the value holder. In the work of Abelson and Carroll (1965), where beliefs are structurally modelled in a knowledgebase, the orientation of a new sentence is compared against sentences in the knowledge base in a process called credibility testing.

Considering the research aims and objectives, the value model and methodology described so far has several drawbacks. The major drawback is the recurring need for human intervention (content analysis and empirical surveys) in practically all the stages. Empirical surveys and content analysis are applied in the identification of the subject, perspective and scenarios. Value items are identified from surveying value holders or enumerated through a knowledge of the domain. The choice of value categories, computation of value orientation and deducing sentence sentiment from orientation are all dependent on human effort making the value modeling process described above unfit for this research. Therefore, this chapter elaborates on a value model formalization making comparison with the five-stage process described above. The model described, is premised on the notion that when people speak or express themselves on any subject matter they ultimately seek to express and convey some functional value concept linguistically. As such, given a large collection of utterances/sentences made by a value holder, the model, on analysing the document should be able to deduce the value holder's orientation on any subject given a particular scenario.

Several challenges exist in this modeling task. How can the contexts, subject and perspective of the value be deduced from the collection of utterances? Similarly, how can the value items, inventories be determined considering that utterances could have diverse meanings and the subject of utterances could be expressed in diverse ways. Most challenging of all, how can the inferred value of a sentence be determined and compared against a value orientation without any human input? To this end, this section describes a model that automates and formalizes the five-stage process. It identifies a structural representation of values drawn from a review of value definitions and conceptualizations. This structural representation called value decomposition plays a vital role in mapping utterances to values.

5.3 Value Model – Decomposition of Values

The conceptual definition of values and the characteristics hold the key to decomposing values. Values decomposition involves identifying the constituent parameters which make up values in observed text. The necessity for value decomposition arises chiefly from the challenging need to not introduce human input in the value modeling process. Therefore, there is a need for a formal approach applicable to any sentence so that the elements which make up the values i.e. the value items and categories are extracted and made ready for aggregation and sentiment prediction. In addition, the need for decomposition also arises from the fact that value sentences can be authored creatively and do not always explicitly reveal their items or orientation. In prior works such as Abelson and Carroll (1965) and Carbonell (1978), the input sentence structure is set to a pattern '[Person] says [C P]' where [C] and [P] refer to a concept and a predicate, making these systems limited and rigid. Values decomposition provides the theoretical framework for the extraction of value items and their categories. It also supports the ability to map utterances which are vocalized sentences to physical representations of values.

To achieve this objective, clues, and recurring patterns in the definition and characteristics of values are identified and applied to formalize the decomposition. Some of the definitions explored in chapter 3 are:

Values are -

"The criteria through which people use to evaluate actions, people and events." (Schwartz, 2006, p.1)

"Abstract coordinators of behaviour." (Rokeach, 1973)

"Latent variables that have explanatory value for the choices people make." (Verplanken and Holland, 2002).

"The belief that a specific mode of conduct or end state is personally or socially preferable to an opposite mode of conduct or end state." (Rokeach, 1973)

"Determining factors for choices and a guide for determining what is desirable." (Guth and Taguiri, 1965; Kluckhorn, 1951).

"Concepts that point out why a behaviour is acceptable or which state or behaviour is most acceptable from a set of options." (Hutcheon, 1972).

"What is important to an individual" (Friedman et al, 2006).

"Principles encompassing abstract goals in life and modes of conduct that an individual or a collective considers preferable across contexts and situations" (Braithwaite and Blamey, 1988).

From these definitions, when values are portrayed as beliefs, observe that the definitions refer to the 'end state' or 'mode of conduct'. As concepts, it refers to 'determining what is desirable' which reflects a question of choice i.e. selecting from a choice of states, features or possible events etc. While as a motivation, reference is made to 'what is important' or the 'motivation behind an action'. Here 'what is important' and the phrase 'an action' could refer to an end state, a state of existence or an event. It is observed from this that any value, as a belief, concept or motivation is directed at an object or entity, which could be abstract or real and would have a set of states[18], properties or features. The value could also be directed at the state or aspect of the event/entity. This relationship between values and the object is expressed as the belief that an outcome, end state or specific mode of conduct is

[18] State refers to a mode which the object can exist in

personally or socially preferable to a converse mode of conduct, state of existence. Therefore, a value is made up of the following constituent parameters:

- Value Holder (H) – Values are developed and applied by value holders. Value holders could be individuals or groups such as societies, clubs, political parties. Hoyland et al (1953), reinforces the importance of the value holder stating that the 'effects produced by persuasive communication are critically dependent upon the characteristics of the communicator'.

- Subject of the Value (θ) – According to Schwartz (2006) values are criteria through which people use to evaluate actions, people and events. Therefore, a value must refer to a subject. The subject of the value is the object, entity, event, person, item, place that is referred to in the expression of the value. It is what the value is about. The subject of the value could be a real or abstract entity.

- State (S) - In the definitions, above, phrases such as 'end state', 'mode of conduct', 'what is desirable' reflect a question of choice and preference. A person's value for a subject will be his preference for a state of the subject; where state refers to a position, marked feature or property of the subject that is preferred amongst a set of features. For instance, on the subject '*house prices*', states could include expressions such as '*high*', '*low*', '*expensive*', '*exorbitant*', '*affordable*', '*stable*' etc. (Use of these words and phrases are a form of linguistic grading). A value holder's value refers to his preference for one state over another.

- Action (A) – For any subject, the preference of the value holder also extends to the preferred action or activity to be carried out or performed by or on the subject. Reference to phrases like 'motivation', 'specific mode of conduct', 'specific mode of action' and 'end state' suggest that the essence of the value could also refer to an action, a conduct or an activity to be performed. In fact, the preferred state (S) can also be described as the 'state of being' preferred by the value holder while the preferred action refers to the 'action, activity or process' undertaken on the subject or by the subject that is most preferred by the value holder. Usually, the preferred action would lead to the subject existing in the preferred state. For instance, in the sentence, "*We have a plan to destroy terrorist groups across the region*", the preferred action on the subject '*terrorist group'*, is '*the plan to destroy*'. Using the '*house prices*' analogy, actions could include expressions such as '*increased*', '*reduce*', '*rocketed*', '*risen*', '*plummeted*', '*rebounded*', '*raised*'.

- Context (C) – Context refers to existential factors that are usually outside the control of the value holder. These factors include elements such as time and place, background knowledge of the speaker or hearer, the expectations of people, the location and the nature of the subject matter. This list of existential factors is not exhaustive. They could be static - factors that are fixed and unchanging such as date of birth, address etc. or dynamic – factors that are constantly changing e.g.

temperature, preference, desires, social environment etc. (Henricksen et al, 2002). The value conceptualization of Braithwaite and Blamey (1988) which states that values are "principles encompassing abstract goals in life and modes of conduct that an individual or a collective considers preferable across contexts and situations" reinforces the notion that the applicability of a value is dependent on the context. Experiments conducted in Germany and Israel to determine the value priorities of adolescents between the ages of 9 and 18 showed that the values prioritized by the subjects were dependent on the context, where context referred to the subjects' environments of school and home. As such it was concluded that, "Value priorities will differ to adapt to contextual demands."[19]

In summary, values are made up of five parameters: the value holder (H), the subject of the value (Θ), the preferred state (S), preferred action (A) and the context (C). Together, these parameters are called Value Components (VC) and represent a formalism for translating value laden sentences into structures ready for identifying value orientations. Thus, given a large corpus of utterances by a value holder, the value model would be a function (f), that takes a text and maps it to a value representation:

$$V = f(H, \Theta, S, A, C) \ldots (1)$$

Following this decomposition of values, figure 3 is modified to portray the VCs H, Θ, S, A, C and outcomes of the VSM. To this end, the first stage 'Identifying the perspective, motivation and subject of the value', the outcome are the parameters $\{H, C, \Theta\}$ because:

- At this point, all the value holders must be enumerated (H).
- Secondly, since the source of values are a large corpus of observed texts and utterances, it is highly likely that numerous subjects would be referred to in diverse contexts and so in addition to the value holders (H), all the subjects (Θ) and relevant contexts (C) must also be enumerated, hence the outcomes, $\{H, C, \Theta\}$. This is depicted in figure 4 which is a modified version of the five-stage VSM illustrated in figure 3.

The next stage in the VSM involves the identification of value items which are a list of concepts that encapsulate the value and its enumerated goals. Parameters, Action (A) and state (S) constitute the outcome of this stage because they encapsulate the make-up of the value subject encompassing its properties or features (state of being) and the actions performed on or by the subject. Critically, it is evident that all the parameters of the decomposed value feature are outcomes in the first and second stages of the VSM. Thus, the VSM is modified further to consist of four stages where the first two stages - Identifying the perspective, motivation and subject of the value and identifying value items are merged into one stage called parameter identification - since the outcome of both stages entail the

[19] http://www.migration.uni-jena.de/project4/about_the_project/index.php: Last accessed 23/05/2015

identification of value parameters as seen in figure 4. By relating the outcomes of the decomposition process to the VSM, an updated VSM with clearly defined outcomes is obtained meaning that the methodology for identifying the items can be focused on in the next section.

5.4 VSM Parameter Identification of Value-Laden Utterances

In the absence of human input, VCs are identified by analysing the corpus of utterances[20] for clues and patterns that are consistent with the properties of each VC. One justification for this approach can be attributed to Chomsky (1963) who suggested that language and particularly its grammatical organization can provide an especially clear window into the structure of the human mind. Therefore, a study of the grammatical constructs and linguistic units used by value holders under certain contexts can provide clues as to the nature of their values.

A second justification is founded on the relationship between abstract values and words. Values reflect a pragmatic aspect of communication. Pragmatics are derived from semantics and the compositional meaning of utterance are derived from structurally and syntactically correct sentences. Therefore, one way of understanding values is by understanding the semantics of the entire sentence, by breaking down the syntactic aspects into parts and analysing them. Consider again the example of the subject '*house prices*' and a value holder who is a '*first time buyer*' and presented with the statement "*House prices will rise by 5%*". The perlocutionary effect of the utterance on the '*first time buyer'* is likely to be a negative sentiment because the end state of the subject matter expressed as '*rise by 5%*' is un-preferred. If the statement is reversed to *"House prices will fall by 5%"*, the sentiment of the value holder is likely to be positive because the end state - *'fall by 5%'* - of the subject matter - *'House prices'* - connotes a state – a drop-in house prices meaning cheaper houses and high possibility of purchasing one - that is preferable and in line with the value holder's values. While the change in the value holder's perlocutionary effect is brought about by unobserved values, the invocation of the values is triggered by expressions in the utterance i.e. the change from *'rise'* to *'fall'* in both sentences triggers different values which consequently results in different perlocutionary acts. In conclusion, VCs can be extracted through an analysis of the corpus because the application of a value is triggered by the intent of the utterance which though abstract is expressed by a series of trigger words or expressions.

Corpus analysis involves identifying common patterns, structures, examining linguistic properties and features for VCs. Several disadvantages exist when using clues embedded in the text. One such downside lies in the inherent structure of the text which is bound to vary widely across a diverse range of speakers, subjects and situations. The implication of is that multiple models would be required in identifying clues for diverse sentence types, thereby making the method inflexible and lacking uniformity. This problem is magnified

[20] By utterances, this research refers to sentences, documents, speeches, statements which are a sequence of words structured and composed together to express and convey meanings and intents.

as the size of the text collection increases. Secondly, because of the large collection of utterances and the diversity in styles, there is no clear criteria or specification that stipulates what a subject or action is. Therefore, the model must address two issues: It must identify VCs from sentences and secondly must provide a methodology for extracting them. In the next sections, different VCs in sentences are identified. As a precursor, two simplifying assumptions are made:

- The granularity of an utterance[21] expressing a value is a sentence. Thus, a corpus will be a collection of sentences called Value Laden Sentences[22] (VLS).
- Each sentence can be expressed as a sequence of words.

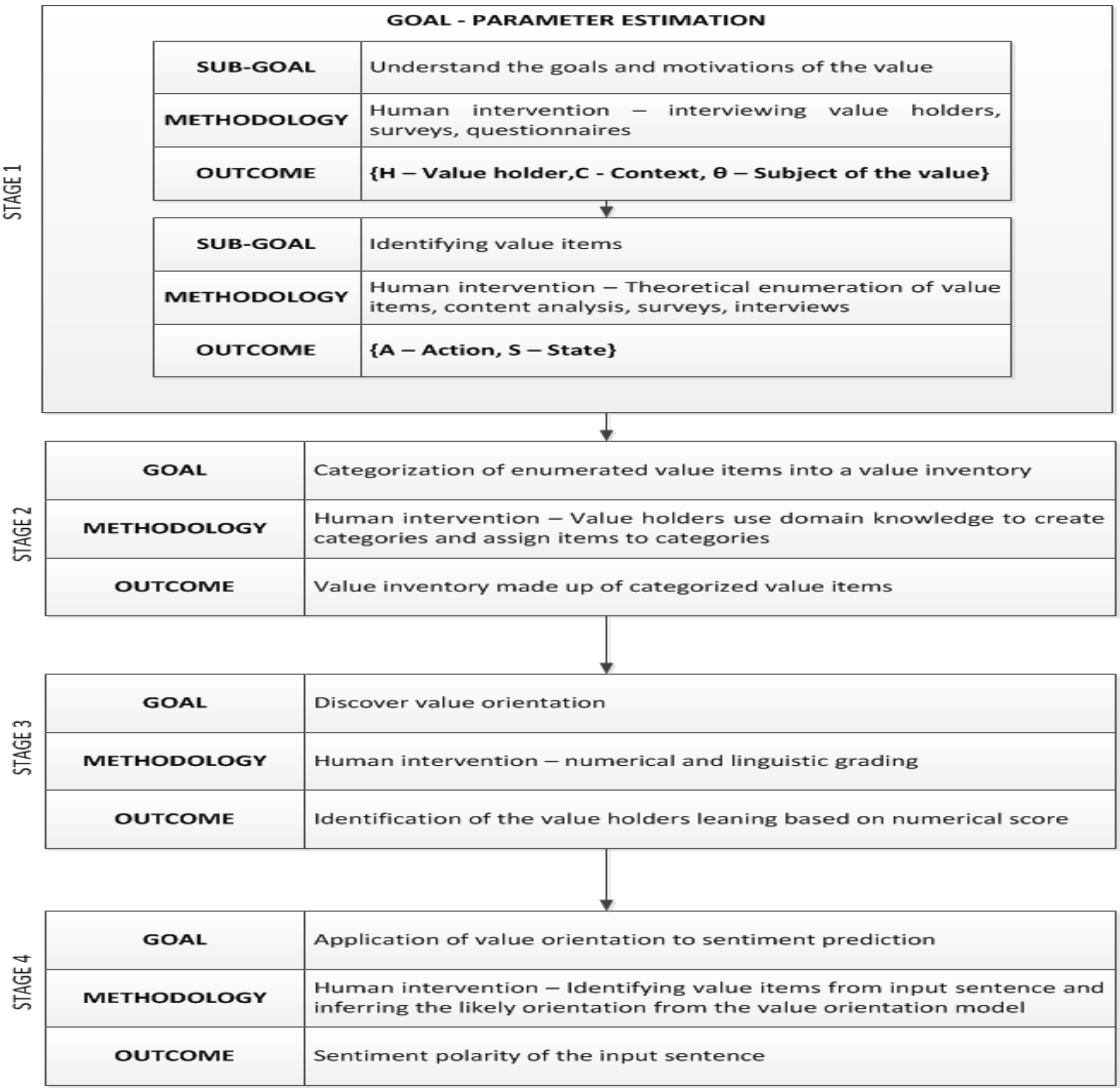


Figure 4: Modified Four-Stage VSM

[21] Utterance and sentence are used interchangeably. Utterances are vocalized sentences.

[22] Value laden sentences (VLS) are statements that impart a personal value that may not be true in the strictest sense but are based on personal opinions or values. They reflect the bias of an author or the speaker while also reflecting the priorities and ideas of the speaker.

From literature, several inferences which are necessary for parameter identification are made:

- Abstract values can be expressed using words. Therefore, to model values, the words and expressions used by value holders in certain contexts can be observed.

- Values are triggered if the effect of the action on the subject or the state of the subject bears some significance to the hearer.

- Subjects, actions and preferred states are semantically relevant trigger words or expressions that encapsulate the intent and meaning of the sentence. In English grammar, such words are called content words.

- As content words, subject, actions and states are open class words because new instances can be added to the lexicon. The model must therefore be flexible enough to accommodate new content word additions.

- There is a relationship between the subject, states and actions since the state refers to a property or feature of the subject and the action conveys an action to be carried out on or by the subject.

From these, a VLS is a sequence of words $\langle w_1 \ldots w_n \rangle$ (where n is an integer and $n > 1$), which express the preferred action w_a or preferred state w_s of a subject w_θ. Subjects, actions and states could be multiple words or phrases which function as a single unit. For example, the sentence *"We will reduce taxes by 2 percent in our first year in Government"* contains the subject *'taxes'* and a multi-word action *'will reduce'*. In the sentence *"EU Taxes will be reduced by 2 percent to take it to an all-time low"* contains a multi-word subject *'EU Taxes'*, an action *'will be reduced'* and a preferred state *'all time low'*.

How then can these value parameters be identified, without explicit human input? This is the subject of the next section. The method described in the next section are based on heuristics and are not exhaustive. However, they are generic enough to be applied to any document or content type.

5.4.1 Identifying Value Subject

As expressed earlier, the subject of a value refers to what the value is all about or what is being valued. In sentences, the subject will typically represent the 'who' or 'what' of the utterance and this could be an object, entity, event, person, item, place or concept. Considering this, a grammatical indicator that a word or expression is a subject is if its POS is a noun. This is because nouns denote concepts such as names, places, concepts or things. For example, in the sentence *"Our Government will say no to ever present union with the EU"* [23], the subject of the expression is the noun *'The EU'*. The value holder, - '*Our Government*' - expresses their preferred state *'no to ever present union'* on the subject.

[23] Culled from UKIP EU manifesto, 2015

Subjects could be expressed as single words, phrases and clauses as in the previous example sentence – proper nouns such as *'Mercedes Benz'*, *'Oslo Peace Accord'*, *'Global warming'* – and generally anything that can be referred to with a proper name. A common class of multi-word subjects are Noun Phrases (NP) which are a sequence of words that occur together and behave as a single unit called constituents. NPs could also consist of a single word or multiple words, with examples ranging from words like *'Services'*, simpler phrases like *'The UK'* and *'The EU'* to more complex series of expressions like *'British involvement in Eurozone bailouts'*.

A simple clue for identifying subjects in sentences is to identify the expression that is acted upon by a verb or the expression described by an adjective. For instance, in the sentence, *"We will curb immigration"*, the action *'will curb'* acts on the word *'immigration'* making *'immigration'* the subject of the sentence. In this example, the simplicity of the sentence, and the fact that the subject is a single word makes this subject identification quite simple.

However, sentences are not always this simple, as sentences could have multiple subjects (this is explored later), or NPs instead of single word expressions. To identify NPs a common clue is to look out for a main noun positioned before (pre-modifier) or after (post-modifier) a modifying expression. These modifying expressions are commonly adjectives, prepositions, determiners, quantifiers and possessive nouns. For example, in the sentence *"We will not deploy any British ships"*. The action *'not deploy'* acts on the noun *'ships'* which is also modified by the adjective *'British'*. The modifier essentially acts as a descriptor for a type of *'ships'* - *'British ships'*, therefore it is a NP and subject of the sentence. Other examples of NPs include expressions such as:

- *'The Prime Minister's office'* where the possessive noun *'minister's'* modifies the noun *'office'*.
- *'The EU'*, *'The Ashes'* where the determiner *'the'* pre-modifies the noun *'EU'* or *'Ashes'*.
- In the snippet, *'We will eliminate all entries …'*, the quantifier *'all'* expressing quantity and size modifies the word *'entries'*.

Similarly, NP post modifiers are words or phrases positioned after the word and normally include prepositional phrases, adjectival clauses and infinitives. For instance, in the sentence, *"Migrants on benefits will be made to pay their fair share"* the prepositional phrase *'on benefits'* modifies the term *'Migrants'*. Finally, in the sentence, *"We will toughen the process for non-EU migrants to enter the UK"*, the infinitive *'to'* post-modifies the subject expression, *'migrants'*.

Other NP patterns include, pre-modifiers (nouns are in bold font) - [determiners + adjective + noun] e.g. *'our American **allies**'*, [quantifiers + adjective + adjective + noun] e.g. *'some young British **soldiers**'*, post-modifier - [noun + prepositional phrase] - e.g. *'**soldiers*** *at the front'*. Despite the patterns identified here, the number of possible NP patterns are innumerable. In summary,

- Subjects are nouns or NPs
- Subjects are identified by locating the expression that is acted upon or described
- NP subjects are identified by locating the main noun and spotting any pre- or post-modifiers.

It is the nature of noun subjects to perform functions and have properties/features. One way of thinking of features/properties is to ask questions such as 'How can it be described?', 'What does it look like?' or 'What could it look like?' The answer to these questions represent the possible states of existence of the subject. To understand the functions of the subject is to consider the actions that can be performed on or by the subject, that is 'what can it do' or 'what can be done with it'. Answering these reveals the possible actions of the subject. These questions form the premise for the discussion on clues for describing actions and states.

5.4.2 Identifying Value Actions

In sentences, actions are words or expressions that denote specific processes and activities to be performed. From a linguistic perspective, actions are commonly expressed as verbs or subcategories of verbs such as verb phrases (VP), phrasal verbs, transitive verbs etc. Like subjects in VLSs verbs can be expressed in numerous ways and in this section a selection of verb types and clues for identifying them in relation to a subject are illustrated.

Like subjects, actions could be single words or multiple word expressions. In the sentence, *"We will deploy British ships"*, the action on the subject *'British ships'* is the verb *'deploy'*. In identifying the action, a reasonable clue is to ask what is being done on or by the subject? or what has happened to the subject? In the sentence above, the subject *'British ships'* are *'deployed'* making *'deployed'* the action. Actions could also be expressed as constituents called verb phrases (VP) which could also be single words.

To identify and extract actions from sentences, verb types that could occur in VLSs and some clues for their identification are considered. A class of verbs common in VLSs are transitive verbs. In English grammar, they are verbs which take a direct object. Direct objects are quite easy to identify in English sentences which take the grammatical form, [subject-predicate-object][24]. For example, in the sentence, *"UKIP will reduce immigration into the UK"*, the grammatical subject is *'UKIP'* the predicate is the transitive verb *'reduce'* and the object is *'immigration'*. Thus, identifying actions in such sentence structure is as simple as identifying the grammatical predicate and asking the question what or whom.

[24] In English grammar, subject-predicate-object, represents the structure of most English sentences. Subject and object in this instance is different from the value subject. To make this differentiation, the terms grammatical subject, grammatical object and grammatical predicate are used to refer to the subject-predicate-object structure.

So, in the previous example if the grammatical predicate is *'reduce'*, then ask the question *'reduce what'*? If there's an answer from the sentence, in this case *'immigration'*, then the grammatical predicate is a transitive verb and thus the action. It is important to note that the grammatical subject, grammatical object and grammatical predicate are different from the value subject/object (value subject and object are synonymous). While value subject refers to the subject of the value or what the value is all about, the term grammatical subject refers to the noun or noun phrase that occurs before a verb and it represents the agent or doer performing the verb. In VLS, the grammatical subject is sometimes the value holder. Grammatical object is usually the object that is acted upon by the grammatical predicate and in value-laden sentences, this is sometimes denoted as the subject/object of the value.

Another common verb type used in value laden sentences to express intent, make declarations and express restrictions are auxiliary verbs and infinitives. A useful clue for identifying these verbs is that they are normally paired with main verbs to construct a meaning. Auxiliary verbs, (also called helping verbs) include modals – can, could, may, might, must, will, would, shall, and should. Example sentence snippets such as *"Our Government will deploy troops …"* uses the modal *'will'* along with a main verb *'deploy'* in expressing an intention. Other auxiliary verbs, *'be'*, *'do'* and *'have'* can also act as main verbs in any of their morphological forms. For instance, in the snippet, *"We have cut net migration from …"*, the auxiliary *'have'* combined with *'cut'* is a verb phrase that expresses an action on the subject *'net migration'*. In describing actions that are presently occurring or actions that have occurred in the past, a common pattern for auxiliaries involves two auxiliaries followed by a main verb. Expressions such as *'have been announcing'* and *'has been announced'* respectively illustrate these use cases. Infinitives on the other hand will usually begin with the word *'to'* followed by a verb. An example is the phrase *'to reduce'* in the sentence, *"UKIP expects **to reduce** the number of EU-Migrants coming into the UK"*, where the action *'to reduce'* is associated with the subject *'EU-Migrants coming into the UK'*.

The final class of actions to be considered in this section are verb types called phrasal verbs. Phrasal verbs are constituents usually consisting of a main verb and a particle whose presence in the constituent expression completely alters the meaning of the expression. In the sentence, *"UKIP will not back down on our promise to take the UK out of Europe"*, *'back down'* is a phrasal verb representing an action synonymous with terms such as *'withdraw'* and *'surrender'*, and completely different from the literal meaning of each individual term. Phrasal verbs could also be separated, for example, in the sentence, *"We will not be leaving our veterans behind"*, *'leaving'* and *'behind'* though separated by two words connote the same concept as *'leave behind'*. Table 2 shows a list of sample actions associated with two subjects EU and Immigration.

5.4.3 Identifying Value States

States generally refer to a property or feature of the subject. In addition, they also include conditions and qualities of a subject. Based on this, in VLSs, they would typically be adjectives, adverbs adjectival or adverbial phrases. This is because adjectives describe properties or qualities while adverbs modify verbs and as such are good indicators of state. This section begins by considering states expressed as adverbs.

Table 2: Sample of possible action expressions for subjects EU and Immigration extracted from Conservative, Labour, Lib Dem and UKIP manifestos

Associated Actions for Subject EU
by not remaining, change, change in, committed to improving, creating opportunities for, creating opportunities within, get us out of, grant access to, have taken action in, have taken action on, including, intend to leave, lead out of, leave, leaving, must leave, negotiate with, negotiating with, not remain in, not stay, opposed to, promoted, promoting, reform, reformed, reforming, supports, to campaign for, to challenge, to improve, to leave, to remain in, to stabilize, to stay in, to stay within, to support, will promote, will reform, will scrap, will work to change, will work with, withdraw from, work with
Associated Actions for Subject Immigration
apply, bringing an end to, cap on, cut, cutting, employ, enforce, ensure, increase, integrated, introducing, monitor, not stigmatise, not to blame, not to do down, not to employ, opposed to, permitted to remain, provide, reduce, reinstate, restrict access to, restricting, secure, stop, take back control of, to cap, to close, to control, to create, to cut, to deport, to employ, to encourage, to monitor, to oversee, to prevent, to reform, to restrict, to tackle, uphold, want to control, will abolish, will cap, will establish, will limit, will train

These could take several forms. Adverbs of manner describe the manner of some action or process and will normally be positioned before a verb (an action). For example, in the sentence, *"We are certainly opposed to Turkey's membership of the EU"*, the adverb of manner *'certainly'* emphasizes the speaker's level of opposition for the subject phrase *'Turkey's membership of the EU'*. A good syntactic clue for the identification of adverbs of manner is that the adverbs normally end with the suffix '*-ly*'. Adjectives of manner – adjectives which end with the suffix '*-ly'* - are an exception to this rule for instance, in the sentence, *"The Government has become too friendly with the EU*", *'friendly'* is an adjective of manner and not an adverb even though it describes a state.

Adverbs of frequency express a state through an expression of the frequency of an event or action. For instance, in the sentence, *"We will never support the EU's migration policy"*. The frequency adverb *'never'* which precedes the verb or action – *'support'* expresses a state for which there will be zero support for the subject *'EU migration policy'*. Frequency adverbs could also precede adjectives as in the sentence, *"We say never to closer union with the EU"*. The phrase *'never to closer'* expresses a state of zero closeness to the subject matter *'Union with the EU'*. Another category of adverbs used in VLSs are mitigators and intensifiers which are degree adverbs that convey intensity. Intensifiers increase or boost the intensity of an adjective or phrase and gives additional emotional context to the adjective or phrase. Intensifiers will normally be positioned behind the phrase or word that they intensify apart from the word *'enough'* which is usually positioned after the adjective. Examples of intensifiers include words such as *'really'*, *'extremely'*, *'incredibly'*, *'exceptionally'*, *'very'*. For example, in the sentence, *"We will provide visas for exceptionally talented individuals"*, the intensifier *'exceptionally'* modifies the subject by expressing an intense degree or level of *'talented individuals'*. The syntactic clue for identifying such a state is by looking out for patterns such as [adverbial intensifier + NP]. Mitigators are the opposite of intensifiers and their goal is to express a lower intensity by making the adjective less strong e.g. *'rather'*, *'a little bit'*, *'a bit'*, *'just a bit'*, *'a little'* etc. For instance, in the snippet *"We will rather increase tuition fees …"*, the state of the subject *'tuition fees'* is the expression *'rather increase'* which takes the pattern [adverb + adjective + NP].

A common syntactic clue for identifying mitigators is that the adjectives or phrases that they modify are normally comparative adjectives. For instance, in the sentence snippet, *"We believe that the situation in the region is a little bit better than …"*, the comparative adjective better is preceded by the mitigator *'a little bit'*. Finally, not all adverbs are good state expressions. For instance, temporal adverbs such as *'yesterday'*, *'Monday'* can be mistaken for nouns and subjects.

As for adjectives, they modify the subject (Nouns or NPs) and are good expressions of measure and preference. For instance, in the sentence, *"We will introduce lighter regulations for small businesses"*, the adjective *'lighter'* represents the preferred state for the subject *'regulations for small businesses'* while in the sentence snippet, *'We will build a stronger economy…'*, the comparative adjective *'stronger'* expresses the preferred state of the subject *'economy'*. Adjectival expressions of state could also be preceded by a linking verb such as *'is'*, *'be'* or *'looks'*. Linking verbs help to link subjects to a descriptive state or an action. In VLSs, these are typically used to describe states, forms or actions. They could take the any of the following patterns:

- [Noun + linking verb + adjective] e.g. *"Our proposed immigration bill is ideal"* the adjective *'ideal'* which comes after the linking verb *'is'* - which is a form of the auxiliary verb *'be'* – describes the state of the subject *'immigration bill'*.

- [Noun + linking verb + NP] e.g. *"The UK has become the defender of Western democracies"*. The noun phrase *'the defender of Western democracies is a state of the subject'*, *'The UK'*.

Another class of adjectives used to describe the qualities of a subject are called qualitative adjectives and examples include words such as small, happy, sad, large etc. In the snippet, *'We are not happy about our open borders…'*, the value holder expresses a state of being *'happy'* about the event subject *'open borders'*. There's also a class of adjectives which resemble verbs because of their syntactic form as they end with the suffices *'-ed'* and *'-ing'*. A case in point – *"UKIP will introduce a modified UK Single Farm Payment (SFP) scheme of*

£80 per acre for lowland farms, with comparable arrangements for lower grades of land, capped at £120,000." The adjective *'modified'* describes the state of the subject *'UK Single Farm Payment (SFP)'*. Other examples of adjectives which end in *'-ing'* and *'-ed'* are *'interesting'*, *'terrifying'*, *'terrified'*. Usually they will be positioned just before the NP such that they take the pattern [adjective + NP].

The subject of a VLS can be associated with several states. Thus, given a very large collection of sentences, containing subjects, each subject can be mapped to a set of possible states to form a collection of relevant states. For instance, table 3 shows a sample of states associated with the subject matter Immigration as drawn from the manifestos of four major UK political parties.

So far, simple sentences involving single subjects, states and actions have been considered. However, because sentences can be creative, complex sentence structures involving multiple subjects, actions and states also exist. These sentence types are considered in the next section.

Table 3: Sample of States associated with the subject matter – 'Immigration' from UK Party Manifestos

States Associated with the Subject Matter Immigration
Affordable, befitting, better, better off out, better placed, controlled, fair, fairer, fairness, first, Hated, is bad for, illegal, legal, lighter, maximum, modern, modified, more fair, more into, sensible, severely, significantly, stealth, sustainable, thriving, tightly, unnecessary, unviable, vibrant

5.5 Complex Value Laden Sentences

A VLS could include more than one subject, state or action in its expression. Consider the sentence, *"We will take action on Europe to make you better off"*. Two subjects are mentioned *'Europe'* and *'you'*. The preferred action for *'Europe'* is the phrase *'take action'* and the state expressed for *'you*[25]*'* is *'to make better off'*. The implication of this is that a VLS could have more than one subject where each subject could be associated with either multiple actions or multiple states. Sometimes, the complexity might involve the added task of pronominal resolution as in the previous sentence which requires resolving the antecedent noun of the subject pronoun *'you'*.

What then are complex sentences? In English grammar, there are three types of sentences simple, complex and compound sentences. Simple sentences are usually short, take the form 'subject-predicate-object', and are expressive of an independent completed thought. While they can be made longer by prepositional phrases, their characteristic independence means that they can be conjoined with other expressions to form part of a longer compound sentence. Most of the examples considered so far have been simple sentences

[25] 'you' refers to 'The British people'

e.g. *"UKIP will close the borders"*, the grammatical subject is *'UKIP'*, the predicate is *'will close'* and the grammatical object is *'the borders'*. The latter two, complex and compound form part of complex VLSs and they are considered in the next section.

5.5.1 Compound Sentences

Compound sentences consist of two independent clauses or simple sentences joined together by a coordinating conjunction or a connective e.g. *"We will reduce tuition fees so that British students can enjoy the University experience"*. The coordinating conjunction, *'so'* links the two independent sentences *"We will reduce tuition fees"* and *"British students can enjoy the University experience"*. Both independent sentences contain subjects which have associated states i.e. subject *'tuition fees'* is associated to action *'will reduce'*, while subjects *'British students'* and *'University experience'* are associated with the state *'enjoy'*. To Identify the VCs in this kind of sentence, the coordinating conjunction is isolated so that the independent sentences are separated. Subsequently, each sentence can be treated by uniquely identifying its subject, action or state by using some of the earlier mentioned clues for simple sentences.

The presence of the coordinating conjunction or connectives in the sentence relates both independent sentences and transitions the reader from the concept or idea expressed in one clause to the concept or idea expressed in the other clause. For example, the use of the coordinating conjunction *'and'* in a sentence suggests meanings of addition, expressing that both sentences or clauses are similar, equal or without contrast. The words *'but'* and *'yet'* express contrasting relationships between sentences, while *'or'* and *'nor'* express positive and negative alternatives or options. Finally, *'so'* expresses consequence, for instance in the example sentence, *"We will reduce tuition fees so that British students can enjoy the University experience"*, the word *'so'* suggests that the consequence of *'reducing tuition fees'* is '*British students can enjoy the University experience'*. Connectives include prepositional phrases and connective adverbs which perform the same functions as coordinators. Examples of functions performed by connectives include expressing additions e.g. *'furthermore'*, expressing conclusions e.g. *'as a result'*, *'for this reason'*, *'therefore'*.

Based on the relationships between sentences linked by coordinators and connectives, a reasonable inference is made that the VCs of individual sentences linked by a coordinator or connective share the same relationship that exist between the sentences. For example, in the example sentence *'We will reduce tuition fees so that British students can enjoy the University experience'*, it can be inferred that the subject *'tuition fees'* is related to *'British students'* and the action *'reduce tuition fees'* is related to the state *'British students enjoy University experience'*. These relationships are not synonymous, that is, in the sentence, one cannot theoretically replace the other even though it can be inferred that they are semantically related. This inference is made based on the principle of expressibility (Searle, 1970), which says that *"for any meaning X and any speaker S whenever S means (intends to convey, wishes to communicate in an utterance, etc.) X then it is possible that there is an* *expression E such that E is an exact expression or formulation of X."* So, in formulating a meaning, a speaker would use exact words and expressions which relate to that meaning even though they might have different forms. Since a speaker is unlikely to use unrelated expressions in the sentence, it is fair to conclude that the expressions which make up the sentence are semantically related. Therefore, in VLS the elements of the VCs are part of a semantic field of related concepts. Related concepts refer to a set of words or phrases in the VLS which depend on one another to make sense and when outlined together within the context of some rational background knowledge form a semantic field of related concepts. The example in the simple sentence below buttresses this point *"We will make a powerful statement to reduce and tighten immigration controls"*, the actions *'reduce'* and *'tighten'* are connected by the additive coordinator *'and'*, therefore it can be inferred that both actions are related. In addition, since the direct action on the subject *'immigration controls'* is *'tighten'* and because *'tighten'* has an additive relationship with *'reduce'* it can be inferred that *'reduce'* is also an action on the subject *'immigration controls'*. Both actions constitute a semantic field of related actions associated with the subject *'immigration controls'*. Similarly, it can also be said that the subject *'immigration controls'* has two preferred actions *'reduce'* and *'tighten'*. Consider another example. In the sentence *'UKIP will repeal EU regulations and directives that stifle business growth'*, there are two subjects here *'EU regulations'* and *'directives'*. The action *'repeal'* is directly connected to the first subject *'EU regulations'*, and since both subjects are connected by the coordinator *'and'*, it can be inferred that the action *'repeal'* impacts both subjects *'EU regulations'* and *'directives'*.

5.5.2 Complex Sentences

In English grammar, complex sentences are made up of an independent clause connected to one or more dependent clauses. Unlike an independent clause, a dependent clause is not a complete sentence, for example, *'…when we are elected into Government.'* The coordinating conjunctions used to link the dependent and independent clauses also convey relationships between the sentences. As in compound sentences, isolating the coordinating conjunction in the sentence can aid in identifying each sentence and identifying the embedded VCs. Some of the coordinating conjunctions used in this class of sentences include *'after'*, *'although'*, *'as'*, *'because'*, *'before'*, *'even though'*, *'if'*, *'since'*, *'though'*, *'unless'*, *'until'*, *'when'*, *'whenever'*, *'whereas'*, *'wherever'* and *'while'*.

Consider two sample sentences: *"UKIP will continue to negotiate and work with Commonwealth nations while leaving the EU"*. In this instance, the independent clause, *'UKIP will continue to negotiate and work with Commonwealth nations'* precedes the independent clause *'leaving the EU'*. Observe also that both sentences are separated by the word *'while'* and each half of the sentence could consist of several value components. For instance, the first half includes components such as '*UKIP*', '*negotiate*', '*Commonwealth Nations*' while the second half consists of '*leaving*' and '*The EU*'. In the concept of 'the EU', these linguistic components are thematically related.

"After leaving the EU, UKIP will negotiate new trade agreements with Commonwealth Nations". In this example, '*after*' is a conjunction connecting two clauses: the dependent clause *'leaving the EU'* comes before the independent clause *'UKIP will negotiate new trade agreements with Commonwealth Nations'*. It can be inferred that both clauses contain related value components since they are linked by the conjunction. The first half consists of VCs '*leaving*' and '*The EU*' while the second half consists of VCs like '*negotiate*', '*trade agreements*', '*Commonwealth Nations*'.

5.5.3 Other Sentence Types

This section considers other complex sentences with multiple subjects linked by contrary coordinators or connectives. Consider the following (Note that concepts representing subjects are in bold font).

1. *We will reduce* ***UK entry level FOR Secondary schools****, when we are elected into Government.*
2. *We will not support* ***British involvement IN Eurozone bailouts****.*
3. ***Visa requirements*** *have been tightened by our Government for* ***migrants coming to the UK*** *from* ***outside Europe****.*

In the first sentence, the main subject is *'UK entry level for Secondary Schools'* and the action applied is expressed as '*reduce'*. The main subject is made up of two subjects *'UK entry level'* and *'Secondary Schools'*. *'UK entry level'* is the primary subject because it is qualified by the prepositional phrase *'for Secondary Schools'* which acts as an adjective and thus provides additional meaning and context to the phrase. *'Secondary Schools'* is called the secondary subject. Since both subjects combine to form another subject, they are called related subjects.

The second sentence is similar to the first. The action *'not support'* applies to the subject *'British involvement in Eurozone bailouts'*. The main subject consists of two subjects – A primary subject *'British involvement'* modified by the prepositional phrase *'in Eurozone bailouts'* which acts as an adverb.

The third sentence is a complex sentence consisting of one independent clause - *'Visa requirements have been tightened by our Government'* - and a dependent clause – *'for migrants coming to the UK from outside Europe'*. The preferred state on the subject *'Visa requirements'* is *'tightened'*. Since the dependent clause is dependent on the independent, it can be inferred that there's a relationship between the subjects of the dependent clause and the independent clause. In the example above, there is a relationship between *'visa requirements'* and the clause *'migrants coming into the UK from outside Europe'*. The latter clause also consists of three nested subjects, *'migrants'*, *'UK'* and *'Europe'*. The concept *'migrant'* is modified by the prepositional phrases *'coming to the UK'* which acts as an adjective, while the concept *'UK'* is modified by the prepositional phrase *'from outside Europe'* which acts as an adverb. Thus, the subjects *'Visa requirements'*, *'migrants'*, *'UK'* and

'Europe' are concepts bound together by a theme and thus belong to the same field of related concepts.

Several conclusions are drawn from this section about the nature of actions, states and subjects. They include the following:

- Subjects, Actions and States in VLSs can be expressed as expressions or words where actions are typically verbs, states are typically adjectives and adverbs and subjects are normally nouns or NPs.

- VLSs, actions, states and subjects form a related field of concepts, so that for any subject or collection of subjects there exists a set of conceptually relevant and related actions and states.

- Subjects can be nested i.e., consist of multiple subjects.

- Structural cues in value laden sentences can be used in identifying VCs, but these cues are not exhaustive as the structure of written and spoken Language is flexible, open ended, creative and constantly evolving.

Based on these points, in the next section, a formal structure of VLSs is introduced.

5.6 Structural Representation of Value Laden Sentences

In the previous section, several clues for identifying VCs from sentences were identified. However, it was noted that the pointers did not satisfy all possible sentence types. To address this issue, a generic formal structure for value laden sentences is discussed in this section.

This section commences with a definition of a VLS as a sequence of words made by a value holder (H) under a particular context (C). Since the objective of the VLS is to express the preferred state or action on a given subject matter/s, the sequence of words which constitute the VLS must include at least one value subject (Θ) and express at least one action (A) and/or preferred state (S). For example, the sentence *"We will reduce taxes by 2 percent in our first year in Government"* contains one main subject *'taxes'*, an action *'reduce'* and a preferred state *'by 2 percent'*. These core words are priority words called content words.

Assuming the value holder, - the speaker - is a known entity, the remaining words in the sentence (including the value holder expressed by the pronouns *'we'* and *'our'*) belong to a second category of words called function or helper words whose aim is to connect the actions, states and subjects in a manner that conveys the intended meaning. Function words are a class of words in English grammar that bear no semantic relevance (Fries, 1952). They are classified as closed vocabulary words because it is quite uncommon to create new ones and are generally restricted to 9 grammatical classes as shown in table 4. It must be said that in grammar the distinction between function words and content words is not always so clear. This is because some function words can also act as content value laden words depending on its role in an expression. For instance, in the sentence *"No to the EU"*, *'No'* is a content word while in *'We have no more money to spend'*, *'no',* is a function word expressing a negative particle. Making this differentiation is not a trivial task and for simplicity, it is assumed that all function words perform the role of function words except for Negations, exclamations and expletives because they are semantically relevant influencing value actions, subjects and states. For this reason, they are treated as content words.

Table 4: Function word classes and examples

Word Class	Example
Auxiliary verbs	Am, are, be, is
Conjunction	Or, and, but, while
Determiner	A, the
Exclamation	Yes, No
Interjection/Disfluencies	Uh, em, huh, duh
Modals	Could, would
Particles	No, not, then, if, thus
Preposition	Of, in, at, between
Negation	Not, never, no

Assuming actions, subject and preferred states – the lexical components of the value – are represented as λ_A, λ_Θ, λ_S and function words as F, a value sentence could take any of the following sample formats -

Sentence 1 –$\langle START, F, F, \lambda_A, F, \lambda_\Theta, F, STOP \rangle$

Sentence 2 –$\langle START, F, \lambda_\Theta, F, \lambda_A, F, \lambda_A, F, \lambda_\Theta, F, STOP \rangle$

Sentence 3 –$\langle START, F, \lambda_\Theta, F, \lambda_S, F, F, F, STOP \rangle$

The '$START$' and '$STOP$' signal are used to mark the beginning and end of the sequence.

Consider the sentence, "*We will implement more apprenticeship programmes for young people in the winter".* Assuming the value holder is known, the sequence of words expressed above can be represented as the sequence

$\langle START, F_1, \lambda_A, \lambda_S, \lambda_\Theta, STOP \rangle$, where,

'We' = F_1

'will implement'= λ_A

'more' = λ_S

'apprenticeship programmes for young people in the winter' = λ_Θ

The subject *'apprenticeship programmes for young people in the winter'* is a long phrase consisting of nested related subjects. So, to make processing and analysis easier, it is split further to show each of the related subjects and the function words linking them. Thus, the concept *'apprenticeship programmes for young people in the winter'* can be expressed as the sequence

$\langle \lambda_{\theta 1}, F_2, \lambda_{\theta 2}, F_3, F_4, \lambda_{\theta 3} \rangle$, where,

'apprenticeship programmes' = $\lambda_{\theta 1}$

'for' = F_2

'young people' = $\lambda_{\theta 2}$

'in' = F_3

'the' = F_4

'winter' = $\lambda_{\theta 3}$

The full sentence *"We will implement more apprenticeship programmes for young people in the winter"* can thus be expressed as the sequence

$\langle START, F_1, \lambda_A, \lambda_S, \lambda_{\theta 1}, F_2, \lambda_{\theta 2}, F_3, F_4, \lambda_{\theta 3}, STOP \rangle$

Following these examples, a formal structure of VLSs emerges. It is proposed that a VLS consists of a sequence of strings composed of one or more subject expressions drawn from a countably infinite vocabulary of subjects, one or more semantically relevant actions or states which are also drawn from a countably infinite vocabulary of actions and states and at least one function word drawn from a finite set of function words. This thesis also proposes that the vocabulary of content words/subjects is countably infinite because they could be about literally anything. Finally, all words and expressions drawn belong to a universal set of expressions.

Thus, the generation of a VLS is defined by the following parameters:

$\sum$ = A countably infinite set of Subjects, Actions, State, Function words and Value holder entity names and expressions (Vocabulary of vocabularies)

λ_θ = A countably infinite set of subject

λ_S = A countably infinite set of States

λ_A = A countably infinite set of Actions

F =/ A finite set of function expressions and words

START = Marker representing the start of a sentence

STOP = Marker representing the end of a sentence

Where, $(\lambda_\theta, \lambda_S, \lambda_A, F) \in \sum$ and $\{START, STOP\} \in \sum$

Following the derivation of the formal structure, how can sentences be generated? To elucidate the generation of VLSs using the above parameters, consider the example of a machine programmed with pro open-source software values. Imagine that this machine needs to generate a VLS that responds adequately to a statement or question. For instance, a statement that challenges the 'benefit of open source software as compared to proprietary software' e.g. *"Should companies developing high precision software use proprietary software or open source software?"*. The goal of the machine is to satisfy three conditions –

- It must generate a response that is grammatically correct i.e. syntax
- It must generate a sentence with actions and states that are semantically relevant and finally i.e. semantically appropriate,
- It must generate a sentence that fits its programmed value i.e. pragmatically appropriate.

Assume the machine has access to the sets λ_Θ, λ_S, λ_A, F, $\{START, STOP\}$.

To generate the sequence of words, the machine could commence with the base subject expressions *'proprietary software'* and *'open source software'* drawn from the vocabulary λ_Θ. Since the resulting sentence is a sequence of words and expressions, assume that the machine prefers the expression '*open source software'* to precede *'proprietary software'* in the resulting sequence, therefore, the goal of the machine is to generate a set of words to fill the empty slots (slots marked '?') as seen in figure 5.

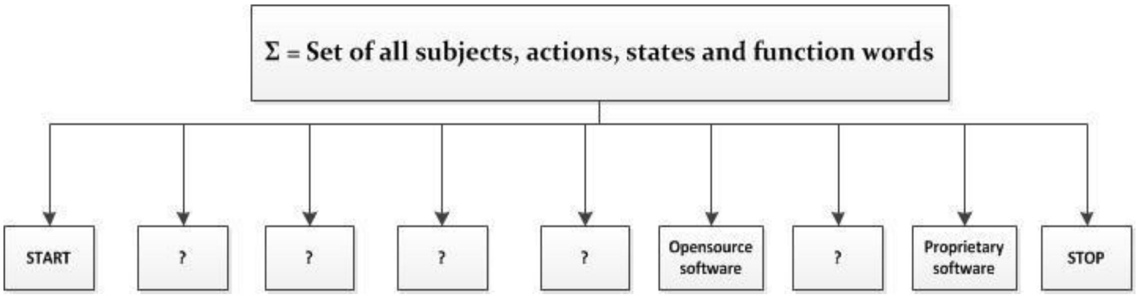


Figure 5: Illustration of Sentence generation by a machine and showing empty word slots

Already it is known that the slots containing actions and states must be filled with expressions that are thematically related to the subjects. So, the machine generates each word in the sequence by picking semantically relevant words from each vocabulary in $\sum$ as seen in figure 6.

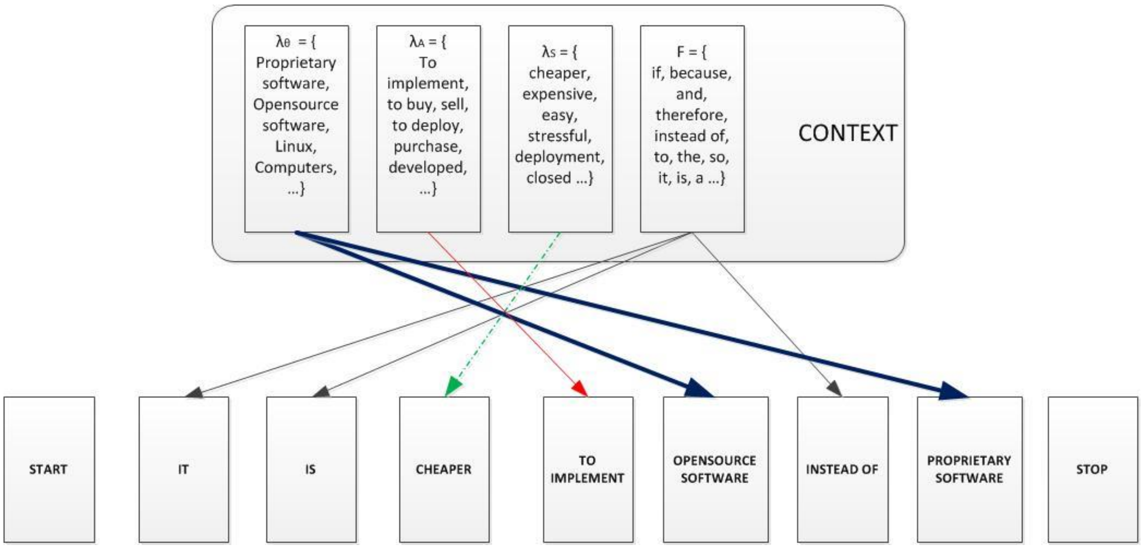


Figure 6: Illustration of Sentence generation by a machine with filled word slots

Following the *START* sign, the machine selects the words *'it'* and *'is'* from the vocabulary of function words. To convey the notion that open-source software is preferable to proprietary, it associates the expressions *'cheaper'* and *'to implement'*, which are drawn from the state and action vocabulary. The functional expression *'instead of'* plays the role of a comparator, bridging the two subject expressions. Once the machine completes the sentence generation sequence it returns the *'STOP'* sign to mark the end of the sentence. Thus, the final sentence sequence becomes,

⟨*START, It, is, cheaper, to implement, open-source software, instead of, proprietary software, STOP*⟩.

This sentence is grammatically correct and satisfies the values of the machine because the actions and states associated with the subjects are semantically and contextually relevant. However, it is possible for the sentence to be grammatically correct but have actions that fail to satisfy semantic correctness. For instance, an element in the set of actions could be the expressions *'to marry'*, which when substituted for the action *'to implement'* forms a grammatically correct sentence that makes no sense semantically – *"It is cheaper to marry open-source software instead of proprietary software"*. This reemphasizes the importance of the semantic relationship between subjects, actions and states, and that each subject in the vocabulary can be mapped only to a subset of semantically relevant states and actions.

In addition, the context and value holder further constrains the choice of action and state. Assuming the machine was programmed to have proprietary values, then the choice of actions and state would have been different. Another way of putting this is that by programming the machine to have proprietary values the likelihood of it generating the sentence *"It is cheaper to implement open-source software instead of proprietary software"* is likely to be considerably lower than the chances of generating the sentence *"It is cheaper in the long run to implement proprietary software instead of open source software because of our dedicated customer service"*.

Therefore, in generating each word for a sentence, the machine determines the likelihood of each word/event by considering several factors including the nature of the subjects, the grammatical structure of the sequence of events etc. Mathematically, each word event generated has certain probabilities associated with it, and the probability of each word or expression is consistent with the value in question, grammatical correctness and the context. To explain this point further, assume the machine tasked with completing the sentence:

"Open source software X" -- (replacing ***X*** with a valid expression)

For a machine with open-source values, the probability of '*X*' being *'is cheaper'* is likely to be higher than it being *'expensive'*. In a different context, say, one in which a machine with open-source values is required to explain the risks of open-source software to a high-valued client, the probability of '*X*' being *'is not always perfect' ('open-source software is not always perfect'*) is higher than the expression *'is rubbish' ('open-source software is rubbish'*). Following this illustration, the model development is about developing a function capable of estimating these probabilities so that for any subject under any context and for a particular value holder, a sentence can be generated.

In conclusion, the process of generating sentences from values is a generative process, where words as events are generated from a vocabulary, and the probability of each event occurring in the sequence is a function of the grammatical relationship between the events, the context, the level of semantic relevance between the events and the event generator that is the value holder. The next section shows how this is modelled mathematically as language models before relating it to sentiment analysis through the unique concept of value fields.

5.7 Expression of Value Models

Having shown that the generation of VLSs is a generative process, the goal is to learn a value model from all the VLSs made by a value holder (H). This model must be able to estimate the probability of a sequence of words in a VLS. To accomplish this task, it is assumed that there exists a large corpus of VLSs made by H, and that each sentence in the corpus represents a distribution of possible utterances that can be made by the value holder. Mathematically, this distribution (f) can be computed by estimating the joint probability of sequences in the corpus so that when presented with a totally new sentence, the estimated distribution can be used to compute the likelihood of the value holder making the sentence. Furthermore, assuming sentences in the corpus can be grouped by context, then separate distributions can be estimated for groups of contextually related sentences such that there could be a distribution (f_1) for context 1, a distribution (f_2) for context 2 and so on and so forth. Consequently, by having multiple distributions, for different contexts, the additional benefit of being able to estimate the likelihood of a sentence made by a speaker across different contexts is accrued. This benefit is discussed further in the section on value fields.

The estimation of these distributions can be broken down into two related tasks.

1. Using the example of the generative machine mentioned earlier (figures 5 and 6), how can the machine correctly generate each word in the sequence so that it forms a correct sentence.

2. How can sentence probability be estimated?

To answer these questions, each sentence in the corpus represents a sequence of random variables, $\langle W_1,\ W_2, W_3, \dots, W_n \rangle$ where each random variable can take any value in a vocabulary of possible words. For instance, in the sentence, *"We will reduce taxes next year"*, the sentence represents a sequence of six random variables, where the first random variable (W_1) takes the value 'We', the second, W_2 is '$will$' etc. The probability of any such sequence of random variables can be expressed as:

$$P(W_1 = w_1, W_2 = w_2, W_3 = w_3, \dots, W_n = w_n) \dots (2)$$

where $n \geq 1$ and w_i is an element of the vocabulary for $i = 1 \dots n$.

Based on the chain rule of probabilities, (1) above can be expressed as,

$$P(W_1 = w_1) \prod_{i=2}^{n} P(W_i = w_i | W_1 = w_1, \dots, W_{i-1} = w_{i-1}) \dots (3)$$

This answers the second question.

As for the first question, it can be reformulated as a word prediction task. Assume that the machine in figures 5 and 6 was asked to generate a word to complete the sentence

'*We will reduce* ___'.

Assuming the machine had the option of three words, '*taxes*', '*fishes*' and '*riches*'. The machine could reference the corpus of sentences to find how frequently the sentences "*We will reduce taxes*", "*We will reduce fishes*" and "*We will reduce riches*" occur. In other words, the machine makes an estimation of the likelihood of a word by looking at some reference history. This analogy can be represented mathematically as the conditional probability of a word given its history (h), that is $p(w|h)$.

The estimations discussed above, - A joint and conditional probability estimated from models of word sequences - falls under a category of statistical models called Language Models (LMs). LMs belong to a class of models called generative models that have been applied to a wide variety of applications such as hand writing recognition (Russell and Norvig, 2002), spelling correction (Kukich, 1992), text prediction and machine translation. Traditionally, they are applied towards estimating the conditional probability of a word given its history $p(w_i | w_i \dots w_{i-1})$, such that given a sequence of words it estimates a distribution over the word that appears in the i^{th} position. In the following sections, LMs are elucidated further and a demonstration of how they are applied towards value modeling and sentiment prediction.

5.8 Basic Formalizations and Assumptions

This section begins with several basic introductory formalizations.

A set Σ is defined as a vocabulary of words and expressions. It includes the sets (θ) of all subject expressions, set (A) of actions, set (S) of all states, set (F) of function words and the binary set $\{START, STOP\}$.

$$\{\theta, \mathrm{A}, \mathrm{S}, \mathrm{F}\} \text{ and } \{\mathrm{START}, \mathrm{STOP}\} \in \Sigma$$

An utterance W is a sequence of random variables $W_1, W_2, W_3, \ldots, W_n$ that can take any value in the set Σ. Since a VLS must contain at least one subject, at least one action or state and at least a function word, then in the sequence, $n \geq 3$.

An example of a sentence sequence could be,

$$START, W_1 = w_{f1}, W_2 = w_{f2}, W_3 = w_{\theta 1}, W_4 = w_A, W_5 = w_{f3}, STOP$$

In the example above, the first random variable in the sequence W_1, takes the form of a function word w_{f1}, the second W_2 takes the form of a function word, w_{f2}. W_3 is a subject while W_4 and W_5 are actions and function expressions respectively. Since it is possible for the random variables in a sequence to take any possible value in the expression, it can be inferred that the number of possible sentences will be significantly large. Thus, a set Σ' is introduced to represent all possible sentences that can be constructed from the set Σ, made by a value holder H under a context C.

Thus, the language model is expressed as a function $f(W_1, W_2, \ldots W_n)$,such that for a value holder H, a context C and any sequence of words in $\sum'$, that $P(W_1, W_2, \ldots W_n) \geq 0$. For any,

$$\langle w_1, w_2, \ldots w_n \rangle \in \Sigma', p(w_1, w_2, \ldots w_n) \geq \cdots (4)$$

The function f, can be simplified by rewriting the probability of a random variable W_1 which takes a value w_1 i.e. $P(W_1 = w_1)$ as $p(w_1)$.

5.9 LM Estimation

To estimate the function f, equation 3, which estimates the joint probability of a sentence by multiplying conditional probabilities is revisited. This formulation is still quite complicated and difficult to resolve, so a simplifying assumption is made based on the Markov assumption which allows us estimate the probability of each event by conditioning only on words in its immediate past. Based on this, instead of conditioning on all the previous words, the probability estimate can be conditioned on the preceding word (called a bigram language model - also called 2-gram) or on the last two preceding words (trigram language model[26] – also called 3-gram). The result of this simplification means that equation 3 becomes:

Bigram language model equation,

$$p(w_1 \dots w_n) \approx \prod_{i=1}^{n} p(w_i|w_{i-1}) \dots (5)$$

Trigram language model equation,

$$p(w_1 \dots w_n) \approx \prod_{i=1}^{n} p(w_i|w_{i-2}, w_{i-1}) \dots (6)$$

So, for instance, given a sample VLS sequence ⟨*START, We, will, reduce, taxes, STOP*⟩, the estimation of the joint probability of the sentence for the bigram case would be

$$\begin{aligned} p(START, We, will, reduce, taxes, STOP) \\ &\approx p(We|START) \times p(will|We) \times p(reduce|will) \times p(taxes|reduce) \\ &\times p(STOP|taxes) \end{aligned}$$

In the trigram case, because the probability of a word is dependent on the previous two words, an assumption is made that the *START* symbol is preceded by a '*' symbol, so that the sequence is represented as[27]:

$$\langle *, START, We, will, reduce, taxes, STOP \rangle,$$

hence, the trigram estimate would be:

$$\begin{aligned} p(START, We, will, reduce, taxes, STOP) \\ &\approx p(We|START,*) \times p(will|We,\ START) \times p(reduce|will,\ We) \\ &\times p(taxes|reduce,\ will) \times p(STOP|taxes, reduce) \end{aligned}$$

These examples show how the probability of each event is estimated by conditioning on a history. With these formulations, a distribution for all events in the corpus can be estimated either for the bigram or trigram case. Although the trigram language model has been shown to produce good model estimates[28], the above model does not fully capture all the properties of the VLS. For instance, the above model only captures local dependencies and does not incorporate the semantic aspects and contexts of each word. In addition, the relationship between the aspects (subject, action, state) of VLSs are not captured at all as each word is only dependent on a word or words seen only in a short window span or prior context (one or two words for the bigram and trigram case respectively).

Consider the longer sentence sequence *"UKIP will implement more apprenticeship programmes for young people in the winter"*, the action *'will implement'* is related to the value holder *'UKIP'* and the subjects *'apprenticeship programmes'*, *'young people'*. Both

[26] See Appendix 3 for formal definitions of trigram LMs

[27] It is common practice to pad the beginning of the sentence with a distinguishing token i.e. *, so that the probability estimation makes sense for $i = 1$. The end of the sentence is also padded with the token 'STOP' to mark the end of the sentence.

[28] Jurafsky and Martin (2009), showed that a trigram model produced the best result when they trained a unigram, bigram and trigram on a Wall Street Journal corpus made up of a vocabulary of 19979 words and 38 million tokens.

subjects *'apprenticeship programmes'*, *'young people'* are related subjects and share a semantic relationship demonstrated through the state word *'more'*. These relationships are not captured by the trigram model. Other types of language models like skip N-gram[29] - where the context skips over some words so that the probability estimate becomes $p(w_i|w_{i-1}, w_{i-3})$ and variable length N-gram which support conditioning on additional contextual information might aid in addressing the problem of long distance dependencies and local context but fail to capture the semantic relationships between the value components of the sentence (Ney et al, 1994; Kneser, 1996). To modify the LM for VLSs, the LM must capture the syntactic and semantic relationships between actions, states and subjects. It must also be tailored to the context and value holder. For now, it is assumed that the value holder (H) and context (C) are known entities. H and C can be incorporated into the model by also conditioning the probability of each word in addition to the history (h) on (H) and (C). The estimation equation becomes,

$$p(w_1 \dots w_n) \approx \prod_{i=1}^{n} p(w_i|h, H, C) \dots (7),$$

where h is a history in the trigram (w_{i-1}, w_{i-2}) or bigram case (w_{i-1}) for a VLS made by a value holder H under a context C.

Having captured the value holder and context in equation (7), capturing the relationships between the value components which make up the sentence follows. These relationships are syntactic (grammatically correct), semantic (meaningful) and pragmatic (acceptable – The inclusion of C and H in equation 7 captures this). In the discussion on the structure and theory of values, four classes of grammar were identified, Subject expressions (θ), Action expressions (A), State expressions (S) and function words (F). Using the earlier example in Figure 5 and 6, for the machine to generate an action expression or state for the subject, it must generate expressions that are semantically relevant to the subject. Since these words represent the main substance of the VLS, they have priority status as the most important events in the sequence. Based on their importance and the semantic relatedness between A, S and θ, their estimations would require additional conditional parameters. Therefore, based on semantic relatedness, the assumption is made that for the VLS to make sense, each Action, State or subject occurring in the sentence is additionally dependent on any other A, S and θ, that is in direct semantic relationship with it.

So, for an expression of a subject, action or state in a VLS, its probability is not only dependent on the history, holder and context, but also on any thematically relevant A, S, θ. The implementation, demonstrates how the semantic relationships are modelled to estimate the probabilities. For now, a simplifying assumption is made that each priority expression, in addition to h, H, C is dependent on all other priority expressions in the sentence. As for the function words, since their primary function is to connect priority expressions, they are estimated from their history alone. Thus, the value language model takes the form,

[29] Language models are also called N-grams

$$p(w_1 \dots w_n) \approx \prod_{i=1}^{n} \begin{cases} p(w_i|h, H, C) & if\ w_i\ is\ a\ function\ word \\ p(w_i|h, H, C, A, S, \theta) & if\ w_i\ is\ any\ of\ A, S, \theta \end{cases} \dots (8)$$

Equation (8) captures a richer informational and semantic context and the net effect is a high order LM which though captures long distance dependencies, increases the complexity of the model's parameter estimation. The next chapter, demonstrates how these probabilities are estimated and implemented.

In conclusion, the value model is a LM which models a distribution of VLSs that can be made by a value holder under a context. Unlike regular LMs, the probability of each word (w) in the VLS is $p(w_i|h, H, C, A, S, \theta)$ if $w \in (A, S, \theta)$ or $p(w_i|h, H, C)$ if $w \in F$. The next section introduces the concept of value fields which plays a role in the application of values for sentiment prediction.

5.10 Value Fields

The influence of values on sentiment is dynamic. In the previous section, it was shown that the generation of a sentence representative of a person's values is dependent on several factors such as the context, the subject matter etc. This same variability applies also to the sentiment or behaviour of a person when presented with an utterance. In effect, the value holder's behaviour or sentiment towards an utterance is predicated on the value invoked which in turn is premised on trigger words in the sentence i.e. features such as the context of the sentence, the proposed action and other VCs. As the subject of a sentence could have multiple actions and states, and perceived from diverse contexts by a value holder, an utterance can be seen to be impacted upon by a field of multiple conflicting values which tend to push sentiment in either positive negative direction. These field of values which influence the sentiment of a value holder is called a Value Field (VF). VF as a concept shares some similarity with Stamper's Information Field. Information fields are a set of shared social norms that governs the behaviour of a group member in an organised fashion (Stamper et al, 2004). People belong to different social groups, which have a set of shared norms. Given a situation, these norms act as a force field which overlap and interact.

To elucidate further on this concept of Value Fields, consider a value holder is presented with an utterance (W). Depending on a value, his/her sentiment could tend towards the positive or negative direction. This phenomenon is presented in figure 7: when W is placed in the value holder's value fields (VF) shown as different coloured circles and labelled VF1, VF2, VF3, VF4. Each field could move the polarity of the utterance to the left (negative polarity), to the right (positive polarity) or not at all (neutral polarity). The fields could also determine how far the orientation of the polarity moves in either direction and this represents the intensity of the sentiment.

VFs consists of distinct value functions that is,

$VF_i = \{vf_1,\ vf_2,\ vf_3, \dots, vf_k\}$, $i =\ 1\ \dots\ k$, and k represents the number of fields.

Each field is made up of distinct contexts, subjects, actions and states. For a field to influence a subject/s in an utterance, the subject of the field must be the same as the subjects expressed in the utterance or be closely related to the subject. 'Closely related' refers to another subject that has similar characteristics and behaviours but might not necessarily have the same word form. The field must also have semantically related state and/or action. So, for an utterance W, decomposed into $(H, \Theta_W, S_W, A_W, C)$, the VFs, that can act on it must have related subject, action, state.

Given that a subject could have more than one action or state and that the set of subjects is countably infinite, the set of a value holder's value fields will be quite large. In addition, the dynamic nature of contexts means that, changes in recipient sentiment over different contexts can be observed. Action and state can also vary as long as they are semantically and pragmatically related to the subject. This concept plays an important role in this research as it offers a way of determining likely sentiment and sentiment intensity of a person given an utterance under diverse contexts. This is a major advantage of the value field concept as it supports the modeling of different contexts. Having established the importance of value fields and described the value model as a LM, the next section brings both concepts together to address the problem of recipient sentiment prediction.

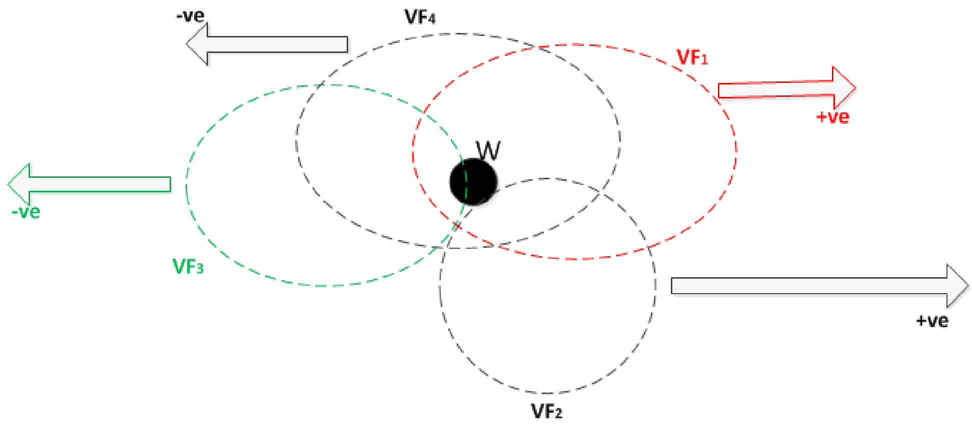


Figure 7: Illustrating the effect of Value Fields on the Sentiment of an Utterance W where coloured circles are different value fields acting on W

5.11 Applying Value Model to Sentiment Prediction

To apply the value model towards predicting the sentiment of a recipient, it is assumed that the utterance/sentence in question, satisfies the formal structure of VLSs. From value theory, for the utterance to evoke either negative or positive sentiment, the state or action on the sentence's subject must be in line with the recipient's preferred state or action or vice versa. Assuming the states or actions in the sentence are identified, the sentiment of the recipient can be manually determined by comparing the states and actions in the utterance against previous statements or utterance made by the recipient i.e. compare the expressed state or action against a history of the recipient's utterance (This is synonymous to a language model). The key question then is 'what is the likelihood of the recipient generating an utterance with the states or actions expressed in the utterance?'. Therefore, assuming all utterances are based on people's values, the behaviour of the recipient to a new utterance (w_1) can be predicted as a measure of how likely it is for the recipient to make the utterance $p(w_1)$. However, this estimate does not tell us anything about the sentiment of the recipient i.e. if the utterance is positive or negative. A simple approach could involve estimating $p(w_1)$ using the generative LM of diverse value holders and setting a probability threshold below or above which the sentence is estimated as positive or negative. However, setting such a threshold is arbitrary and the estimation of sentiment intensity is difficult because of the absence of another probability estimate to measure $p(w_1)$ against. To accomplish some sort of measurable reference for the estimate, a second assumption is introduced. Since the state (S_1) and action (A_1) of the subject (θ) in w_1 are known and given a universal set of expressions $\sum$, and a value model of the recipient, a second utterance w_2 can be generated such that the action (A_2) or state (S_2) is completely opposite in sentiment to the state or action in the original sentence w_1. This state/action of the new sentence must be drawn from a finite subset of $\sum$, $\sum'$ containing semantically relevant state/actions associated with the subject (θ) and the newly generated sentence must also be syntactically correct. By expressing contrary action or states the methodology attempts to generate a syntactically correct and relevant sentence that connotes sentiments opposite to the original sentence w_1. To this new sentence w_2, the likelihood of the recipient making the statement by applying the value language model on it can be estimated.

The outcome of this process is two probability estimates ($p(w_1), p(w_2)$) where utterances w_1 and w_2 portray opposite sentiments. As such, if $p(w_1) > p(w_2)$, then it is inferrable that the recipient is more likely to make the statement w_1 and so, it means he is more likely to prefer the state and actions expressed in w_1 which will most likely result in positive behaviour or sentiment. Conversely, if $p(w_1) < p(w_2)$, the recipient is less likely to make the statement w_1 and so the sentiment is likely to be negative, because he is less likely to utter the original statement and more likely to utter or make the new sentence w_2, with the opposite sentiment. In other words, the recipient is more likely to prefer the state or action expressed in the new sentence w_2. Based on this, the sentiment (Ψ) is a measure of the difference between the probability estimate of the actual utterance (w_1) and the probability estimate of a new utterance (w_2) depicting a state or action that is opposite in sentiment to the initial utterance.

In addition to this, another key aspect of value fields is the ability to detect the intensity of the sentiment as a measure of how far to the left (-ve) or right (+ve) of the sentiment orientation scale that the field moves the sentiment of the utterance (see figure 7). This intensity can be determined by observing the difference between $p(w_1)$ and $p(w_2)$. The margin of the difference between the probability likelihoods in either direction can allow us infer sentiment orientations like extremely negative, negative, positive, extremely positive. The steps in algorithm 1 provide a stepwise guide to predicting the sentiment of an utterance.

This section concludes with a brief comparison of the approach taken in modeling values as LMs and its application in sentiment prediction against the existing approach described at the start of this chapter. The benefits of modeling values as LMs as compared to existing approaches include:

- Elimination of the need to manually enumerate and categorize value items: In the manual modeling process, the vocabulary of all value items is scanned, itemized and manually grouped into categories, after which they are manually rated and made ready for value orientation classification. This process of manually grouping items into categories is essentially a process of mapping semantically related items or in the case of this research's approach, VCs. In this model, relevant VCs are identified using linguistic clues (content or function words), and the LM generates semantically related words from a prior distribution of words in the vocabulary.

Algorithm 1 Steps for Calculating Sentiment Orientation of an Utterance

---

Inputs

(i) Utterance (w_1) as a sequence of words

(ii) Value model (V) of a recipient (H)

Identify

(i) Subject (Θ)

(ii) Action (A_{w1}) of (w_1)

(iii) State (S_{w1}) of (w_1)

Generate: Utterance (w_2)with Action (A_{w2}) and State (S_{w2})

Estimate: $p(w_1)$

Estimate: $p(w_2)$

If $p(w_1) > p(w_2)$

Ψ = positive for w_1

Else

Ψ = negative for w_1

- It also renders redundant the need for determining any explicit value orientations because the LM is a representation of what the value holder is likely to say or not say (Pragmatically, this is a measure of how acceptable or unacceptable the sentence is). This makes the approach in this research more flexible as it is not constrained to a particular value orientation category. Based on this observation, the VSM illustrated in figure 4 is updated to portray the research design. In the updated VSM the goals 'Categorization of enumerated value items into a value inventory' and 'Discover value orientation' are eliminated and replaced with the respective goals 'Build values Model' and the methodology 'Create Language Model'. In addition, the methodology of the final goal, 'Application of value orientation to sentiment prediction' is updated to reflect this research's approach which is based on comparing $p(w_1)$ and $p(w_2)$. Finally, the methodology of the sub-goal, 'Identifying value items' is modified to portray the approach which is based on identifying and using syntactic and grammatical clues embedded in the sentence towards the identification of VCs.

Figure 8 portrays the updated VSM.

STAGE 1

GOAL - PARAMETER ESTIMATION

SUB-GOAL	Understand the goals and motivations of the value
METHODOLOGY	Human intervention – interviewing value holders, surveys, questionnaires
OUTCOME	**{H – Value holder,C - Context}**

SUB-GOAL	Identifying value items
METHODOLOGY	Syntactic and Grammatical clues
OUTCOME	**{A – Action, S – State, θ – Subject of the value}**

STAGE 2

GOAL	Build Values Model
METHODOLOGY	Create Language Model
OUTCOME	Language model (LM)

STAGE 3

GOAL	Application of value orientation to sentiment prediction of a sentence (w)
METHODOLOGY	Use LM to estimate p(w). Generate contrary sentence (w_2) and check if $p(w) > p(w_2)$
OUTCOME	Sentiment polarity of input sentence

Figure 8: Illustration of Sentiment Prediction Process using value LM

5.12 Conclusion

The primary outcome of this chapter is the design of the value model and its application towards recipient sentiment prediction. The structure and makeup of values called VCs, derived from identifying commonalities in value conceptualizations has been represented.

This chapter has shown that linguistic clues embedded in value laden sentences can be mapped to these VCs without the need for human input. In addition, the value model has been shown to be a language model of a recipient's vocabulary, and the unique benefit of this is that it eliminates the need for manual identification of value items or categories. In addition, this chapter has shown that modeling values as LM, allows for greater flexibility towards determining whether an utterance satisfies the values of a particular recipient since the LM, also incorporates the context of the utterance. Finally, this chapter has shown how recipient sentiment can be estimated as a function of two probability estimates derived from the recipient's LM. This unique design eliminates any human input and does not require annotations. In the next chapter, the model implementation is described.

6. Model Implementation

This chapter discusses the implementation of the model described in chapter 5. It details the implementation of the value language model along with the implementation of recipient sentiment prediction. It is divided into three sections.

The first section describes the implementation of a set of processes designed to prepare the data for building the model. The second part addresses the extraction of value components from sentences. Chapter 5, showed how syntactic clues can aid the identification of value components and highlighted the importance of capturing and introducing the relations between VCs in the model. This section, describes how these relationships are captured and represented by introducing the concept of Dependency Grammars (DG) - a grammar formalism that plays a major role in how subjects, actions and states are identified and the relationships between them. The third part addresses the implementation of the value model and its adaptation for sentiment analysis. This section is divided into two: The first part focuses on the implementation of the value language model, illustrating how the rich context of the model is implemented combining language models and a maximum entropy (maxent) classifier. The second part demonstrates the implementation of sentiment prediction via the LM and maxent classifiers through a method termed feature switching. Figure 9, acts as an illustrative stepwise guide for the implementation beginning with the document preparation and terminating with the prediction of recipient sentiment.

6.1 Data Preparation Implementation

The objective of data preparation is to extract, compile and process value laden sentences to be applied in building the value model. This process is important because the relevant data would be drawn from a diverse array of source documents e.g. speeches, commentary and debates, made by different value holders H[30].

In figure 9, data preparation is the second process while the first process is labelled as 'Domain based document pre-processing of data'. In reality, the latter process is a pre-pre-processing stage that is entirely dependent on the domain or document type. It makes ready each unique document or domain type for data preparation (see figure 9). The implementations discussed in chapter 7, show how the different domain documents are pre-processed and made ready for data preparation.

Figure 10 is an exposition of data preparation pipeline which transforms the data to a format that is ready for processing and analysis. Its design is implemented in a way that is independent of the domain, so that it is independent of the structure or format of the

[30] Where $H = \{H_1 \; \dots \; H_n\}$, n is an integer representing the number of value holders.

document but on the linguistic units that make up each sentence. The next section proceeds with a description of the 'Data Preparation' pipeline[31].

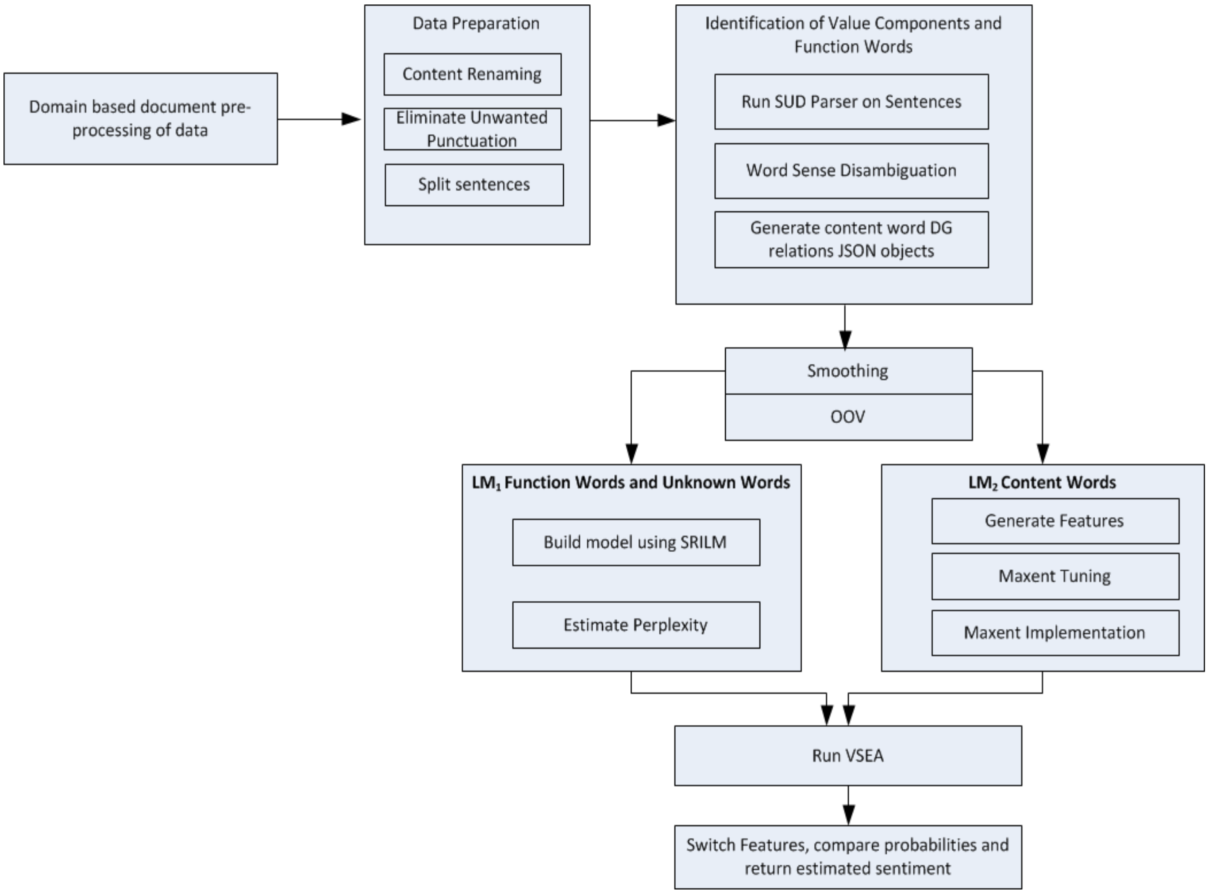


Figure 9: Stepwise Implementation Process of Value Sentiment Model

6.1.1 Content renaming

Content renaming is the first stage of the data preparation pipeline. The goal of content renaming is to identify and rename rare or unseen semantically relevant lexical units[32]. These units are typically VCs such as names, titles, locations and product types etc. The rationale behind content renaming stems from the fact that during model implementation, there would be a set of relevant words that would be observed during testing but unseen during model training or vice-versa because they are simply rare. In the sentence *"Andre Pitovsky will be addressing the Shandong community"*, the likelihood of observing the expression *'Andre Pitovsky'* and *'Shandong'* in the training or test set is likely to be low. Using a process of smoothing such rare words could be omitted completely from the vocabulary of content words, but since they are important, content renaming is applied by assigning pseudo-words to a class of commonly occurring linguistic units.

---

[31] Domain based document pre-processing of data' is left out because it is dependent on the document or domain. For the test cases used in this research the pre-pre-processing in chapter 7 is described.

[32] Lexical units refer to grammatical word forms that make up the sentence. These include words and punctuation.

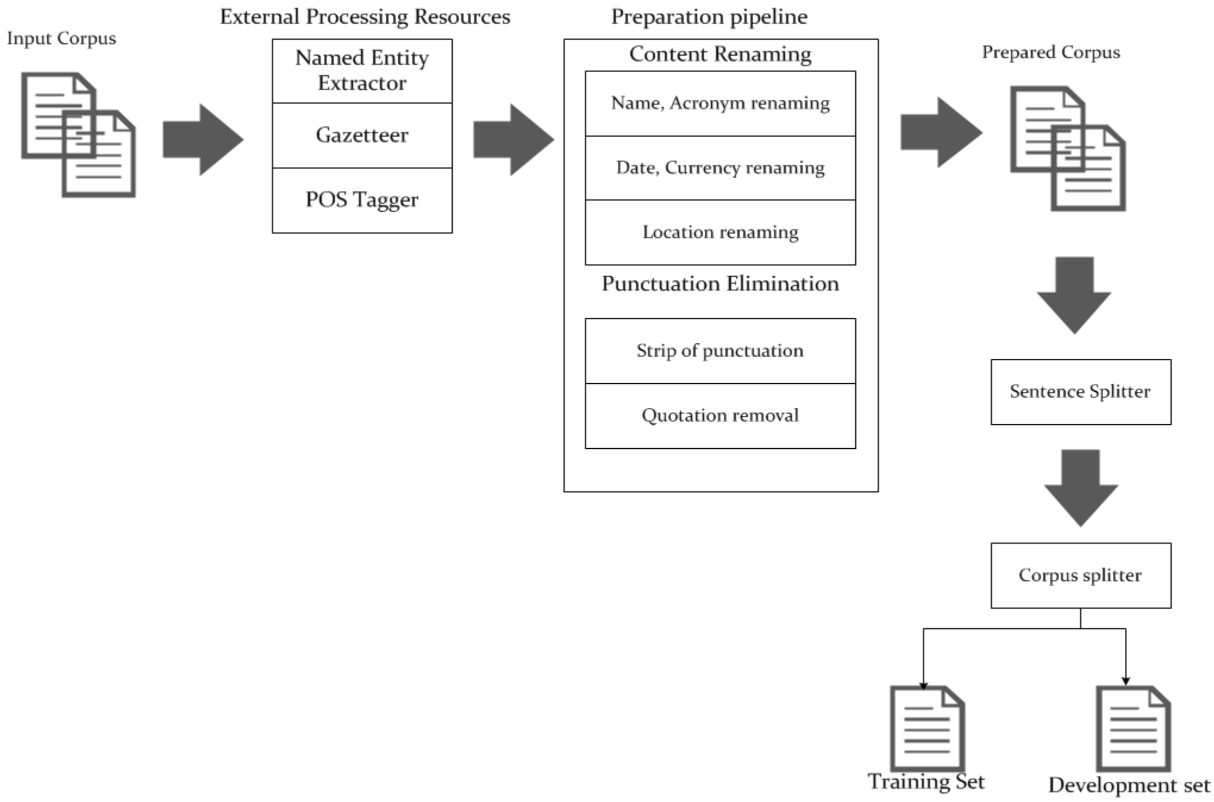


Figure 10: Illustration of the Data Preparation Pipeline

The renamed linguistic units include:

- **Person Names** – In this implementation, GATE's (General Architecture for Text Engineering[33]) named entity transducer which is a processing resource for identifying named entities including person names is used. In figure 10 it forms part of the external processing resources seen in the architecture. In the implementation, GATE is quite good at identifying patterns such as first names e.g. '*John*', [first name + last name] e.g. '*John Smith*', [first name + middle name +last name] e.g. '*John Doe Smith*'. However, it sometimes fails to identify other non-descript name patterns such as names preceded by titles e.g. '*President Thambo Mbeki*' or even names that are not Germanic, Latin or Hebrew like '*Thambo Mbeki*'. To address this, a unique processing resource was implemented in GATE to identify such name using regular expressions. All person names identified are mapped to the pseudo-word 'PERSONNAME'.

  In addition to person names, all acronyms/abbreviations are identified and reverted to their abbreviated form. As part of the external processing resource seen in figure 10, an external gazetteer of acronyms (a list of acronyms) is compiled to aid in the identification of popular acronyms and abbreviations. In addition to

[33] GATE – gate.ac.uk: Last accessed 12/12/2016

acronyms, alpha numeric characters which could be product names e.g. '*Audi A3*' or locations e.g. *'M40'* are also identified. Seeing as the gazetteer's coverage cannot always be extensive, regular expressions were developed to augment the process of identifying abbreviations, acronyms and alphanumeric entities[34]. Abbreviations and acronyms are mapped to the pseudo-word 'ABBREVIATIONNAME' while alphanumeric characters are mapped to the pseudo-word 'ALPHANUMERICNAME'.

- **Dates, Numbers and Currency** – Since instances of numbers and currencies are important as quantifiers, it was vital to represent them. Like names, GATE's processing resource was applied. However, GATE is incapable of identifying date entities like *'In the 20th century'*, and so a regular expression engine is implemented to identify these. The following pseudo-words are assigned:
    - **Dates** - All date expressions are assigned the pseudo-word 'FIGUREDATE'. These include expressions such as '*20th century*', '*12th of August*', '*July*', '*12-11-2001*'. For instance, in sample sentence 1 below, drawn from the research data, '*2017*' is replaced with the expression 'FIGUREDATE' as seen in sample sentence 2.

      Sample sentence 1 – "*Only the Conservative Party will deliver real change and real choice on Europe, with an in-out referendum by the end of 2017.*"

      Sample sentence 2 – "*Only the Conservative Party will deliver real change and real choice on Europe, with an in-out referendum by the end of FIGUREDATE.*"

    - **Currency, Numbers and Percentages** – This category includes numbers expressed either as words or Arabic numerals. Since numeric elements can range from $-\infty\ to + \infty$, a generic pseudo-word is assigned for all numbers, currencies and percentages that are less than 0 to be 'FIGURENEGATIVENUMBER' for numeric expressions, 'FIGURENEGATIVEPERCENT' for percentages and 'FIGURENEGATIVEMONEY' for currencies. For numeric expressions, greater than 0 a naming convention is used which combines the expression 'FIGURE' followed by the number of digits expressed in words and the category 'PERCENT', 'MONEY' or 'NUMBER'. Table 5 shows samples of these mappings.

[34] Regular expressions - '$\hat{}(?\,!\,[0-9]*\$)([A-Z0-9](\backslash.)?)\{2,\}\$$' and '$\backslash b[A-Z\backslash.?]\{2,\}\backslash b$' - are used in identifying abbreviations and alphanumeric entities.

Table 5: Samples of Mappings from Linguistic Numerals to Pseudo-words

Linguistic Expression	Pseudo-word Expression
5000 pounds	FigureFiveDigitMoney
10.45%	FigureTwoDigitPercent
2 trillion dollars	FigureThirteenDigitMoney
75987	FigureFiveDigitNumber
0.45	FigureOneDigitNumber

This sort of numeric or monetary entities were identified generically by GATE's Semantic Tagger (GST). However, in making the distinctions seen in table 5, a second parser called a post parser is implemented and applied such that for each monetary, date or numeric entity identified by GST, the post parser parses it to see if it contains a decimal, determine digits before the decimal and then assigns a pseudo-word to the entity.

- **Locations** – Locations like names are proper nouns. The reason behind separating locations from names was to capture the semantic reality that persons are different from locations. The GATE gazetteer was extended with a list of countries, cities and towns around the world. Locations identified in the sentence are mapped to pseudo-word 'LOCATIONNAME'.

The order of the content renaming pipeline commences with the renaming of persons, abbreviations and acronyms. This is followed by renaming locations and finally Dates, Numbers and Currency. The reason behind this ordering is that, it is possible for a person's full name or a substring of the name to also be the name of a country. For instance, the last name in the person name *'Johnny England'* is also the name of a country – *'England'*. Therefore, to avoid assigning the wrong pseudo-word to the entity, it is reasoned that it is more likely for such expressions to be names than countries and so names are converted before locations. In conclusion, the job of this aspect of the pipeline is to map named entities to pseudo-words**.** The second part of the pipeline seen as 'Punctuation Elimination' in figure 10, is discussed in the next section.

6.1.2 Eliminating Unwanted Punctuation

This process involves eliminating unwanted punctuation as well as replacing content enclosed in brackets and quotation marks. Punctuation elimination varies from dataset to dataset. It could be as simple as removing all punctuation and as extensive as eliminating disfluencies like *'uh'*, *'em'* etc. which are not uncommon in speeches and informal documents, like reviews and comments. Some documents contain hashtags and smiley faces and in most 21st century documents, they have semantic meanings. Quotations can also be eliminated because it can be difficult to attribute them to the actual speaker.

With the data preparation process completed, the modified documents in the corpus are converted to a corpus of sentences. This is because the unit of analysis is a sentence. As seen in figure 10, this conversion is implemented using GATE's sentence splitter. The final process involves splitting the corpus of sentences into training and test sets so that the process of identifying value components can commence.

6.2 Identifying Value Components

With the data prepared, the next stage involving the identification of value components is described (refer to figure 9). Chapter 5, showed that A, S, and θ are parameters which describe an action, a state, character or nature of existence of a subject and can thus be grouped under the grammatical class of words called content words. In addition, it was pointed out that in VLSs, these words are linked together by a class of words called function words. Since content words could have multiple semantic meanings - in other words they could serve as actions, states and subjects - all content words are treated as belonging to a generic set of content words and a distinction is not made between actions, states and subjects.

It is assumed that the parameter (H) is known at the point of data collection because the owners of the data or the speakers, authors or commentators are typically the value holders. This leaves the content words and the context. As for context, this thesis has expressed that it consists of a non-exhaustive range of existential factors which is determinable on a domain or case by case basis. So, at this point, it is assumed that a set of different contexts $\{C_1 \ldots C_n\}$, where n is an integer exists. In chapter 7, where a test case implementation is described, the applied contexts are further elucidated. For now, assume that context is known. Thus, the goal in this section focuses on the identification of the function and content words.

A naïve approach to identifying content and function words would have entailed identifying all function words in the corpus leaving only the content words and vice-versa. However, simply identifying these elements was not informative enough for the model implementation since it did not reflect the relationship or dependency between any of the linguistic units nor portray the primacy of the content words either as independent units or when compared to other units.

To accomplish the task of content and function word identification, English language parsers based on the Dependency Grammar (DG) formalism are applied (Tesniere, 1959; Nivre, 2005). DG formalism stems from the work of the French linguist Lucien Tesniere (1959). It is a formalism that provides theoretical approaches for formalizing sentence structure. The central tenet behind DG is that linguistic units in a sentence are connected to each other by asymmetric binary relationships and consequently there exists a type of dependence between them. In fact, the notion of dependence between words in a sentence was also reinforced by Saussure who said that "Language is a system of interdependent terms in which the value of each term results solely from the simultaneous presence of others" (Manjali, 1994). A detailed discussion of DG, dependency parsers, parsing algorithms and how the dependencies between words are learnt is beyond the scope of this thesis.

Before expounding on DG implementation, an explanation of why DGs are applied to VSM implementation is provided and they are as follows:

- **Reflects the core subject of the sentence** – VLSs express the intents of a person. Although these intents are expressed as a sequence of words at the communicative level, humans (the speaker and recipient) can cognitively deduce and summarize the substance of the intent into one or two expressions. Consider the sentence *"The Government will ban head scarves"*. The central intent of this sentence describes an action – *'will ban'*. All the adjourning linguistic units, though necessary for sense making cannot exist without the main intent of the sentence expressed as *'will ban'*. Thus, the verb *'ban'* is linguistically the most important unit in the expression as it semantically represents the act or process that is expressed by the grammatical subject or object. Consequently, it is inferrable that the intent of the speaker or primary thought which is an expression of his values is encapsulated in the expression *'will ban'*.

  One of the reasons for using DG formalism, is that it captures the primary thoughts of a sentence as a concept called Head Word[35]. In DG, Head word refers to the highest-level word in a sentence (usually a verb[36]) for which all other words depend (Manjali, 1994; De Marneffe et al, 2008). Such a word semantically represents the process or act that is expressed by the grammatical subject or grammatical object in the sentence. Therefore, with a dependency parser, the central thought of any value laden sentence or head word can be identified.

  Another justification for using DG formalism that extends beyond identifying the core subject of a sentence is that it also aids in identifying other secondary relations and units which make up the sentence. These relations represent the function and role played by the linguistic units in the sentence. For example, with DG formalisms, the agents (a semantic concept representing the initiator of an event), patients (semantically represents the target or participator in an event) and themes operational in the sentence (These are value components) can be identified. By

[35] Head word is also called the root word.

[36] Note that the root word is the most linguistically relevant word in the sentence and so there are cases where the head word is not a verb. For instance, 'She could have been sick'. The root here is 'sick'. In the sentence, 'All these three books', the root is 'books' a noun. (Source: Stanford Universal Dependencies, http://universaldependencies.org/docsv1/u/overview/syntax.html: Last accessed 08/01/2016)

identifying these elements, the central core/intent of a sentence can be mapped to a value component. For instance, in the sentence *"The Government will ban head scarves"*, a dependency parser will identify the content words *'ban'*, *'Government'*, *'head scarves'*, and return a mapping featuring the head word *'ban'* and the arguments *'Government'* and *'head scarves'* i.e. $ban('Government', 'head\ scarves')$. In this example, '*Government*' is the agent and '*will ban head scarves*' is a phrase representing the event or theme. This mapping encapsulates the relationship between the three expressions.

- **Captures Linguistic Priority** –DG formalism is applied because it expresses linguistic priority which is the notion that one linguistic unit is superior to another. This principle is in fact not new. According to Jespersen (1924, p96),

  "In any composite denomination of a thing or person, we always find that there is one word of supreme importance to which the others are joined as subordinates. The chief word is defined (qualified, modified) by another word, which in turn may be defined (qualified, modified) by a third word, etc. We are thus led to establish different 'ranks' of words according to their mutual relations as defined or defining. In the combination, 'extremely hot weather' the last word 'weather', which is evidently the chief idea, may be called primary; 'hot', which defines 'weather', secondary, and 'extremely', which defines 'hot', tertiary."

  Also, Zwicky (1985, p.2) notes that, "in syntactic constructs, one constituent characterizes or dominates the whole". The notion of priority words (some terms are more important than others) in the value model is in concord with this principle. DG formalism reflects this notion of priority in that the asymmetric relationship between linguistic units link a superior term or governor to an inferior term called the dependent (Kubler et al, 2009; Aydin & Torusdag, 2008). Based on this notion of priority, DG formalisms can support hierarchically ranking the class of content words in a sentence where the primary word is the head-word verb, followed by its actants which are usually nouns. Next in line in this hierarchy are the modifiers of the nouns which are typically adverbs and adjectives as illustrated in figure 11.

- **Captures relationships between Linguistic Units** – The value model proposes to employ the relationships between linguistic units. DG formalisms reflect dependency relationships between lexical units. These relationships could be morphological, syntactic or semantic (Polguere and Mel' cuk, 2009). Semantic roles are difficult to categorize and determine, requiring significant human input (de Marneffe et al, 2014) as such syntactic roles for which exist considerable literature and implementations are used. In addition, dependency structure is not defined by word order implying that the relationships that exist between content words regardless of how they are positioned or their proximity to one another can be identified. This means that the likelihood of a word can be predicated on not only words in its vicinity but words in a DG relation.

In summary, DG presents a reasonable theoretical formalism for identifying not just content words (value components) and function words but also identifying the relationships between words and the priority of each linguistic expression. In the next section, the DG implementation used is described.

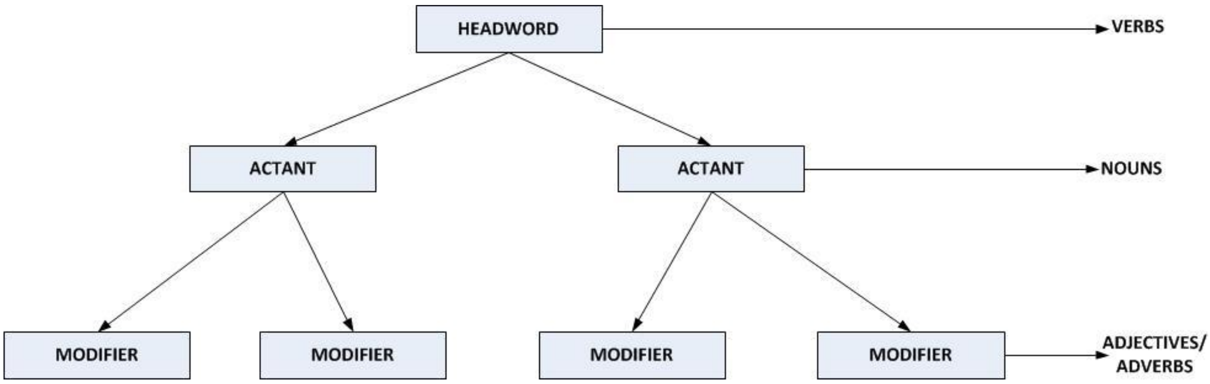


Figure 11: Illustration of Word Class Priority Hierarchy

6.2.1. Dependency Parser

The task of dependency parsing entails taking an input sentence (w) and returning a parse tree of typed dependencies. A parse tree is simply a tree-like representation of the structure of a sentence. Several computational implementations of DGs exist such as Minipar, RASP, SUPPLE, and while the details of these implementations are beyond the scope of this research, Stanford Universal Dependencies implementation (SUD) (de Marneffe et al, 2014a; de Marneffe et al, 2014b) is used.

SUD is a typed dependency implementation. It is called typed because each link in the parse tree is labelled with a dependency type. For example, figure 12 shows the labelled link or dependency relationship between *'Europe'* and *'Union'* to be the expression *'compound'* for the sentence *"We will definitely be leaving the European Union"*. For the linguistic units *'European'* and *'Union'*, the governor or superior unit is *'Union'* while the dependent is *'European'*. SUD is used specifically for its popularity and accessibility but most importantly because of its broad universal taxonomy of relations which have been designed to represent the grammatical function of each linguistic unit relative to its dependent. For instance, in figure 12, the dependency relationship between *'definitely'* and *'leaving'* expressed typographically as $advmod(\text{'}leaving\text{'}, \text{'}definitely\text{'})$ states that the dependent *'definitely'* is the adverbial modifier of the governor *'leaving'* while in figure 13, the dependent word *'British'* is the adjectival modifier of the governor *'players'* i.e. $amod(\text{'}players\text{'}, \text{'}British\text{'})$.

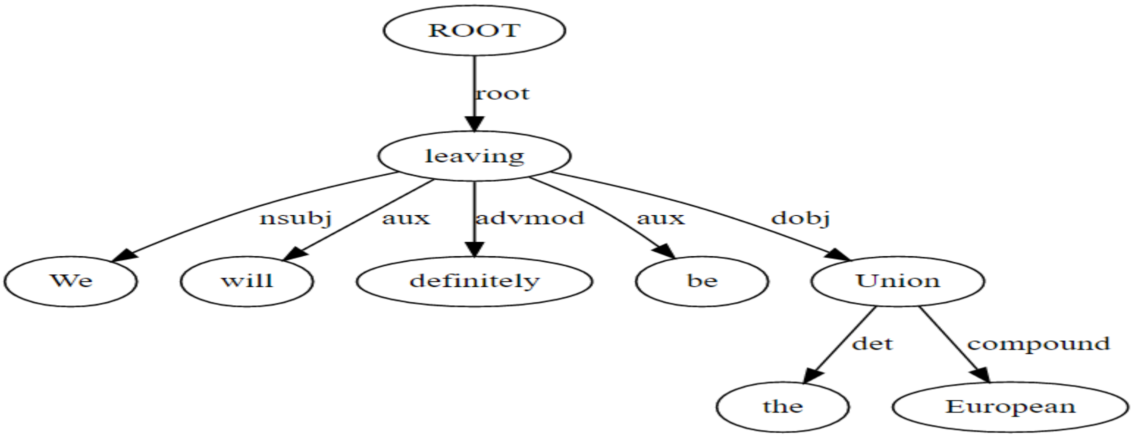


Figure 12: Illustration of a Typed Dependency for sentence 'We will definitely be leaving the European Union'

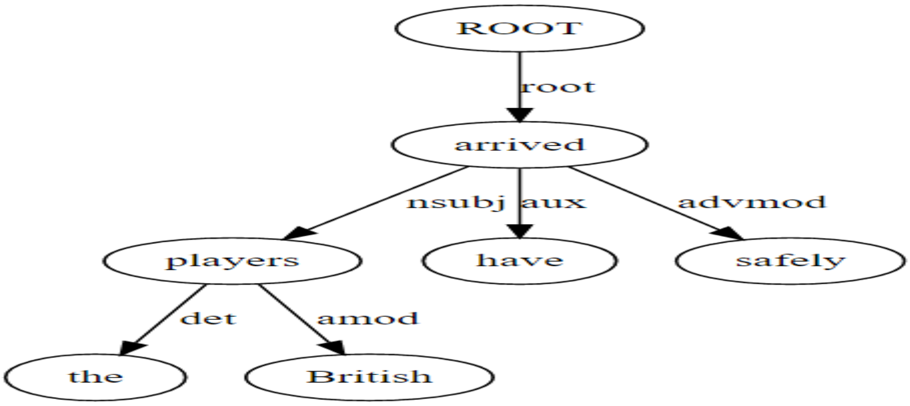


Figure 13:Illustration of a Typed Dependency for sentence 'The British players have arrived safely'

SUD consists of 42 relations centred around core arguments - These are the dependency relations that the sentence predicate partakes in - and non-core arguments - which generally covers the dependency relations of modifiers, nouns and function words. Although there are 42 relations in the SUD, not all relations are applicable to all grammar types. Some relations tend to occur more commonly in non-Germanic languages for instance the relation '*clf*[37]' occurs predominantly in Asian languages. In addition, the evolving nature of language means that additional relations are being discovered with newer relations emerging from older ones. Nevertheless, SUD clearly delineates core dependency relations - which are relations between the verb e.g. the root verb and subjects, objects or clausal complements in the sentence – and other dependency relations such as modifiers. Using the word class priority illustration in figure 11 as a reference, the relations used in this research include all 8 core dependency relations (based on figure 11, the relations between the headword and verbs in the sentence and the actants). These relations

[37] http://universaldependencies.org/u/dep/clf.html: Last accessed 19/04/2017

are listed in table 6. The second category of relations are oblique relations or modifiers which relate the actants and modifying expressions as seen in figure 11.

Table 6: Core Dependency Relations (Source: SUD v1.0)

Dependency Relation	Full Meaning
nsubj	Nominal subject
csubj	Clausal subject
nsubjpass	Passive nominal subject
cubjpass	Passive clausal subject
dobj	Direct object of a predicate
ccomp	Clausal complement of a verb or adjective
xcomp	Clausal complement of a verb
iobj	Indirect object of a verb

Table 7: Modifier Relations of Nouns and Clausal Predicates (Source: SUD v1.0)

Dependency Relation	Full Meaning
nmod	Nominal modifier
advcl	Adverbial clause modifier
advmod	Adverbial modifier
neg	Negation
acl	Adjectival clause modifier of a nominal
amod	Adjectival modifier
appos	Appositional modifier of a noun
nummod	Numeric modifier of a noun

SUD's 8 non-core and modifying relations are used and outlined in table 7. The third category are relations which exist between content words and function words and relations which express coordination, disfluencies, multiword expressions and punctuation. Although this third category of relations are important they are less regular and their occurrence can be dependent on the type of document or language. For example, the relation '*expl*' relates expletives, and is commonly seen in informal documents. Therefore, 14 relevant content-function word relations are used and outlined in table 8.

In addition to these relations, SUD includes a relation called *'root'* which reflects the root of the sentence. The governor of this relation is a fake node *'ROOT'* which marks the start of the dependency tree and its dependent is the root word. In figures 12, 13 and 14, the root relations are $root(ROOT, leaving)$, $root(ROOT, arrived)$, $root(ROOT, defeated)$.

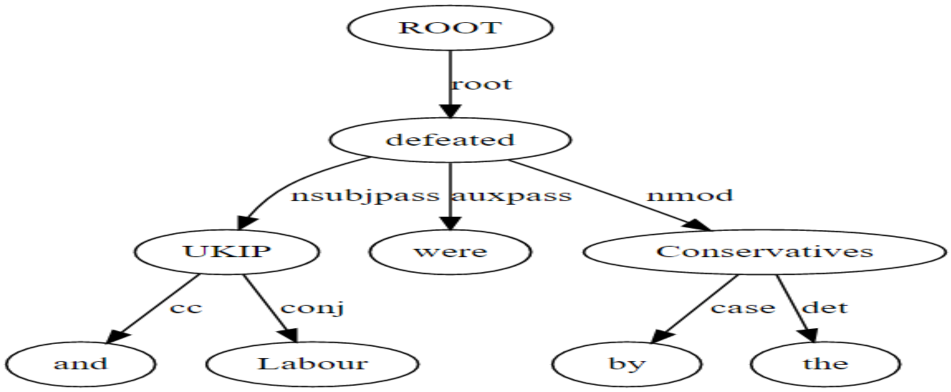


Figure 14: Illustration of a Typed Dependency for sentence "UKIP and Labour were defeated by the Conservatives"

Table 8: Content-Function Word Relations (Source: SUD v1.0)

Dependency Relation	**Full Meaning**
det	Determiner relation between a nominal and a determiner
mwe	Relates multiword expressions
goeswith	Links two parts of a word that are separated in text that is not well edited
name	Used to relate names e.g $name(Cameron, David)$
foreign	Used to label sequence of foreign words
list	Used to relate chains of comparative items
aux	Relates an auxiliary of a clause
auxpass	Relates passive auxiliary of a clause
mark	Relates a marker, which is the word introducing a finite clause subordinate to another clause
cop	Captures relation between the complement of a copular verb and the copular verb '*to be*'
cc	Relates an element to a coordinating conjunction
conf	Relates two elements connected by a coordinating conjunction
case	Used for any case-marking element which is treated as a separate syntactic word (including prepositions, postpositions, and clitic case markers)
compound	Relates compound expressions e.g. $compound(warrant, arrest)$

6.2.2 Axioms for Identifying Content and Function Words

In applying SUD towards the extraction of content words, the following assumptions are introduced:

1. Function words are words which do not partake in modifying relations or core dependency relations. Table 4, outlines the class of function words. Typically, these words do not have any dependents of their own and are normally dependents of content words.

   The class of relations between function words and content words consists of the relations in table 8. For instance, in figure 12, the function words *'will'*, *'be'* and *'the'* are respective dependents of the words *'leaving'*, *'leaving'* and *'Union'*. Therefore, in a sentence, a function word (x) is a word that:

   a. Belongs to the class of words in table 4.

   b. Is a dependent in a relation (R) where R belongs to one of the relations in table 8. In this research, the exception to this is negation relation and negation words, which are viewed as content words.

2. **The Unit of a Content Word** - Since SUD provides dependency labels that differentiate compound expressions from modifications, the unit of content words is formulated as singular lexical units and not expressions. For example, the noun phrase expression *'EU Commission'* is split into two content words *'EU'* and *'commission'* so that when parsed results in the relation $compound(Commission, EU)$. Thus, in a sentence, a content word (x) is a singular word.

3. Based on SUD literature, core dependency relations and modifying relations (table 6 and 7) exist only between content words. Therefore, in a sentence, a word x is a content word if it is in a relationship R with another word y such that the relation R is a core or modifying relationship. It can also be inferred that y is a content word since only content words can partake in the core or modifying relations of table 6 and 7.

4. If two words x and y are linked by a coordinating conjunction e.g. '*and*', '*or*', '*but*', then both x and y must belong to the same class (content or function word). Therefore, if in a compound relation, a known content word (x) is related to a word (y), then (y) is also a content word. Similarly, if a known function word (x) is related to a word (y) through a compound relationship, then (y) is also a function word.

   In figure 12, a compound relation exists between *'Union'* and *'European'*. In addition, *'European'* does not partake in any of the relations in table 6 and 7, but because it is in a compound relation with *'Union'*, it is inferred that it is also a content word. Another instance of this is seen in figure 14 where the word *'UKIP'* is a known content word because of its status as the passive nominal subject of the root word *'defeated'*. *'UKIP'* is related to the word *'Labour'* - which has no dependents - through the joining conjunction *'and'* i.e. $conj('UKIP', 'Labour')$ thereby making

'*Labour*' a content word. This rule applies to function words and it is one of the exceptions to point (1) above where a function word can exist as a governor. For instance, in the sentence snippet '*This or that …*', two relations where function words exist as governors $cc(this, or)$ and $conj(this, that)$.

Therefore, given a sentence, the output of the dependency parser is a set of tuples consisting of a dependent (d), governor (g) and its relation (r). From the axioms discussed above, if r is a core, modifying or negation relation, then d and g belongs to any of the VCs A, S, θ. Otherwise, the words are not mapped to a VC. Following these axioms and a description of the SUD parser, the implementation of a pipeline for extracting content and function words is discussed. This step is highlighted in figure 15 (red bordered box) which is a snippet of the design implementation process shown in figure 9.

6.2.3 Content and Function Word Pipeline

The content and function word pipeline consist of an implementation of the SUD parser, a part of speech tagger and a word sense disambiguation engine (see figure 15[38]). The job of this pipeline is to extract a list of content words and function words from sentences in the training corpus. Figure 16 illustrates the pipeline's two stage process flow. A pre-parsing stage which involves word sense disambiguation subsequently followed by the application of the SUD parser.

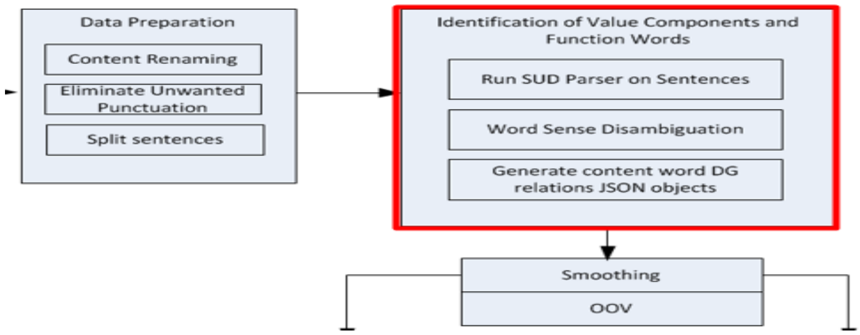


Figure 15: Snippet of Stepwise Implementation Process of Value Sentiment Model focusing on Content and Function Word Identification

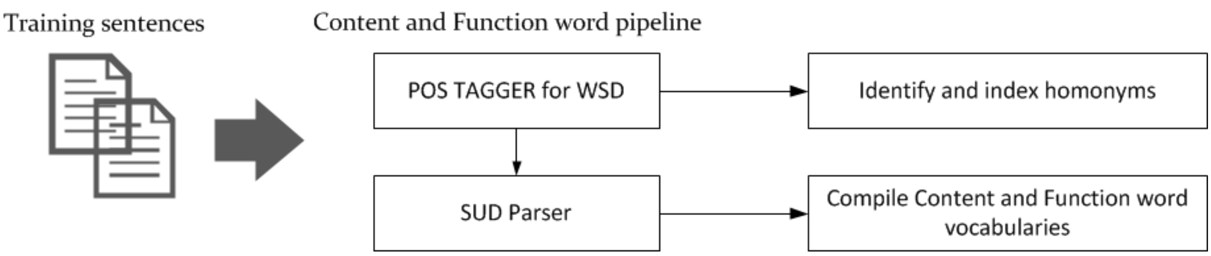


Figure 16: Process Flow showing content and function word pipeline

[38] This snippet of the design process in figure 10 is included on this page to ensure that the reader does not have to continually switch back to or refer to figure 10.

Pre-Parsing Stage – Here, a POS tagger is applied on the training corpus as a first step towards identifying homonyms of candidate content and function words. Stanford NLP POS Tagger[39] is used.

To elucidate the importance of this stage, consider these sentences -

Sentence 1 – *"The people will it"*.

Sentence 2 – *"John's will was destroyed"*.

Sentence 3 - *'Has Will abused his position?'*

Sentence 4 - *'We will leave the UK.'*

In sentence 1, *'will' ('want or desire'*) is a verb and the root of the sentence and so it is a content word. In sentence 2 '*will*' is a type of document and in 3 '*Will*' is the name of a person. In these 3 instances, *'will'* is a noun and a content word albeit with diverse senses. In sentence 4, *'will'* (*'An assertion or intent'*) is a modal verb and a function word. Clearly, all four occurrences of *'will'* are homonymous expressing different senses and it is for this reason that the senses are differentiated when compiling content and function words.

To identify these variations in expressions, a POS tagger is applied on all the sentences, to identify and index all nouns, adjectives, verbs and adverbs. Following POS tagging and by a process of elimination, there are three separate classes for the word *'will'*. The occurrence in sentence 4 is indexed as a function word seeing as it is a modal. This leaves the instances in sentence 1, 2, and 3. The instance in sentence 1 is indexed as a verb content word, leaving the two noun instances. Seeing as the remaining two instances are nouns with different senses a Word Sense Disambiguation algorithm should ideally be applied to detect the difference, however making this distinction is consigned to future work. In effect, the outcome of this stage is that all homonymous words in the training data are differentiated.

In implementing this, each sentence is assigned a unique identifier and mapped to a JSON object that is representative of the sentence. The sentence JSON object encapsulates the sentence ID, and a nested ordered list of words which make up the sentence. Each word in the list comprises of valuable information such as the word sense, the POS and the start and end position of the word. These identifiers are used in subsequent processes in identifying and mapping content word instances. Figure 17 illustrates a snippet of such an object.

Parsing Stage – As mentioned earlier, function words can be identified from their grammatical word class (see table 4). Based on the POS tagging exercise in the pre-parsing stage, all the function words are identified and indexed. The SUD parser is applied on each sentence in the corpus and the outcome of the parsing is represented as a JSON object. This JSON object consists of a list of relevant relationships, a list of content words identified based on their participation in content word relations – relations in table 6 and 7 - (see

[39] http://nlp.stanford.edu/software/tagger.shtml: Last accessed 12/12/2016

figure 18) and an array of object relations representing the relations between content words. The content words relationships are extracted from each sentence's JSON object (In figure 18, '`listOfObjects`' key contains the array of content word relations - that is the details of the relations between two content words) and stored as unique JSON objects in a Mongo Database. A list of content words is subsequently populated from this.

Sentence: UKIP will leave the UK


```
{
        "analysisType": "string",
        "sentenceString": "UKIP will leave the UK",
        "ID": "1",
        "listOfWords": [{
                "word_original": "UKIP",
                "isAlphaNumeric":false,
                "isCapitalized":true,
                "isNamedEntity":true,
                "isNumeric":false,
                "isLowercase":false,
                "word_index": "0",
                "word_lower_case": "ukip",
                "word_POS": "Noun",
                "POS_CATEGORY": "NNP",
                "word_startposition": "0",
                "word_endposition": "5",
                "wordsense_index": "1",
                "wordsense_POS": "Noun"
        }, …
```


Figure 17: Snippet of a JSON representation of a sentence

Since the pre-parsing stage is likely to reveal homonymous words, and as a consequence, duplicate word entries in the final list of content words, each content word is differentiable based on the unique '`wordsense_index`' field in figure 17. So that, in the content word vocabulary the word *'will'* would be expressed with the pattern [*word - word sense index*] e.g *'will - 1'*, *'will -2'*, *'will - 3'* where 1 is a noun, 2 is a modal verb and 3 is a verb.

The outcome of this stage is not only a vocabulary of content and function words for each value holder's training corpus but also JSON objects which encapsulate the DG relations between content words in the training sentences. With these outcomes, the development of the LM is discussed.

6.3 Value LM Implementation

Following the identification of components, the goal is to implement value LMs for the training data. In other words, the goal is to create a LM for each value holder's corpus. To this end, the LM implementation is premised on trigrams which have been shown to produce better estimates than bigrams or unigrams (LDC, 1993; Chen and Goodman, 1999) and so the implementation commences with the identification of trigrams in each corpus. This followed with a reformulation of the LM.

Recalling the LM model in chapter 5, its goal is to model value laden sentences from a vocabulary of words $\boldsymbol{V}$ comprising of both content and function words.

```
{
"analysisType": "string",
"sentenceString": "As a former airline manager, I totally support the Prime
Minister’s determination to get full access to airline passenger name records,
which would be to the advantage of Governments in both preventing terrorist
movements and protecting young and vulnerable UK nationals.",
"listOfRelationsAsString": ["amod", "compound", "nmod:as", "advmod", "root",
"compound", "nmod:poss", "dobj", "acl", "amod", "dobj", "compound", "compound",
"compound", "nmod:to", "nsubj", "acl:relcl", "nmod:of", "advcl", "amod", "dobj",
"advcl", "conj:and", "amod", "conj:and", "amod", "compound", "dobj"],
"listOfVitalWords": ["protecting", "Prime", "access", "manager", "preventing",
"records", "young", "advantage", "nationals", "determination", "movements",
"former", "terrorist", "Governments", "vulnerable", "passenger", "UK", "get",
"name", "airline", "totally", "support", "Minister", "full"],
"listOfObjects": [{
             "word_gov": "manager",
             "word_dep": "former",
             "word_gov_lemma": "manager",
             "word_dep_lemma": "former",
             "word_gov_stem": "manager",
             "word_dep_stem": "former",
             "relationship": "amod",
             "word_gov_POS": "NN",
             "word_dep_POS": "JJ",
             "word_gov_POS_root": "Noun",
             "word_dep_POS_root": "Adjective",
             "word_gov_index": 5,
             "word_dep_index": 3,
             "word_gov_startposition": 20,
             "word_dep_startposition": 5,
             "word_gov_endposition": 27,
             "word_dep_endposition": 11
      },
```


Figure 18: Snippet of a JSON representation of a SUD parsed sentence

$$p(w_1 \dots w_n) \approx \prod_{i=1}^{n} \begin{cases} p(w_i|h, H, C) & if\ w_i\ is\ a\ function\ word \\ p(w_i|h, H, C, A, S, \theta) & if\ w_i\ is\ any\ of\ A, S, \theta \end{cases}$$

Since no distinction is made between content word classes (A, S and θ), a set $\boldsymbol{O}$ is defined to be the set of all content words. In addition, because the relationship between the content words is also a factor in the model, the set $\boldsymbol{R}$ of all representative relationships between content words is also defined. The set $\boldsymbol{F}$ of all function words is also introduced. The LM implementation is premised on trigrams. Unigrams or bigrams are unused because they do not capture enough context and perform poorer than trigram LMs (Jurafsky and Martin, 2009). Trigrams are also known to be very popular (Jurafsky and Martin, 2009). Conversely, using higher order n-grams like 4,5-grams result in the estimation of more parameters and can make the solution impractical, hence trigrams are commonly used (Brown et al, 1992; Chen and Goodman, 1999; Manning and Schutze, 1999)[40]. Therefore, the history h is expressed in terms of the prior context, i.e. the two words preceding w_i i.e. $w_{i-2},\ w_{i-1}$. Thus equation 8 in chapter 5 can be rewritten as,

$$p(w_1 \dots w_n) \approx \prod_{i=1}^{n} \begin{cases} p(w_i|w_{i-2},\ w_{i-1}, H, C) & if\ \{w_i: F\} \\ p(w_i|w_{i-2},\ w_{i-1}, H, C, O, \boldsymbol{R}) & if\ \{w_i: O\} \end{cases} \dots (9)$$

[40] Higher order n-grams like 4-grams are generally useful when training tens of millions of words of data (Manning and Schutze, 1999, Cui et al, 2006).

Since the value holder and context are already known, the modified value LM for sentences made by a value holder H_1 given a context C_1 can for instance be expressed as:

$$p(sentence) \approx \prod_{i=1}^{n} \begin{cases} p(w_i|w_{i-2},\ w_{i-1}, H = H_1, C = C_1) & if\ \{w_i: F\} \\ p(w_i|w_{i-2},\ w_{i-1}, H = H_1, C = C_1, \boldsymbol{O}, \boldsymbol{R}) & if\ \{w_i: O\} \end{cases} \quad \text{...(10)}$$

Consider the first condition in the modified model i.e. if w_i is an element of F. A natural approach to estimating $p(w_i = function\ word|w_{i-2},\ w_{i-1})$[41] is the Maximum Likelihood Estimation (MLE) expressed as the ratio of the trigram count (the frequency of occurrence of the sequence of words $\langle w_{i-2}, w_{i-1}, w_i\rangle$ in the training corpus) to the frequency of occurrence of the sequence of words $\langle w_{i-2}, w_{i-1}\rangle$ in the training corpus. That is,

$$p(w_i|w_{i-2}, w_{i-1}) = \frac{c(w_{i-2}, w_{i-1}, w_i)}{c(w_{i-2}, w_{i-1})} \text{... (11)}$$

where $c(x)$ represents the count of x.

However, because of the large set of likely trigrams[42] and comparatively small trigram count, it is inevitable that most of the MLEs will be zero (a significant proportion of the likely trigrams would not be seen in training) and the net effect will be poor probability estimates. This problem is called the sparse data problem and it is addressed in the next section with the aid of a technique called smoothing.

6.4 Smoothing

Smoothing is applied in addressing the data sparsity problem. It is the process of redistributing some of the probability mass from high probability ngrams to low probability or unseen ngrams so that zero probabilities observed for unseen events are eliminated and more reliable estimates generated.

Smoothing has been shown to improve the overall performance of ngram models. (Chen and Goodman, 1999). Several smoothing techniques have been applied in LM including Witten-Bell smoothing (Bell et al, 1990), Katz back-off (Katz, 1987), Good-Turing estimation (Good, 1953), Absolute discounting (Ney & Essen, 1991; Ney et al, 1994), Jelinek-Mercer Smoothing (Jelinek and Mercer, 1980) and Kneser-Ney smoothing (Kneser and Ney, 1995).

---

[41] The value holder and context variable are omitted in the condition because they are known entities and because each value-holder/context corpus is treated as an individual LM. So, each value-holder/context probability estimate can be rewritten as $p_{conservative-EU}(w_i = function\ word|w_{i-2}, w_{i-1})$, representing the probability of a trigram function word by a conservative under the EU context.

[42] Given a vocabulary V, the number of likely trigrams is V^3. For instance, in the example discussed in chapter 7, vocabulary size of the EU-Conservative corpus is 15756 words, thus the number of possible trigrams is V^3 or 15756^3 (approximately 3.9 trillion trigrams). However, in training, the trigram count (which is the number of trigrams seen in the training data) is 3304053, meaning that a significant fraction of trigrams will be unseen thereby resulting in the sparse data problem.

In the model development, the intuition was to apply the interpolated Kneser-Neys smoothing which has been shown to yield better results than most smoothing techniques (Chen and Goodman, 1999; Ney et al, 1997). However, because of variations in data type and unique nature of the model, 4 other known smoothing techniques are compared in the implementation described in chapter 7– Absolute discounting, linear interpolation, Good-Turing and Witten-Bell smoothing – against interpolated Kneser-Neys smoothing. For the dataset used in this research, the interpolated Kneser-Neys smoothing produces the best probability estimates. Kneser-Neys (KN) smoothing has its origins in absolute discounting (see appendix 4) and it aims at combining information from lower order ngrams towards improving the estimate of higher order ngrams. A description of interpolated KN is described in appendix 5. Finally, the interpolated KN LM was implemented using SRILM toolkit (Stolcke, 2002; Stolcke et al, 2011) primarily because it supports a wide variety of LM implementations.

Following the description of the LM i.e. the equations for estimating function word probability and content word probability and an explanation of the smoothing techniques and the toolkit to be used in building the LM, the next two section describes the estimation of function word and content word probabilities. They are implemented as two separate LMs, LM_1 and LM_2, depicted in figure 19, which is a snippet of the process diagram in figure 9.

6.5 LM Experimentation for Function Word Probability

The objective of this section is to implement a LM for the entire vocabulary of words in the training data set and apply it in estimating the probability of function words for any given test sentence (Let's call this model LM_1). Later, the implementation of LM_2, used in estimating content word probability will be described. Both models will be subsequently applied towards sentiment prediction.

The methodology involves implementing a base-line Kneser-Neys smoothed model which is tuned by modifying the discount and lower order bigram and unigram. Other smoothed LMs are implemented and measured against this baseline model. The performance of the models is compared intrinsically by calculating perplexity (See Appendix 6 for a description of perplexity). As part of the implementation, two vital issues associated with LMs are addressed: Out of vocabulary words and cut-offs.

6.5.1 Out of vocabulary (OOV) words

One of the challenges in LM involves out of vocabulary words. These are words that appear in the test data but not the training set and because they do not appear in the training set it is impossible to estimate their probabilities. Therefore, because insufficient information about the words exists, their probabilities are inestimable. Usually, low frequency words (words with frequency of 1 or 2) in the training set are considered as OOV words because they do not have enough information to make reasonable estimates.

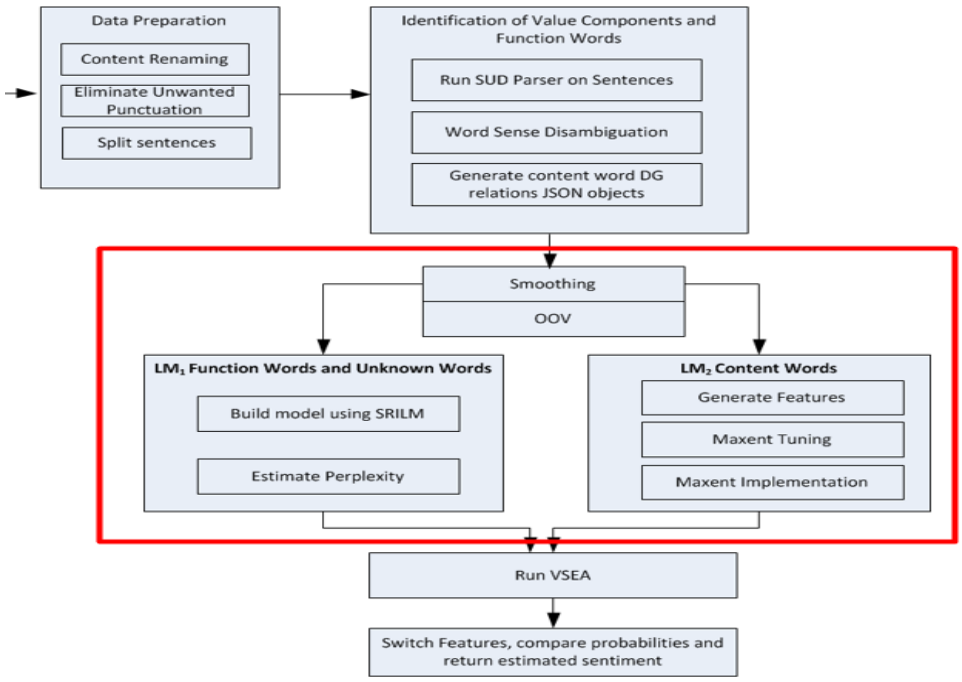


Figure 19: Snippet of Model Process Portraying the development of LM1 and LM2

This thesis makes the open vocabulary assumption which means that the model accounts for unseen/OOV words. The reasoning behind this is because new words (formal or colloquial) and expressions are constantly being added to the English vocabulary. Traditional approaches to dealing with OOV typically involve estimation of pseudo-word probabilities via the following steps:

1.Pre-selecting a fixed vocabulary of words

2.Identifying and converting words in the training set that are not in the fixed vocabulary to the pseudo-word 'UNK'[43] (Bell et al, 1990; Jurafsky and Martin, 2009).

3.Finally, estimating the probability of 'UNK' like any other word.

The problem with this approach is expressed in the question: What is the criteria for compiling the fixed vocabulary of words? How is the size of this fixed vocabulary determined, since it is possible for the vocabulary of words in the corpus to be larger? A second approach entails replacing just the first occurrence of every word type in the training data with 'UNK' (Bell et al, 1990; Jurafsky and Martin, 2009). A third approach involves estimating the probabilities for the most common *k* words, while all others are mapped to the token 'UNK' (Manning and Schutze, 1999). This work proposes a methodology that uses clues and patterns in the corpus towards identifying OOV words. The aim is to gather enough probability information for unseen words that are likely to occur with high frequencies in the test set and low frequency unseen words. This methodology is based partly on research carried out by Muller and Schutze (2011) who

[43] UNK' is short form for unknown word.

showed that OOV words are typically short words, names, acronyms and words containing special characters. Therefore, the first instance of high-frequency words of such types in the corpus can be converted to 'UNK'. Since these words are high frequency words, the assumption is that not much is lost in the way of probability mass because of its high frequency. To this end, the following substitutions are made in the training corpus:

- Returning to the original training text, i.e. the unmodified unprepared text. Mentions of alpha-numeric text, locations, names and organizations are counted. For instances that occur more than 5 times[44] in the text, the very first instance is replaced with 'UNK' in the mirror sentence. This accounts for low frequency content words that are typically discarded in training but are likely to occur in the test set. For instance, in this implementation, an important word 'C4-logistics' occurred just twice in the Conservative-EU test set.

- The first occurrence of other high frequency content words (words with a count that is greater than 15[45]) is replaced with 'UNK'. Since these words are high frequency words it is expected that modifying just one of it would not have a considerable effect on the probability estimation. In making this modification, some probability weight is borrowed from high frequency word and assigned to the unknown word representation 'UNK'. Examples of such words and their frequencies include: 'administration' – 152 times, 'government'- 1974 times, 'party' – 468 times in Conservative-EU training set.

In the next section, cut-offs are considered.

6.5.2 Cutoffs

Cutoffs are a way of restricting the size of the LM by cutting off or ignoring infrequent ngrams. "The count below which the ngrams are discarded is called cutoffs" (Clarkson and Rosenfeld, 1997, p. 1). While cutoffs generally tend to reduce the size of the LM, they have also been shown to slightly reduce the performance of the model (Chen and Goodman, 1998; Clarkson and Rosenfeld, 1997). Thus, the decision about applying cutoffs is about weighting the benefit of a very large model against the slight loss in performance incurred from cutoffs.

In deciding, the experiments of Chen and Goodman, (1998) which compared the effect of cutoffs on several smoothing trigram models on a Wall-street Journal (WSJ) corpus was considered. Their findings were considered because it provides results over a reasonable range of sentences (from 100 sentences to over a million). Their findings show that for KN

[44] This number is actually a function of the training set. In the case study, 5 is used. Other data types might require less.
[45] This number is actually a function of the number of named entities in the training set. In the case study, 15 was used. Other data types might require less.

models, cutoffs lead to a loss in performance as seen in figure 20[46], where cross-entropy[47] is greater for 0-1-1 and 0-0-2 ngrams and increases with increase in corpus size up to 100000 sentences (see figure 20). To this end, 0-0-1 cutoffs were applied in this implementation.

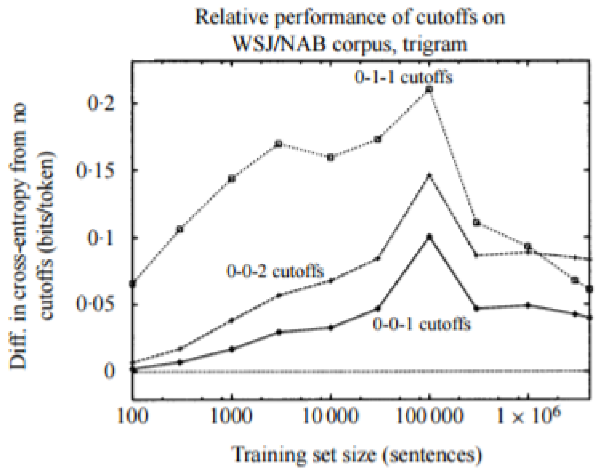


Figure 20: Comparing performance of trigram models with KN smoothing with cutoffs and KN smoothing without cutoffs (Source: Chen and Goodman, 1998).

Following smoothing, cutoffs, determination of OOVs, the implementation of interpolated KN language model (LM_1) for each value holder's training corpus using the SRILM toolkit is described. This concludes the implementation of LM_1.

6.6 Estimating Content Word Probability

Although the interpolated KN model implemented in the previous section can estimate the probability of content words, the value model proposes that such an estimation fails to encapsulate the necessary parameters that make up the value. This is because the KN model restricts itself to a very small history, does not reflect the relationship between the content words in the sentence nor the additional semantic properties that reflect the purpose of the content word.

To elucidate, consider the sentence "*We will not accept the actions of the EU*". For this sentence, LM_1 would estimate the probability of the content word '*EU*' from the bigram '*of the*' i.e. $p(EU|of, the)$. However, the value model suggests that while the trigram probability is a valid feature in estimating the likelihood of '*EU*', it alone, is insufficient. For the model to truly be a value model, it must reflect the semantic properties and function of each content words while capturing syntactic, lexical and sentiment features. For instance, it must reflect the association between the content word '*EU*' and other content words '*accept*' and '*actions*', the fact that the nominal head '*the EU*' modifies the phrase '*the actions*' giving the estimate $p(EU|EU\ nominal\ modifier\ of\ accept)$. In addition, as a value

[46] The terminology 0-0-1 for a trigram model means unigrams with 0 or fewer counts are ignored, bigrams with 0 or fewer counts are ignored and trigrams with 1 or fewer counts are ignored.
[47] Cross entropy, $H = log_2\ Perplexity$

model it must also reflect the positioning of each content word, for instance, '*accept*' and '*actions*' precede the word '*EU*' in the sentence. The VM must also reflect the nature of the content word capturing some of its syntactic and semantic features like its POS e.g. is it a noun $p(EU|noun)$, is it capitalized $p(EU|EU\ is\ capitalized)$, is it the root of the sentence $p(EU|root)$, is it connected to a negation $p(EU|EU\ is\ connected\ to\ a\ negation)$. Thus, $p(EU)$ is computed as a product of a set of conditional probabilities which incorporate the relationship between the word '*EU*' and other content words, the syntactic and grammatical positioning of the word, its history and finally its relationship to sentiment bearing words. $P(EU)$ is estimated by taking the product of these conditional probabilities. It is observed that the histories or conditions required in estimating content word probabilities is quite extensive. That is, by capturing such conditional estimates, properties which reflect the expressed value are garnered. However, incorporating all these features in the interpolated LM results in a formulation that becomes unwieldy[48]. To resolve this problem, log linear models are adopted as they are renowned for their robustness and ability to incorporate large feature sets.

This section discusses the implementation of a log linear distribution as well as the generation of the feature set for each content word. Ultimately, it shows how $p(w_i|h,H,C,A,S,\theta)$ is estimated for any content word w_i. The log linear distribution implemented here represents LM_2, which is applied in estimating content word probability. In describing this implementation, two input parameters are required: training sentences and a vocabulary of content words. In the next section, the application of these parameters in building the maxent model for content word probability estimation is described. First, maxent log linear models are introduced, before delving into content word feature selection, followed by a description of the weight learning process for the features. All these descriptions culminate in the estimation of the content word probability in the form of an implemented maxent model.

6.6.1 Maximum Entropy (Maxent) Model for Estimating Content Word Probability

The objective of this section can be described as follows: what is the probability of a content word (c) given a set of observed features $x_1 \dots x_n$ i.e. $p(c|x_{1\dots n})$ where n is an integer and $n > 1$. As a precursor to this section, maxent models are briefly described.

Maximum entropy or multinomial logistic regression classifiers are a popular multi-class probabilistic classifier that has been applied in many aspects of speech and language processing (Berger et al, 1996; Della Pietra et al, 1997; Ratnaparkhi, 1996). They are known to be quite flexible towards the addition of new features, scalable and parameter estimation

---

48 Imagine that 10 features exist such that the estimation of a content word w is given as $p(w|1,2,3,4,5,6,7,8,9,10)$. In a trigram model this becomes $\lambda_1 P(w|1,2) + \lambda_2 P(w|2,3) + \lambda_3 P(w|3,4) + \lambda_4 P(w|4,5) + \dots$. Estimating these λ can become extremely complicated especially for large feature sets.

is relatively easy (Audhkhasi et al, 2012). In addition, they have been shown to perform quite well in document classification tasks (Nigam et al, 1999). For the purposes of this research maxent models are used in estimating content word probability primarily because it allows for the expression of a richer feature set thereby resulting in a richer representation of the classes (content words) to be modelled. It does this by linearly combining a set of relevant features associated with a discrete set of classes and producing a probability distribution across all classes.

For instance, given a content word (c) in a sentence: A feature (x_1) for recognizing c might be that it is the root of a sentence. Other features could include the fact that it is preceded by a bigram w_{1-2}, w_{1-1} ($feature\ x_2$), that it modifies a noun ($feature\ x_3$), that it is preceded by a negation ($feature\ x_4$) etc. Each feature is associated with some weight (w), so that given each feature and weight, the maxent model estimates the probability of the content word by taking the dot product of w and the value of x. In Natural Language Processing (NLP), features usually take a binary value of 0 or 1 and so they are called indicator function. With this brief explanation, the requisite parameters for maxent content word estimation are described.

Definition: A maxent model for the estimation of content words consists of the following components:

- A set X of observed input features i.e. $X = \{x_1, \ldots, x_N\}$ where $N > 1$.
- A finite set C of content words C = $\{c_1, \ldots, c_n\}$ where n is the size of the content word vocabulary.
- A positive integer N specifying the number of input features in the model
- An indicator function $f: X\ x\ C \rightarrow \mathbf{R}^N$ which maps any (x, c) pair to a feature vector $f(x, c)$. Each indicator function takes any (x, c) pair and maps it to a real value, either 0 or 1, such that $f(x, c) \in \{0,1\}$. These features are concatenated to produce a feature vector.

Thus, for any $x \in X, c \in C$, the model estimates a conditional probability,

$$p(c|x) = \frac{exp(\sum_{i=1}^{N} w_{ci} f_i(c,x))}{\sum_{c' \in C} exp(\sum_{i=1}^{N} w_{c'i} f_i(c',x))} \ldots (12)$$

Where

$f_i(c, x)$ is the indicator function for the i^{th} feature of a class c for a given observation x. w_{ci} is the weight associated with the i^{th} feature of class c.

Given the maxent equation, the implementation task now involves: the selection of features, learning feature weights and finally computing content word probability estimations using equation 12.

6.6.2 Content Word Feature Selection

The goal is to generate set of relevant features which encode information enabling content word prediction. An ideal feature would be one that encapsulates the uniqueness of each content word, its context and semantic relationship with other content words around it. Therefore, the selected feature set comprised of two classes, the first captures the word history and the second captures the relationships between content words – i.e. word priority, dependency relationships and any sentiment clues associated with the content words.

1. **Features of Word History:** These features consist of the set of all the distinct trigrams seen in the data i.e. one trigram feature for every content word's trigram seen in the training data. That is for all trigrams (p, q, r), seen in training, a feature,

$$f_{N(p,q,r)}(x, c) = \prod_{i=1}^{n} \begin{cases} 1 & if\ c = r\ and\ w_{i-2} = p, w_{i-1} = q \\ 0 & otherwise \end{cases}$$

   is created where,

   N is the number of unique trigrams,

   $N(p, q, r)$ is a function that maps each (p, q, r) trigram to a unique integer,

   r is a content word.

   However, this feature set only captures the backward history of the content word i.e. w_{i-2}, w_{i-1}, and so additional contextual history is included in the form of forward trigrams, where for the content word w_i, the contexts (w_i, w_{i+1}, w_{i+2}) is captured. In doing this, the features would be made up of all observed content word backward trigrams and all content word forward trigrams. It is important to note that unseen trigrams are not included because they would give rise to too many features. Secondly, because such trigrams are unseen there is no basis for estimating them in the model.

2. **Features of Content Word Relationships**: Having created a content word vocabulary from training data and parsed all the training data with the SUD parser, the dependency relationships for every content word in the sentence can be identified (see figure 18). Each observed relationship is converted to a feature. Features are expressed through an if-else condition template: $\begin{cases} 1\ if\ 'condition' \\ 0\ otherwise \end{cases}$, which means if the '*condition*' is satisfied the function returns a value of 1, else it returns 0. A sample feature would be - if the content word w_i is the $root$ of a sentence, then the indicator function returns 1 else 0. Later, additional examples of these feature types are provided.

Summarily, three features types are provided for a content word: the set of all observed forward and backward trigrams which attempts to capture lexical and syntactic clues associated with each content word. The third feature is the set of all observed dependency relations encapsulating the relationships between the content words. However, during implementation it was observed that for training sets with over 10000 sentences the number of unique forward and backward content word trigram features were in the range of two to four hundreds of thousands and while this is not uncommon for log-linear models, it presented 2 problems.

The first arose from the fact that the number of features was almost equal and in some cases larger than the number of observations thus the potential risk of overfitting the loglinear model. There are mathematical approaches for addressing this issue that are beyond the scope of this research. The second issue was the absence of the computational power required in training models with this feature size. To ameliorate this, the set of features was reduced. Instead of making all observed forward and backward trigrams features, a smaller feature set that captured peculiar semantic clues ranging from the syntactic structure of the content word, its history, its relationship with other words and its meaning was developed. A justification for this alternative feature set is found in the methodologies reviewed in chapter 2, which showed that linguistic features such as POS and discourse patterns augmented with social information improved the performance of sentiment prediction models. Thus, a set of features encapsulating the following categories were generated:

- **The nature and character of the content word** – This feature encapsulates the physical properties of the content word as it appears in the training sentence. For instance, the first letter of a content word could be capitalized in one sentence, while in another it could be all lowercase. These content word characteristics are clues describing the function it performs in the sentence. Features used include: the POS of the word, if it is capitalized or in lowercase, if the word is alphanumeric. For instance, feature 1 in equation 13 returns a value of 1 if the word is a verb past tense and 0 if it is not. Equation 14 returns a value of 1 if the content word is completely capitalized so that acronyms can be differentiated from regular words. The total number of features is 24. See table 9 for a list of features associated with the nature and character of the content word.

Feature 1 - $f_1(x,c) = \begin{cases} 1\ if\ POS\ of\ c = VBD \\ 0\ otherwise \end{cases}$ ... (13)

Feature 2 - $f_2(x,c) = \begin{cases} 1\ if\ c\ is\ all\ uppercase \\ 0\ otherwise \end{cases}$ ... (14)

Feature 3 - $f_3(x,c) = \begin{cases} 1\ if\ c\ ends\ with\ 'fy' \\ 0\ otherwise \end{cases}$ ... (15)

Feature 4 - $f_4(x,c) = \begin{cases} 1\ if\ c\ starts\ with\ 'mid' \\ 0\ otherwise \end{cases}$ ... (16)

Feature 5 - $f_5(x,c) = \begin{cases} 1\ if\ c\ is\ converted\ to\ base\ form \\ 0\ otherwise \end{cases}$ ... (17)

Table 9: List of features associated with the nature and character of the content word

Content word (c) Features (Return 1 if feature is satisfied else return 0)	
If POS of c is NN	If POS of c is JJ
If POS of c is NNP	If POS of c is JJR
If POS of c is NNPS	If POS of c is JJS
If POS of c is NNS	If POS of c is RB
If POS of c is NP	If POS of c is RBR
If POS of c is NPS	If POS of c is RBS
If POS of c is VB	If POS of c is MD
If POS of c is VBZ	If c starts with uppercase
If POS of c is VBP	If c is all uppercase
If POS of c is VBD	If c is alpha numeric
If POS of c is VBG	If c contains punctuation
If POS of c is VBN	If c is converted to its base form

- **The context of the content word** –This feature captures contextual history associated with the content word. Instead of making each observed content word trigram a feature, features were developed using the POS of words in the backward trigram ($POS_{i-2}, POS_{i-1}, w_i$) and forward trigram ($w_i, POS_{i+1}, POS_{i+2}$) of the content word w_i. Table 10 shows a description of these features. An example of one of the features is 'if w_{i+1} is a noun' and the reverse feature 'if w_{i-1} is a noun'. In addition to using POS features of the surrounding words, features relating to the position of the word in the sentence were also included. With this approach, the contextual feature set was reduced from potentially millions to a handful (70 features). Given considerable computational power, potential, future work would explore and compare the implementation of forward and backward context word trigram features.

Table 10: List of features associated with the content word context

Content word (c) Features (Return 1 if feature is satisfied else return 0)	
If c is first word in sentence	If POS of w_{i+2} is conjunction
If c is last word in sentence	If POS of w_{i-2} is conjunction
If POS of w_{i+1} is preposition	If POS of w_{i+2} is pronoun
If POS of w_{i-1} is preposition	If POS of w_{i-2} is pronoun
If POS of w_{i+1} is interjection	If POS of w_{i+2} is WH determiner
If POS of w_{i-1} is interjection	If POS of w_{i-2} is WH determiner
If POS of w_{i+1} is adjective	If POS of w_{i+2} is monetary text
If POS of w_{i-1} is adjective	If POS of w_{i-2} is monetary text

If POS of w_{i+1} is noun	If w_{i+2} is numeric
If POS of w_{i-1} is noun	If w_{i-2} is numeric
If POS of w_{i+1} is adverb	If w_{i+2} is date
If POS of w_{i-1} is adverb	If w_{i-2} is date
If POS of w_{i+1} is TO	If w_{i+2} is measure
If POS of w_{i-1} is TO	If w_{i-2} is measure
If POS of w_{i+1} is modal	if w_{i+1} is a function word
If POS of w_{i-1} is modal	if w_{i-1} is a function word
If POS of w_{i-1} is verb	if w_{i+2} is a function word
If POS of w_{i+1} is verb	if w_{i-2} is a function word
If POS of w_{i+1} is conjunction	If POS of w_{i-2} is modal
If POS of w_{i-1} is conjunction	If POS of w_{i+2} is modal
If POS of w_{i-1} is numeric	If POS of w_{i-2} is verb
If POS of w_{i-1} is pronoun	If POS of w_{i+2} is verb
If POS of w_{i+1} is WH determiner	If POS of w_{i-2} is TO
If POS of w_{i-1} is WH determiner	If POS of w_{i+2} is TO
If POS of w_{i+1} is monetary text	If POS of w_{i-2} is adverb
If POS of w_{i-1} is monetary text	If POS of w_{i+2} is adverb
If POS of w_{i+1} is pronoun	If POS of w_{i-2} is noun
If POS of w_{i+1} is numeric	If POS of w_{i+2} is noun
if w_{i+1} is a content word	If POS of w_{i-2} is adjective
if w_{i-1} is a content word	If POS of w_{i+2} is adjective
if w_{i-2} is a content word	If POS of w_{i-2} is interjection
if w_{i+2} is a content word	If POS of w_{i+2} is interjection
If w_{i+1} is date	If POS of w_{i-2} is preposition
If w_{i-1} is date	If POS of w_{i+2} is preposition
If w_{i+1} is a measure	If w_{i-1} is a measure

- **Dependency relations which capture word relations** – This feature is designed to capture the semantic relevance and role of a content word in the sentence. This is accomplished by deriving features from the DG relations between a content word and linguistic units within its vicinity. Earlier in this implementation (section 6.2) DGs were used to identify content words and the result of the parse process included a compilation of JSON objects representing the relationships between content words (‘`listOfObjects`’ array in figure 18). From these JSON objects, extracted content words (governor and dependent) and relationships, were converted into a feature. Sample features include: if the word is the root of the sentence, if the word is an adjectival modifier of another content word, if the word is the subject of the sentence, if the word belongs to a compound relation etc. Discourse relations which exists between content word and valence shifters are also captured. Valence shifters are linguistic units which tend to alter the semantic orientation of the term they are referring to (Polanyi & Zaenen, 2006; Musat & Trausan-Matu, 2010). In this section, the referenced valence shifters are connectors. Connectors are conjunctive words which link similar elements in a sentence. They include words such as *'and', 'or', 'but', 'however', 'although', 'moreover', 'therefore'* etc. Dependency features such as '*cc*', '*conj*' and '*mark*' capture relationships between content words and such connectors (see table 8). Therefore, the following sample features are included in the feature set illustrated in table 11: *if c is in a dependency relation* $cc(c, y)$, *if c is in a dependency relation* $conj(c, y)$ *or* $conj(y, c)$ *where y is a content word*, *if c is in a dependency relation* $mark(c, y)$. Other sentimental valence shifters like negations are featured in a later part of this section. Additional semantic features which capture named entities are also included. Such features ask questions such as 'is the content word a location', 'is the content word an organization' and 'is the content word a person'. In total, 52 features, are used in capturing dependency relationships between content words in the sentence.

Table 11: List of Dependency Relation and Named Entity Features

Content word (c) Features (Return 1 if feature is satisfied else return 0)
If c is in a dependency relation R where R is a compound relation
If c is the root word of the sentence dependency graph
If c is the object of the root word of the sentence dependency graph
If c is the nominal subject of the root word of the sentence[49]
If c is the clausal subject of the root word of the sentence[50]
If c is the passive clausal subject of the root word of the sentence[51]
If c is the passive nominal subject of the root word of the sentence[52]
If c is a clausal component of a verb[53]
If c is a governor verb in a clausal component relationship with a dependent[54]
if c is in a dependency relation $dobj(y, c)$ where c is a verb and y is not the root word
if c is in a dependency relation $nsubj(y, c)$ where c is a verb and y is not the root word
if c is in a dependency relation $dobj(c, y)$ where c is not the root
if c is in a dependency relation $nsubj(c, y)$ where c is not the root
if c is in a dependency relation $aux(c, y)$ where y is an auxiliary
if c is in a dependency relation $det(c, y)$ where y is a determiner
if c is in a dependency relation $case(c, y)$ where y is a clitic

[49] This can also be rewritten as, if the relationship between c and the root is nsubj or nsubj(root,c)
[50] This can also be rewritten as, if the relationship between c and the root is csubj or csubj(root,c)
[51] This can also be rewritten as, if the relationship between c and the root is csubjpass or csubjpass(root,c)
[52] This can also be rewritten as, if the relationship between c and the root is nsubjpass or nsubjpass(root,c)
[53] The word c is the dependent of a verb (y) where the relationship is xcomp or ccomp or xcomp(y,c)
[54] xcomp(c,y) or ccomp(c,y) where y is a dependent content word

if c is in a dependency relation $amod(c, y)$ where c is a verb and y POS JJ
if c is in a dependency relation $amod(c, y)$ where c is a noun and y POS JJ
if c is in a dependency relation $amod(c, y)$ where c is a noun y is not an adjective
if c is in a dependency relation $amod(c, y)$ where c is a verb y is not an adjective
if c is in a dependency relation $amod(c, y)$ where c is not a noun or a verb, y POS JJ
if c is in a dependency relation $amod(c, y)$ where c is a verb and y POS JJR
if c is in a dependency relation $amod(c, y)$ where c is a noun and y POS JJR
if c is in a dependency relation $amod(c, y)$ where c is not a noun or a verb, y POS JJR
if c is in a dependency relation $amod(c, y)$ where c is a verb and y POS JJS
if c is in a dependency relation $amod(c, y)$ where c is a noun and y POS JJS
if c is in a dependency relation $amod(c, y)$ where c is not a noun or a verb, y POS JJS
if c is in a dependency relation $amod(y, c)$ where y is a noun and c is the dependent
if c is in a dependency relation $amod(y, c)$ where y is a verb and c is the dependent
if c is in a dependency relation $advmod(c, y)$ where c is a verb[55] with POS VB
if c is in a dependency relation $advmod(c, y)$ where c is a verb with POS VBD
if c is in a dependency relation $advmod(c, y)$ where c is a verb with POS VBG
if c is in a dependency relation $advmod(c, y)$ where c is a verb with POS VBN
if c is in a dependency relation $advmod(c, y)$ where c is a verb with POS VBZ
if c is in a dependency relation $advmod(c, y)$ where c is a verb with POS VBP
if c is in a dependency relation $advmod(c, y)$ where c is not a verb
if c is in a dependency relation $advmod(y, c)$ where c is an adverb with POS RB
if c is in a dependency relation $advmod(y, c)$ where c is an adverb with POS RBR
if c is in a dependency relation $advmod(y, c)$ where c is an adverb with POS RBS
if c is in a dependency relation $nmod(c, y)$
if c is in a dependency relation $nmod(y, c)$
if c is in a dependency relation $acl(c, y)$
if c is in a dependency relation $acl(y, c)$
if c is in a dependency relation $advcl(c, y)$
if c is in a dependency relation $advcl(y, c)$
if c is in a dependency relation $cop(c, y)$
if c is in a dependency relation $cc(c, y)$
if c is in a dependency relation $conj(c, y)$ or $conj(y, c)$ where y is a content word
if c is in a dependency relation $mark(c, y)$
if c is a location
if c is an Organization
if c is the name of a person

- **The meaning of a Word from its Prefix and Suffix** – In English language, prefixes and suffices connote particular meanings. For instance, words ending with *'-able'* like *'excitable'*, *'portable'*, connote ability. Words ending with the suffix *'-age'*

[55] In this feature c is modified by an adverb

like *'voyage'*, *'pilgrimage'* connote process or action. As for prefixes, words starting with *'ante-'* like *'antenatal'*, *'antecedent'*, connote before or prior while words starting with *'anti-'* like *'antislavery'*, *'antidepressant'*, *'antibiotic'* connote opposition or standing against something. Due to the semantic relevance associated with prefixes and suffices, features for words containing meaningful prefixes and suffixes are introduced. A list of prefixes, suffices and their meanings are compiled from online resources[56]. In using this feature, an attempt is made at capturing a property of the word's inherent meaning. For example, equation 15 returns a value of 1 if the word ends with *'-fy'* (words ending with *-fy* connote *'to make'* or *'become'* or *'to cause'* e.g. justify, amplify) while equation 16 returns a value of 1 if the word begins with *'mid-'* (words beginning with *mid-* connote middle e.g. *'midsummer'*, *'midway'*). In this implementation, not all prefixes and suffixes available in the online resources were found in training sentences. In fact, some prefixes and suffices were more commonly used than others, and it was concluded that the prevalence of some prefixes and suffices over others was a function of the text or corpus domain. Table 12 illustrates a list of the prevalent prefix and suffix features used. It must be said that for a different domain such as medical or legal, the prefix and suffix features are likely to be different. Based on this, it was surmised that although this feature type has certain recurring generic prefixes and suffices it has specific prefixes and suffices that are dependent on the corpus and domain of interest. The total number of features is 88.

Table 12: List of prefix and suffix features

Content word (c) Features (Return 1 if feature is satisfied else return 0)	
if c starts with $ante -$	if c ends with $-ic$
if c starts with $anti -$	if c starts with $im -$
if c starts with $circum -$	if c ends with $-y$
if c starts with $co -$	if c ends with $-ling$
if c starts with $de -$	if c ends with $-ular$
if c starts with $dis -$	if c ends with $-ure$
if c starts with $em -$	if c ends with $-tude$
if c starts with $hyper -$	if c ends with $-pathy$
if c starts with $homo -$	if c ends with $-ous$
if c starts with $fore -$	if c ends with $-th$
if c starts with $extra -$	if c ends with $-ing$
if c starts with $ex -$	if c ends with $-ile$
if c starts with $il -$	if c ends with $-ious$ or $-ous$
if c ends with $-hood$	if c ends with $-ish$

[56] For a list of prefixes and suffices see https://www.myenglishteacher.eu/blog/prefixes-suffixes-list/, http://www.bbc.co.uk/skillswise/factsheet/en17suff-l1-f-what-is-a-suffix, https://www.learnthat.org/pages/view/suffix.html (Last accessed on 28/01/2017)

if c ends with $-tic$	if c ends with $-ive$
if c ends with $-est$	if c ends with $-less$
if c ends with $-ess$	if c ends with $-ty$
if c ends with $-ese$	if c ends with $-ly$
if c ends with $-ency$	if c ends with $-or$
if c ends with $-ee$	if c ends with $-ory$
if c ends with $-ade$	if c ends with $-ward$ or $-wards$
if c ends with $-cracy$	if c starts with $en-$
if c ends with $-ac$	if c ends with $-ful$
if c ends with $-able$ or $-ible$	if c ends with $-ence$
if c starts with $uni-$	if c ends with $-ise\ or\ -ize$
if c starts with $un-$	if c ends with $-ify$ or $-fy$
if c starts with $inter-$	if c ends with $-sion$ or $-tion$ or $-ation$
if c starts with $micro-$	if c ends with $-ship$
if c starts with $mid-$	if c ends with $-ness$
if c starts with $mis-$	if c ends with $-ment$
if c starts with $non-$	if c ends with $-ity$ or $-ty$
if c starts with $para-$	if c ends with $-ist$
if c starts with $post-$	if c ends with $-ism$
if c starts with $pre-$	if c ends with $-er$
if c starts with $re-$	if c ends with $-en$
if c starts with $semi-$	if c ends with $-ed$
if c starts with $super-$	if c ends with $-dom$
if c starts with $sub-$	if c ends with $-cy$
if c ends with $-ion$	if c ends with $-cide$
if c ends with $-ness$	if c ends with $-ate$
if c starts with $in-$	if c ends with $-ant$
if c starts with $ir-$	if c ends with $-ance$
if c ends with $-ical$	if c ends with $-an$ or $-ian$
if c is a disfluency	if c ends with $-al$

- **Clues of Sentiment** – Since the ultimate objective of this approach is to capture sentiment, features that influence or sway the sentiment of a content word are also introduced. Negations like the expression *'not'* is a sentiment feature that has been shown in previous work to be good valence shifters of content words (Das and Chen, 2001; Pang, Lee and Vaithyanathan 2002; Polanyi & Zaenen, 2006). Other simple negations like *'no', 'never', 'neither', 'nobody', 'nowhere', 'nothing'* which belong to diverse word classes have also been shown to be good indicators of sentiment expression (Polanyi and Zaenen, 2006). Additionally, some words are generally seen to be negative and their presence in sentences, will typically suggest the expression of some sort of negativity. For instance, the word *'unsuccessful'* is generally considered to be negative. In summary, the sentiment features captured include:

- *'if the word has prior negative polarity'*. MPQA subjectivity lexicon of Wilson et al (2005) is used, so that content words with negative prior polarity are assigned a value of 1. Similar features are also included to cover cases of the word having prior neutral polarity and prior positive polarity. In this case study implementation it is discovered that a significant proportion of the content words have neutral prior polarity e.g. *'Immigrant'*, *'People'*. However, their polarity can be altered by their proximity to a negation or having a dependency relation with a word with prior negative polarity. To this end, additional features outlined below are introduced.

- *'if the word is in a relationship with a word with prior negative polarity or prior positive polarity or prior neutral polarity'*. That is if the word w is in a dependency relationship with a word y, where y's prior polarity is negative.

- *'if the word is a negation'* e.g. *'never'*, *'not'* or if the word starts with a suffix or prefix that connotes negativity such as *'anti-'*, *'non-'*, *'in-'*, *'dis-'*, or *'ill-'*.

- *'if the word is in a dependency relationship with a negation'*

Table 13 itemizes the list of sentiment features.

The total number of features used in building the maxent model is 241.

Table 13: List of Sentiment Features

Content word ($\boldsymbol{c}$) Features (Return 1 if feature is satisfied else return 0)
if c is in a dependency relation $neg(c, y)$ where y is a negation
if c has prior negative polarity or word is a negation or word connotes negativity
if c has prior positive polarity
if c has prior neutral polarity
if c is in a relation $\boldsymbol{R}$ with a word with negative connotation
if c is strongly subjective
if c is weakly subjective

Although the class of features selected encapsulate the implicit properties of the value model described, an associated research limitation lies in the absence of additional tests to determine which set of features would provide the best model. In fact, theoretically, the likely set of feature classes is almost inexhaustible. The response to this limitation is that, since part of the research objective was to build a model without human annotation, the feature set generated had to be organic. In addition, without any human annotated data, it was necessary to implement the model with some baseline feature sets. This means that future work can be devoted towards engineering and tweaking these baseline feature sets to obtain the optimum features.

6.6.3 Capturing Alternate Word forms and Meanings

Before describing how feature weights are learnt a problem relating to the occurrence of multiple word forms encountered during the compilation of content words is described. An observation made after the compilation of content words was the existence of multiple word forms with the same meaning. For instance, in the implementation (discussed in chapter 7), the content word vocabulary contained the word *'accompany'* and other inflectional morphemes - *'accompanied'*, *'accompanying'*, *'accompanied'*. The dilemma here revolved around the choice of treating the morphemes as distinct linguistic units even though they invoked the same meaning and intent. Alternatively, to treat them as the same word by converting all morphemes to the base form of the word. For the latter case, this would have meant reducing every inflected word morpheme to its base form. For instance, the sentence *"We accompanied the Government"* becomes *"We accompany the Government"*. The problem with this was that the resulting sentence would lose its syntactic correctness. To this end, the goal was to implement the capture of the word's syntactic function in the sentence while maintaining its semantic intent. To illustrate this, consider the following scenario, involving a value holder who makes the following comments:

Sentence 1: "*We will fight hard to stay in the EU*"

Sentence 2: "*We have fought to stay in the EU*"

Sentence 3: *"We must leave the EU"*

Assuming all the words in the sentence represent a bag of words and all the sentences constitute a corpus. The total number of words N in this sample corpus is 22[57]. If the words *'fight'*, *'fought'* and *'leave'* are treated as independent words in the sentences then, the probability of picking the word *'fight'* would be the same as the probability of picking the word *'fought'* or the word *'leave'*. That is,

$$p(fight) = p(leave) = p(fought) = \frac{1}{22}$$

However, since word meanings are important, then clearly *'fight'* and *'fought'* have the same meaning, albeit different forms, so the actual likelihood of picking a word that means *'to fight'* should be $\frac{2}{22} = \frac{1}{11}$ since *'fight'* and *'fought'* have the same meaning. As such, by recognizing that *'fight'* and *'fought'* have the same meaning but different word forms, the probabilities can be re-estimated to conclude that the speaker is twice more likely to *'fight to stay in the EU'* instead of *'leaving the EU'*.

---

[57] Count of words in sentence 1 = 9, Count of words in sentence 2 = 8, Count of words in sentence 3 = 5

Therefore, the goal was to model each inflectional content word in such a way that capture its syntactic form and meaning through its root or base form was captured. To this end, given a content word c, such that c is not in its base form, the feature vector for c is represented as follows:

1. c is converted to its base form c' so that the class or outcome of the feature vector is c' not c. This way the base form of the word is captured.

2. The indicator function for the feature representing the suffix of c e.g. ends with -*ing*, ends with -*s*, ends with -*ed*, returns a value of 1. This way the original content word form and some semantics from its suffix are captured.

3. A value of 1 is returned for the feature in equation 17 (or the last feature in table 9), essentially representing that the base form c' was not the original word in the sentence. The outcome of all other features categories is determined using c.

The next section, describes how the weights are learned for the features.

6.6.4 Weight Learning

Having derived feature vectors for all content word, the weights w_{ci} is estimated in the maxent equation $p(c|x) = \frac{exp(\sum_{i=1}^{N} w_{ci} f_i(c,x))}{\sum_{c' \in C} exp(\sum_{i=1}^{N} w_{c'i} f_i(c',x))}$.

Typically, this involves estimating weights that maximize the log likelihood of the training data using some convex optimization algorithm whose detail is beyond the scope of this research. Thankfully, there was no need to reinvent the wheel since several software implementations of this process already exist. In implementing the maxent model, the R[58] maxent implementation of Jurka (2012) was used as it is designed to reduce the memory consumption required in estimating weights by using three state of the art parameter estimation techniques. The maxent implementation also provides a function for tuning the model by estimating new weights using different regularization parameters. It is also quite fast.

6.7 Maxent Implementation

Following the identification of features, this section describes the implementation of the maxent model. For every content word instance observed in the training set, a feature vector is computed from all the feature templates discussed in the previous section. Thus, for each value holder's training data, a matrix X of features and a vector Y of content words are obtained. Since the feature size is 241, and assuming the number of content word instances seen in training is n, a $[n, 241]$ matrix is obtained, while the vector Y is of size n. This is expressed as:

---

[58] https://www.r-project.org/: Last accessed 28/01/2017

$$\begin{bmatrix} x_{1,1} & \cdots & x_{1,241} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \cdots & x_{n,241} \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

where n is the number of the content word instances. Part of the task involved in building the model is estimating the weights (or coefficients) $w_1 \dots w_{241}$ for each feature (the process was described in section 6.6.4). The matrix X turns out to be a sparse matrix, where most of the elements are equal to 0.

In implementing the model, the coefficients $w_1 \dots w_{241}$, were derived using a slightly modified version of Jurka's maxent implementation. This modification removed the limit on the maximum number of features that can be used in the model. This original limit was set at 255 features and according to Jurka (personal communication, August 2016), the model performs best with 255 features with gradual depreciations in performance when over 255 features are used. This performance depreciation according to Jurka is computational and does not impact the estimation of coefficients and so it was completely removed form the implementation[59]. With a shrunken feature set, all the features were applied in the implementation of the model and so L2[60] regularization was used instead of L1[61]. Consequently, the maxent tuning function (Jurka, 2012) was applied, varying the value of the L2 parameter and tuning with a 10-fold validation (The default is 3-fold). All other parameters are left as their default. The resulting model is applied in estimating content word probability.

6.8 Dealing with Unseen Content Words

Having implemented a model for content and function words, how is the probability of unseen content words estimated i.e. content words observed in testing but unobserved during training. In cases where the word is unseen, the probability of the word was estimated from the LMs estimation (LM_1) of 'UNK'. The justification for this is that since the word is unseen, it is assumed to be a rare word for which the only estimate of rare words available in the model is 'UNK' $p(UNK)$. Further research should explore estimating unseen word probability from observed synonyms or word substrings which resonate same meaning. For instance, if the word *'uncharacteristically'* is unseen in training, its probability can be estimated from the related word *'uncharacteristic'* as long as it is observed in training. With the implementation of the maxent model LM_2 which enables content words estimation, the next section shows how sentiments can be estimated from the two models.

---

[59] This modification was carried out to support the inclusion of additional features. As part of our tests carried out in chapter 7, we enhanced the model with additional corpus dependent semantic to observe if semantic enhancements improve our model's precision. This is detailed in chapter 7.

[60] L2 regularization shrinks the estimates by getting them as close to zero as possible but not making them zero.

[61] The effect of L1 regularization is that it forces some of the coefficients or feature weights to be exactly equal to zero and as such it also acts as a variable selector. This is not required in this implementation, since all the features are deemed to be important.

6.9 Sentiment Estimation – Value Sentence Estimation Algorithm (VSEA)

Given a test sentence sequence of words $W = \langle w_1, w_2, \ldots, w_n \rangle$ where n is the number of words in the sentence. In estimating the likelihood of the sentence, two distributions are applied. The first distribution, LM_1 estimates the probability of any function and unknown words in the sentence, while the second distribution LM_2 estimates the probability of content words. The estimation of sentence probability can be outlined as seen in algorithm 2.

Algorithm 2 Value Sentence Estimation Algorithm (VSEA)

Inputs

(i) $V = \{v_1, v_2, v_3, \ldots, v_N\}$ where V is the vocabulary of content words and N is the size of the vocabulary.

(ii) The Interpolated KN Smoothed LM model LM_1

(iii) Maxent model for content words LM_2

(iv) Sentence $W = \langle w_1, w_2, \ldots, w_n \rangle$

(v) Context $C = \{C_1, C_2, \ldots, C_k\}$, where k is the number of contexts

Result: Product (P)
Initialize: $P = 1$
Process

1. **for** each word, w_i, in $W = \langle w_1, w_2, \ldots, w_n \rangle$, where i is an integer **do**
   - (a) **Identify** the word type of $w_i (i = 1 \ldots n)$, i.e. is it a content word or a function word
   - (b) **If** w_i is a function word estimate $p_f = p(w_i | w_{i-2}, w_{i-1})$ from LM_1, **Compute** $P := P \times p_f$
   - (c) **Else if** w_i is a content word
     - (i) **If** $w_i \in V$ **do**
     - (ii) **Build** feature vector X_i for w_i using features $f_1 \ldots f_{241}$
     - (iii) **Feed** vector X_i into LM_2 to estimate the probability p_{w_i} of w_i i.e $p(w_i | f_1 \ldots f_{241})$
     - (iv) **Compute** $P := P \times p_{w_i}$
   - (d) **Else if** w_i is an unknown content word **do**
     - i. **Estimate** the probability p_{UNK} from the LM_1
     - ii. **Compute** $P := P \times p_{UNK}$
2. **Return** product P

In the VSEA algorithm, p_f refers to the probability of the function word, while p_{UNK} refers to the probability of unseen words 'UNK'. Features $f_1 \ldots f_{241}$ refers to all the content word features derived in section 6.6.2. The probability P in algorithm 2 answers the question how likely is it for a value holder (H) i.e. the recipient, under a context C_k to utter a sentence $W = \langle w_1, w_2, \ldots, w_n \rangle$. To determine the sentiment, a contrary question is asked: How likely is it for the same value holder to utter a sentence W' that expresses a sentiment that is opposite to the sentiment in W.

Concretely, if the estimate of the contradictory statement is greater than W, then it is inferred that the sentiment of the recipient is negative for W since H is more likely to express sentence W' - a sentence that has an opposite sentiment. By estimating the likelihood of making a contrary statement as compared to the given statement, sentiment is inferred. The question then is how can the sentence W' be constructed?

Since the abstract expression of values in the sentence revolves around the content words particularly the root, which based on DG theory conveys the meaning and intent of the sentence, an initial idea was to reconstruct the sentence W by positioning a negation *'not'* behind the root word. This solution was too naïve and had several flaws. The main flaw was that by including an extra word in W', the size of the word sequence is increased and consequently the number of probability estimates to be made for the sentence's words. For instance, if the sequence W contains 7 words, the probability of seven words is estimated. However, W' would have 8 words because of the inclusion of the negation and this would bias the results considerably. Another concern was that by positioning a negation behind the root, the sentence loses its linguistic and grammatical sense. For instance, if W is the sentence *"LD say yes to the EU"*, by appending a negation like *'not'*, *'no'* or *'never'* before the root *'say'*, the sentence becomes grammatically incorrect – *"LD not say yes to the EU"*.

An additional problem was, what if the sentence W, already contained a negation for instance, *"We say no to the EU"*. A logical solution would have been to simply remove the negation from the derived sentence W' if the sentence contained a negation. This again leads to the same problem in that the resulting sentence might not read well e.g. *"We say no to the EU"* becomes *"We say to the EU"*. Also, the number of words to be estimated for W' becomes $W - 1$.

Another explored solution was to substitute the root with its antonym. The problem with this approach was that the size of the set of likely antonyms though likely to be relatively small cannot be determined without first considering the word sense and its impact on the sentence's grammatical sense, in other words would the newly formulated sentences make sense. Secondly, what determines the choice of antonyms to select? In addition, the presence of multiple antonyms in the training set, presents the risk of bias where the most frequent antonyms are selected. Similarly, the added complication of not seeing the antonyms in training also exists? To circumvent these issues, an innovative solution based on altering the polarity of the features in table 13 called Feature Switching (FS) was implemented.

6.10 Feature Switching (FS)

Given a word in a sentence, the features in table 13 are designed to capture sentimental aspects observed in the sentence. For instance, consider the sentence: *"We will support the EU"*. The sentiment feature vector for the word *'support'* is expressed in table 14. To capture a sentiment opposite to that expressed for *'support'*, the sentiment feature for the word is switched. This is done by changing the value from '1' to '0' if the original sentence had a feature '0' and vice versa i.e. substitute the binary feature value for its compliment. Feature switching produces the effect of constructing features for an unseen imaginary sentence that is opposite in meaning to the original and this is illustrated in table 15[62].

In table 15, the model proposes that for the opposite sentiment of the word *'support'*: It is in a dependency relation with a negation, has prior negative polarity and is also a word with a negative connotation. The value of the feature f_4 is unaltered because the word cannot have more than one sentiment orientation. If the original value of f_4 was 1, that is the word had prior neutral polarity, then for the opposite feature vector, f_4 will remain 1, f_2 and f_3 will remain as 0. This is in line with the notion that the word can have only one polarity type at a time.

However, the value of f_1 and f_5 are altered to say that even though the word is neutral, it could also be a part of a neg relation as well as in relation with a word with a negative polarity. Conversely, table 16, illustrates a case where the original sentence was *"We will not support the EU"*, where *'support'* is in direct relation with a negation. Considering this the Algorithm for sentiment estimation for a sentence W is illustrated in algorithm 3 below.

Table 14: Sentiment Feature Vector for the content word support in 'We will support the EU'

Sentiment Features	value
if c is in a dependency relation $neg(c, y)$ where y is a negation (f_1)	0
if c has prior negative polarity or c is a negation or c connotes negativity (f_2)	0
if c has prior positive polarity (f_3)	1
if c has prior neutral polarity (f_4)	0
if c is in a relation $\boldsymbol{R}$ with a word with negative connotation (f_5)	0
if c is strongly subjective (f_6)	0
if c is weakly subjective (f_7)	1

Table 15: Comparing Sentiment Feature Vector for the content word support in 'We will support the EU'

Sentiment Features	Original features value	Opposite feature value
f_1	0	1
f_2	0	1
f_3	1	0
f_4	0	0
f_5	0	1
f_6	0	0
f_7	1	1

62 In table 14, the sentiment feature the verb 'support' has a weak subjective polarity

Table 16: Comparing Sentiment Feature Vector for the content word support in 'We will not support the EU'

Sentiment Features	Original features value	Opposite feature value
f_1	1	0
f_2	0	1
f_3	1	0
f_4	0	0
f_5	1	0
f_6	0	0
f_7	1	1

Algorithm 3 Recipient Sentiment Prediction using VSEA and Feature Switching

Inputs

(i) Sentence W

(ii) $V = \{v_1, v_2, v_3, \ldots, v_N\}$ where V is the vocabulary of content words and N is the size of the vocabulary.

(iii) Interpolated KN LM – LM_1

(iv) Maxent model LM_2

Process

1. **Apply** VSEA on W and estimate a probability P.
2. **Apply** VSEA on W' such that for all content words in W we assign opposite features and estimate P'.
3. **Determine** $max(P, P')$

Output

If $max(P, P') = p$, **infer** sentiment Ψ of W (Ψ_W) is positive

Else, infer sentiment Ψ of W (Ψ_W) is negative

For instances of double negation such as '*can't not support*', the presence of '*can't not*' is treated as a positive since two negations theoretically equate to a positive. As such the content word 'support' is assumed to not be in a relation with a negation i.e. feature $f_1 = 0$, for support while the opposite feature will be its complement 1.

Consider an example where the objective is to predict the sentiment of a value holder H over the statement: "*We will support the EU*" when the context is C_k. Given the root word '*support*', content words '*EU*' and '*will*', the probability of the sentence can be estimated as:

$$\boldsymbol{p_{(H,C_k)}(sentence)} = p_{(LM-H-C_k)}(we|\ *,*) \times p_{(maxent-H-C_k)}(will|f_{1\ldots241}) \times$$

$$p_{(maxent-H-C_k)}(support|f_{1\ldots241}) \times p_{(LM-H-C_k)}(the|will, support) \times$$

$$p_{(maxent-H-C_k)}(EU|f_{1\ldots241}) \times p(STOP|the, EU)$$

Where, $p_{(LM-H-C_k)}$ is the probability estimate of the holder's (H) LM for a context C_k and $p_{(maxent-H-C_k)}$ is the probability estimate of *H*'s maximum entropy distribution under context C_k.

Next, the probability of *H* making a statement with the opposite polarity is estimated as:

$$\boldsymbol{p_{(H,C_k)}(sentence)'} = p_{(LM-H-C_k)}(we|\ *,*) \times p_{(maxent-H-C_k)}\left(will\middle|f_{1\ldots241}{}'\right) \times$$

$$p_{(maxent-H-C_k)}\left(support\middle|f_{1\ldots241}{}'\right) \times p_{(LM-H-C_k)}(the|will, support) \times$$

$$p_{(maxent-H-C_k)}\left(EU\middle|f_{1\ldots241}{}'\right) \times p(STOP|the, EU)$$

Where $f_{1\ldots241}{}'$ connotes sentiment features that have been switched.

Finally, $p_{(H,C_k)}(sentence)$ and $p_{(H,C_k)}(sentence)'$ are compared.

If, $p_{(H,C_k)}(sentence)' > p_{(H,C_k)}(sentence)$ then it is inferred that the speaker is less likely to make the original statement and hence the sentiment is negative. If, however, $p_{(H,C_k)}(sentence)' < p_{(H,C_k)}(sentence)$ then it is inferred that the speaker is more likely to make the original statement and hence the sentiment is positive. In addition, the effect of value fields can be portrayed by asking the question, what will be the sentiment of the same value holder H on the same utterance if the context was C_j. Based on value field theory, this is estimated by applying *H*'s LM and maxent models built under context C_j. That is:

$$\boldsymbol{p_{(H,C_j)}(sentence)} = p_{(LM-H-C_j)}(we|\ *,*) \times p_{(maxent-H-C_j)}(will|f_{1\ldots241}) \times$$

$$p_{(maxent-H-C_j)}(support|f_{1\ldots241}) \times p_{(LM-H-C_j)}(the|will, support) \times$$

$$p_{(maxent-H-C_j)}(EU|f_{1\ldots241}) \times p(STOP|the, EU)$$

And,

$$\boldsymbol{p_{(H,C_j)}(sentence)'} = p_{(LM-H-C_j)}(we|\ *,*) \times p_{(maxent-H-C_j)}\left(will\middle|f_{1\ldots241}{}'\right) \times$$

$$p_{(maxent-H-C_j)}\left(support\middle|f_{1\ldots241}{}'\right) \times p_{(LM-H-C_j)}(the|will, support) \times$$

$$p_{(maxent-H-C_j)}\left(EU\middle|f_{1\ldots241}{}'\right) \times p(STOP|the, EU)$$

This concludes the description of the model's implementation.

6.11 Conclusion

In conclusion, a detailed description of the value language model and its application to recipient sentiment prediction has been described.

The value model has been shown to be two language models: One for predicting or estimating function words while the second estimates content word probability. It has also been shown that the LM incorporates all the vital characteristics representative of a mapping from abstract values to observed text. The use of DGs in capturing and relating priority has also been illustrated. In addition, it shows how features which capture both semantic and syntactic properties of content words are captured and incorporated into the value model.

Also described was an innovative approach called feature switching which embodies the creation of a sentence that is contrary to the sentence presented to the recipient. As for sentiment prediction, it can be estimated as a function of two probability estimates which depict opposing and contrary sentiments. Ultimately, three artifacts are implemented in this section: Two minor artifacts (an algorithm and a method) consisting of the VSEA algorithm and feature switching method. These two artifacts are combined in implementing the core artifact of this thesis which is the Recipient Sentiment Prediction Algorithm (RSPA) seen in algorithm 3. Finally, the implemented approach is completely devoid of human input or annotations and is readily applicable to any document space or domain. In the next chapter, an implementation carried out on a political corpus is described.

7. Implementation of Sentiment Prediction for UK Political Data

In DSR, the evaluation of an artifact includes the integration of the artifact within a technical environment or use-case of a business environment (Hevner et al, 2004). As such, this chapter discusses the implementation of VSM on data from the political sector culminating in the implementation of several VSEA algorithms which are subsequently applied in testing the RSPA. Thus, it acts as a precursor to the next chapter which discusses the tests carried out on the built models.

This chapter begins with a description of the purpose and set-up of the implementation, including the reasons behind the choice of political documents. Before delving into the actual implementation, a description and background of the data is provided followed by the 'domain based document pre-processing' stage required to ready the data for processing and modeling. Finally, the machine specification used in this implementation is as follows: 64bit Windows 10 pro, intel i7 processor -2.7Ghz, 16 GB RAM.

7.1 Reasons for Using Political Data

In implementing this research, a corpus of spoken or written content is required as the source of values. To this end, with politics as the domain of interest, data from political parties including speeches, policy manifestos and debates were used in compiling the training and test data. The main reasons are as follows:

- Policy documents represent a coherent depiction of the views and values held by individuals or groups in political parties. As such they are a rich source of VLSs.
- Political debates, party conference reports and newspaper interviews are publicly available information sources, thus, easily accessible.
- Policy design is a value-laden process (Fischer, 1980).
- Political debates and manifestos are structured in such a way that it is easy to identify what the subject matter of the debate or discussion is. In addition, the players or speakers are also easily identifiable and can thus be mapped to their ideologies.
- The values of most political parties are common knowledge. Therefore, it is not especially complicated for most political observers to predict the behaviour of value holders in political parties. This allows for relative ease in validating the results of the model with known human knowledge.

The next section describes the implementation setup.

7.2 Implementation Setup

The subject of this implementation revolves around two timely topics in UK politics and they are *'Immigration'* and the *'European Union (EU)'*. These subjects stir up diverse views amongst political observers and parties. Due to the broad nature of these subjects, political parties could hold similar or contrasting views on aspects. For instance, a party might have values that encourage 'EU migration', the same party might oppose 'EU investment in Britain'. Both 'EU migration' and 'EU investment in Britain' are aspects of 'EU' subject. In another respect, while parties might have a set of values and express some sentiment on aspects of subjects, they generally display an overarching value on subjects. For instance, the United Kingdom Independent Party (UKIP) frames itself as a pro-British/anti-EU party, and most of its policies stems from a stand point of getting Britain out of the EU. Therefore, the model implementation in this domain is required to capture the values and sentiment parties associate with subject with the added capability of modeling the overall value orientation of the party. The goal is to build a model of values for these two subjects such that the model is representative of three major UK political parties (value holders). Afterwards, these models are applied towards predicting the sentiment of the value holders on sentences. The political parties are The Conservative party, Labour party and Liberal Democrats (LD).

Up until now context has been described as ambiguous, able to take an innumerable number of forms. As a result, the contexts of interest in this research are the domain subjects *'EU'* and '*Immigration*'. These domain subjects or contexts can also be viewed as conditions or scenarios. For example, the United Kingdom Independence Party (UKIP) manifesto, suggests that in the context of the EU, they are opposed to unlimited relationship and uncontrolled migration with and from the EU but open to migration and more relationship with the Commonwealth. Whereas, in the context of Immigration, they advocate limited migration into the UK from anywhere (Europe and outside Europe). So, in one topical subject context, they support migration and in another they are opposed to it.

In implementing this model, a collection of relevant documents authored by UK political parties were downloaded. This comprised of reports, debates and policy statements made by parties. A large proportion of the data was sourced from Hansard which is a transcript of Parliamentary debates. Although the implementation focuses on value holders from the three parties mentioned, documents associated with UKIP were also downloaded to evaluate the model of the 3 parties. The idea behind this evaluation is that UKIP views on the EU and Immigration are very well-established and not as unclear as the three main parties and as such a comparison of the similarities in the value orientation and sentiment of each of the three political parties in relation to UKIP policies and sentiment is carried out.

Finally, for this implementation, an expression 'context-party' pair is introduced, referring to the corpus of training data for a party or value holder under a context. For example, the expression 'EU-Labour' pair refers to data by a Labour value holder under the EU context. 'EU-Labour' model, refers to a model built for a Labour value holder under the EU context.

7.3 The Implementation Dataset

The implementation dataset was drawn from four major UK political parties – Conservative Party, Labour Party, Liberal Democrats (LD) and UKIP. The data can be divided into two categories.

The first category consists of policy documents, manifestos and reports issued by the political parties. Unfortunately, these documents typically comprise of a plethora of topics and themes and since the focus was on the subjects Immigration and EU, each document was manually filtered to extract content (content refers to sections and subsections of the document) relevant to the subjects. This process was carried out manually because most of the documents were not too long (ranged from about 10-250 pages) but most importantly, because they were all structured, having clearly labelled titles, subtitles and a table of content[63], thus making it easy to identify relevant sections. However, the size of the extracted content was significantly small and insufficient for any meaningful test. Therefore, a second category drawn entirely from Parliamentary debate transcripts - Hansard - covering the periods between 2010 and 2015 was crawled. Since the average number of debates held yearly in the UK Parliament is about 145[64], it was expected that a considerable amount of data would be extracted which in addition to the manifestos and policy documents would comprise the test and training set.

As value holders, the authors of reports or speakers in debates belong to a party, and so there was a requirement to associate each speaker or author to a political party as a precursor to compiling the final corpus. Additionally, some of the data sources were authored by multiple individuals, and the potential implication of this was the development of individual models for each speaker, which was not feasible. So, since all the speakers or contributors belong to a political party, all their utterances i.e. speakers/value holders' utterances were grouped under the umbrella of their political party on the assumption that they all share the same values. In reality, this is not always the case, as there is always the odd case of a party member going against the main stream values of his/her party. However, as most party members tend to speak along the same lines, this assumption was made to simplify the process of building single models for the entire party and not individuals. The next section, describes each data type and the approaches used in generating the corpus. This description also includes the domain based document pre-processing required to generate candidate documents for training and test set (see figure 9 in chapter 6).

[63] In addition, most of them include a table of content containing subsections and titles whose names act as reasonable inferable clues for the identification of relevant content.
[64] Obtained by taking an average of UK Parliamentary debates since 2007

7.3.1 Manifestos, Policies, Reports and Newsletters

Appendix 7 outlines a sample of reports and policy documents used in this implementation and table 17 shows a count of documents downloaded and sentences extracted post-processing. For naming purposes, all documents in this category will be called manifestos. As mentioned earlier, this category of documents normally contain content that cuts across several subjects and because their layout is structured, the first step of domain based data pre-processing was to identify contents (sections and subsections) that are relevant to either contextual subjects. The approach was to use the structure of each document in identifying and extracting relevant subsections towards building a corpus for each context party pair. In doing so, each extracted content became an individual document or mini-document[65].

Table 17: A Summary of Manifesto and Policy Documents Downloaded and Sentences Extracted Post-Processing

Party	Number of Extracted Documents	Number of sentences post-processing	
		EU	Immigration
Conservative	73	227	177
Labour	97	196	202
Liberal Democrats	131	149	162
UKIP	114	216	208

The following steps describe the domain-based pre-processing for generating the mini-documents in this category.

1. For each downloaded document, a manual examination of the table of content was carried out to identify titles relevant to the subjects i.e. 'EU' and 'Immigration'. In figure 21 shows a screen shot of the 2015 Conservative party manifesto's table of content with a black rectangle for subsections related to Immigration and a red rectangle for areas related to the EU. Figure 22 shows a sample of subsections (see black bordered section) in the section on Immigration that are converted to mini-documents.

   If the table of content (toc) was not comprehensive enough or if the document did not contain a toc e.g. (debate and speeches from European House of Parliament), the entire document was scanned manually for sections relevant to the context. Manual search also included 'Forwards' and 'Introductions'.

[65] Mini documents could be single sentences or entire paragraphs.

2. Sub-sections deemed to be relevant became mini-documents and mapped to the appropriate context-party pair before storing in named directories. The directory name takes the form '*context_party*' e.g. '*EU_Labour*'. If the entire document is about a context, mini documents were simply generated from each of the sub-sections in the document. The reasoning is that since the document is about say '*Immigration*' then everything must be relevant to '*Immigration*' regardless of its subtitle. Although the final unit of analysis would be sentences, at this point the aim was simply to gather as many relevant mini-documents as possible.

3. Pronominal Resolution – In some of the data, mostly speeches from European Parliamentary debates, named entities like persons and locations were referenced using their pronoun form. For instance, consider the snippet (Pronominal references are in bold font),

   > "*UKIP Leader **Nigel Farage** said: '**I**'m not against immigration. Far from it. Migrants have qualities we all admire. Looking for a better life. They want to get on. **I** like that'…*[66]"

   '*I*' in the sentence references the person entity '*Nigel Farage*'.

1. AN ECONOMIC PLAN TO HELP YOU AND YOUR FAMILY
A strong economy to help you and your family 7
Better roads, trains and modern communications 14

2. JOBS FOR ALL
Jobs for all 17

3. CUTTING YOUR TAXES, MAKING WELFARE FAIRER AND CONTROLLING IMMIGRATION
Cutting your taxes and building a fairer welfare system 25
Controlled immigration that benefits Britain 29

4. THE BEST SCHOOLS AND HOSPITALS FOR YOU AND YOUR FAMILY
Giving your child the best start in life 33
Protecting and improving our National Health Service 37
Enabling you to enjoy our heritage, creativity and sports 41
Helping you build the Big Society 45
Making government work better for you 47

5. SECURING YOUR HOME AND YOUR NEIGHBOURHOOD
Helping you to buy a home of your own 51
Protecting and enhancing our natural environment 54
Guaranteeing you clean, affordable and secure energy supplies 56
Fighting crime and standing up for victims 58
Preventing terrorism, countering extremism 61

6. DIGNITY IN YOUR RETIREMENT
Dignity in your retirement 65

7. KEEPING OUR COUNTRY SECURE
Stronger together: a Union for the 21st century 69
Real change in our relationship with the European Union 72
A Britain standing tall in the world 75
A stronger voice for our nation on the world stage 76
Keeping Britain safe 77
Tackling global challenges to make you safer and more prosperous 78

Figure 21: Screen shot of Conservative Party Manifesto 2015, TOC

[66] http://blogs.spectator.co.uk/2013/09/nigel-farages-speech-full-text-and-audio/ - Last accessed 03/11/2015

The task of reference resolution was to identify named entities referred by pronominal linguistic units. To identify the pronouns mentioned, a pronominal resolution algorithm was applied on each mini-document, to resolve the following pronouns:

a. She, he, her, him, his, himself, herself

b. It, its, itself

c. I, me, my, myself

will have to be earning here for a number of years before they can claim benefits, including the tax credits that top up low wages. Instead of something-for-nothing, we will build a system based on the principle of something-for-something. We will then put these changes to the British people in a straight in-out referendum on our membership of the European Union by the end of 2017. At the same time, we will continue to strengthen our borders, improve the enforcement of our immigration laws and act to make sure people leave at the end of their visas. Across the spectrum, from the student route to the family and work routes, we will build a system that truly puts you, your family and the British people first.

Our plan of action:

We will regain control of EU migration by reforming welfare rules

Changes to welfare to cut EU migration will be an absolute requirement in the renegotiation. We have already banned housing benefit for EU jobseekers, and restricted other benefits, including Jobseeker's Allowance. We will insist that EU migrants who want to claim tax credits and child benefit must live here and contribute to our country for a minimum of four years. This will reduce the financial incentive for lower-paid, lower-skilled workers to come to Britain. We will introduce a new residency requirement for social housing, so that EU migrants cannot even be considered for a council house unless they have been living in an area for at least four years. If an EU migrant's child is living abroad, then they should receive no child benefit or child tax credit,

We will tackle criminality and abuse of free movement

We will negotiate with the EU to introduce stronger powers to deport criminals and stop them coming back, and tougher and longer re-entry bans for all those who abuse free movement. We want to toughen requirements for non-EU spouses to join EU citizens, including with an income threshold and English language test. And when new countries are admitted to the EU in future, we will insist that free movement cannot apply to those new members until their economies have converged much more closely with existing Member States.

We will continue to cut immigration from outside the EU

We have already capped the level of skilled economic migration from outside the EU. We will maintain our cap at 20,700 during the next Parliament. This will ensure that we only grant visas to those who have the skills we really need in our economy. We will reform the student visa system with new measures to tackle abuse and reduce the numbers of students overstaying once their visas expire. Our action will include clamping down on the number of so-called 'satellite campuses' opened in London by universities located elsewhere in the UK, and reviewing the highly trusted sponsor system for student visas. And as the introduction of exit checks will allow us to place more responsibility on visa sponsors for migrants who overstay, we will introduce targeted sanctions for those colleges or businesses that fail to ensure that migrants comply with the terms of their visa.

Figure 22: Mini-documents from subsection on Immigration in Conservative Manifesto 2015

The named entity types include: Persons, locations and Organizations. GATE's Annie Pronominal Coreferencer version 8.1 (Cunningham et al, 2015) was applied with the *'resolve it'* runtime parameters for Annie Pronominal Coreferencer changed to *'true'* to resolve *'it'* pronouns. After pronominal resolution, a list of sentences containing resolved pronouns was maintained. For such sentences, key value pairs were created for each pronoun and its resolved entity. This was used to update the sentence's JSON object after sentence splitting.

4. It was apparent that there was a high possibility that content relevant to the contexts were made outside of explicitly labelled relevant subsections. For instance, sentences relevant to Immigration were identified in a sub-section titled *'Building an Economy that works for people'* in the Conservative Party Manifesto. To identify these, the initial approach was to parse the remainder of the document looking for

sentences containing a seed set of semantically relevant words and expressions. For instance, on the EU, documents were parsed for mentions of expressions like *'The EU'*, *'European Union'*, *'Europe'*, *'The Continent'*. Similarly, on immigration, documents were parsed for mentions of terms like *'Immigrant/s'*, *'Immigration'*, *'migrant'*, *'asylum'*, *'refugee'*, *'foreigner'*, *'work permit'* and *'visa'*. However, it became clear that this seed set of expressions did not encapsulate the entirety of the subject matter thus the risk of not identifying all the relevant sentences. In other words, low recall. To increase the seed set of relevant terms, for each context, all the mini documents obtained for all the parties were compiled into a single corpus. In doing this, it is assumed that regardless of party affiliation each document would contain words or expressions that are relevant to the subject domain so that the goal becomes extracting domain relevant terms (single or multiword word expressions). Two techniques were employed:

a. A bootstrapping language neutral extraction technique based on KYOTO scoring (Bosma and Vossen, 2010) which computes a domain-relevance score as a measure of how well connected a candidate expression is to other terms in the document (the number of hyponyms associated with the expression) and the document frequency of the term. A unique benefit of this approach is that it supports the extraction of multiword expressions. A domain relevance score $R(t)$ for a term t is expressed as follows:

   $$R(t) = ||doc(t)|| \cdot (1 + ||hypo(t)||) \ldots (18),$$

   where t is the term or expression, $||doc(t)||$ is the document frequency of t, and $||hypo(t)||$ is the number of hyponyms of t. Since $R(t)$ lies between 0 and ∞, a normalized relevance score $R_{norm}(t)$ is used and expressed as:

   $$R_{norm}(t) = 1 - (1 + log(1 + R(t)))^{-1} \ldots (19)$$

   Following this, words and expressions in the top 25% of scores were extracted.

b. TF-IDF (Term Frequency-Inverse document frequency) scores were calculated over all the terms in the corpora, setting a threshold at 25% for words with the highest tf-idf score. TF-IDF a term t is expressed as:

   $$tfidf_t = tf_t \,.\, idf_t \ldots (20)$$

   Where,

   tf_t is the term frequency of t

   idf_t is the inverse document frequency of t (Sparck, 1972) expressed as

   $$idf_t = log_2 \frac{N}{df_t + 1} \ldots (21)$$

Where, N is the total number of documents in the corpus and df_t is the document frequency of term t.

The terms and expressions extracted using both techniques were merged into a set. Terms deemed to be too generic e.g. '*school*', '*fine*', '*deal*' were also manually eliminated from the list, leaving a list of relevant terms. Table 18 shows a sample of terms extracted. In the next step, sentences in the domain corpus that contain any of the seed set of extracted words are extracted.

5. Some of the documents contained bullets and ordered list of sentences. For these, all bullets or ordered list were extracted followed by a manual inclusion of relevant bulleted or listed sentences in the corpus.

The outcome of this process is a small corpus containing mini-documents (including sentences) relevant to the subject matter (see table 17).

Table 18: Sample of Extracted words using KYOTO relevance score

Europe
Barrier, Brexit, British passport, Brussels, Citizen, Country, EU, EU Citizen, EU Convention, deal, election, benefits, Juncker, free movement, European budget, Merkel, Credit, family, European Parliament, Europe, loan, market, migrant arrival, migrant camp, minister, bloc, budget, Eastern, EU country, EU Commission Headquarters, EU Commission, policy, policy position, single market, reform, policy, support, student study, trade deal agreement, Single Market Access, EU quota, EU reform, EU trade, member state, transition, membership, movement, negotiation, trade, transition, travel, vote
Immigration
Migrants, Immigrant, people, entering Europe, North Africa, Australian Point based system, EU migrants, coming, stay, visa, student, student visa, UKBA, border police, criminals, child, clearance, policy, Home Office, slave, work permit, change, denied, enforce, migration policy, benefits, method, fine, Europe, gangs, movement, freedom, kidnap, holiday, NHS, Eastern Europe, Partnership, house, allowance, passport, citizenship test, English, British, English language

7.3.2 Hansard Transcripts

Hansard transcripts are semi-structured. Figure 23, provides a screenshot of a typical Hansard document with emphasis on its structure. Each debate transcript consists of several sub-debates. Its structure includes a title followed by an ordered sequence of contributors and their contributions (speaker-contribution pair). The debate types used in this thesis include:

- Debates and Oral Answers to Questions: Includes Commons debates on bills, oral statements made by Ministers and issues raised during ministerial question times.

- Westminster Hall: Features a range of subjects raised by MPs in adjournment debates and during consideration.

- Written Statements: Written statements on policy or government.

Creation of mini-documents from Hansard also involved domain pre-processing. The first part involved associating each contributor or speaker to his/her party, so that contributions could be grouped and organized based on party affiliation. Using a compiled gazetteer of MP names and their party affiliations, contributors mentioned in the transcript were mapped to a party so that each debate became a collection of party-contribution pairs. For instance, if the contributor was '*David Cameron*' and his contribution was '*We are working towards securing Britain's status in the EU.*', a JSON object was created with the following key-value pairs, as seen in figure 24. For referencing purposes, this object is called speaker-contribution-object.

Figure 23: Screenshot Snippet of a Structural Illustration of Hansard


```
{
'speaker':'David Cameron'
'party':'Conservative party'
'contribution':'We are working towards securing Britain's status in the EU'
'context':'EU'
'debate title': 'Latest Immigration figures'
}
```


Figure 24: JSON representation of Hansard Debate Contribution

The second process involved in corpus creation was to group debate contributions by contexts. Unlike Policies and manifestos which are considerably smaller, the number of debate subjects covered in Hansard is considerably large meaning that manually mapping each debate to a context was impractical. To address this, an unsupervised clustering approach, Latent Dirichlet Allocation (LDA) (Blei et al, 2003) was adopted to enable identification and grouping of different debates based on their theme (which in this case is the context). LDA is used because it is a topic model, and since the task is to identify latent topics or themes in the corpus, it suffices. In addition, LDA has also been shown to outperform a similar topic model Latent Semantic Analysis (LSA) (Griffiths et al., 2007). LDA's application is described in a later section. For now, it is assumed that the subject category to which each debate belongs to is known.

The next objective was to generate a corpus of mini-documents for each context-party pair. Speaker-contribution-object was also modified by including an additional map specifying the context of the contribution e.g. '*context': 'EU'* (see figure 24).

For Hansard, a mini-document was defined to be a contribution. A simplifying assumption was made that for every debate, each contribution is independent[67], so that all the contributions made by members of a party can be grouped into a corpus. From the collection of speaker-contribution-objects, all contributions made by speakers from the same party on a context are extracted. This process is illustrated in figure 25. Finally, pronominal resolution was performed on contributions. As illustrated in figure 25, before a contribution is added to the corpus, irrelevant contributions and snippets were excluded like:

- One-word contributions or uniquely Parliamentary expressions like – '*Order*', '*I give way for the Honorable Gentleman*'.
- All references to quotations made by people other than the current contributor, because it is difficult to identify the original speaker.
- Contributions made by the Deputy Speaker and the Speaker because they are meant to be neutral parties.
- Sentences enumerating a list of MPs, votes (ayes or nays), costings and summaries. These are quite common during voting, the reading of motions and petitions.

Finally, a unified corpus is created by merging the mini-documents derived from manifestos with the Hansard mini-documents compiled for each context-party. Following this, data-preparation commences.

7.4 Data Preparation

The domain dependent pre-processes carried out in the previous section was designed to compile a corpus. With the corpus generated the process of preparing it for analysis is described. The preparation steps follow the steps described in figure 9 of chapter 6.

Step1: Content Renaming

This involved assigning pseudo-words to named entities in the corpus. Pseudo-words were assigned to a class of commonly occurring linguistic units. This implementation focused

[67] In reality, this is not true as a speaker's contribution is usually dependent on what was previously said.

on the three named entity categories described in chapter 6 - names of persons or organizations, locations and numbers/currencies.

First, a duplicate of each corpus was created. This was done because this processing stage involved replacing and modifying content in the corpus. So, in order not to lose the original content which would be needed later in context generation and for referencing, a copy was created while content renaming was performed on one copy, leaving the other unmodified.

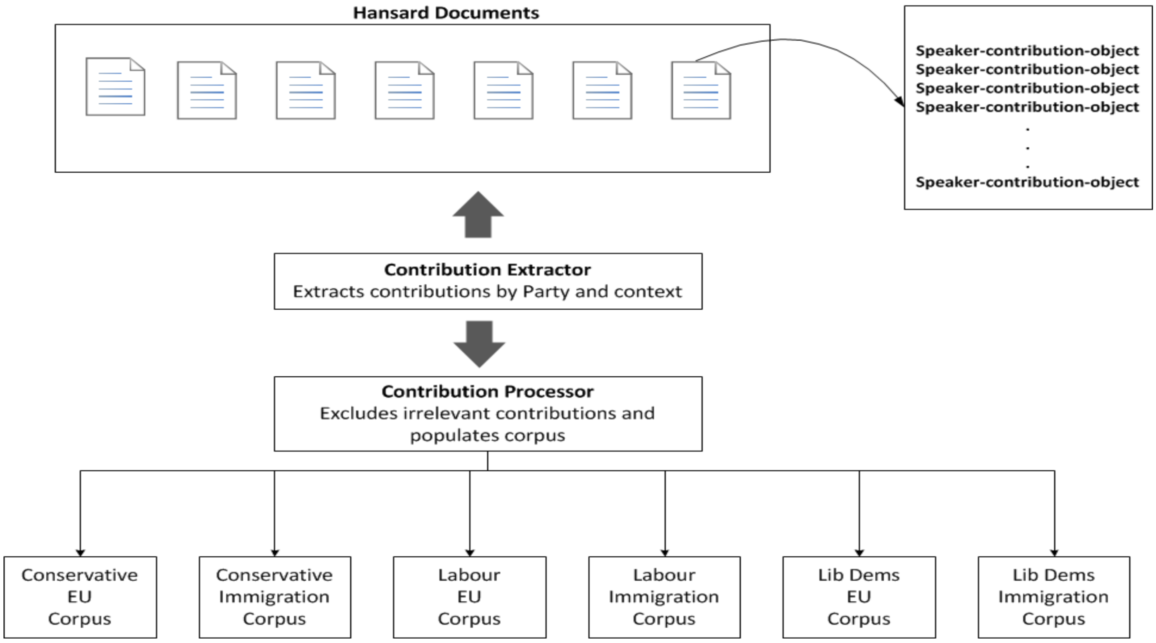


Figure 25: Process Flow for Converting Hansard Documents to Corpus

- **Person Names** – Names are commonly mentioned in this corpus. In identifying and resolving names, GATE's named entity transducer was used. Given the nature of the domain, external resources and gazetteers were incorporated in resolving names. A gazetteer of MPs[68] names and the names of world leaders were compiled for resolving names that were not identified by GATE's default named entity transducer. In addition, in some sentences, persons are referenced through their position, for instance, in the sentence snippet: *'The Prime Minister's response is …'* or '*The Honourable MP for Berkshire …*'. In these examples, the expression *'Prime Minister'* is a position occupied by a person (during this period, it was *'David Cameron'*) and the latter refers to a person (during this period, it was *'Theresa May'*). To make such a reference, an MP's ontology was implemented with data populated from www.theyworkforyou.com. Using this ontology, ontological identification of positions, titles as well as semantic annotation[69] of mentioned positions was performed. In summary, identified names (including resolved person

[68] A list of UK Parliament MPs since 2007 till 2015 is obtained using https://www.theyworkforyou.com/api/ - Last accessed 20-12-2015

[69] The process of annotating or modifying in the texts all mentions of instances relating to concepts in an ontology.

pronouns) and mentions of positions were replaced with the pseudo-word 'PERSONNAMES'[70]. Additionally, all abbreviations/acronyms and alphanumeric expressions were converted to the pseudo-word 'ABBREVIATIONNAME' and 'ALPHANUMERICNAME' respectively[71].

- **Dates, Numbers and Currency** – The implementation technique described in chapter 6 is applied in renaming 'Dates, numbers and currency'.

- **Locations** – Countries and locations were renamed by mapping them to the pseudo-word 'LOCATIONNAME'[72].

Step 2: Eliminating Unwanted Punctuation

In formal documents like manifestos and policies, it is not uncommon to find bracket enclosed phrases and expressions inside a sentence. By simply stripping off the brackets, the resulting sentence might lose its grammatical meaning. To address this, a condition was applied where if the bracket enclosed expression was a sentence[73], it was stripped of its opening and closing brackets, removed from the original sentence and made a completely new addition to the corpus. Otherwise, the bracket enclosed expression including the brackets was eliminated so that the containing sentence was rewritten without the bracketed text. In addition, sentences enclosed in quotations were removed since it is often difficult to attribute it to a speaker. GATE's ANNIE Pronominal Coreferencer processing resource was applied in identifying quoted segments since it includes a quoted speech submodule for identifying quoted text segments (Cunningham et al, 2015).

With the data preparation process completed, each document corpus was converted to a corpus of sentences using GATE's sentence splitter module. The training and test set[74] were

---

[70] In a later section, when testing for the impact of semantic enhancement a different content renaming strategy is adopted to encapsulate the semantic relevance of a particular name. For instance, in this second test, the pseudo-word 'PERSONNAMELABOURMP' was assigned to mentions of Labour MPs and assign the pseudo-word pattern '[Country Name]PRESIDENT' e.g. 'GERMANYPRESIDENT' to mentions of Presidents. The intuition is that certain names have more relevance and impact in certain domains and contexts and capturing this diversity might improve the model's performance.

[71] As part of the test to observe the effect of semantic enhancement of the model, unique acronyms that are dominant in the domain like 'NATO' or 'TTIP' were renamed. Again, the assumption here is that these acronyms have unique semantic relevance and by generically lumping them together as was done in this stage, some performance might be lost. This is described in a later stage

[72] As with previous named entities, the effect of capturing the semantics of location (location semantic enhancement) by differentiating between locations e.g. differentiating mentions of EU countries from mentions of non-EU countries etc is shown in a later test.

[73] The enclosed expression is checked to see if it includes a grammatical subject, object and predicate.

[74] In Machine Learning and statistics, the data is split into two sets. The training set is used to train the statistical parameters of the model. The trained model is subsequently used to estimate probabilities on the test set. An additional set called a dev set or development set is set aside to tune the parameters of the model.

made up of a collection of sentences. Table 19 shows the original number of documents prior to pre-processing while table 20 shows the total number of sentences in the corpus associated with each party and domain after data cleansing and preparation. As seen in table 19, the size of the UKIP document set is quite small, and so a model was unimplementable from it. However, it is used later in evaluating the model. Also, table 20, shows that the data is split into a training, test and development set using a ratio of 7:2.5:0.5. Table 21 portrays the total number of tokens (N) in each training corpus. The comparative difference in corpus and token size between Conservative/Labour and the Liberal Democrats is accounted for by the comparative size of the parties and the number of MPs. Since Labour and Conservative party have more MPs, it is expected that they would make more contributions, thus a larger data set.

Table 19: Number of documents extracted for Parties

YEAR	Conservative	Labour	LD	UKIP
2010	25556	37108	1001	
2011	28003	19011	3103	
2012	30965	17882	2782	
2013	32517	18429	2840	
2014	30151	16407	2298	
2015	29532	14182	1048	22
Total	**176724**	**123019**	**13072**	**22**

Table 20: Number of sentences post-data preparation

Data	Cons. EU	Cons. Immigration	Lab. EU	Lab. Immigration	LD EU	LD Immigration
Train	238135	194838	165769	135628	20132	16470
Test	85050	69585	59203	48439	7190	5883
Dev	17008	13918	11840	9688	1439	1177
Total	**340193**	**278341**	**236812**	**193755**	**28761**	**23530**

Table 21: Total Number of Tokens across training corpus

Party	EU	Immigration
Conservative	2974827	2577704
Labour	2488436	2355246
Liberal Democrats	385171	356076

With the sentences identified, the next implementation stage was pre-parsing and it involved applying a POS tagger on the training sentences in other to disambiguate words and convert each sentence to a sentence JSON object. Figure 26 shows a snippet of the resulting sentence JSON object. It is slightly different from the sentence JSON object described in chapter 6 with the inclusion of key value pairs associated with pronominal resolution of words.

7.5 Value Components (VC) Identification Implementation

So far, the training set of value sentences have been identified, prepared and compiled in a corpus. The next step involved the identification and extraction of VCs. VCs include value holder (H), the action (A), state (S), subject (θ) and Context (C) where A, S, and θ are content words. The first part of this section focuses on identifying the value holders H. Following this, the implementation for identifying action, states, subjects and context is described.


```
{
      "analysisType": "string",
      "sentenceString": "UKIP will leave the UK.",
      "ID": "1",
      "listOfWords": [{
            "word_original": "UKIP",
            "isAlphaNumeric":false,
            "isCapitalized":true,
            "isNamedEntity":true,
            "isNumeric":false,
            "isLowercase":false,
            "word_index": "0",
            "word_lower_case": "ukip",
            "word_POS": "Noun",
            "POS_CATEGORY": "NNP",
            "word_startposition": "0",
            "word_endposition": "5",
            "wordsense_index": "1",
            "wordsense_POS": "Noun",
            "isPronominalResolution":false
            "pronounWord":""
      }, {
            "word_original": "will",
            "isAlphaNumeric":false,
            "isCapitalized":false,
            "isNamedEntity":false,
            "isNumeric":false,
            "isLowercase":true,
            "word_index": "1",
            "word_lower_case": "will",
            "word_POS": "Verb",
            "POS_CATEGORY": "MD",
```


Figure 26: Snippet of Sentence JSON object for Political Data

Value holders H in this implementation are party members or persons associated with them. For manifesto type documents described in section 7.3.1, the parties associated with the documents were known from the moment they were collated and so no name or party resolution was performed. Thus, focus was given towards identifying H in Hansard.

Using a compiled gazetteer of MPs and MP ontology, each Hansard contributor was mapped to a party. In addition, some of the internal structure and patterns in Hansard were applied in identifying a contributor's political party. For example, 3 naming patterns were observed in Hansard as illustrated with the screen shot in figure 27. The first follows the pattern:

[Contributor full-name (Contributor constituency) (Abbreviated Contributor party)]

Examples include, '*Stuart Andrew (Pudsey)(Con)*' or '*Mr David Heath (Somerton and Frome) (LD)*'. For these naming templates, the value holders are easily identified as Conservatives and Liberal Democrat MPs respectively. The second naming pattern is:

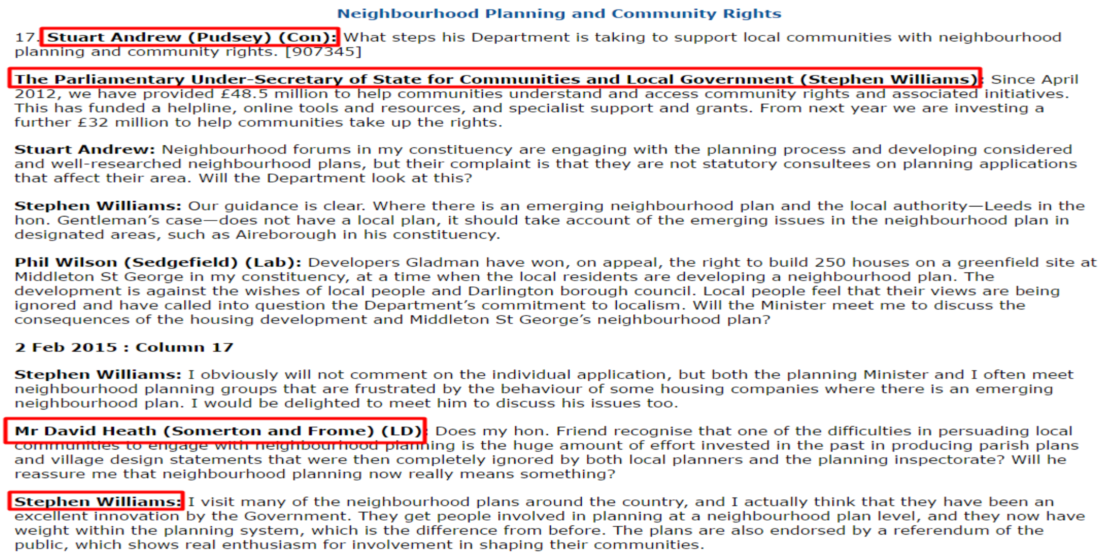
[Contributor position or title (Contributor full name)]

An example from figure 27, is *'The Parliamentary Under-Secretary of State for Communities and Local Government (Stephen Williams)'*. For such naming patterns, the party of the value holder was resolved by querying the MPs ontology or gazetteer of MP names. The third naming pattern takes the form:

[Contributor title or position]

E.g *'Mr Speaker'* or *'The Prime Minister'*. This pattern is usually accorded to persons with the position of Speaker, Deputy Speaker, Prime Minister or Deputy Prime Minister. Since the Speaker and Deputy Speaker are supposed to be neutral, their contributions were ignored. The party of the Prime Minister or Deputy Prime Minister was resolved by querying the MPs ontology.

With H resolved, the following sections discusses the implementation of context C, the identification of content and function words (A, S, θ).



Neighbourhood Planning and Community Rights

17. **Stuart Andrew (Pudsey) (Con):** What steps his Department is taking to support local communities with neighbourhood planning and community rights. [907345]

The Parliamentary Under-Secretary of State for Communities and Local Government (Stephen Williams): Since April 2012, we have provided £48.5 million to help communities understand and access community rights and associated initiatives. This has funded a helpline, online tools and resources, and specialist support and grants. From next year we are investing a further £32 million to help communities take up the rights.

Stuart Andrew: Neighbourhood forums in my constituency are engaging with the planning process and developing considered and well-researched neighbourhood plans, but their complaint is that they are not statutory consultees on planning applications that affect their area. Will the Department look at this?

Stephen Williams: Our guidance is clear. Where there is an emerging neighbourhood plan and the local authority—Leeds in the hon. Gentleman's case—does not have a local plan, it should take account of the emerging issues in the neighbourhood plan in designated areas, such as Aireborough in his constituency.

Phil Wilson (Sedgefield) (Lab): Developers Gladman have won, on appeal, the right to build 250 houses on a greenfield site at Middleton St George in my constituency, at a time when the local residents are developing a neighbourhood plan. The development is against the wishes of local people and Darlington borough council. Local people feel that their views are being ignored and have called into question the Department's commitment to localism. Will the Minister meet me to discuss the consequences of the housing development and Middleton St George's neighbourhood plan?

2 Feb 2015 : Column 17

Stephen Williams: I obviously will not comment on the individual application, but both the planning Minister and I often meet neighbourhood planning groups that are frustrated by the behaviour of some housing companies where there is an emerging neighbourhood plan. I would be delighted to meet him to discuss his issues too.

Mr David Heath (Somerton and Frome) (LD): Does my hon. Friend recognise that one of the difficulties in persuading local communities to engage with neighbourhood planning is the huge amount of effort invested in the past in producing parish plans and village design statements that were then completely ignored by both local planners and the planning inspectorate? Will he reassure me that neighbourhood planning now really means something?

Stephen Williams: I visit many of the neighbourhood plans around the country, and I actually think that they have been an excellent innovation by the Government. They get people involved in planning at a neighbourhood plan level, and they now have weight within the planning system, which is the difference from before. The plans are also endorsed by a referendum of the public, which shows real enthusiasm for involvement in shaping their communities.

Figure 27: Hansard screenshot snippet Illustrating Hansard Naming Templates

7.5.1 Content and Function Word Pipeline Illustration

As mentioned in section 6.2, action, states and subjects are treated as content words. Table 22 shows the number of unique content words retrieved for each party and context, this includes homonymous words with different word sense. The disparity in content word vocabulary size between the three parties, particularly between the Liberal Democrats and the other parties is quite significant. It was observed that over 90% of the content words used by the Liberal Democrats were expressed by the other parties. There were several reasons for this, the obvious been that there are considerably more Labour and Conservative MPs hence more contributions. In addition, on manual review of the content words, a portion of the content words in Conservative and Labour vocabulary were diverse word forms of the same lemma. Figure 28, shows a screen shot of the vocabularies viewed in the open-source software Beyond-Compare[75] to enable the comparison. It shows a sample of Conservative content words on the left and Liberal Democrats content words on the right. For instance, observe that both vocabularies contain the word '*accompanied*', however, they also contain additional word forms – '*accompanies*', '*accompanying*', '*accompaniment*'[76].

Table 22: Vocabulary Size of Content Words across all Parties and Contexts

Party	EU	Immigration
Conservative	35756	32925
Labour	31077	28582
Liberal Democrats	15629	14684

7.5.2 Context Identification

So far, Hansard documents have been associated to either '*Immigration*' or '*EU*' context without explaining how this was accomplished. This determination is presented in this section.

Seeing as the topical domains in Hansard are not explicitly mentioned and considering its large size, an unsupervised topic modeling algorithm - LDA (Blei et al, 2003) - was applied towards assigning topics to Hansard debate content. This was not applied to the manifestos and policies since their topical domains were already known. Although a brief description of LDA is provided in appendix 8, a detailed theoretical explanation is outside the scope of our research.

The identification of context involved two processes: Preparing Hansard for Context Identification, followed by the LDA implementation of the prepared data.

Preparing Hansard for Context Identification

The identification of document context was one of the first tasks performed in this implementation even though up until now it is assumed that the context of each document is known. Using Hansard's structure, the unit of analysis (the document) was extracted, by concatenating all the contributions made in each sub-debate into one contiguous piece of text (Contributors were not included) (see figure 29). The reasoning behind this compilation hinged on two principles:

- Using all the contributions, avails a much larger document and consequently capturing a richer vocabulary set that reflects the associated topics[77].

[75] http://www.scootersoftware.com/: Last accessed 17/12/2016
[76] This observation justifies the implementation of section 6.6.3, which addresses word forms with the same meaning.

[77] Individual contributions could be as bare as a single phrase or sentence and thus not as rich and expressive as joining multiple contributions.

- LDA is based on the bag-of-words assumption which means that word order is irrelevant.

Stop words and punctuation were also eliminated from the corpus. Contributor name was also excluded, so that a document comprised of all the contributions made in a sub-debate. Hansard debates from the period between 1st of January 2007 and 20th of June 2013 were collected bringing the total number of documents to 16933. The size of the smallest document was 1KB containing 8 sentences and the size of the largest document was about 497KB containing 2636 sentences.

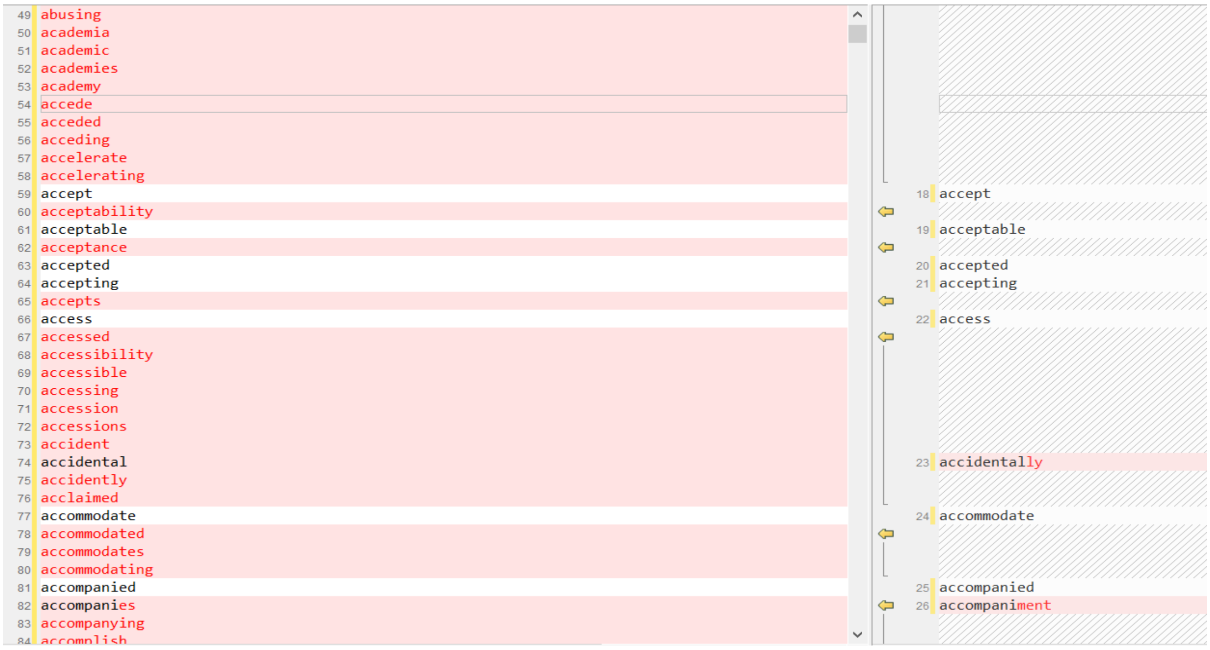


Figure 28: Screen shot comparing content words from different vocabularies. Conservatives on the left, LD on the right.

New Magna Carta

Q15. [907581] **Mr Graham Allen (Nottingham North) (Lab):** If he will commission a new Magna Carta to renew democracy in the UK as part of the celebrations of the 800th anniversary of Magna Carta; and if he will make a statement.

The Prime Minister: We should be proud that in Magna Carta our country established rules of justice and freedom that, 800 years later, still inform our constitution and resonate around the world. While there is a long-standing debate over the issue, there are no plans at present for a written constitution.

Mr Allen: I note that the Prime Minister says "at present". Does he agree, though, that there are unacceptably high levels of voter disengagement, with more people staying at home than voted Labour and Conservative at the last election? Would he commit his Government, now, to preparing an all-party constitutional convention, in order to give every UK citizen a copy of our society's rulebook—either a statute of the Union or a written constitution—as a part of electors feeling once again that they own our democracy?

The Prime Minister: Obviously, I always look at the hon. Gentleman's suggestions very carefully, because he has made a number of sensible cross-party interventions over recent years, but I have my doubts as to whether another talking convention is the answer. I think we need to look at some of the constitutional issues that leave people feeling left behind, not least English votes for English laws, and make sure that we put those things in place. The disappointment I have with the Labour party is that it is prepared to talk about all-party talks on Wales, Scotland or Northern Ireland, but when it comes to empowering English people and making sure that they have rights in this House, it is completely absent from the debate.

Jacob Rees-Mogg (North East Somerset) (Con): Article 39 of Magna Carta contains the origins of our right to trial by jury. In a recent report, Sir Brian Leveson, not satisfied with undermining the right to a free press, wants to restrict the right to trial by jury. Will my right hon. Friend, as long as he is Prime Minister, defend our historic rights?

The Prime Minister: I am a great supporter of jury trial. I think it is one of the very important things we have in this country that safeguard people's rights and freedoms, and I do not want to see it reduced.

11 Feb 2015 : Column 779

Francis Report: Update and Response

12.36 pm

The Secretary of State for Health (Mr Jeremy Hunt): With your permission, Mr Speaker, I would like to make a statement on the Government's response to today's report on NHS whistleblowing by Sir Robert Francis, and on progress to date in implementing previous recommendations from his public inquiry into the failures of care at Mid Staffordshire NHS Foundation

Figure 29: Screenshot of Full Hansard sub-debate document enclosed in red box

The LDA Implementation

LDA was implemented using MALLET (McCallum, 2002) and it involved:

1. Importation of data: This involves importing the documents into MALLET. Stop words were removed from the input document.

2. Build topic model: MALLET's '`train-topics`' option was used to build the topic model by setting the following parameters:

    a. The Number of topics – Topic size determination involved domain experts who iteratively assigned user friendly topic names to LDA output topic clusters. An LDA model is built with a small topic space e.g. 5 (MALLET's default topic size is 10), domain experts assign general topic names to each output word cluster and as the topic size is iteratively increased, the topic names assigned become less general and more specific to each generated word cluster. In this research, the iterative process of topic reduction was carried out in increments of 5. After 40 topics, participants were unable to assign additional specific names to the generated clusters. This led to the conclusion that 40 was the optimum topic size for the dataset and the derivation of 40 user friendly topic names.

    b. The number of iterations was set at 3000 iterations.

    c. The hyperparameter optimization parameter was set to 10[78], a reasonable value recommended in the MALLET documentation.

    d. The alpha and beta parameters are smoothing parameters for document-topic distributions and topic words respectively. Out of the box values were used i.e. α = 5.0, β = 0.01.

Following training of the LDA model, a document type threshold was set at 51% (A document was identified as being about the '*EU*' only if at least 51% of it contained the topic '*EU*'). Only documents relating to Immigration and EU, were extracted and these documents formed the foundation of our training and test data. This concludes the identification of VCs i.e. value holders (H) was a known entity, Context (C) was implemented as a function of the debate's topical theme using LDA and actions (A), state (S) and subject (θ) were extracted as content words using an implementation of dependency grammar formalism. Function and content word relationships were also captured.

Therefore, the output of this section is such that, in each corpus, for each word that makes up a sentence, the following information is maintained–

- The status of the word – Content or Function word

---

[78] This option turns on hyperparameter optimization, which allows the model to better fit the data by allowing some topics to be more prominent than others. (http://mallet.cs.umass.edu/topics.php: Last accessed 02/12/2016

- A distinguishing identifier number based on its part of speech (POS), which enables us distinguish homonyms.
- The base form or lemma of the word e.g *'came → come'*, *'ran' → 'run'*
- A list of grammatical relations for which each word in the sentence partakes in, and its dependents or governors.

Following this, the LM implementation is discussed.

7.6 LM Implementation

LM implementation is premised on trigrams. Table 23 illustrates the count of trigrams identified in each context-party pair.

The implemented model is an interpolated Kneser-Neys smoothed LM. Initially, a base-line model was implemented to tune the interpolated parameters. Baseline interpolated parameters were $\lambda_1 = 0.5$, $\lambda_2 = 0.5$ and $\lambda_3 = 0$ and these were varied to tune the model until perplexity was minimized.

Other smoothed LMs were implemented and measured against this baseline model just to verify the superiority of the interpolated KN model. These models include absolute discounting LMs for which the discount d was varied, Good-Turing, linear, Witten-Bell, back-off[79] Kneser Neys. The performance of these models was compared intrinsically by calculating perplexity.

The models were implemented on the training sentences. Table 24 shows the perplexity of each implemented LM for the EU-Conservative pair and it also shows that the interpolated KN model produced the best performance by returning the lowest perplexity.

Table 23: Trigram count for each value-holder/context pair

Value-Holder/Context	Trigram Count
Conservative/EU	3304053
Conservative/Immigration	3182559
Labour/EU	2169023
Labour/Immigration	1932922
Lib Dems/EU	548845
Lib/Dems/Immigration	417157

In tuning the interpolated KN, 400 iterations were initiated, modifying the interpolated weights λ_1 (trigram weight), λ_2 (bigram weight) and λ_3 (unigram weight) to obtain the best performance. Figure 30 shows a scatter plot of the interpolated weights and perplexity for

[79] Backing-off is an approach for estimating the probability of an ngram in that if the sequence of words $\langle w_{i-2}, w_{i-1}, w_i \rangle$, is unseen some approximate estimate can be reached by recursively backing off to the $n-1$ gram $\langle w_{i-1}, w_i \rangle$ (Katz, 1987).

the interpolated KN model on EU-Conservative data after 400 iterations. Tuning the interpolated weights converged at the following values λ_1 = 0.00077, λ_2 = 0.632, λ_3 = 0.36723 (see scatter plot in figure 30) and a perplexity value of 118.63. As seen in table 24, the Kneser-Neys back-off trigram perplexity is 133.6883, thus asserting that the interpolated model performs better than other models.

The results for all the other context-party pairs mirror this pattern as seen in appendix 9 which shows that for the best performance, interpolated weights consist of very small λ_1 weights, λ_2 are observed to lie between 0.5 and 0.7, while λ_3 weights lie in the range of 0.2 and 0.4. The estimation and tuning of the LM models for the context-party pair concludes the implementation of LM_1 models. This model would be used later in estimating the probability of function word in test sentences.

Table 24: Model Perplexity for Conservative-EU

Model	Perplexity
Good-Turing	137.81
Linear	158.91
Witten-Bell	134.73
Absolute	136.67
Kneser-Neys (Back-off)	133.68
Kneser-Neys - Interpolated	118.63

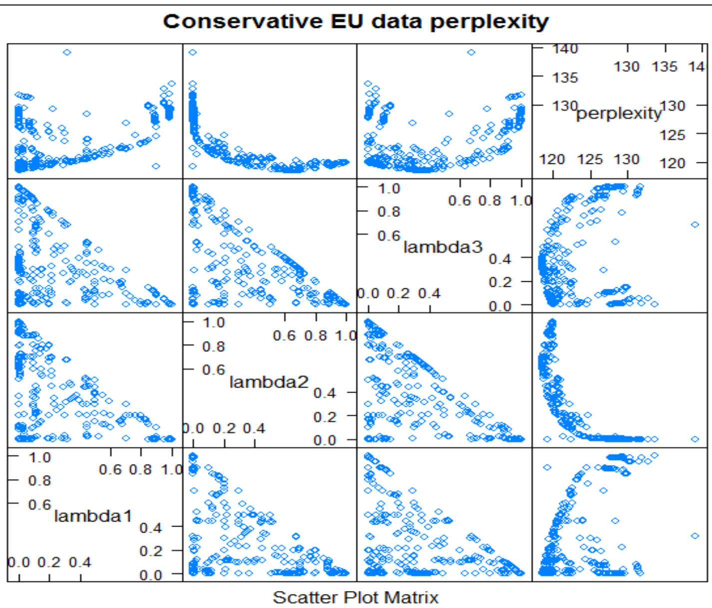


Figure 30: Scatter plot of the Interpolated KN model weights for Conservative-EU

7.7 Model Implementation for Content Words

Content word implementation involved the implementation of a maxent model. This is preceded by the generation of the feature set for each content word (see chapter 6 for a list of features). In this section, in addition to discussing the estimation of content word probability, the probability estimates of rare content words observed in the test set is also described.

7.7.1 Model Implementation for Observed Content Words

For every content word instance observed in the training set, a feature vector was built from all the feature templates illustrated in chapter 6. Thus, for each context-party pair, a matrix X of features and a vector Y of content words is obtained. Since feature size is 241, and assuming the number of content word instances seen in training is n, X becomes a $[241, n]$ matrix, while the vector Y is of size n. Table 25 shows the size n of each context-party pair. X turns out to be a sparse matrix. Despite this, computing the weights for the maxent algorithm was quite slow requiring considerable computational power. To ameliorate this issue, a novel procedure was implemented which involved grouping the vocabulary of content words into clusters, and then building a maxent classifier for each cluster. The net effect was to reduce the dimensionality of the entire training set making it computationally possible to calculate the weights of each word. This clustering procedure is briefly described but a more detailed explanation is found in appendix 10.

Table 25: Size (n) of Content Words across each context-party pair

Context-Party Pair	Size n or number of content word instances
EU-Conservative	500093
Immigration-Conservative	311386
EU-Labour	289629
Immigration-Labour	246101
EU-LD	82066
Immigration-LD	62114

The objective of the clustering approach is to reduce the horizontal dimension of a large matrix of feature vectors, so that it is computationally easier to construct multiple maxent models for each of the constituent matrices. This is accomplished by partitioning the content words into disjoint buckets. Each bucket consists of words that share similar quantitative properties, in this case frequency. Each bucket is treated as a separate distribution and a maxent model is built for each. For instance, the matrix $\begin{bmatrix} x_{1,1} & \cdots & x_{1,241} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \cdots & x_{n,241} \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$, where n is the number of the content word instances is partitioned into:

$$\left\{ \begin{bmatrix} x_{1,1} & \cdots & x_{1,241} \\ \vdots & \ddots & \vdots \\ x_{k^1,1} & \cdots & x_{k^1,241} \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_{k^1} \end{bmatrix}, \begin{bmatrix} x_{1,1} & \cdots & x_{1,241} \\ \vdots & \ddots & \vdots \\ x_{l^2,1} & \cdots & x_{l^2,241} \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_{l^2} \end{bmatrix}, \dots , \begin{bmatrix} x_{1,1} & \cdots & x_{1,241} \\ \vdots & \ddots & \vdots \\ x_{z^p,1} & \cdots & x_{z^p,241} \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_{z^p} \end{bmatrix} \right\}$$

Where, in y_{k^i}, k represents the number of content words and i represents the bucket. So k^1 represents the k^{th} content word of the first partition (bucket 1). Similarly, l^2 represents the l^{th} content word of the second partition (bucket 2). Different alphabets are used for the content word size because the buckets have different sizes.

With the words partitioned into buckets, maxent models were created for each partition/cluster. To estimate the probability of a word w' observed in a test sentence, the word's bucket/partition is identified first and then the maxent model for the word's bucket is used in estimating its probability. This approach enabled the computation of maxent models for each group because of the reduced size of the vector Y. Table 26 shows the total number of buckets identified for each context-party pair, while algorithm 4 illustrates the steps involved in estimating the probability of a content word.

Table 26: Total number of Clusters for each Party-Context Pair[80]

Context-Party Pair	Number of Content Word Partitions
EU-Conservative	50
Immigration-Conservative	42
EU-Labour	42
Immigration-Labour	38
EU-LD	18
Immigration-LD	18

Algorithm 4 Illustrates process of maxent estimation for a Content word from Clusters

Inputs

(i) Sentence, $W = \langle w_1, \ldots, w_n \rangle$

(ii) Maxent models of Clusters, $M = \{M_1, \ldots, M_z\}$, where z is the number of clusters.

Process

1. **Extract** content word w from sentence.
2. **Check** if word can be converted to its base form w'.
3. **Identify** the cluster B which w or w' belongs to and its associated maxent model M_b.
4. **Estimate** the probability of w or w' from M_b.

7.7.2 Rare-Word Estimation

Since rare words do not provide enough information to estimate their probabilities the initial inclination was to convert them to the expression 'UNK'. However, substituting with 'UNK', increased the risk of skewing the model since the matrix Y could have high occurrences of 'UNK'. Conversely, eliminating them brought about the risk of not been able to estimate the probability of rare but relevant words. Therefore, an approach was adopted where low frequency words that were word forms of already existing root words (words in their base form) would be converted to their base form on the assumption that

[80] The number of clusters is equal to the number of maxent models

the speaker expressed both words in the same context. This process is called lemmatization. For instance, the word *'phrases'* occurred just once in EU-LD data, however the word *'phrase'* occurred 3 times. For this, it was assumed that the context in which the speaker used both expressions was the same and so *'phrases'* was reduced to its base form giving the effect of increasing the count of *'phrase'* (the base form of *'phrases'*) by 1, and consequently pretending that *'phrases'* never existed[81]. On converting low frequency words (words occurring once or twice) to their base form, it was observed that on average 31% of words occurring once were alternate word forms of already existing content words. The remaining 69% are converted to 'UNK' and their probabilities are determined by the LM LM_1.

Due to the addition of the clustering process as a means for reducing the dimensionality of the matrix, VSEA also underwent some modification to account for the clusters. It is important to note that this modification is specific to this implementation. The modified VSEA algorithm is described in appendix 11, and the sentiment prediction algorithm with the modified VSEA is described in appendix 12.

This concludes the implementation of the model. In total, 6 models are implemented for the following context-party pairs: EU-Conservative, Immigration-Conservative, EU-Labour, Immigration-Labour, EU-LD, Immigration-LD models. Each model consists of a language model LM_1 and a maxent model LM_2. The next chapter, describeS the tests carried out on the implemented model.

7.8 Conclusion

This section has described the implementation of a use case featuring data from the political domain. The data sources have been drawn from manifestos and debates because they are good sources of implicit and explicit enumeration of party values.

The steps involved in developing the model are in sync with the processes described in chapter 6. Context in this implementation is described as the subjects of the documents or the utterance. This chapter also suggests that considerable computational power is required to implement the model. The lack of power was ameliorated using a modified VSEA algorithm featuring a word clustering algorithm designed to reduce the dimensionality of the training set matrix. As part of proposed future tasks, the unmodified VSEA algorithm should be implemented and its performance compared against the modified VSEA implementation.

Finally, the features of the model implemented were deemed to be quite generic, corpus and domain independent. Theoretically, it is expected that if the features are semantically enhanced to suit the domain of interest and tailored to the document set, then potentially, the model would perform better. This assumption is tested in the next section, where a

[81] Note that this transformation is also captured in the feature set.

semantically enhanced model is implemented on the same training corpus and compared against the performance of the model built in this chapter.

8. Testing and Evaluation

This chapter describes relevant tests carried out, followed by an evaluation and discussion of the results. Established performance metrics - precision, recall F-score, accuracy and misclassification rate are applied in evaluating the test results. Definitions of these evaluation metrics and relevant formulas can be found in appendix 18.

This section begins by describing the relationship between the artifacts described so far: VSM, VSEA, Feature switching (FS) and Recipient Sentiment Prediction Algorithm (RSPA). VSM refers to the methodology describing the journey from abstract values to full sentiment prediction. Within VSM is the VSEA, which is an algorithmic formalization of the values embedded in value laden sentences. The outcome of VSEA on a sentence is a probability estimate. RSPA refers to the algorithmic model used in estimating the sentiment of a recipient by combining VSEA and FS. It is the eventual outcome of the VSM.

Testing was carried out on five cases:

1. To determine the VSM's overall performance, evaluated by comparing the predictions of VSM on a test set against ground truth of the same test set annotated by actual value holders.
2. **Semantic Enhancement**: So far, the data preparation processes and the features used in estimating content word probabilities are generic and corpus independent. However, there are potentially additional features and data preparation methods that can be applied in the VSM to capture additional semantics, embedded in the sentences. This process is called semantic enhancement. In this test case, the existing model is compared against a separate semantically enhanced model. The data preparation of the semantically enhanced model is less flexible allowing for more domain dependent processes and the inclusion of additional features relevant to the named entities. The outcome here is to determine if the model performs better with generic features versus semantically enhanced corpus dependent features.
3. The sentiment prediction model is compared against contemporary sentiment analysis Implementations.
4. The VSM's ability to estimate the sentiment of objective and subjective sentences.
5. Does the VSM reflect the value of actual value holders on subjects.

The criteria described above are summarized using the following test and evaluation scenarios.

Test/Evaluation Scenario 1

- Test: The overall performance of the VSM (both domain independent VSM and semantically enhanced VSM) in predicting recipient sentiment. Because the model in this scenario is referenced in subsequent sections, for ease of referencing, it is called the domain independent model 'm1' (short for model 1) while the semantically enhanced model is called 'm2' (see test scenario 2).

- Evaluation: Using the evaluation metrics, the model's performance is evaluated against gold standard ratings.

Test/Evaluation Scenario 2

- Test: The effect of semantic enhancement on the VSM. Model m2 is built using the same data in 'Test Scenario 1', however in this scenario, the data preparation process is modified to capture relevant semantic features while expanding on the feature set used in estimating content words.

- Evaluation: Model (m2), is evaluated against a gold rated test set and its performance is compared to m1.

Test/Evaluation Scenario 3

- Test: The objective of this test is to compare the model's performance against traditional SA implementations which classify the sentiment expressed in the sentence and not the sentiment of the recipient. Expectedly, the result of this test should show a performance drop for traditional models.

- Evaluation: Compare the results of the VSM in predicting the recipient's sentiment against the performance of contemporary SA implementation – Sentiwordnet (Baccianella et al, 2010).

Test/Evaluation Scenario 4

- Test: VSM's Performance on Objective and Subjective Sentences. One of the benefits of applying a value model in SA is its ability to capture implicit and explicit sentiments expressed in objective sentences. In this test, the test set is separated into objective and subjective sentences and the performance of m1 and m2 on each set is observed.

- Evaluation: Compare the m1 and m2 classification against gold rated test set and determine performance using evaluation metrics.

Test/Evaluation Scenario 5

- Test: Since the VSM is a value based model, this test determines if the VSM models (m1, m2) reflect the values expressed by the value holders.

- Evaluation: Compare the result of gold rated sentences by value holders against the classifications made by VSM.

The next section, describes the test set and the method used in its generation.

8.1 Determining the Test Set Methodology

The normal evaluation method in NLP is to build a model from a training set and test the performance of the model against a test set. Using this approach would result in a biased evaluation. To understand why, the goal of this research is reiterated: To predict the sentiment of a recipient given a sentence (the recipient is also a value holder). Under this objective, applying a VSM model built from a given context-party pair on a test set drawn from the same context-party pair will most likely produce higher probability estimates than a feature switched version of the same sentence. Based on the sentiment methodology for which this thesis is founded upon, this is simply evaluating how likely it is for the value holder to make the sentences in the test set versus how likely it is to make a contrary sentence. Since the test sentences are owned by the value holder, such results are already biased. To avoid this bias, the test set had to be sentences that were unseen by the value holder (i.e. sentences that are not made by the value holder). This strategy lead to a second issue, where the value orientation of the value holders on the unseen tests sentences are also unknown. To address this, the following strategies were used:

1. Since a corpus of test sentences for each context-party pair existed, instead of testing a context-party pair model on its associated context-party pair test set, it was tested against the context-party test set of opposing parties. This is done on the assumption that the same core expressions will be used across all parties as the contexts are the same. For instance, instead of testing, *'EU-Conservative'* model on test data drawn from the Conservatives, the model was tested against a test set drawn from the other EU- '*party*' pairings i.e. *'EU-Labour'* and *'EU-LD'*.

2. Volunteer value holders were required to judge and rate selected test sentences from parties' other than theirs. So, given a Conservative value holder, he/she would rate test sentences from Labour and LD, and this would form the ground truth. This addresses the question of the value holders' orientation towards unseen sentences. An added advantage to this approach is that, the similarity between the parties could be determined from the gold standard ratings and the results compared against the estimations made by the VSM.

3. By applying context-party pair models on different party test sets, that the process was not uniform because each model was applied to a different data set. To bring about uniformity, volunteers also rated the UKIP sentences compiled earlier to be used as a baseline model for all value holders.

4. In addition to human rated test sentences, additional test sentences were generated from actual value holders and their sentiments determined using the VSM. These additional sentences were derived through a group discussion involving the judges.

Thus, a key part of the setup involved soliciting the aid of value holders who would rate/judge sentences and participate in a focus group discussion. The core criteria for participation were:

- A participant should be registered with one of the four UK political parties mentioned in this research (Conservative, Labour, LD, UKIP)[82]. Difficulty in finding registered respondents meant lowering this criterion to anyone who had voted for the same political party in the last two general elections (2010, 2015).

- Participants were required to have voted in the last general election[83] and the EU referendum 2016. This research was unable to recruit anyone with UKIP leanings.

Initially 10 people responded but 3 withdrew. The make-up of the remaining 7 was:

- 3 Conservatives, 2 Labour and 2 LD.
- Their ages ranged from 31 to 55 and the average age was 40.7[84].
- All the participants had University level qualifications ranging from a Bachelor's degree to a Doctorate.
- All the participants were employed.
- All the participants were British and married/civil partnership.
- All the participants had voted for the party of their affiliation in the last two general elections and claimed they would vote the same way if an election were announced at the time of this documentation.

Appendix 13, shows a summary of the participants. Participants offered their informed consent by signing an information sheet as required by the University's ethical approval procedure.

8.2 Test Set Generation

Two test set types were generated: The first was drawn from the corpus (let's call this Corpus-based test set) while the second is generated from focus group discussion (this is called focus-group test set). This section, discusses each test set generation and how gold standard ratings were obtained.

---

[82] Unfortunately, only 3 out of all respondents were registered with a political party. 2 were registered conservatives and 1 was registered Labour. Unfortunately, the registered Labour respondent pulled out for personal reasons.

[83] UK General Election 2015

[84]In conducting the tests, it was assumed that participants' age made no difference in their judgements. It is recommended that future experiments should consider the diversities in judgement observed amongst the different age groups.

8.2.1 Creation of Corpus Based Test Set

Table 20 in chapter 7 shows the number of test sentences. Due to the volume of this set, rating all the sentences by the judges was impractical and so a sample was taken. 250 sentences were randomly selected to be rated from each context-party test set. Each generated sentence was designed to contain at least 3 content words.

In addition, the UKIP documents collected earlier (see table 19) were prepared. Using the same preparation processes applied on previous documents, sentences were derived from the UKIP documents. A total of 75 Immigration and 95 EU sentences were extracted. Appendix 14 shows a sample of some of the sentences. The next objective was to compile a test corpus for each context-party pair. To illustrate corpus creation, consider the creation of a test corpus for EU-Conservative. The following were merged: All 95 UKIP sentences, the 250 randomly selected sentences from the EU-LD and EU-Labour, bringing the total number of test sentences for EU-Conservative corpus to 595. Observe that sentences from the Conservative test set are not in this corpus. This process was repeated for Immigration-Conservative, Immigration-Labour, Immigration-LD, EU-Labour and EU-LD pairs. The immigration-[party] pairs comprised of 575 documents. By taking such a sample across context-party pairings, variations in the VSM's performance could be observed.

With the test corpus defined, the next task involved ground truth rating of sentences by the judges. The major reason behind this rating exercise was to identify sentences for which the judges agreed on their ratings, and consequently attain a subset of correctly rated test sentences. In addition, this process enabled the elimination of unnecessary or definitional sentences. Thus, participants were asked to rate sentences in the newly created context-party pair corpus associated with their party. For example, Conservative participants were asked to rate EU-Conservative corpus and Immigration-Conservative corpus test sentences. They were required to rate each sentence by considering if it was in line with the policies and overall views of their party and its members. They were instructed to make their considerations based on their party's policy and its members' overall perception of EU/Immigration policy. The need for participants to make their ratings not based on their own personal views but based on their knowledge and judgement of party members, values and policies was emphasized.

Participants were not informed of the provenance of the sentences instead they were informed that the sentences were extracted from a selection of political transcripts and interviews. This was done to not induce biases associated with being aware of the source of the utterance.

Sentences were rated according to the ratings outlined in table 27. It was determined not to use explicit ratings of sentiment to protect against participants making any direct judgements that might mask their biases. In addition, ratings were designed to solicit their judgements on their party's values which were translated to sentiment prediction such that if a sentence is 'in-line with a party's policies and views of most members' then it is inferred that the sentiment of the sentence is positive and vice-versa. Unrated sentences were assigned a rating of '0' and attributed this to uncertainty or a judgement of '*it could go either way*'.

Table 27: Participant Sentence Rating Score and Meaning

Rating	Meaning
-2	Definitely not in-line with party's policies and views of most members
-1	not in-line with party policies and views of most members
0	It could go either way or sentence does not invoke any sentiments
1	In-line with party Policies and view
2	Definitely and strongly in-line with party's policies and views of most members

Overall, the agreement using the fine-grained rating in table 27 was quite low across board. For instance, out of 595 sentences in Conservative-EU test corpus, the 3 participants assigned the same rating to 186 sentences. Of these 186 sentences, they agreed on ratings other than 0 to 102 sentences, resulting in percentage agreement of 19.9% (the total number of unrated or 0 rated sentences was 84). For Conservative-Immigration test corpus the three participants assigned ratings other than 0 to 100 sentences, resulting in agreement of 19.08% on all non-zero rated or unrated sentences. Appendix 15 illustrates this break down.

Given the low agreement between Conservative participants, ratings were merged such that sentences assigned [-2 or -1] ratings were jointly rated [-1] and sentences assigned [1 or 2] ratings became [1]. In reducing the coarseness of the ratings, the agreement between the participants increased significantly to 41.68% for Conservative-EU non-zero or blank rated sentences and 54.96% for Conservative-Immigration non-zero rated or blank sentences. Ignoring the 0 rated test sentences, this increased the total number of test sentences to 213 and 288 respectively[85].

As for the ratings made by Labour and LD participants, the levels of agreement are tabulated in appendix 15. A higher level of agreement was observed primarily because the number of participants unlike the Conservatives is 2 and not 3. The coarser level of agreement between Labour and LD participants on their test sets was between 54% and 69% and the number of $[-101]$ rated sentences was sufficiently large enough to evaluate the model. Table 28 outlines, the total number of gold rated sentences for each context-

[85] Using this rating system i.e. $[-1,0,1]$, this research also considered the number of sentences for which at least two of the participants assigned the same rating and one participant assigned a rating of 0. Under this condition, it was assumed that since at least two of the participants agreed on a rating, with one unsure, then the overall judgement could be accepted. Here, the agreements for Conservative-EU and Conservative-Immigration increased to 61.4% or 314 sentences and 74.61% or 391 sentences respectively.

party pair as rated by the value holders. It highlights the attainment of the objective which was to generate a subset of correctly rated sentences.

Table 28: Total Number of Gold Rated Test Sentences

Context-Party	# Non-UKIP Sentences	# UKIP Sentences	Total	Focus Group Sentences	Total plus focus group sentences
EU-Conservative	168	45	**213**	11	**224**
Immigration-Conservative	250	38	**288**	26	**314**
EU-Labour	205	62	**267**	11	**278**
Immigration-Labour	153	64	**217**	28	**245**
EU-LD	305	82	**387**	12	**399**
Immigration-LD	308	65	**373**	32	**405**

8.2.2 Creation of Focus Group Test Set

In addition to the participants/judges, the setup involved, a moderator/facilitator and 2 observers tasked with taking notes. The facilitator was provided with a set of opening questions designed to engage the participants.

The facilitator's first question was based around a recent event: '*The Trump Muslim ban*'. The moderator posed the question to the participants in a way that elicited the views of their party and its members and not the participants themselves: "*How do you think members of your party would view this policy*". This evoked comments and discussion amongst the participants which were recorded. Subsequent questions by the facilitator followed a similar pattern e.g. "*Will members of the Conservative party vote for …*", "*Can the Labour party support…*", "*How do you think most LDs will react …*" etc. Due to the cross-over in issues related to Immigration and the EU, the questions veered from one concept to the other. The facilitator also introduced 5 questions drawn from UKIP's manifesto to observe the response of the participants. The participants were not informed that they were UKIP proposals[86].

After reviewing the transcript and notes, a total of 38 comments which the participants responded to were identified. Appendix 16 provides a list of the comments and appendix 17 summarises the responses made by the value holders. Each of the comments were manually

[86] Although they weren't informed that the questions were drawn from the UKIP manifesto, all of them were not oblivious to this fact, as they referred to this in the course of their exchanges.

classified into one of the contextual categories, so that of all 38 comments, 25 were solely on Immigration, 4 focused solely on the EU and 9 were classified as belonging to both the EU and Immigration. This brought the total number of Immigration comments to 34 and EU comments to 13. It was observed that of the 5 UKIP proposals, the Conservatives agreed with 3 of the proposals saying that most Conservative MPs were likely to agree and express such views, one was rejected out-rightly, and they were unsure of the last one. As for Labour, they disagreed with 4 of the five UKIP comments and were unsure of one[87]. LD participants disagreed with all the UKIP related comments. Some of the comments were restructured into sentences, since they were outlined by the facilitator as questions and the responses of the participants were recorded as gold standard ratings. Three ratings were derived following the comments made by the participants: These were:

'u': If the participants were unsure of their party's values and policies

'y': if the comment was in line with the party's values and policies

'n': if the comment was not in line with the party's values and policies

On reviewing the ratings ascribed by participants, there were sentences where participants from the same party did not agree. For example, Conservative participants failed to agree about the '*UK closing its borders to people coming from outside the EU*'[88]. Similarly, and surprisingly, the LD participants failed to agree on the question of supporting migrants coming into the UK from Syria and Turkey. Most of the disagreement came from Conservative participants. There were also cases where all the participants were unsure or simply disagreed and such sentences were assigned a rating of 'u'.

To determine which sentences from the focus group to include in the test set, all sentences where there was universal disagreement or uncertainty were excluded i.e. sentences with ratings of 'u' - only sentence 8 in Appendix 16. Sentences, for which there was disagreement within a group were not included in the group's test set. For example, in the case of the Conservatives, comments 2, 4, 21, 22, and 37 are excluded from the Conservative test set. Similarly, for LD and Labour test set, comments 21 and 37 respectively were excluded. In addition, comments for which one group was unsure while others assigned ratings were excluded for the group that rated it 'u'. For example, comment 10 "*Migrants have the right to make visa appeal after appeal and this must be stopped*" was excluded from the Labour test set. To this end, the test set sentence breakdown for the focus group is shown in table 29.

---

[87] UKIP view: "*Migrants should not have the right to make visa appeal after appeal*'. The Labour participants felt that "*Labour's policies towards migrants are reasonable but appeal after appeal might bug down the Judicial process*" and so their uncertainty was hinged on not knowing the overall effect on the Judicial process.

[88] Reframed as: "*Surely, the UK must not close her borders to people coming from outside the EU*"

Table 29: Focus group test set distribution

Context-Party	Number of Sentences
EU-Conservative	11
Immigration-Conservative	26
EU-Labour	11
Immigration-Labour	28
EU-LD	12
Immigration-LD	32

8.3 Test/Evaluation Scenario 1

8.3.1 Test: Overall Performance of Model VSM - 'm1', 'm2'

The goal of this test was to apply model 'm1' and 'm2' in predicting the sentiment polarity of the test sets identified in tables 28 and 29, and then compare the assigned polarity to the ground truth ratings assigned by judges/participants. Model 'm2' is described in greater detail in section 8.4. Using VSEA algorithm each context-party model is applied on its corresponding test set (Multiplying small probability scores can lead to numerical underflow, therefore log probabilities are added up as this is equivalent to multiplying in linear space[89]). A sample of confusion matrices illustrating the results of this test is presented in table 30 and 31. The confusion matrix has two classes: '-1' – Which translates to negative polarity because the sentence is not in line with party policies or views of most members and '1' - which translates to a positive polarity because the sentence is in line with party policies or views of most members. If the model's probability estimate of the sentence is greater than the feature switched probability estimate, it is classified as '1', else '-1'. From all the confusion matrices, the accuracy, precision, recall and misclassification rate for each class is estimated. Table 32 presents these results including the macro average precision and recall of the classifier. The next section evaluates and discusses the results of this test.

Table 30: Sample Confusion Matrix for EU-Conservative (Non-UKIP Sentences)

Total = 168	Assigned (-1)	Assigned (1)
True (-1)	55	28
True (1)	23	62

Table 31: Confusion-Matrix for all Sentences and Contexts by Model 'm1'

Total = 1865	Assigned (-1)	Assigned (1)
True (-1)	619	337
True (1)	179	730

[89] $p_1 \times p_2 \times \ldots \times p_n$ *becomes* $\log_2 p_1 + \log_2 p_2 + \cdots + \log_2 p_n$

8.3.2 Evaluation 1: Overall Model Performance

In applying DSR, an evaluation must describe the observations made in the test and a rigorous analysis of the limitations/benefits/conclusions that can be drawn. Therefore, each evaluation section, describes observations made in the test and provide a rigorous analysis of any limitations/benefits/conclusions that can be drawn. Overall, the evaluation metrics applied in evaluating the research are relative to the sentiment prediction classes positive (1) and negative (-1).

Table 32 shows the precision ($prec$), recall (rec) and F_1 score of each class for particular context-party pair test sets. They also include the macro average precision and recall of both classes, the overall accuracy, misclassification rate and average F_1 score derived by averaging the F_1 values of each class. In addition, the last row in the tables shows the result of the overall evaluation metrics which is calculated from all classified test sentences regardless of context or party.

Observations

The model m1 which is the model implemented without semantic enhancement attained an overall precision of 68.4%, a recall of 80.3% and an F_1 score of 73.8% for class 1 or positive predictions while it attained a precision of 77.5%, a recall of 64.7% and an F1 score of 70.5% for class -1/negative sentiment prediction. The overall average precision, recall, accuracy and F1 of m1 were 72.9%, 72.5%, 72.3% and 72.2% respectively. In total, m1 misclassified 516 sentences of the 1865 test sentences, correctly classifying 1349. These figures are illustrated in table 32, where the last row provides a computation of the overall average precision, recall, accuracy, misclassification rate and F1 score. Table 33 presents the evaluation results for model m2 which features the inclusion of corpus dependent data preparation and semantically enhanced features. A slight improvement in the overall average precision and recall is observed, as m2 yields an average precision of 73.3% and recall of 72.7%. This improvement is further evidenced in the slight drop in the misclassification rate from 27.6% in m1 to 27.4% in m2. In total, it misclassified 512 sentences, correctly classifying 1353. Accuracy increased from 72.3% in m1 to 72.5% in m2. On the surface, it appears that m2 performs better than m1, however further examination reveals that this is not entirely the case. The reasons for this is discussed in section 8.4.

Using the overall accuracy and average F_1 score as an overall metric, it is concluded that m1 and m2 attain an F1 score that lies in the range of 72.2% and 72.4% and an accuracy which lies between 72.3% and 72.5%.

From the results obtained, the observed performance of VSM is compared to existing models. Table 34 illustrates the overall performance of VSM against six existing models. The overall accuracy across all models ranges from 64% to 88.4%, and VSM's performance in comparison with other systems falls within this range. In addition, the VSM methodology is compared against the methodology applied in the models illustrated in table 34.

Table 32: Results of Model 'm1' Evaluation on Ground Truth Test Sentences

Context-Party Pair	# sentences	Class (1)			Class (-1)			Macro Prec	Macro Rec	Accuracy	Misclassified	Avg F_1
		Prec	Rec	F_1	Prec	Rec	F_1					
EU-Con (Non-UKIP)	168	0.688	0.729	0.708	0.705	0.6626	0.683	0.697	0.696	0.696	0.303	0.696
EU-Con (UKIP)	45	0.428	0.75	0.54	0.875	0.6363	0.736	0.651	0.693	0.666	0.333	0.638
Focus EU-Con	11	0.888	1	0.941	1	0.666	0.799	0.944	0.833	0.909	0.09	0.870
Total EU-Con	224	0.658	0.752	0.702	0.75	0.655	0.699	0.704	0.703	0.70	0.299	0.700
Immigration-Con(Non-UKIP)	250	0.733	0.818	0.773	0.739	0.633	0.682	0.736	0.726	0.736	0.264	0.728
Immigration-Con (UKIP)	38	0.458	0.647	0.536	0.571	0.380	0.457	0.514	0.514	0.5	0.5	0.496
Focus Immigration-Con	26	0.888	0.8	0.842	0.5	0.666	0.571	0.694	0.733	0.769	0.230	0.706
Total Immigration Con	314	0.714	0.8	0.754	0.703	0.597	0.645	0.708	0.698	0.710	0.289	0.700
EU-Lab (Non-UKIP)	205	0.797	0.844	0.820	0.661	0.585	0.621	0.729	0.715	0.756	0.243	0.720
EU-Lab (UKIP)	62	0.545	0.923	0.685	0.975	0.795	0.876	0.76	0.859	0.822	0.177	0.781
Focus EU-Lab	11	0.6	0.75	0.666	0.833	0.714	0.769	0.716	0.732	0.727	0.272	0.717
Total EU-Lab	278	0.758	0.848	0.801	0.787	0.674	0.726	0.772	0.761	0.769	0.230	0.763
Immigration-Lab (Non-UKIP)	153	0.705	0.827	0.761	0.758	0.611	0.676	0.731	0.719	0.725	0.274	0.719
Immigration-Lab (UKIP)	64	0.406	0.726	0.52	0.843	0.586	0.692	0.625	0.654	0.625	0.375	0.606
Focus Immigration-Lab	28	0.625	0.769	0.689	0.75	0.6	0.666	0.687	0.684	0.678	0.321	0.678
Total Immigration-Lab	245	0.629	0.803	0.705	0.784	0.601	0.680	0.706	0.702	0.693	0.306	0.693
EU-LD (Non-UKIP)	305	0.821	0.809	0.815	0.591	0.610	0.601	0.706	0.710	0.747	0.252	0.708
EU-LD (UKIP)	82	0.333	0.888	0.484	0.982	0.780	0.870	0.658	0.834	0.792	0.207	0.677
Focus EU-LD	12	0.75	0.75	0.75	0.875	0.875	0.875	0.8125	0.8125	0.833	0.166	0.813
Total EU-LD	399	0.77	0.811	0.79	0.743	0.693	0.717	0.757	0.752	0.759	0.24	0.754
Immigration-LD (Non-UKIP)	308	0.55	0.828	0.661	0.865	0.619	0.721	0.708	0.724	0.694	0.305	0.691
Immigration-LD UKIP	65	0.481	0.764	0.59	0.894	0.708	0.79	0.688	0.736	0.723	0.276	0.690
Focus Immigration-LD	32	0.666	0.428	0.521	0.652	0.833	0.731	0.659	0.630	0.656	0.343	0.626
Total Immigration-LD	405	0.546	0.781	0.643	0.846	0.65	0.735	0.696	0.715	0.696	0.303	0.689
Overall Evaluation	**1865**	**0.684**	**0.803**	**0.738**	**0.775**	**0.647**	**0.705**	**0.729**	**0.725**	**0.723**	**0.276**	**0.722**

To make this comparison, the models in table 34 are grouped based on their approaches and compare each group's methodology to the VSM. The two types of recipient SA approaches mentioned in chapter 2 form the groupings: Text and knowledgebase and socio-theoretic models.

Starting with the text and knowledge base dependent approaches i.e. (Sridhar et al (2014), Lin and Chen (2008) and Tang and Chen (2011). A summary of the steps taken in these works involves: the identification, collection and processing of data, followed by annotation and classification of the data by domain experts to identify sentiments, emotions and obtain a gold-standard corpus of ground truths. Subsequently, features are determined and extracted, a model is trained based on these features and finally evaluated and applied. In comparison, the VSM's methodology, illustrated in figure 4 and 8, also follows slightly similar steps: Data is collected and processed, value decomposition and feature selection is performed, followed by the implementation of the VSEA/Feature Switching and finally the application and evaluation. Figure 31 provides a comparison of both approaches. One significant difference exists.

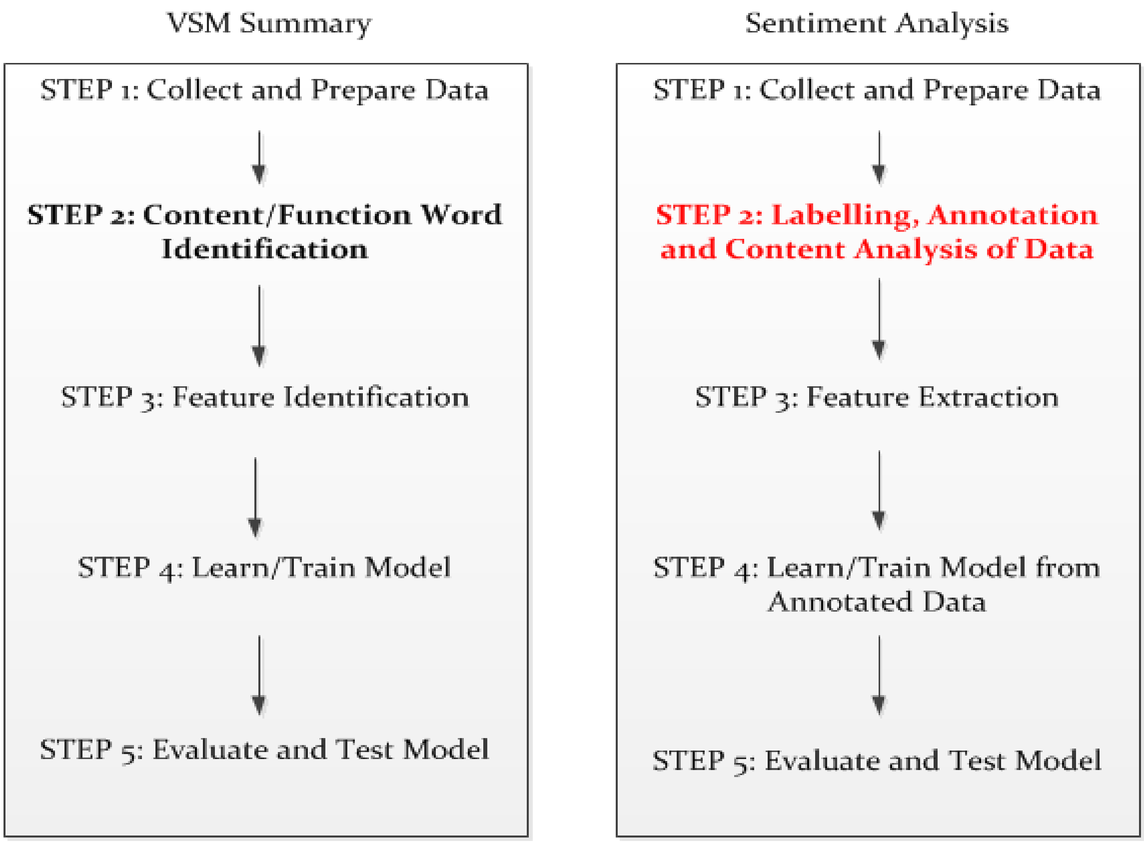


Figure 31: Comparative Summary of VSM Methodology and Recipient SA Methods

Observe that the VSM does not include the second step in the text/knowledgebase approaches requiring the annotation and classification of training data by human or domain experts. For instance, in Tang and Chen (2011), recipients are required to annotate and assign labels to sentences. The model of Lin et al (2001) requires a human tagged and labelled corpus to learn a model while the model of Sridhar et al (2014) required training set of hand annotated sentences. While this is not uncommon in SA research, the annotation and content analysis of training data by humans is an expensive and time-consuming exercise (Wissler et al, 2014; Takayama et al, 2014). It is also subjective because annotators or content analyst are required to make assignations based on their interpretations of the content, hence annotators are prone to bias. Furthermore, to prepare annotators, the researcher needs to provide guidelines and coding schemes. This requires some effort and still does not negate the fact that annotators are prone to making errors (Ho and Quinn, 2008). Another issue associated with annotation of a corpus lies in the difficulty associated with recruiting annotators, particularly in cases where the type of annotators required might not be easily accessible or have unique or specialist knowledge. Therefore, by been devoid of human annotations or content analysis, the VSM circumvents these deficiencies. In conclusion although the VSM's accuracy falls slightly below (Sridhar et al (2014), Lin and Chen (2008) and Tang and Chen (2011), by not been dependent on human annotation or content analysis, it provides an approach that negates all the deficiencies associated with human annotations and content analysis while still performing within the acceptable range of existing systems.

As for socio-theoretic models i.e. Ahothali and Joey (2015), Mejova (2012) and Bhomick et al (2009) (see table 34). The performance of these approaches shows accuracies in the range of 64.2% to 85.7%, precisions in the range of 68% to 82%. The approach also falls within this range. Nevertheless, as these approaches do not explicitly make use of human annotations, the socio-theoretic principles used are compared against the VSM and this thesis argues that at their core is a dependence on human input. Consider ACT: ACT has its foundations in the work of Osgood et al (1951), which proposes a method for determining the affective sentiment of cultures through surveys requiring individuals with a knowledge of a culture to rate concepts expressed as words on a numerical scale. This approach is not so different from the conventional approach to modelling values used by social scientist. In figure 3 'Illustration of the Five-Stage VSM', the process of obtaining the ACT scores is actually like the third goal, 'Categorization of enumerated value items into a value inventory' and the outcome which is the value inventory is not dissimilar to the ACT database.

The value inventory is an inventory of concepts derived through surveys by value holders while the ACT lexicon is also a database of concepts derived via surveys on groups or cultures. Essentially, models which make use of ACT are not different from the design methodology shown in figure 3, because the derivation of the sentiment is drawn from humanly derived scores of concepts. Like value inventories, a downside to the ACT database of scores/equations lies in its rigidity; If a database of scores for a particular culture or group does not exist then SA models cannot be built for such groups. Unlike ACT, the VSM is flexible in that content words do not have any fixed scores and are treated based on the surrounding words in the sentences. Thus, value models can be built and implemented for any group or culture. Finally, since socio-theoretic approaches are dependent on human inputs, they are also prone to been expensive and time consuming. These deficiencies are negated in the VSM.

As for limitations, although the VSM includes a shallow representation of the recipient, it is shown later that the inclusion of features capturing the relationship between the recipients and their social status played a major role in determining the sentiment expressed by judges and that its non-inclusion might have influenced the model's performance. In fact, the best performing model in table 34, Tang and Chen (2011) with an overall accuracy in the range of 80.67% - 88.37%, is worth considering. The features used in this system like the VSM consists of linguistic units.

However, Tang and Chen's work also captures additional features like user behaviour, social information and relationship of the hearers which was not captured in the VSM. This research proposes that these features be incorporated in future work.

In conclusion, VSM's performance falls within the range of existing systems and while it does not perform as well as some of the works illustrated in table 34, by been devoid of human influence in the form of annotation and content analysis, it offers an approach that is more flexible and devoid of the deficiencies associated with content analysis and annotation.

8.4 Test/Evaluation Scenario 2

8.4.1 Test Scenario 2: Generic Features vs Domain Dependent Features

Model m2 is implemented by modifying the data preparation process and introducing semantically enhanced features and data preparation techniques for estimating content word probability. It involves two parts: A modified data preparation stage and inclusion of domain dependent features in the maxent model. This model is applied on the test set and its performance is compared against m1. It is hypothesized that a semantically enhanced model will perform better than model m1.

Modified Data Preparation: In m2, the test sentences were prepared differently in order to assign more semantically relevant pseudo-words to the linguistic units. Domain relevant words are identified in the sentence and instead of using the generic pseudo-words used previously in section 7.4, more specific pseudo-words are used. For example, instead of using a generic pseudo-word as a replacement for all identified locations, unique pseudo-words which differentiate EU countries from non-EU countries are used. One benefit of this modification is that it further reduces the vocabulary size and hence the parameters to be estimated for the trigram model. Pseudo-word modifications were applied to:

Names – Since names are inexhaustible, the primary focus is on modifying the group of names which occur the most in the corpus. These are the names of UK MPs, position holders and the names of world leaders, presidents or Prime-Ministers.

Table 33: Results of Model 'm2' Evaluation on Gold rated sentences using Semantically Enhanced Domain Dependent Features

Context-Party Pair	Number of sentences	Class (1)			Class (-1)			Macro Prec	Macro Rec	Accuracy	Misclassified	Avg F_1
		Prec	Rec	F_1	Prec	Rec	F_1					
EU-Con (Non-UKIP)	168	0.747	0.764	0.755	0.753	0.734	0.743	0.750	0.749	0.75	0.25	0.749
EU-Con (UKIP)	45	0.36	0.75	0.486	0.85	0.515	0.641	0.605	0.632	0.577	0.42	0.563
Focus EU-Con	11	0.888	1	0.941	1	0.666	0.799	0.944	0.833	0.909	0.09	0.870
Total EU-Con	224	0.677	0.780	0.725	0.776	0.672	0.720	0.727	0.726	0.723	0.276	0.723
Immigration-Con(Non-UKIP)	250	0.704	0.811	0.754	0.714	0.580	0.640	0.709	0.695	0.708	0.292	0.697
Immigration-Con (UKIP)	38	0.52	0.764	0.619	0.692	0.428	0.529	0.606	0.596	0.578	0.421	0.574
Focus Immigration-Con	26	0.888	0.8	0.242	0.5	0.666	0.571	0.694	0.733	0.769	0.230	0.706
Total Immigration Con	314	0.698	0.805	0.748	0.696	0.561	0.621	0.697	0.683	0.697	0.302	0.684
EU-Lab (Non-UKIP)	205	0.841	0.903	0.871	0.783	0.671	0.723	0.812	0. 787	0.824	0.175	0.797
EU-Lab (UKIP)	62	0.423	0.846	0.564	0.944	0.693	0.8	0.683	0.77	0.725	0.274	0.682
Focus EU-Lab	11	0.6	0.75	0.66	0.83	0.71	0.769	0.716	0.732	0.727	0.272	0.717
Total EU-Lab	278	0.772	0.894	0.829	0.843	0.682	0.754	0.807	0.788	0.798	0.201	0.791
Immigration-Lab (Non-UKIP)	153	0.687	0.814	0.745	0.736	0.583	0.651	0.712	0.699	0.705	0.294	0.698
Immigration-Lab (UKIP)	64	0.451	0.777	0.571	0.878	0.630	0.734	0.665	0.704	0.671	0.328	0.652
Focus Immigration-Lab	28	0.625	0.769	0.689	0.75	0.6	0.66	0.687	0.684	0.678	0.32	0.678
Total Immigration-Lab	245	0.629	0.803	0.705	0.784	0.601	0.680	0.706	0.702	0.693	0.306	0.693
EU-LD (Non-UKIP)	305	0.835	0.823	0.829	0.622	0.642	0.632	0.729	0.732	0.767	0.232	0.730
EU-LD (UKIP)	82	0.333	0.888	0.484	0.982	0.78	0.870	0.658	0.834	0.792	0.207	0.677
Focus EU-LD	12	0.75	0.75	0.75	0.875	0.875	0.875	0.8125	0.8125	0.833	0.166	0.813
Total EU-LD	399	0.782	0.825	0.803	0.762	0.710	0.735	0.772	0.767	0.774	0.225	0.769
Immigration-LD (Non-UKIP)	308	0.523	0.792	0.630	0.835	0.593	0.694	0.679	0.693	0.665	0.334	0.662
Immigration-LD UKIP	65	0.454	0.882	0.6	0.937	0.625	0.75	0.696	0.753	0.692	0.307	0.675
Focus Immigration-LD	32	0.66	0.428	0.521	0.652	0.833	0.73	0.659	0.630	0.656	0.343	0.626
Total Immigration-LD	405	0.519	0.767	0.619	0.830	0.615	0.707	0.674	0.691	0.669	0.330	0.663
Overall Evaluation	**1865**	**0.682**	**0.816**	**0.743**	**0.785**	**0.639**	**0.704**	**0.733**	**0.727**	**0.725**	**0.274**	**0.724**

Table 34: Comparing overall performance of VSEA against existing models

Author	Evaluation Metric	Methodology Type	Data Type Used	Application
Ahothali and Joey (2015)	Precision: 68% - 82%	Socio-Theoretic	2080 news headline from news websites archives	Computing reader sentiment of news headlines using ACT
Mejova (2012)	Overall accuracy: 71.9% - 80.3% Accuracy of positive class: 64.2% - 85.7% Accuracy of negative class: 77.5% - 78.1%	Socio-Theoretic	Political discussions on YouTube, Twitter Posts	Computed sentence sentiment in political data using ACT lexicon and equations
Bhomwick et al (2009)	Overall F1 score – 82.1%	Socio-Theoretic	1305 news sentences, emotion classes considered include disgust, fear, happiness and sadness	Applied Framenet in classifying the emotion of readers
Sridhar et al (2014)	Overall F1 score – 74.0%	Used linguistic, structural features and human annotated data	Annotated collection of 109533 social and political forum posts from www.4forums.com	Implemented a stance detection in classifying the stance on gun control and gay marriage
Lin and Chen (2008)	Overall accuracy – 76.8%	Applied Linguistic features and a manually tagged corpus	37416 News Articles dating from January 24, 2007 to August 7, 2007	Estimating the emotions of readers.
Tang and Chen (2011)	Overall accuracy – 80.67% - 88.37%	Linguistic features, user behaviour and social information of readers	50000 Conversations from Plurk[90] platform, from June 21, 2008 to November 7, 2009	Estimating the emotion of readers in online forum
VSM	**Overall accuracy – 72.3% - 72.5% F1 score – 72.2% - 72.4%**	**Linguistic features, context and shallow knowledge of value holders**	**Hansard transcripts, manifestos and policy documents**	**Estimating the sentiment prediction of recipients**

[90] https://www.plurk.com/portal/: Last accessed 06/11/2017

A reason behind the semantic enhancement of the selected names is hinged on the notion that these group of people are known to have specific views on the subject matters considered in this research and so in renaming them with relevant pseudo-words, the objective is to associate them with value words and expressions that are representative of their values. External resources and gazetteers are used in resolving names. Using a list of MPs[91] and their political affiliation as a gazetteer, mentions of MP names are identified and replaced with the pseudo-word pattern 'PERSONNAME+[party]+MP'. For instance, mentions of 'Edward Miliband' become 'PERSONNAMELABOURMP'. Using a Gazetteer of World leaders, mentions of any Presidents or Prime Ministers are replaced with the pseudo-word pattern, '[Country Name]PRESIDENT'. '*Angela Merkel*' the chancellor of Germany becomes 'GERMANYPRESIDENT' and '*Vladimir Putin*' of Russia becomes 'RUSSIAPRESIDENT'. Conversion of world leader names to pseudo-words is limited only to the actual mention of the names. Expressions such as the 'German Chancellor' or 'the French President' are left unchanged. In addition to person names, unique pseudo-words are assigned to the most dominant acronyms and abbreviations (exclude alphanumeric). High frequency acronyms (acronyms with a frequency of 10 or more), were assigned the pseudo-word pattern [acronym]+ACRONYM e.g. 'NATO', 'NHS', 'TTIP' were substituted for 'NATOACRONYM', 'NHSACRONYM' and 'TTIPACRONYM' respectively. Low frequency acronyms like 'UNESCO' were treated in the same way described in section 7.4 (assigned the pseudo-word 'ABBREVIATIONNAME'). In adopting this technique, better probability estimates are attained for low frequency acronyms while reaping the benefits of treating high frequency acronyms as unique linguistic entities.

- **Locations** – Mentions of Countries were differentiated from mentions of cities. Similarly, locations in the UK were differentiated from locations outside the UK. A database and a gazetteer of UK cities and towns, was used as a reference for the renaming process. Given the nature of the context, two distinct gazetteers were compiled for cities and towns in the EU and the rest of Europe. By a process of elimination, cities identified to be outside Europe were simply renamed as 'NONEUROPEANCITYTOWN'. Mentions of EU and Non-EU cities and towns in the text were thus renamed using the convention, [Continent name] + 'citytown', where 'continent name' is substituted for 'EU' (for EU cities) or 'NonEU' (if the city is European but outside the EU). So, the location 'Berlin' would be renamed as 'EUCITYTOWN'. During the implementation, it was established that cities and towns across continents could have the same name, so based on the assumption

[91] A list of UK Parliament MPs since 2007 till 2015 is obtained using https://www.theyworkforyou.com/api/ - Last accessed 20-12-2015

that UK cities would be mentioned more than any other cities in the world, cities were renamed using the following steps:

1. Rename all occurrences of UK cities and town first.
2. It is assumed that the next most frequent mention of cities after the UK would be EU cities. So, all occurrences of EU cities were renamed.
3. Using the assumption in '2', it is assumed that Non-EU cities would be next in frequency and finally cities in the rest of the world.

Countries are not renamed because they are distinct and are treated as unique words. Hence, the approach taken in section 7.4 where countries were mapped to the pseudo-word 'LOCATIONNAME' is not adopted. The semantics of countries will be captured using additional context word features. In conclusion, location renaming is limited to places, cities and towns, thus concluding the first part of the semantic enhancement modification.

Additional Features for Content Word Prediction: The second part involves the inclusion of additional maxent model features which capture additional politically relevant named entities for countries. This feature, encapsulates some characteristics of countries such as if it is an EU country, a country in the British Isles or a country in the Middle East. Table 35 shows a list of 13 additional named entity features used bringing the total number of content word features to 254. This concludes the feature modifications. LM_1 and LM_2 models were reimplemented from this domain dependent processes, and applied on the test set. Table 33 details the result of the test.

Table 35: List of Additional Semantically Enhanced Named Entity Features

Content word (c) Features (Return 1 if feature is satisfied else return 0)
if c is an EU country
if c is a North-American Country
if c is a South-American Country
if c is a non-EU European country
if c is a non-EU non-European country
if c is an African country
if c is an Oceanian country
if c is a Middle-Eastern country
if c is an Asian country
if c is England
if c is Wales
if c is Northern Ireland
if c is Scotland

8.4.2 Evaluation 2: Effect of Domain Dependent Preparation and Semantic Enhancement ('m1' vs 'm2')

Observations

To observe the effect of semantic enhancement and corpus based data preparation, the results of m1 and m2 are compared. For the null hypothesis H_o there is no statistically significant difference between the performance of m2 and m1. The alternative hypothesis assumes that m2 will perform better than m1 i.e. semantic enhancement will improve the performance of the model. A paired t-test was run on 18 of the F1 scores of the sample context-party pair test data. P-value was calculated to be 0.4878. Since $p > 0.05$, there was no statistical significance, hence this research failed to reject the null hypothesis. The effect of this test is visualized in figures 32.

The topmost graph in figure 32, is a plot of differences between the F1 scores of m1 and m2. Points below the blue line indicate observations where m1 performs better than m2. Similarly, the bottom graph in figure 32, is a plot of paired values, m1 and m2 where circles below or to the right of the blue line (intercept=0, gradient = 1) indicate observations with higher values for m1 than m2. In both figures, the number of points above and below the blue line is almost the same - 5 and 6 respectively -, thus suggestive of negligible performance change. Figure 33, reinforces this, comparing individual F1 scores for m1 and m2. The 10th column in figure 33 shows that misclassification rate falls slightly from 27.6% in m1 to 27.4% in m2. To determine the reason for this negligible difference, this research began by comparing the precision, recall and F_1 score of context-party pairings in m2 which had better F1 scores than m1.

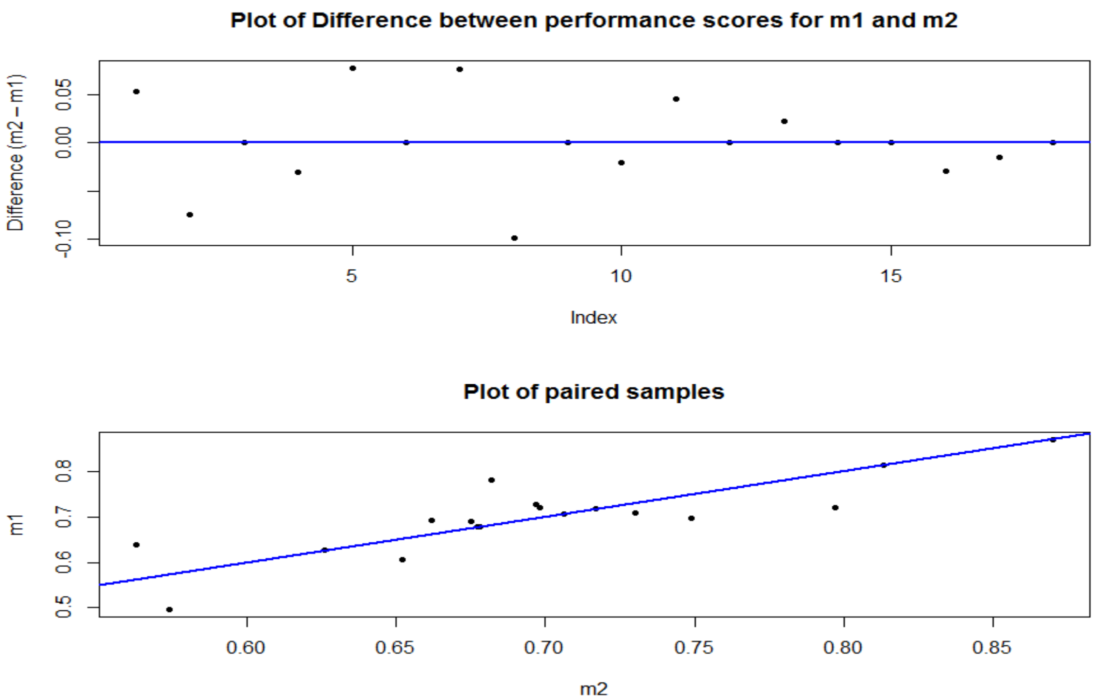


Figure 32: Point plot of F1 Scores for models m1 and m2

It was observed that there were 6 context-party pairings classified by m1 that attained higher F_1 scores as compared to 5 context-party pairings classified by m2. This is visualized in figure 34 which is a comparative bar graph comparing the overall F1 score for each context-party pair that was tested (including the overall F_1 score when all the data for each context party pair was aggregated). It was observed in figure 34, that model m2 performs better than m1 on data from the EU context, specifically the non-UKIP and non-focus group EU data (also observe that for aggregated context-party pair data, m2 performs better than m1 on EU-Con, EU-Lab and EU-LD). On Immigration type data, the reverse is observed except for Immigration-Lab (UKIP) and Immigration-Con (UKIP). This led to the conclusion that the modified data preparation in m2 targeted at certain named entities types appears to have had a greater influence on the EU data space than the Immigration data, and might have accounted for the slight improvement in performance. To confirm this, two related tasks were performed:

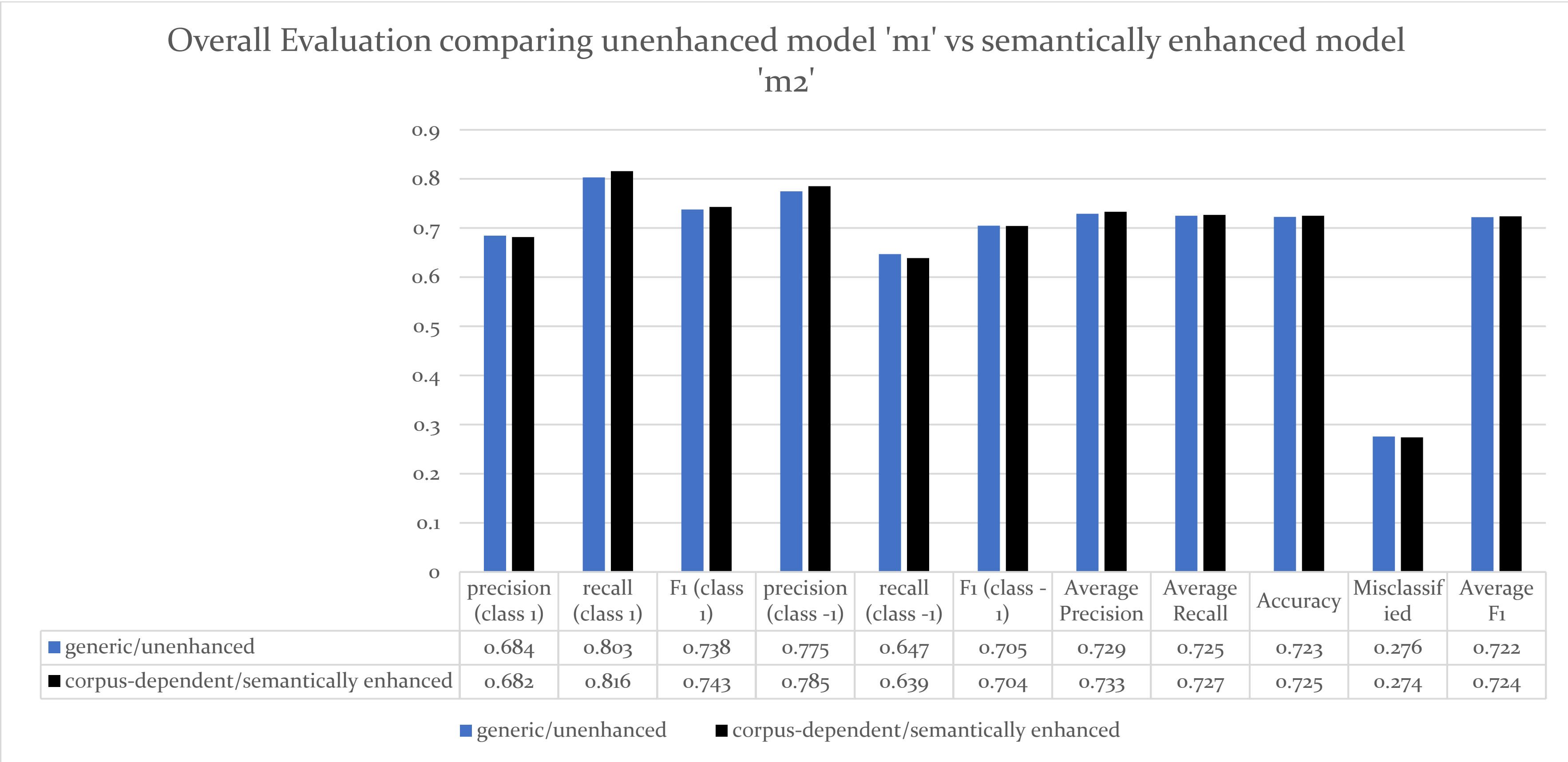
- First, the count of named entities in each EU training set was compared against the count of named entities in each Immigration training set. The named entities compared consisted of the named entities applied in the pre-processing task of m2 – MP Names, World Leader Names, UK Cities and Towns, EU Cities and Non-European Cities. Overall, there were considerably more mentions of world leader names, abbreviations/acronyms, cities in the EU training set than the Immigration training set, which meant that there were more sentences in the EU set that were affected by the preparation and modifications done by the m2 model. This is illustrated in figure 35 and appendix 19 which shows a wide variation in frequency for a sample of named entities taken for both the EU set, and the Immigration set.

- In the second task, training set sentences containing named entities modified in m2 are selected and the classifications assigned by both models observed. For example, in the test set belonging to EU-Conservative (Non-UKIP), of the 168 sentences, there were 24 sentences containing named entities that were pre-processed using the approaches in m1 and m2. Of the 24, 16 were wrongly classified in m1, while just 2 were wrongly classified in m2. Consider a sample sentence: "*The right of self-determination of Gibraltar must surely be respected above all*". The sentence was made by Conservatives and therefore a part of the Labour and LD test set. It was assigned positive polarity by Labour and LD judges. As for the models, the m1 model for both Labour and LD assigned a negative polarity to it whereas, the m2 model for both parties assigned a positive polarity. One of the reasons for the misclassification was that in changing '*Gibraltar*' to '*LOCATIONNAME*' in m1, some of the word association between the word '*Gibraltar*' and some of the semantically relevant words in its neighbourhood are lost. For instance, the word '*self-determination*' in the corpus was associated with the following countries (in order of frequency) – 'Palestine', 'Kashmir', 'Falklands', 'Western Sahara', 'Kosovo', 'Ukraine'. Table 36, shows the frequency of co-occurrence of the word 'self-determination' and the mention of a country. In m1, this is interpreted as 73 and 68 co-occurrences of '*LOCATIONNAME*' and '*self-determination*' pair in the EU-Labour and EU-LD sets respectively. When the feature vector for each of the sentences containing the expression pair '*LOCATIONNAME/self-determination*' was resolved, 60% (84 out of 141) of the occurrences of '*LOCATIONNAME'* or '*self-determination*' was in a dependency relation with a negation or in a relationship with a word with prior negative polarity. Since m1 associaties '*LOCATIONNAME*' and '*self-determination*' with negative features, it assigned a negative polarity to the sentence. Conversely, in m2, '*Gibraltar*' is treated independently as a unique entity and the generalization assumed in m1 does not occur. When the feature vectors were resolved for the 29 occurrences of '*Gibraltar*' in both 'EU-Labour' and 'EU-LD', just 5 negative associations between '*Gibraltar*' and '*Self-determination*' were identified. Thus, the over generalization applied in m1 was curtailed in m2 by not resolving the country. The conclusion reached was that semantic enhancement leads to a general improvement in sentiment prediction for sentences containing named entities, as is the case with the observed improvement in performance for EU content as compared to Immigration content.

Table 36: Frequency of Co-Occurrence of 'Self-Determination' and Countries

Countries	EU-Labour	EU-LD
Ukraine	19	21
Gibraltar	16	13
Palestine	14	9
Israel	8	11
Kosovo	7	6
Kashmir	4	4
Western Sahara	3	3
Falklands	2	1

Some downsides to m2 was observed. The additional data preparation and semantic enhancement increases the sensitivity of m2 making it more biased towards sentences containing named entities. The net effect is that because of the increase in named entity terms and their features, higher weights are assigned to them by the maxent model, and weights are taken away from other content words like adjectives or verbs. Consequently, the overall effect is that some sentences that were previously assigned a classification of -1/negative polarity are assigned a classification of 1/positive polarity, making the model more biased towards assigning positive ratings. For example, in Immigration-Conservative (Non-UKIP), with the addition of semantic enhancement, the model gets more biased towards non-feature switched sentences. It is observed that the number of sentences assigned class 1/positive polarity increases from 154 to 159 in m2, yet there is a drop in the number of correctly classified sentences from 113 to 112 in m2.



	precision (class 1)	recall (class 1)	F1 (class 1)	precision (class -1)	recall (class -1)	F1 (class -1)	Average Precision	Average Recall	Accuracy	Misclassified	Average F1
generic/unenhanced	0.684	0.803	0.738	0.775	0.647	0.705	0.729	0.725	0.723	0.276	0.722
corpus-dependent/semantically enhanced	0.682	0.816	0.743	0.785	0.639	0.704	0.733	0.727	0.725	0.274	0.724

Figure 33: Overall Evaluation comparing unenhanced model 'm1' vs semantically enhanced model 'm2

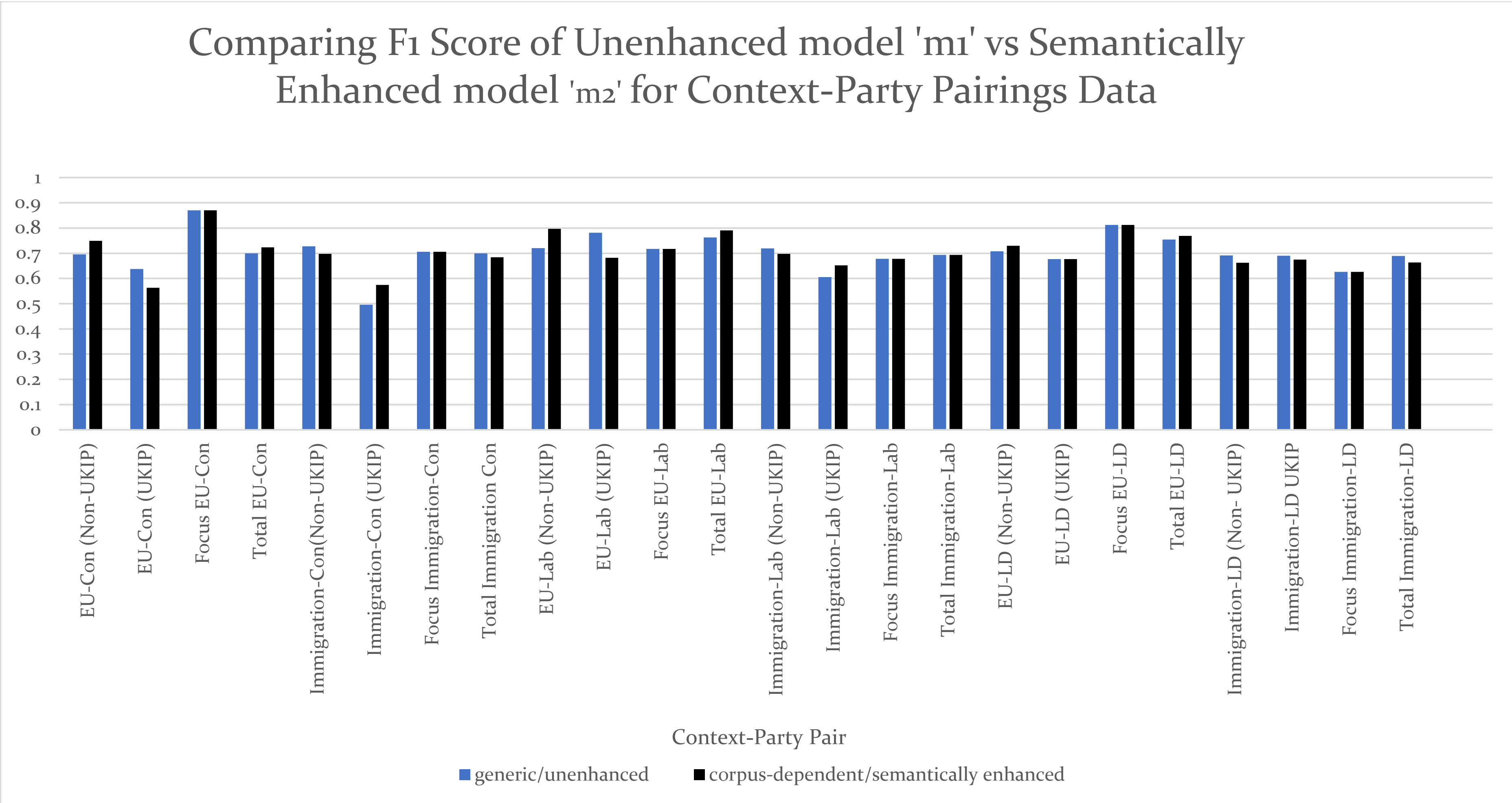


Figure 34: Comparing F1 Score of Unenhanced model 'm1' vs Semantically Enhanced model 'm2' for Context-Party Pairings Data

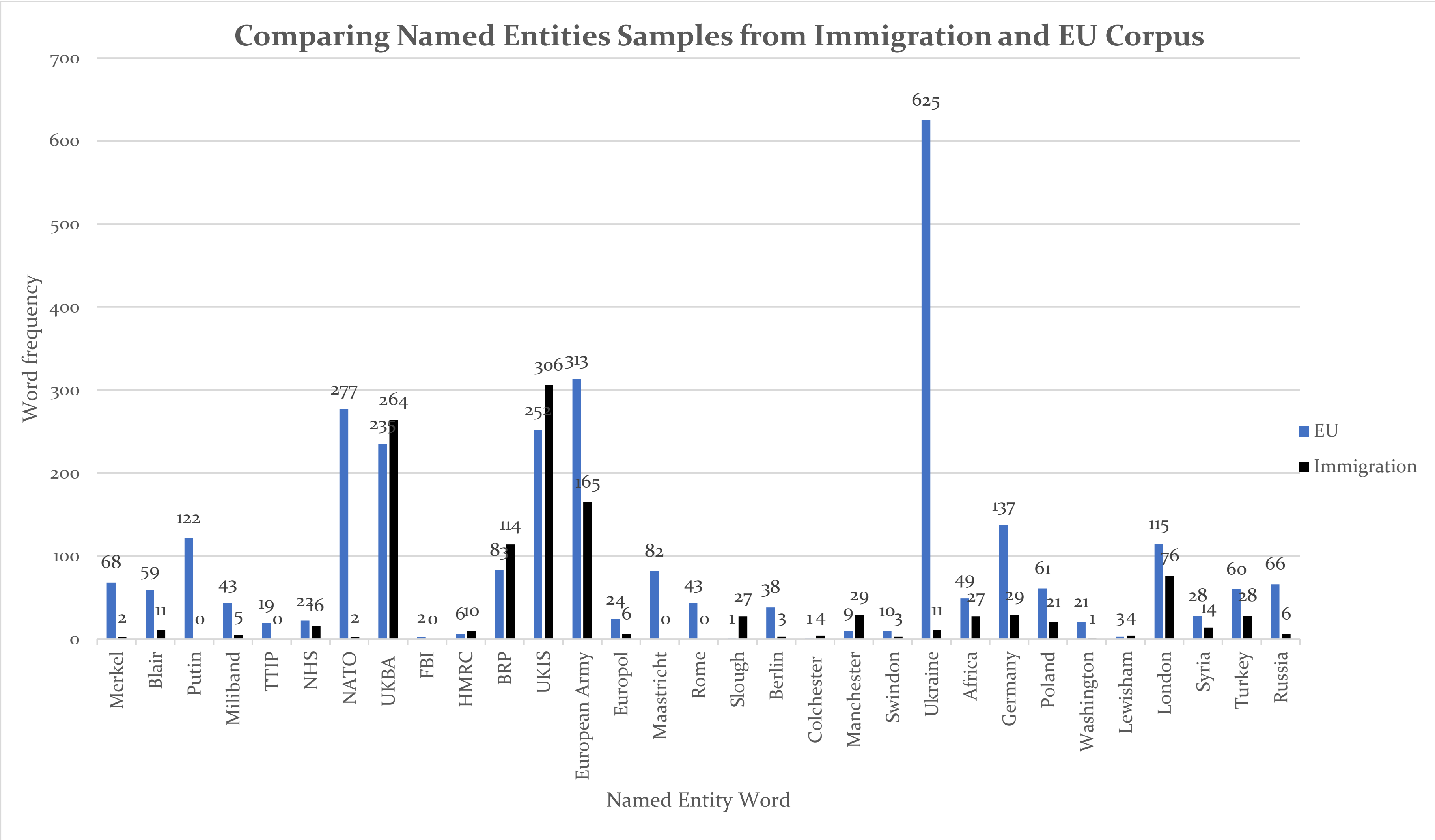


Figure 35: Comparing Named Entities Samples from Immigration and EU Corpus

This decrease in precision behaviour is also observed in EU-Con UKIP and Immigration-Labour context-party pairings, where correctly classified sentences assigned negative polarity were wrongly classified as positive by model m2. Thus, the semantic enhancement and corpus based data preparation process tries too hard to improve the system. Even sentences that did not contain locations, names or places were classified wrongly as compared to the previous class.

In conclusion, semantic enhancement and corpus/domain based data preparation results in:

- No statistically observed difference between the unenhanced and semantically enhanced model.
- An improvement in the model's ability to predict the sentiment of sentences containing the named entities that were semantically enhanced. This is observed as an increase in the F1 score for test corpus containing named entities. A 2.3%, 2.8%, 1.5%, 7.8%, 4.6% increase in F1 score is observed for EU-Conservative, EU-Labour, EU-LD, Immigration-Conservative (UKIP), Immigration-Labour (UKIP) corpus respectively.
- A slight drop in the model's precision in classifying negative sentences brought about by the model becoming sensitive and biased towards positive sentences.

8.5 Test/Evaluation Scenario 3

8.5.1 Test Scenario 3: Comparing VSM implementation against Contemporary SA Implementations

This test observes how poorly contemporary SA models perform when applied to recipient sentiment analysis. This is accomplished by comparing the performance of an Opensource SA implementation - Sentiwordnet - when applied to recipient sentiment prediction.

SentiWordNet (Esuli and Sebastiani, 2006; Das and Bandyopadhyay, 2010; Baccianella et al, 2010; Mejova, 2012) is based on the opensource lexical resource Wordnet (Miller, 1995), has been applied in classifying financial news (Devitt and Ahmad, 2007) and news headlines (Chaumartin, 2007). SentiWordNet provides a lexical resource of Wordnet synsets annotated based on the notion of semantic orientation. Each wordnet synset is associated to three numerical scores $Pos(s), Neg(s), Obj(s)$ which indicate how positive, negative and neutral the terms contained in the synset are (Baccianella et al, 2010). The scores are derived via a semi-supervised model. This test uses SentiWordNet 3.0 which contains 117658 words and assigns to each word a positive and negative score. For example, it assigns a positive score of 0 to the adjective *'last'*, and a corresponding negative score of 0.25. Since all scores add up to 1, the neutral score is equal to 0.75. While to the noun form of the same word, '*last'* it assigns a score of 0 to positive and negative, therefore, meaning a neutral score of 1. Sentiwordnet has been shown to have an accuracy of about 54% in Mejova (2012) and a precision of 57% in Das and Bandyopadhyay (2010).

In this research, the approach taken in estimating the sentiment of a sentence is by subtracting the Sentiwordnet negative score of each word in the sentence from the positive score to obtain the overall polarity score of the word. The overall polarity of the sentence is a measure of the occurrence frequency of positive and negative words and it is derived by summing over the polarity score of each word in the sentence. If the overall score is negative, the sentence is considered to have a negative polarity, otherwise if it is positive, it is deemed to have a positive polarity. If the score is 0, then the sentence is neutral. For the purposes of this test, neutral classifications are assumed to be wrongly classified while positive or negative classifications are compared against the ground truth annotation. Table 37 shows the overall precision and recall when Sentiwordnet is applied on the test set.

8.5.2 Evaluation 3: Model's Performance Compared to Sentiwordnet Observations

Understandably, the performance of Sentiwordnet is poor because it would always produce the same sentiment classification. Table 38, compares the accuracy and misclassification of VSEA and Sentiwordnet on the sentiment prediction task. The difference in performance and misclassification observed is significant, with Sentiwordnet misclassifying 74.7% of test data. This test also highlights how ill-suited traditional SA approach is to sentiment prediction. Earlier, it was indicated that the accuracy of Sentiwordnet was in the range of 54%-57% for sentiment reporting tasks. Comparing these figures to the accuracy shown in table 39, demonstrates that the accuracy of Sentiwordnet drops from 57% to 25.2% when applied to recipient SA.

In conclusion, compared to the VSM, Sentiwordnet, will assign the same sentiment classification to a sentence. For example, on the sentence, "*Opposed to Turkey's membership of the EU*", Sentiwordnet assigns a neutral sentiment. VSM however, assigns diverse sentiments to different groups – assigning a positive sentiment to Conservatives and a negative sentiment to Labour. This highlights a major limitation of Sentiwordnet which is its dependence on prior probabilities that do not consider contextual factors and recipient values in the sentiment determination problem. The effect of this is that in using traditional SA approaches, only one sentiment class can be assigned, whereas the VSM can predict multiple sentiments for different recipients or contexts. Therefore, traditional SA models are unsuited to sentiment prediction and are likely to experience a significant decrease of about 55% when applied to recipient sentiment prediction. Also, a benefit of the value based SA is its ability to predict different sentiments for the same sentence for different recipients.

Table 37: Evaluation Results showing the application of SentiWordnet on Gold rated sentences

Context-Party Pair	Number of sentences	Class (1)			Class (-1)			Macro Prec	Macro Rec	Accuracy	Misclassified	Avg F_1
		Prec	Rec	F_1	Prec	Rec	F_1					
EU-Con (Non-UKIP)	168	0.209	0.2	0.204	0.218	0.228	0.223	0.214	0.214	0.214	0.785	0.214
EU-Con (UKIP)	45	0.111	0.25	0.153	0.5	0.272	0.352	0.305	0.261	0.266	0.733	0.253
Focus EU-Con	11	0.6	0.375	0.461	0.166	0.333	0.222	0.383	0.354	0.363	0.636	0.341
Total EU-Con	224	0.196	0.219	0.207	0.261	0.235	0.247	0.228	0.227	0.228	0.771	0.227
Immigration-Con(Non-UKIP)	250	0.109	0.072	0.08	0.194	0.276	0.228	0.152	0.174	0.164	0.836	0.158
Immigration-Con (UKIP)	38	0.15	0.176	0.162	0.222	0.190	0.205	0.186	0.183	0.184	0.815	0.183
Focus Immigration-Con	26	0.666	0.4	0.5	0.142	0.333	0.2	0.404	0.366	0.384	0.615	0.35
Total Immigration Con	314	0.170	0.12	0.140	0.193	0.266	0.224	0.182	0.193	0.184	0.815	0.182
EU-Lab (Non-UKIP)	205	0.257	0.133	0.175	0.133	0.257	0.175	0.195	0.195	0.175	0.824	0.175
EU-Lab (UKIP)	62	0.075	0.230	0.113	0.545	0.244	0.338	0.310	0.237	0.241	0.758	0.225
Focus EU-Lab	11	0.166	0.25	0.2	0.4	0.28	0.333	0.283	0.267	0.272	0.727	0.266
Total EU-Lab	278	0.179	0.138	0.156	0.186	0.238	0.209	0.182	0.188	0.183	0.816	0.182
Immigration-Lab (Non-UKIP)	153	0.238	0.197	0.216	0.244	0.291	0.265	0.241	0.244	0.241	0.75	0.241
Immigration-Lab (UKIP)	64	0.25	0.555	0.344	0.666	0.347	0.457	0.458	0.451	0.406	0.593	0.400
Focus Immigration-Lab	28	0.285	0.307	0.296	0.357	0.333	0.344	0.321	0.320	0.321	0.678	0.320
Total Immigration-Lab	245	0.247	0.267	0.257	0.338	0.315	0.326	0.293	0.291	0.293	0.706	0.292
EU-LD (Non-UKIP)	305	0.5	0.309	0.382	0.171	0.315	0.222	0.335	0.312	0.311	0.688	0.302
EU-LD (UKIP)	82	0.058	0.333	0.1	0.806	0.342	0.48	0.432	0.337	0.341	0.658	0.290
Focus EU-LD	12	0.125	0.25	0.166	0.25	0.125	0.166	0.187	0.187	0.166	0.833	0.166
Total EU-LD	399	0.365	0.309	0.334	0.266	0.318	0.290	0.315	0.313	0.313	0.686	0.312
Immigration-LD (Non-UKIP)	308	0.175	0.270	0.212	0.408	0.284	0.335	0.292	0.277	0.279	0.720	0.274
Immigration-LD UKIP	65	0.125	0.294	0.175	0.52	0.270	0.356	0.322	0.282	0.276	0.723	0.265
Focus Immigration-LD	32	0.214	0.214	0.214	0.388	0.388	0.388	0.301	0.301	0.312	0.687	0.301
Total Immigration-LD	405	0.168	0.267	0.207	0.422	0.288	0.343	0.295	0.278	0.281	0.718	0.275
Overall Evaluation	**1865**	**0.226**	**0.222**	**0.224**	**0.276**	**0.281**	**0.278**	**0.251**	**0.251**	**0.252**	**0.747**	**0.251**

Table 38: Comparing the accuracy and misclassification of VSEA and Traditional SA models

Model	Accuracy	Misclassification
VSEA	72.3% - 72.5%.	27.4%-27.6%
Sentiwordnet	25.2%	74.7%

8.6 Test/Evaluation Scenario 4

8.6.1 Test Scenario 4: VSM's Performance on Objective and Subjective Sentences

Evaluation of VSM on objective evaluative sentences, involved first identifying candidate sentences followed by reviewing the classifications assigned by the VSM. Since the test set consisted of 1865 sentences, subjective objective sentences were manually separated. Three volunteers including the author of this research were tasked with annotating the test sets into objective and subjective sentences. The option of automating this task was considered but because the dataset was reasonably sized, manual annotation was chosen. By manually annotating sentences, there was a reasonable certainty of obtaining sensible annotations. 604 sentences were classified as objective and evaluative, while 1261 were classified to be subjective by the annotators. Afterwards, the performance of m1 and m2 models in correctly predicting the sentiments of the objective and subjective sentences was observed. Table 39 provides a breakdown of the evaluative metrics for the application of both m1 and m2 models on the objective and subjective sentences.

8.6.2 Evaluation 4: Model's Performance on Objective and Subjective Sentences

Observations

As seen in table 39, overall, VSM performs averagely for objective sentences. A comparatively low precision of 53.5% and 55.1% respectively is observed for m1 and m2 in predicting '1' classified sentences. While both models produce relatively high recall for positive classified sentences, respectively 77.4% and 80.5%. In the case of subjective sentences, a considerable difference in performance is observed. The precision and recall of the m1 model for positive classified sentences is considerably higher than the values attained in the case of the objective sentences. A precision of 71.3% is observed for positive classified sentences using the m1 model and a recall of 78.4%.

Model m2 improves on this, increasing precision slightly to 72.5% and recall to 80%. Good scores are also attained for negative classified subjective sentences, with m1 producing a precision of 71.1% and a recall value of 62.8%. m2 improves this slightly with a precision of 73.1% and a recall of 64.1%. Overall the average F1 score for the model's evaluation of subjective sentences is 70.7% with m1 and 72.2% with m2.

Table 39: Results of Model Evaluation on Objective and Subjective Sentences

Confusion Matrix Name	Number of sentences	Class (1)			Class (-1)			Macro Prec	Macro Rec	Accuracy	Misclassified	Avg F_1
		Prec	Rec	F_1	Prec	Rec	F_1					
Objective Sentences - m1	604	0.535	0.774	0.632	0.815	0.597	0.690	0.675	0.686	0.663	0.336	0.661
Objective Sentences – m2	604	0.551	0.805	0.654	0.839	0.608	0.705	0.695	0.706	0.682	0.317	0.680
Subjective Sentences – m1	1261	0.713	0.784	0.747	0.711	0.628	0.667	0.712	0.706	0.712	0.287	0.707
Subjective Sentences – m2	1261	0.725	0.800	0.761	0.731	0.641	0.683	0.728	0.721	0.727	0.272	0.722

The following conclusions are made from these results:

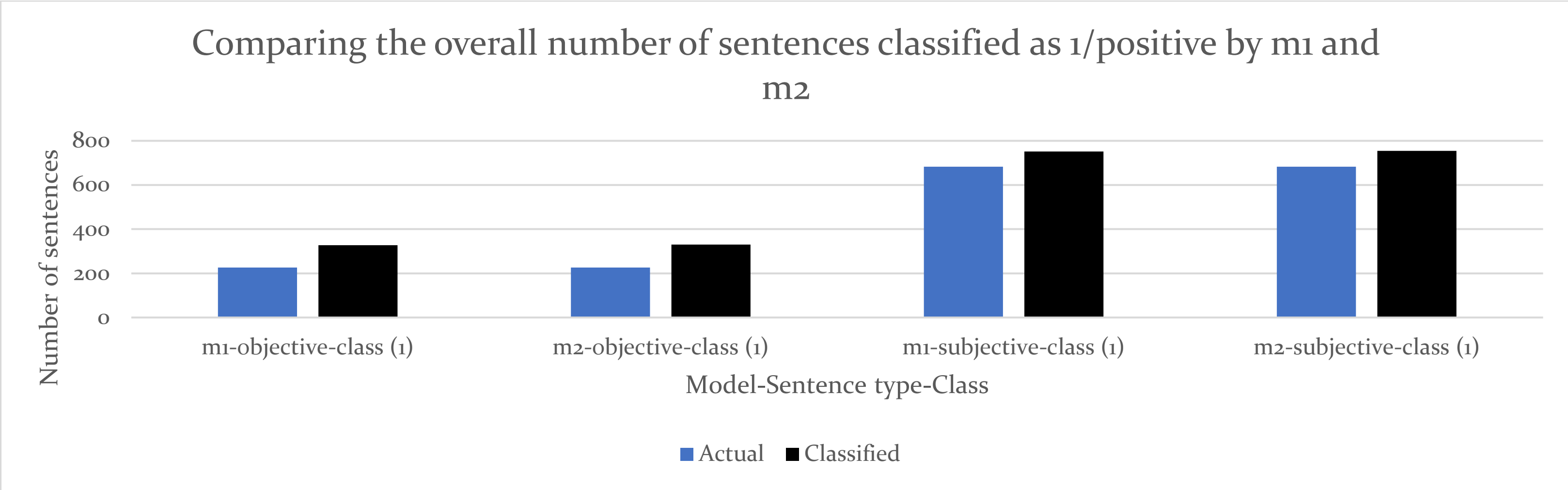
- The VSM (m1 and m2) performs reasonably well on subjective sentences – 70.7% and 72.2% for m1 and m2 respectively.
- Although F_1 score of the VSM on objective evaluative sentences falls just slightly below the 70% mark for m1 and m2, the precision for positive rated sentences is considerably low (53.5% for m1 and 55.1% for m2). However, the recall is high. The converse is the case for negative rated sentences where the precision is relatively good, but recall is low. This behaviour is attributed to the VSM's tendency to assign more positive classifications than negative. This bias for positive rated sentences which are sentences that have not been feature switched is observed in figure 36 (top figure). Figure 36 compares the number of sentences classified as positive or negative by the VSM to the total number of positive or negative sentences classified by the judges. It shows higher recall for positive sentences than negative sentences. This behaviour is as a result of feature switching (FS). When a sentence is feature switched, a new vector is invented for any content word contained in the sentence and so when the weights for this new vector is estimated there's likely to be more evidence for estimating the probability of the original content word than its feature switched version. In other words, the LM is biased towards content it has seen than content it has not seen. This is a limitation of applying lexical features.

8.7 Evaluation 5: Does the model's (m1, m2) values reflect the values expressed by the Judges?

Observations

To address this criterion, this research attempts to draw some conclusions from the classifications made by the judges and then determines if VSM's predictions matches these classifications. To make this judgement, the classifications assigned by judges to the UKIP test data are compared to classifications made by the model. The UKIP test data is used because it was annotated by all the judges and so represents a common denominator from which reasonable conclusions on the values held by the judges can be drawn. Essentially, with this test, inferences can be drawn as to how similar or dissimilar the judges' values are to UKIP sentences and compare this inference against inferences drawn from the model's classification of UKIP sentences. The methodology for this evaluation involves:

1. Identify the percentage agreement and disagreement for each context-party UKIP test set as classified by Conservative, Labour and LD judges. This is accomplished by comparing the percentage of positive and negative classifications made by the judges. It is assumed that the judgement of the judges is representative of their respective parties. However, this assumption has one shortcoming: the number of judges used in experimentation is quite small (7 Judges – 3 Conservative, 2 LD, 2 Labour). This was because of difficulty in recruiting judges.



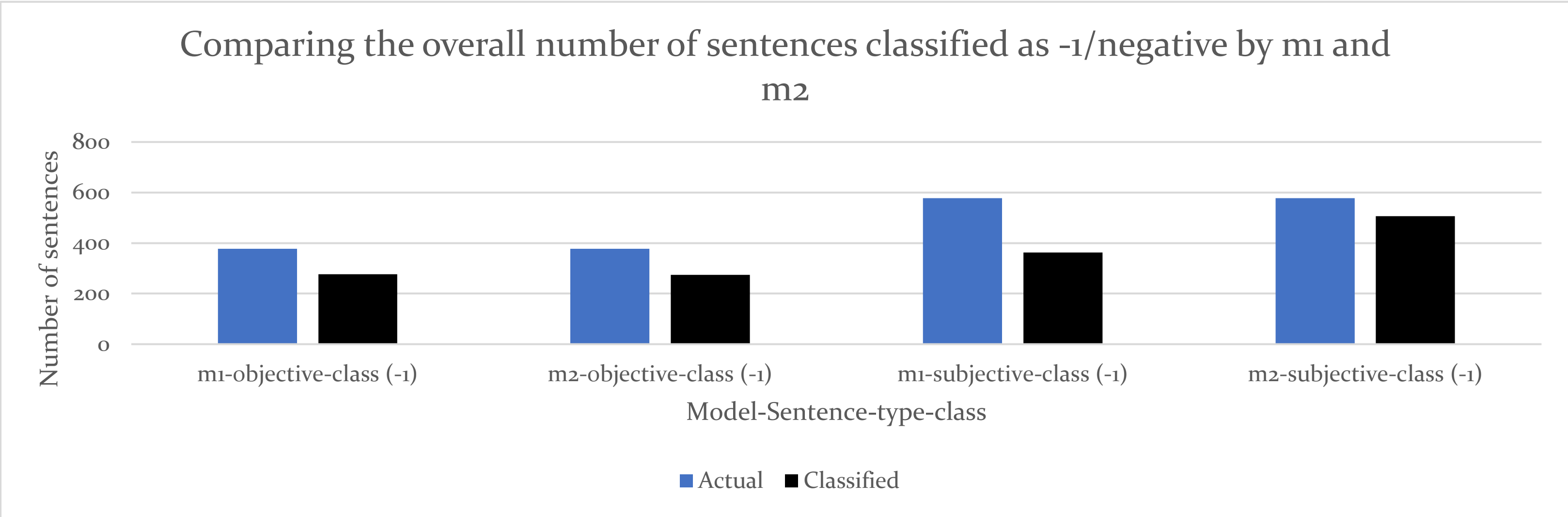


Figure 36: Showing the System's Bias for Class (1) over class (-1)

2. Identify the percentage agreement and disagreement for each context-party UKIP test set as classified by Conservative, Labour and LD judges. This is accomplished by comparing the percentage of positive and negative classifications made by the judges. It is assumed that the judgement of the judges is representative of their respective parties. However, this assumption has one shortcoming: the number of judges used in experimentation is quite small (7 Judges – 3 Conservative, 2 LD, 2 Labour). This was because of difficulty in recruiting judges.

3. Make inferences on the level of agreement and disagreement between the judges' value and UKIP's based on the percentages derived in step 1.

4. Compare the agreement and disagreement level amongst the three parties using the F_1 scores derived from m1 and m2 classification of the UKIP sentences. F_1 score is used because it combines both the precision and recall of the classifier.

5. In carrying out step 3, the objective is to compare the results of the VSM model against the inferences drawn from the judges. Therefore, the inference made in step 2 is compared against the percentage difference in the F_1 score, and a judgement is reached.

Following this description, the next sub-section describes each context-party pair test set.

8.7.1 EU-UKIP Test Judgement

Step 1: Percentage Agreement and Disagreement

Conservative Judges: Of the 95 UKIP EU sentences, the conservative judges assigned a rating of '0' to 11 sentences. Of the remaining 84 sentences, they agreed on 45 classifications and disagreed on 39, thus a percentage disagreement of 46.4%. Since there were 3 Conservative judges, this research also considered classifications where at least two of the participants assigned a positive or negative rating to a sentence, with the third judge assigning a classification of '0'. For this, the percentage disagreement was 39.28%.

Focusing on classifications where all three judges agreed: Of the 45 sentences on which there was agreement, judges assigned a negative classification '-1' to 33 sentences (73.3%) and a positive classification '1' to 12 sentences (26.6%). In other words, out of 45 UKIP-EU sentences, the Conservative judges suggested that at least 26.6% of the UKIP sentences were in line with their policies and at least 73.3% of UKIP utterances went against their values. Therefore, on the EU, this research concludes that Conservatives are more likely to assign negative sentiments to UKIP sentences. Another extension to this inference is that there is a 26.6% chance that Conservatives will agree with UKIP on the EU.

Labour Judges: Of the 95 sentences, Labour judges assigned '0' ratings to 7 sentences, which was lower compared to the Conservatives. The 2 Labour judges, disagreed on 26 sentences leaving a disagreement rate of 29.54%. Of the 62 sentences on which they agreed on the judgement, they assigned a negative classification to 49 sentences meaning a 79%

disagreement with UKIP. Conversely, they assigned positive classifications to 13 sentences, which meant a 20.96% chance of agreement with UKIP on the EU.

LD Judges: Of the 95 sentences, LD judges disagreed on 11 sentences resulting in a disagreement rate of 11.82%. Assigning ‘0’ ratings to 2 sentences. Of the remaining 82 sentences on the EU, they assigned a negative classification to 73 sentences resulting in an 89% disagreement with UKIP on the EU. Conversely, they assigned a positive classification to 9 sentences, thereby resulting in a 10.97% chance of agreement with UKIP.

Table 40 provides a complete breakdown of the judge’s classification.

Step 2: Infer Agreement Levels

Based on the discussion and with table 40 as a reference, the following inferences can be made:

- There appears to be comparatively more agreement and consensus on UKIP-EU amongst LD than Labour and Conservatives. As observed in table 40, Conservative’s disagreement rate is almost 4 times that of LD and Labour’s is almost 2 times LD’s.
- All 3 parties express relatively high negative sentiments towards UKIP-EU, even though comparatively, Conservatives are more likely to agree or assign positive sentiments towards UKIP-EU sentences (26.6%) as compared to Labour - 20.96% and LD - 10.97%.

Step 3: Inference from Model’s Classification Using F1 Score

Table 41 portrays a snippet of the evaluation tables 32 and 33. It illustrates the F_1 scores assigned to positive and negative classifications by the models m1 and m2 for UKIP-EU test sentences. The difference in positive and negative F_1 scores for all three parties is shown in table 41.

Table 40: Summary of Judge's Agreement and Disagreement on EU-UKIP Test data

Party Name	Percentage Disagreement	Number of sentences agreed	Percentage of Positive Sentences	Percentage of Negative Sentences
Conservative	46.4%	45	26.6%	73.3%
Labour	29.5%	62	20.9%	79.0%
LD	11.8%	82	10.9%	89.0%

Table 41: Comparing positive and negative F1 Scores for EU Test data

Context-Party pair (UKIP)	Model Type	Positive F_1 score	Negative F_1 score	F_1 difference (Negative-positive)
EU-Cons	m1	54.0%	73.6%	19.6%
EU-Labour	m1	68.5%	87.6%	19.1%
EU-LD	m1	48.4%	87.0%	38.6%
EU-Cons	m2	48.6%	64.1%	15.5%
EU-Labour	m2	56.4%	80.0%	23.6%
EU-LD	m2	48.4%	87.0%	38.6%

Step 4: Overall Inference and Comparison of Model Classification and Human judgement

The difference between negative and positive F_1 scores as classified by the models and seen in the last column of table 42, suggests that all 3 parties will assign predominantly negative classifications to UKIP EU sentences. This is in line with the inference drawn from human judgement.

In addition, both m1 and m2 models show that compared to Labour and Conservatives, LD will assign comparatively more negative sentiment classifications to UKIP-EU sentences. For instance, m2 produces an F1 score difference of 38.6% for LD, 23.6% and 15.5% for Labour and Conservative respectively. This mirrors the judgements made by the judges who assigned the highest negative sentiment classification to LD, followed by Labour and finally Conservative. m1 deviates slightly from this in assigning fractionally more negative sentiments classifications to Conservatives than Labour. This is observed in table 42 where the F1 difference for Conservatives and Labour is 19.6% and 19.1% respectively.

In conclusion, both m1 and m2 assign proportionally more negative sentiments to UKIP EU sentences than positive sentences for all three parties. This mirrors the classification assigned by the judges, implying that the VSM mirrors the observed values of the judges towards UKIP-EU. Secondly the model's classifications suggest that of all three parties, LD would assign the most negative sentiments to UKIP-EU sentences and as such will show the most opposition to UKIP values. This also mirrors the judges, with the Lib-Dems showing the least disagreement and assigning the highest percentage of negative sentiments to UKIP-EU sentences.

Finally, unlike m1 which suggests that Labour would assign the least percentage of sentiment classifications, m2, mirrors the judges' classification in showing that Conservatives assign the least percentage of sentiment classifications to UKIP-EU sentences. Therefore, model m2 best mirrors the values of the human judges, while m1 mirrors aspects of the human values with slight anomalies.

8.7.2 Immigration-UKIP Test Judgement

Step 1: Percentage Agreement and Disagreement

Conservative Judges: Of the 75 immigration based sentences, Conservative judges left blank or assigned a classification of '0' to 9 sentences, agreeing on 38 classifications out of 66, thus, an agreement score of 57.5% and disagreement of 42.4%. Of the 38 sentences on which there was agreement, 21 sentences were assigned a negative classification, while 17 were assigned a positive classification. Therefore 55.2% of the sentences were assigned a negative classification and 44.7% of the sentences were assigned a positive classification. An implication of this is that there is a 55.2% chance of Conservatives disagreeing with UKIP on Immigration as compared to 44.7% agreement.

Labour Judges: There was a greater agreement level amongst these judges on Immigration-UKIP sentences. They agreed on classifications for 66 sentences of which 2 were unclassified. This meant a disagreement percentage of 12.3% thus implying a greater certainty of views on policy in the light of the sentences provided. Of the 64 classified sentences, 46 were classified as negative while 18 were positive, resulting in a 28.1% positive sentiment and 71.8% negative sentiment. It is inferred from this that on Immigration, Labour appeared to be in greater disagreement with UKIP as compared to the level of negative sentiment or disagreement inferred from the analysis of EU data earlier.

LD Judges: 3 sentences were unclassified and there was disagreement on classifications for 9 sentences, resulting in a disagreement level of 12.16% and an agreement level of 87.8%. Of the 65 sentences on which they agreed, they assigned a negative classification to 48 sentences (73.8%) and a positive classification to 17 sentences. This implied a 26.15% agreement or positive sentiment towards UKIP-Immigration values and a 73.8% negative sentiment or disagreement with UKIP Immigration values. Table 42 provides a complete breakdown of the judge's classification.

Step 2: Infer Agreement Levels

Based on the discussion, with table 42 as a reference, the following inferences were made:

- There's comparatively more disagreement amongst Conservatives on UKIP-Immigration sentences than Labour and LD participants. In table 42, Conservatives disagree on 42.4% of the classifications assigned to sentences as compared to 12.4% and 12.16% for Labour and LD respectively.

- Comparatively, there's a significantly higher percentage of Conservative judges agreeing with UKIP in assigning positive sentiments than Labour and LD participants. Of the classified sentences, Conservatives classify 55.2% of them as negative, which is lower than Labour and LD with 71.8% and 73.8% respectively. This also suggests that both Labour and LD are more likely to assign negative sentiments to UKIP immigration sentences than the Conservatives. In addition, of all 3 parties, LD, are likely to assign more negative sentiments to UKIP immigration sentences followed closely by Labour and finally Conservatives.

Table 42: Summary of Judge's Agreement and Disagreement on Immigration-UKIP Test data

Party Name	Percentage Disagreement	Number of sentences agreed	Percentage of Positive Sentences	Percentage of Negative Sentences
Conservative	42.4%	38	44.7%	55.2%
Labour	12.3%	66	28.1%	71.8%
LD	12.2%	65	26.2%	73.8%

- The difference between positive and negative classifications for Conservatives on UKIP-Immigration sentences suggest that the Conservatives might have significantly closer values and views on Immigration with UKIP than Labour and LD. This is evidenced in table 42 which shows that out of 38 Conservative sentence classifications, there's a 44.7% chance of assigning a positive sentiment. This percentage represents almost half of the sentences classified, and compared to Labour and LD with 28.1% and 26.15% respectively, it is not unreasonable to conclude that Conservatives values on Immigration are not so dissimilar to UKIP's.

Step 3: Inference from Model's Classification

Table 43 portrays a snippet of the evaluation tables - 32 and 33. It illustrates the F_1 scores assigned to positive and negative classifications by m1 and m2 for UKIP Immigration sentences.

Step 4: Overall Inference and Comparison of Model Classification and Human judgement

The difference between negative and positive F_1 scores as classified by the models and seen in the last column of table 43, suggests that both Labour and LD will assign predominantly negative classifications to UKIP sentences. With m1 returning differences of 17.2% and 20% for Labour and LD respectively, while m2 returns differences of 16.3% and 15% respectively. This is in line with the classifications made by judges which suggest that Labour and LD will assign more negative classifications. m2 deviates slightly from the judges in fractionally assigning less negative classifications for LD than Labour, where LD's difference is 15% and Labour is 16.3%.

A major difference is observed in the case of the Conservatives. Table 43 shows that both models would assign more positive classifications than negative to UKIP sentences, with m2 assigning more positive classifications than m1. Based on the model, it is inferred that Conservatives are more likely to agree with UKIP on Immigration than disagree or assign negative sentiments. Although the inference drawn from the human judges suggests that Conservatives would assign more negative sentiments than positive sentiments to immigration sentences, it also suggests that this likelihood is slight seeing as the fraction of positive and negative classifications made by the judges was quite close. In fact, the conclusion reached on the Conservatives judges was that their values were significantly closer to UKIP values than any of the other parties. This conclusion is just slightly different from the inference reached by m1 and m2.

Table 43: Comparing positive and negative F1Scores for Immigration Test data

Context-Party pair (UKIP)	Model Type	Positive F_1 score	Negative F_1 score	F_1 difference (Negative-positive)
Immigration-Cons	m1	53.6%	45.7%	-7.9%
Immigration-Labour	m1	52.0%	69.2%	17.2%
Immigration-LD	m1	59.0%	79.0%	20.0%
Immigration-Cons	m2	61.9%	52.9%	-9.0%
Immigration-Labour	m2	57.1%	73.4%	16.3%
Immigration-LD	m2	60.0%	75.0%	15.0%

In conclusion, the model mirrors the sentiments and values of Labour and LD on Immigration, but fractionally over estimates the level of similarity between Conservatives and UKIP in suggesting that Conservatives and UKIP share similar values contrary to human judgement.

8.8 Contextual Limitation

Before concluding this chapter, an observation made during the focus group discussion is elaborated upon. It was observed that value holders/judges from opposing parties tended to oppose sentences mentioned or supported by judges from opposing parties. This behaviour was observed between Labour and Conservative judges. To test this, the moderator posed a statement that came straight from the Labour party manifesto. The statement focused on Labour's policy pledge to reduce immigration and introduce 'sensible immigration reforms'. Interestingly, the Conservatives made a similar pledge, to restrict immigration to a few thousand. When the moderator posed the statement "*UK's position on immigration should be to reduce it sensibly to a few thousand people*", the Labour judges opposed the statement even though it was a party policy pledge. Conversely both models m1 and m2, assigned a positive classification to the sentence. There was a high degree of certainty that the model's response was correct as there were several instances in the training data where Labour pledged or suggested '*reducing immigration*'. When the moderator questioned the Labour judges further, it became clear that the negative sentiment was not directly related to the subject of ‘reducing Immigration’ but was connected to the fact that ‘reducing immigration’ represents a success claim of the Conservative party. Labour judges viewed the claim to be dubious, and assigned a negative sentiment to the utterance. As such during the discussion, it became clear that the judges’ sentiments were influenced by the presence and opinion of other parties. This feature was not captured in this model and future work should consider capturing the relationship between value holders.

8.9 Conclusion

This chapter has described the test methodology and its evaluation. The test implementation used here entails compiling some test documents, annotating them for ground truth via judges’ familiar with the domain and comparing the human annotations against the model’s.

This chapter has shown that the overall performance of the VSM falls in the acceptable performance range. In addition, the model’s performance compared to the other systems discussed is quite unique considering that it was implemented without human annotated training data. However, the absence of modeling additional human context such as the relationship between value holders might have impacted the performance of the model. It has been shown that this factor played a role in the behaviour of human judges and that Tang and Chen (2011) included this feature in their model resulting in an accuracy of 80.67% - 88.37%. To this end, future work should consider harnessing and incorporating human relationships in the model.

This chapter has also shown that semantic enhancement does not necessarily improve the overall performance of the recipient sentiment analysis task. However, it improves recipient sentiment prediction of sentences containing named entities. This improvement has a converse effect on the classifications assigned to sentences that have not been semantically enhanced.

As expected the model performs significantly better than traditional SA methods in sentiment prediction. It also performs better for predicting the sentiment of subjective sentences than objective sentences even though the F1 score for both types of sentences fall within the acceptable performance range. Finally, this chapter has shown that the behaviour of the value model imitates and mirrors the behaviour of human judges, whose judgements and sentiment assignations was based on their values.

9. Conclusion and Future Work

9.1 Overall Summary

The aim of this thesis was to develop a method of SA that could predict recipient's sentiment without the use of explicit human annotations or empirical surveys. This research achieved this objective by adopting a methodology based on a model of human values in sentiment prediction.

This research began by elaborating on the difference between traditional sentiment analysis and recipient sentiment prediction. It illustrated with examples that depending on the recipient an objective or subjective sentence could invoke diverse sentiments. In addition, it also showed that SA methodologies based purely on the linguistic units in sentences are incapable of making this prediction without the input of human centric features. In chapter 2, socio-theoretic methodologies which tend to capture human centric features were reviewed. The review showed that they required considerable human input in the form of content analysis or empirical survey and were too structured hence lacking the flexibility required to adapt them towards diverse subject domains. Regardless of these deficiencies, most of these systems harnessed a combination of linguistic and socio-theoretic features in their sentiment prediction models and achieved accuracies in the range of 64% - 88% (see table 36). The combination of such features was a reasonable justification for this research's approach in applying values as a socio-theoretic principle and linguistic units as features in the model.

In chapter 3, part of the aim was to relate values to sentiments. To accomplish this, the concepts and definitions of values were reviewed. From the definitions and concepts reviewed, it was concluded that values are abstract coordinators of behaviour, which reflect the measure of preference for one state of existence of an object over another. The expression 'abstract coordinators of behaviour', implies that for any form of behaviour for which sentiment is a type, values represent the primary causal factor. In linking values as a determining factor for the sentiment expressed by an individual, the methodologies for classifying and applying values derived from text were reviewed. This research elaborated on the fact that these methodologies involved textual content analysis (CA) by domain experts, empirical surveys of domain experts or a hybrid approach involving both CA and empirical surveys. All three approaches were shown to be expensive, requiring extensive human input. The outcome is a model called a value inventory (VI) which is essentially a categorised list of words which convey specific value types. This thesis showed that VIs are highly dependent on human input and are designed specifically for a domain or user base. Finally, VIs as a model are incapable of accommodating new words or unseen scenarios.

Having identified the problem with existing value models, in chapter 5, the aim was to create a design for modelling values and apply the modelled value in sentiment prediction. The fundamental principle of the value model was hinged on the notion that when people speak or express themselves on any subject matter, they ultimately seek to express and convey some functional value concept linguistically. To accomplish this, this thesis pictorially represented the journey from abstract values represented textually to sentiment prediction as a five-stage process called the Value-Sentiment Model (VSM) (see figure 3). The VSM, illustrated through a process called values decomposition that based on the definitions of values, the parameters that make up a value could be extracted from sentences. These parameters were identified as the value holder, subject of the value, state, action and context. Having derived these parameters, methods for extracting them from sentences were discussed. Linguistic clues were used to identify subjects as nouns or NPs, states were shown to be typically adjectives or adverbs while actions were shown to be verbs or VPs. In deriving these, it was shown that the words which make up a value laden sentence could be grouped into content and function words, where content words include actions, states and object. Given this structure, it was shown that since abstract values determine the words a person will use in constructing a sentence that expressed a sentiment, then the value model could be expressed as a generative process where semantically relevant words are generated from a vocabulary of content and function words. Consequently, the generative process was modelled as a Language Model (LM). The value LM implemented was quite unique in that parameter estimation differed from content word to function words. Function word probability was estimated using a trigram estimate by conditioning the probability on the two previous words in its history. As for content word probability estimation, it was imperative that the LM capture the syntactic and semantic relationships between actions, states and subjects. In addition, the content word estimate had to be tailored for the value holder and context. This lead to a content word estimation formalization conditioned on five parameters.

Contrary to known value models, the value LM design did not require human input. It was shown that value components unlike value inventories can be identified and extracted by using existing linguistic clues embedded in the sentence. This research also described how recipient sentiment can be determined, as the difference between the probability of a value holder making a sentence and the probability of the same value holder making a sentence opposite in sentiment to the original sentence. The algorithm for implementing this was called VSEA. A unique concept called value fields, which showed the influence of diverse values on sentiments was also introduced.

Chapter 6 described a stepwise implementation of VSM and showed how the semantic relations between content words were captured using DG. The process of estimating the content word probability using a maximum entropy classifier was described. 241 baseline features were generated. These features were generated due to the absence of adequate processing power resulting in a smaller feature set capturing implicit value model characteristics. In addition, the requirement for zero human annotation necessitated this approach. Finally, a technique called feature switching was introduced for estimating the probability of the sentence with the opposite sentiment. A full implementation of the model was described in chapter 7, focusing on political domain and two timely contexts: EU and Immigration.

In chapter 8, the models were tested and evaluated. Based on the tests carried out, it was shown that the model's performance was in the range of existing models. This was an important accomplishment seeing as other models were dependent on human annotated training sets or knowledge bases. In addition, due to the absence of domain specific features, this research's methodology is applicable to other domains. Additionally, semantic enhancement and corpus/domain dependent data preparation has been shown to not improve overall performance and that it influences the sentiment prediction of sentences containing named entities. Finally, it was demonstrated that the sentiment prediction of the VSM matched the sentiment estimation of actual value holders, thus reflecting the values held by real individuals.

In conclusion, this research has implemented a unique methodology and approach for modelling values and applying it to sentiment prediction. This implementation has been carried out free of human annotation, content analysis and empirical surveys.

9.2 Summary of Research Findings

The findings of this research are summarized in this section.

1. Recipient sentiment prediction can be accomplished through the VSM as the VSM's overall evaluation, observed in the precision, recall, accuracy and F-score falls within the performance range of existing models. The accuracy range of known models that are dependent on human annotations and input is 64% to 88%, while the VSM's accuracy is 72.3%-72.5%. Considering that the VSM methodology does not involve human annotation/input, and the VSM's accuracy is over 70%, and in the 70%-80% range (for which the performance of most of the approaches in figure 35 exist), it is surmised that this performance level is acceptable

2. Semantic enhancement does not improve the overall performance of the VSM.

3. The VSM performs reasonably well in predicting the recipient sentiment of subjective sentences with both the semantically enhanced model and the unenhanced model returning F1 scores of 72.2% and 70.7% respectively. However, the VSM's performance in predicting the sentiment of objective sentences is below the 70% mark falling between 66.1% and 68%.

4. The VSM model outperforms Sentiwordnet which focuses on the sentiment expressed in the sentence. Sentiwordnet was observed to have a misclassification rate of 74.7% for the test data, as compared to VSM's misclassification which is between 27.4% and 27.6%. This wide difference in misclassification between the recipient based SA methods which focuses solely on the sentiment expressed in the sentence highlights the benefit of the VSM.

5. In addition to predicting recipient sentiment, the VSM also reflects the actual values of value holders. Therefore, since sentences made by people reflect their

values, the value LM of the VSM represents an implementation of a person's abstract values.

6. Finally, the VSM provides a methodology that assigns different sentiments to a sentence based on the recipient. This is an important finding as current SA approaches will return a single sentiment for a sentence.

9.3 Contributions

The contributions made in this thesis are:

1. The VSM artifact – A major contribution of this research is the VSM artifact and its implementation using a real-life scenario. In addition, in implementing the VSM, this thesis provides a methodology for predicting recipient sentiment that is devoid of human annotation or contribution. The steps taken in building this model, illustrated in figure 8 and 9 represents a methodology that is replicable and applicable to other domains and thus represents a contribution to existing work. Other VSM based practical contributions include the implementation of both the RSPA (Recipient Sentiment Prediction Algorithm) and VSEA (Value Sentence Estimation Algorithm).

2. Two theoretical contributions made in this thesis relate to the approach taken in estimating recipient sentiment. They are:

    a. Conventional approaches to SA is based on gold rated corpus, annotated with positive or negative sentences. Due to the absence of such a corpus, the approach taken in this research was to express the sentiment as a function of how likely it is for a speaker to make a sentence versus the likelihood of the speaker making a contrary sentence opposite in sentiment to the original sentence. This perspective of sentiment is innovative and unique to this research.

    b. Following '2a' above, the introduction and implementation of Feature Switching (FS) in estimating the likelihoods of the two sentences is a unique innovation and a contribution.

3. Another theoretical contribution of this research is the concept of value fields (VF) (section 5.9), which enables the estimation of recipient sentiments, for any combination of contexts, actions, states, objects and value holders. Based on VF, researchers can potentially determine recipient sentiment under diverse contexts and conditions.

4. Finally, this research's contributions also extend to value models:

    a. First, unlike most value modeling research which employ content analysis in identifying the concepts of a value, one of this thesis' theoretical contributions demonstrated that abstract values can be formalized through value decomposition (identifying value parameters from sentences) – values consist of five parameters value holder (H), action (A), state (S), object (θ) and context (C), hence the value model is a function $f(H, \Theta, S, A, C)$. The decomposition of the value model leads to a related contribution – which is the ability to structurally represent values in text (see section 5.2 and 5.5).

b. Based on '4a' above, another practical contribution made in this thesis, lies in the portrayal of values as a generative process and the building of the value model as a language model artifact (see section 5.8).

9.4 Evaluation of Research

In this section, the research is critically evaluated by examining if it satisfies DSR evaluation criteria drawn from Hevner et al (2004). In addition, key assumptions made in the research are also evaluated. Evaluation criteria include:

Conceptualization and Representation of Problem – According to Hevner et al (2004, p.84), "*Design-science research efforts may begin with simplified conceptualizations and representations of problems.*" Having determined the research requirements, an initial conceptualization of the problem was provided by pictorially representing the journey from abstract values to sentiment prediction as carried out by humans in figure 3. This illustration marked the first conceptualization of the VSM and served as a step-wise approach towards its implementation. In addition, pictorially representing each step, meant that problems and challenges associated with each step could be itemized and treated accordingly. For instance, in representing and formalizing abstract values, all known definitions of values were considered, commonalities were identified, and this led to the value components discussed in section 5.3. Unlike other research approaches which define values in terms of an inventory of words associated with a particular domain, the identification of value components meant that for any domain or usecase, if the VCs could be identified then potentially, a VSM could be applied towards recipient sentiment prediction. This formalization gave rise to several follow-up questions: How can the VCs be identified without human intervention or annotations? what are the VCs? how can the VCs be represented in a value model? etc. In addressing these questions in chapter 5, additional artifacts - methodologies and algorithms - were proposed. For example, artifacts include methods - methodologies such as value decomposition/value component identification, models - value language model, feature switching and VSEA, constructs – the structural representation of value laden sentences described in section 5.6. Each artifact built in this research was founded on a prior artifact. For example, value decomposition, was founded on the identification of value characteristics, the analysis and formalization of VLSs was also based on the decomposition of sentences. This satisfied the DSR requirement that the research artifact "is incrementally built by the implementation of new IT artifacts" (Hevner et al, 2004, p85).

Some of the assumptions made are now considered.

For instance, in implementing value decomposition, although it was assumed that actions are verbs, states are adjectives/adverbs and subjects are nouns, these components were grouped under the single umbrella of content words, to allow for flexibility in the model so that it can accommodate new content words. This makes the VSM applicable to any domain or datatype as the value components which make up the sentence are identified by filtering out known function words so that the remaining content words are identified and their probabilities estimated. This assumption thus makes the implementation easier and the model applicable to other document types.

Another simplifying assumption made was reflected in the capturing and representation of the semantic relationship between value components. Actual semantic relationship aims at capturing the meaning representation of the sentence by identifying the who, when, what, why of the sentence predicates (Oepen et al, 2014). This determination is difficult to undertake since the arguments in semantic parsing – agents, patients, instruments, temporal manner - can take a number of roles or frames. In addition, human input is required in annotating the most suitable semantic frames required for the domain type. As a result, dependency relations which are more syntactic than semantic were used. A benefit of syntactic dependency parsing is that there is generally a broad consensus on the DG representations unlike in semantic dependencies where several textual semantic annotations exist without any consensus or linkage between existing resources (Oepen et al, 2014). This research theoretically showed the importance of DGs in capturing these relationships in section 6.2, however, they are still not true representations of meaning as they only cover grammatical structure. Despite the challenges associated with identifying semantic relations, their inclusion in this research is likely to have improved the performance of the model and also reflect the definition of the value model. Consider the test sentence "*Reducing migration to the thousands",* assuming the semantic frame resource FrameNet was applied the word '*reducing*' would have been recognized as belonging to the semantic frame '*Cause_change_of_position_on_a_scale'*, reflecting that the 'frame consists of words that indicate that an Agent or a Cause affects the position of an Item on some scale[92]', here the agent is the value holder '*Conservatives*', the item is '*Immigration*' and the attribute is '*to the thousands*'[93]. What this also connotes is that by recognizing that the Conservatives associates the item 'Immigration' to the semantic frame '*Cause_change_of_position_on_a_scale*', utterances made by the Conservatives can be estimated by observing if the expressions used reflect the semantic frame thereby making the model less dependent on the grammatical relations as is the current VSM

[92] https://framenet2.icsi.berkeley.edu/fnReports/data/frameIndex.xml?frame=Cause_change_of_position_on_a_scale: Last accessed, 20-06.2016

[93] Determining the most appropriate frame is a highly subjective process that requires multiple domain experts agreeing. In addition, some words or expressions might not exist in the vocabulary of semantic frames or resources.

implementation. Therefore, although the grammatical relations used in the implementation suffices the inclusion of semantic frames would likely improve the accuracy of the model and provide a truer reflection of the value model.

Evaluation Methods and metrics: According to Hevner et al (2004), the design artifact must be well evaluated using methodologies available in the knowledge base. Hevner et al (2004, p85) suggests that the evaluation method must be "matched appropriately with the designed artifact and the selected evaluation metrics". Based on this, an experimental approach was adopted in testing and evaluating the VSM artifact. According to Hevner (et al, 2005) this involves, studying the artifact in a controlled environment and executing the artifact with artificial data. This was satisfied in the implementation of the test case which focused on the political domain and two contextual subjects – Immigration and the EU. Several simplifying assumptions were made during this implementation. Two key modifications involved the use of a maxent model in implementing the content word LM and the creation of a reduced feature set in implementing the maxent model.

A maxent model was implemented in estimating content word probability because given the number of features associated with the content word its estimation became quite unwieldy and complicated. Maxent models are flexible and capable of handling many features, thus their use. However, in this research only maximum entropy models were used and future work should explore the use of other models in estimating content word probability. Furthermore, the reduced feature set used in estimating the content word maxent model, was due to a lack of processing power, leading to the construction of several lexical features. However, while these reduced features were based on the stipulated class of features, they are potentially inexhaustible and only represent a baseline of possible features. The question that emerges from this is 'Do the features cover the optimum set of features required in implementing the model'? Since this determination was not carried out, the question around the implementation features used remains an open one and as such, the implemented features can be deemed to be dependent on the current implementation. This downside can be addressed in future research by identifying the optimum feature set required for the implementation.

Finally, according to Hevner et al (2004, p.85), "evaluation of a designed IT artifact requires the definition of appropriate metrics … and when analytical metrics are appropriate, designed artifacts may be mathematically evaluated". Since SA is a NLP classification task, analytical metrics – precision, recall, F1 score, accuracy and misclassification rate were used. The reason for using these metrics was because of their use in related and prior research, hence satisfying the DSR requirement that the evaluation metric be drawn from existing literature.

Additionally, the approach chosen in generating the gold-rated test set for estimating performance, used 7 judges, which is a small number of judges and not reflective of a true representation of all value holders. Although the small number of judges makes it easy to estimate annotator agreement and select test sentences. With a larger number of value holders which cuts across a more diverse range – age group, educational background and race - the VSM could be tested for instance on certain age or social groups. In addition, some of the conclusions drawn from the Brexit vote could have been verified such as the fact that most young people between the ages of 18 and 30 voted to remain as compared to middle aged individuals who voted to leave the EU (Kelly, 2016). Due to the small volunteer group used in this research, such conclusions cannot be reached with great certainty. This is also why we limit the evaluation in section 8.7 to whether the model reflects the values of the value holders tested in this research.

Artifact Style: Although style is subjective, the description proposed by Gelernter (1998), as a marriage between simplicity and power was adopted. The question asked to define this is, how simple is the VSM in applying it to other domains? The theoretical model described in chapter 6 is devoid of domain, corpus or user influences. In addition, the sentence decomposition based on splitting words into content and function words, without explicitly associating words as actions or states makes the implementation applicable and extensible to any dataset. More so, simplifying semantic relations by applying syntactic and grammatical relations eliminates the need for semantic annotations which is also a domain dependent process requiring human influence. However, the implementation described in chapter 7 suggests that in the absence of adequate computational power, domain specific modifications might be required to implement the VSM.

Research Limitations

Following the evaluation above, some of the limitations associated with this research are:

- Tang and Chen (2011) applied features capturing the social relationship and inter-personal exchanges between recipients and obtained the best performing model (overall accuracy in the range of 80.67% - 88.37%). Based on literature and observations made from conducting a focus group of value holders it was established that social and interpersonal relationships play a major role in the determination of expressed sentiment. This relationship is not captured in the VSM implementation.
- Part of the evaluation is accomplished by comparing the percentage of positive and negative classifications made by judges or value holders on the assumption that the judgement of the judges is representative of their respective parties. However, the number of judges used in the experimentations is quite small (7 Judges) and this small sample is not a true reflection of all value holders in the parties.
- The baseline set of features applied in the implementation of the maxent model are potentially inexhaustible. This means that they are not entirely optimized and since they have been applied only to the political sector, they can be viewed to be domain dependent.
- Only a trigram LM and maxent classifier was applied in the model implementation. Due to space and time constraints, other classifiers were not trialed.

- It is arguable that the corpus used in this research in building the LM is quite small compared to similar works. Table 21 shows that the total number of words used in this research ranges from between 356076 to 2974827 as compared to the Wall street Journal corpus of 38 million words and a vocabulary of 19979 words (LDC, 1993), and the Associated Press Newswire corpus used by Church and Gale (1991) consisted of 44 million words (22 million words were used for training – half the text) and a vocabulary of 400653 tokens. Nevertheless, the Berkley Restaurant corpus used by Jurafsky and Martin (2009) consisted of 9332 sentences (considerably less than the sentences used in this research) and the Jane Austen corpus used by Manning and Schutze (1999), was made up of 617091 words and 14583 tokens. Like the corpus used in this research both are deemed to be small but considered as seminal datasets in language modeling.

Following the evaluation and limitations discussed above, the next section itemizes potential future research.

9.5 Future Work

Section 9.4, indicated that the use of 7 judges in providing value ratings was quite small. Thus, it is proposed that future work should use a higher number of judges on this same task. However, recruiting judges/annotators was and still is quite a difficult task and so this research proposes applying the same VSM methodology to a different domain where the criteria for selecting judges is not so difficult to attain.

Future research should introduce features capturing the social relations between recipients. For instance, in the political domain, a possible feature could be the frequency of adjectives used by Conservatives in response to Labour or LD comments etc. The performance of the model should be compared against the implementation in this thesis.

Feature Switching is proposed as a means for deriving alternate sentiments. However, given that generative models were implemented for each value holder, a simpler alternative solution could explore comparing the output probability of each LM on the sentence. However, the challenge with this approach is that a threshold probability would be required to assign positive or negative classifications. Since such a threshold probability determination is arbitrary and the challenge for future tasks would be in implementing a solution which eliminates the arbitrariness and subjectivity inherent in threshold determination when using a single probability estimate $p(w_1)$.

Tweaking and testing Linguistic Features: Given that baseline maxent features have been established, future research can focus on tweaking and reengineering these features. Combining them to establish the optimum set of features. Future research should also explore these features on other datasets and domains.

Automation of Sentiment intensity detection: An aspect of value fields is the ability to detect the intensity of the sentiment as a measure of how far to the left (-ve) or right (+ve) of the sentiment orientation scale that the field moves the sentiment of the utterance. The margin of the difference between the probability likelihoods in either direction can enable sentiment orientations determination like extremely negative, negative, positive, extremely positive. In this research, this intensity can only be determined manually, by subtracting $p(w_1)$ from $p(w_2)$, and inferring the intensity based on the magnitude of the difference. With a large data set, this manual inferencing will be insufficient and so automating this process could prove useful. An added benefit of this task will be the ability to predict neutral sentiments as this research only focuses on positive and negative sentiments.

Backing-off to Semantically relevant words: One of the approaches used in this thesis in estimating unknown word probability was by estimating $p('UNK')$. As part of future work, a semantically effective approach involving backing-off to synonymous words or word substrings that resonate the same meaning could be explored. For instance, if the word *'uncharacteristically'* is unseen in training, its probability can be estimated as $p('uncharacteristic')$ if *'uncharacteristic'* is observed in training.

Future research can explore the addition of semantic frames in addition to the linguistic feature sets used in estimating content word probability. Also, the VSM can also be applied on other domains such as 'Global Warming' to determine both performance and feature sets. Potentially such a task could replicate the approach and features used as well as implement additional features tailored for the domain. Finally, although maximum entropy classifier was implemented in predicting the sentiment of content words, it would be beneficial to explore the use of alternative ML algorithms to determine if the model's performance can be improved.

References

Abelson, R.P. and Carroll, J.D. (1965) 'Computer Simulation of Individual Belief Systems.' *The American Behavioural Scientist,* 3(9), pp. 24-30.

Ahothali, A. and Hoey, J. (2015). 'Good News or Bad News: Using Affect Control Theory to Analyze Readers' Reaction Towards News Articles.' *Human Language Technologies: The 2015 Annual Conference of the North American Chapter of the ACL*, Denver, Colorado, May 31 – June 5, pp. 1548–1558.

Asher, N., Benamara, F. and Mathieu, Y.Y. (2008). 'Distilling opinion in discourse: A preliminary study.' *In Proceedings of Computational Linguistics (CoLing)*, Manchester, UK. Association for Computational Linguistics. Pp. 7–10.

Audhkhasi, K., Sethy, A., Ramabhadran, B. and Narayanan, S.S. (2012) 'Creating Ensemble of Diverse Maximum Entropy Models'. Proc. ICASSP, pp. 4845-4848.

Aydin, I. and Torusdag, G. (2008) 'Dependency Grammar of Luciene Tesniere in the Perspective of Turkish, English and French.' In *Turkish Studies – International Periodical for the Languages, Literature and History of Turkish* or *Turkic.* 8(8), pp. 189-201. (Accessed: 10 October 2015)

Baccianella, S., Esuli, A. and Sebastiani, F. (2010). *SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining.* Available at: http://nmis.isti.cnr.it/sebastiani/Publications/LREC10.pdf (Accessed: 09 June 2013)

Balahur A., Kozareva Z., Montoyo A. (2009) 'Determining the Polarity and Source of Opinions Expressed in Political Debates.' In Gelbukh A. (ed.) Computational Linguistics and Intelligent Text Processing. CICLing 2009. *Lecture Notes in Computer Science*, vol 5449. Springer, Berlin, Heidelberg, pp. 468-480.

Balahur, A., Hermida, J. and Montoyo, A. (2011). 'Detecting Implicit Expressions of Sentiment in Text Based on Commonsense Knowledge.' In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA 2.011)*, Portland, Oregon, June, Association for Computational Linguistics, pp. 53-60.

Bandura, A. (1977). 'Self-Efficacy: Toward a Unifying Theory of Behavioural Change.' *Psychological Review*, 84, pp. 191-215.

Bardi, A. and Schwartz, S.H. (2003). 'Values and Behaviour: Strength and Structure of Relations.' *Personality and Social Psychology Bulletin*. 29(10)., pp. 1207-1220.

Bell, T. C., Cleary, J. G. and Witten, I. H. (1990). *Text Compression*, Prentice Hall, Englewood Cliffs, NJ.

Benamara, F., Chardon, B., Mathieu, Y. and Popesca, V. (2011). 'Towards Context-Based Subjectivity Analysis.' In *Proceedings of the 5th International Joint Conference on Natural Language Processing*, Chiang Mai, Thailand, November 8-13, pp. 1180-1188.

Bengston, D.N., Webb, T.J. and Fan, D.P. (2004). 'Shifting forest value orientations in the United States, 1980-2001: A computer content analysis.', *Environmental Values* 13, pp. 373-392.

Berger, A., Della Pietra, S.A., and Della Pietra, V.J. (1996). 'A Maximum Entropy Approach to Natural Language Processing.', *Computational Linguistics, 22*(1), pp. 39-71.

Berlyne, D. E. (1960). *Conflict, arousal and curiosity*. New York: McGraw-Hill.

Bernthal, W. F. (1962). 'Value perspectives in management decisions.', *Journal of the Academy of Management*, 5(3), pp. 190-196.

Bhomwick, P.K., Basu, A. and Mitra, P. (2009). 'Reader perspective emotion analysis in text through ensemble based multi-label classification framework.', *Computer and Information Science*, 2(4), pp. 64-74.

Bird, F., Waters, J. A. (1987). 'The nature of managerial moral standards.' *Journal of Business Ethics*, 6(1), pp. 1-13.

Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003). 'Latent Dirichlet Allocation.', *Journal of Machine Learning Research,* 3(5) pp. 993-1022.

Bloom, K. (2011). *Sentiment analysis based on appraisal theory and Functional Local Grammars.* PhD thesis, Illinois Institute of Technology.

Bosma, W. and Vossen, P. (2010) 'Bootstrapping language-neutral term extraction.', In *7th Language Resources and Evaluation Conference (LREC)*, Valletta, Malta, pp. 1-6.

Braithwaite, V.A. and Scott, W.A. (1991). 'Values.' In J.P. Robinson, P.R. Shaver, and L.S. Wrightsman (ed.) *Measures of Personality and Social Psychological Attitudes*. New York: Academic Press.

Braithwaite, V.A. and Blamey, R. (1998). 'Consensus, Stability and Meaning in Abstract Social Values.', *Australian Journal of Political Science,* 33, pp. 363-380.

Brown, P.F., Pietra, V., DeSouza, P.V., Lai, J.C., and Mercer, R.L. (1992). Class-based N-gram Models of Natural Language. *Computational Linguistics*, 18(4):467-479.

Callicott, J.B., Crowder, L.B. and Mumford, K. (2000). Normative Concepts in Conservation Biology: Reply to Willers and Hunter. *Conservation Biology* 14:575-578.

Caprara, G. V., Schwartz, S., Capanna, C., Vecchione, M., and Barbaranelli, C. (2006). 'Personality and Politics: Values, Traits, and Political Choice.', *Political Psychology*, 27 (1), pp. 1-28.

Carbonell, J. (1978) 'POLITICS: Automated Ideological Reasoning' *Cognitive Science,* 2, pp. 27-51.

Carbonell, J. (1979) *Subjective Understanding: Computer Models of Belief Systems.* PhD thesis, Yale.

Chaumartin, F.R. (2007). 'UPAR7: A knowledge-based system for headline sentiment tagging.', In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007).* Prague, Czech Republic: Association for Computational Linguistics, pp. 422–425.

Chen, F.S. and J. Goodman, (1999) 'An empirical study of smoothing techniques for language modeling', *Computer Speech and Language*, 13(4), pp. 359–394.

Chen, F.S., Beeferman, D. and Rosenfeld, R. (2017) 'Evaluation Metrics for Language Models', Available at: http://www.itl.nist.gov/iad/mig/publications/proceedings/darpa98/html/lm30/lm30.htm (Accessed: 23/03/2017).

Cheng, A.-S. and Fleischmann, K.R. (2010). 'Developing a meta-inventory of human values.' In *Proceedings of the American Society for Information Science and Technology (ASIST2010)*, 47(1), pp. 1-10.

Cheng, A.-S, Fleischmann, K.R, Wang, P., Ishita, E. and Oard, D.W. (2012). 'The role of innovation and wealth in the net neutrality debate: A content analysis of human values in congressional and FCC hearings.', *Journal of the American Society for Information Science and Technology,* 63(7), pp. 360-373.

Choi, Y. and Claire, C. (2009) 'Adapting a polarity lexicon using Integer linear programming for domain-specific sentiment classification.' In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP-2009)*, pp. 1-9.

Chomsky, N. (1963). 'Formal Properties of Grammars.' In Luce, R.D., Bush, R., and Galanter, E. (ed.), *Handbook of Mathematical Psychology,* Vol. 2, pp. 323-418. Wiley.

Church, K.W. and Gale, W.A. (1991). 'A Comparison of the Enhanced Good-Turing and Deleted Estimation Methods for Estimating Probabilities of English Bigrams.' *Computer Speech and Language,* 5, pp. 19-54.

Clarkson, P.R. and Rosenfeld, R. (1997) 'Statistical Language Modeling Using the CMU-Cambridge Toolkit.', In Proceedings ESCA. Eurospeech 1997. Available at: https://www.cs.cmu.edu/~roni/papers/SLM-TK-V2-eurospeech-97.pdf (Accessed: 22/08/2011)

Crace, R. K. & Brown, D. (1996). *Life Values Inventory.* Chapel Hill, NC: Life Values Resources.

Cui, H., Mittal, V. & Datar, M. (2006). 'Comparative experiments on sentiment classification for online product reviews.', *National Conference on Artificial Intelligence (AAAI),* 21(2), pp. 1-6.

Cunningham, et al. (2015). 'Developing Language Processing Components with GATE Version 8.1.'. University of Sheffield Department of Computer Science, 2015.

Das, A. and Bandyopadhyay, S. (2010). 'Sentiwordnet for Bangla.', *Knowledge Sharing Event-4: Task, 2.*

Das, S. and Chen, M. (2001). 'Yahoo! for Amazon: Extracting market sentiment from stock message boards.', *Asia Pacific Finance Association Annual Conference (APFA)*, pp. 1375-1388.

Dave, K., Lawrence, S. and Pennock, D. (2003) 'Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews.', *WWW'03,* Budapest 20-14, pp. 1-10.

Della Pietra, S.A., Della Pietra, V.J., and Lafferty, J.D. (1997). 'Inducing Features of Random Fields.' *IEEE Transactions on Pattern Analysis and Machine Intelligence, 19*(4), pp. 380-393.

De Marneffe, M.C., MacCartney, B. and Manning, C.D. (2008) 'Generating Typed Dependency Parses from Phrase Structure Parses.', In *Proceedings CrossParser '08 Coling 2008: Proceedings of the Workshop on Cross-Framework and Cross-Domain Parser Evaluation.* Manchester UK, Aug 23-23, pp. 1-6.

De Marneffe, M.C., Dozat, T., Silveira, N., Haverinen, K., Ginter, F., Nivre, J. and Manning, C.D. (2014a) Universal Stanford Dependencies: A cross-linguistic typology. In *Proceedings of LREC,* pp. 1-8. Available at: https://nlp.stanford.edu/pubs/USD_LREC14_paper_camera_ready.pdf (Accessed: 02/06/2015)

De Marneffe, M.C., Ginter, F., Goldberg, Y., Hajič, j., Manning, C., McDonald, R., Nivre, J., Petrov, S, Pyysalo, S., Schuster, S., Silveira, N., Tsarfaty, R., Tyers, F. & Zeman, D. (2014b). 'UD v1.0'. Available at: http://stp.lingfil.uu.se/~nivre/master/intro_ud.pdf (Accessed: 05/11/2015).

Devitt, A. and Ahmad, K. (2007). 'Sentiment polarity identification in financial news: A Cohesion-based Approach.', In *Proceedings of the 45th Annual Meeting of Association of Computational Linguistics (ACL),* Prague, June, pp. 984–991.

Ding, X., Liu, B. and Yu, P.S. (2008). 'A Holistic Lexicon-Based Approach to Opinion Mining.', *In WSDM '08*, February 11-12, Palo Alto, California, pp. 1-9.

England, G. W. (1967). 'Personal value systems of American managers.', *Academy of Management Journal*, 10, pp. 53 - 68.

Entman, R.M. (1993). 'Framing: Toward clarification of a fractured paradigm.', *Journal of Communication*, 43(4), pp. 51–58.

Esuli, A. and Sebastiani, F. (2006) 'Sentiwordnet: A Publicly Available Lexical Resource for Opinion Mining.', In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC-2006)*, May, Genoa, pp. 1-6. Available at: http://www.lrec-conf.org/proceedings/lrec2006/pdf/384_pdf.pdf (Accessed: 02/03/2012).

Feather, N. T. (1995). 'Values, valences. and choice: The influence of values on the perceived attractiveness and choice of alternatives.', *Journal of Personality and Social Psychology*, 68, pp. 1135-1151.

Fillmore, C.J. (1982). 'Frame Semantics'. In the Linguistic Society of Korea (Ed.), *Linguistics in the Morning Calm*, pp 111-137, Seoul: Hanshin.

Finn, A., Kushmerick, N. and Smyth, B. (2002). 'Genre classification and domain transfer for information filtering'. In Fabio Crestani, Mark Girolami, and Cornelis J. van Rijsbergen (ed.), In *Proceedings of ECIR-02, 24th European Colloquium on Information Retrieval Research*, Glasgow, UK. Springer Verlag, Heidelberg, DE.

Fischer, S. (1980). 'Dynamic Inconsistency, Cooperation, and the Benevolent Dissembling Government.', *Journal of Economic Dynamics and Control 2,* pp. 93-107.

Friedman, B., Kahn, P. H., Jr., and Borning, A. (2006). 'Value Sensitive Design and Information Systems', In P. Zhang & D. Galletta (ed.), *Human-Computer Interaction in Management Information Systems: Foundations,* pp. 348-372. New York: M. E. Sharpe.

Fries, C.C. (1952). *The Structure of English*. New York: Harcourt Brace.

Ganter, V. and Strube, M. (2009). 'Finding hedges by chasing weasels: Hedge detection using Wikipedia tags and shallow linguistic features.', In *Proceedings of the ACL-IJCNLP 2009 Conference, Short Papers*. Suntec, August 4, pp. 173-176.

Gelernter, D. (1998) *Machine Beauty: Elegance and the Heart of Technology*. Basic Books, New York,1998

Greene, S. and Resnik, P. (2009) 'More than words: Syntactic packaging and implicit sentiment.', In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the ACL (NAACL-2009)*. Boulder, Colorado, pp. 503-511.

Good, I.J. (1953) 'The Population Frequencies of Species and the Estimation of Population Parameters.' *Biometrika*, 40(3/4) pp. 237-264.

Griffiths, T., Steyvers, M., and Tenenbaum, J. (2007). 'Topics in Semantic Representation.' *Psychological Review*, 114(2), pp. 211–244.

Grunert, S.C. and Juhl, H.J. (1995). 'Values, environmental attitudes and buying of organic foods.', *Journal of Economic Psychology, 16* pp. 39-62.

Gurel-Attay, E., Xie, G., Chen, J. and Kahle, L.R. (2010) 'Changes in Social Values in the United States: 1976-2007 "Self-Respect" Is on the Upswing as "A sense of Belonging" Becomes Less Important' *Journal of Advertising Research*.

Guth, W.D. and Tagiuri, R. (1965). 'Personal Values and Corporate Strategy'. *Harvard Business Review, 43,* pp. 123-132.

Hatzivassiloglou, V. and McKeown, K. R. (1997) 'Predicting the semantic orientation of Adjectives.' In *Proceedings of the 35th Annual Meeting of the ACL and the 8th Conference of the European Chapter of the ACL*, Madrid, July 7-12, pp. 174–181.

Hearst, M. (1992) 'Direction-based text interpretation as an information access refinement'. In Paul Jacobs, (ed.), *Text-Based Intelligent Systems*, pp. 257–274. Lawrence Erlbaum Associates.

Heise, D.R. (1979). *Understanding Events*. Cambridge University Press, 1979.

Heise, D.R. (1987). 'Affect control theory: Concepts and model'. *Journal of Mathematical Sociology*, 13(1- 2) pp. 1–33.

Heise, D.R. (2006), 'Sentiment formation in social interaction', In McClelland, Kent (ed.) Fararo, Thomas, Purpose, Meaning, and Action: Control Systems Theories in Sociology, pp. 189–211.

Heise, D.R (2007). *Expressive order: Confirming sentiments in social actions.* Springer.

Henricksen, K., Indulska, J., & Rakotonirainy, A. (2002). 'Modelling Context Information in Pervasive Computing Systems'. *Pervasive Computing,* Vol. 2414, pp. 79-117, Springer Berlin/Heidelberg.

Henry, P. J. and Reyna, C. (2007), 'Value Judgments: The Impact of Perceived Value Violations on American Political Attitudes.' *Political Psychology*, 28, pp. 273–298.

Hevner, A. R., March, S. T., Park, J. and Ram, S. (2004) 'Design Science in Information Systems Research.', *MIS Quarterly*, 28(1), pp. 75-106.

Hitlin, S. (2003). 'Values as the core of personal identity: Drawing links between two theories of self.', *Social Psychology Quarterly*, 66(2), pp. 118-137.

Hitlin, S. & Piliavin, J. A. (2004). 'Values: Reviving a dormant concept.' *Annual Review of Sociology*, 30, pp. 359-393.

Ho, D.E. and Quinn, K.M. (2008). 'Measuring Explicit Political Positions of Media.' *Harvard Department of Government Working Paper.*

Hofstede, G. (1980). *Culture's consequences: International Differences in Work Related Values*. Beverly Hills, CA: Sage.

Hoyland C. I., Janis I. L., Kelley H. H. (1953). *Communication and Persuasion.* New Haven: Yale Press.

Hu, M. and Liu, B. (2004). 'Mining and Summarizing Customer Reviews',. *In Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004),* Seattle, WA, August 22-25, pp. 168-177.

Hutcheon, P.D. (1972). 'Value Theory: Towards Conceptual Clarification.' *The British Journal of Sociology,* 23(2), pp. 172-187.

Ishita, E., Oard, D.W. and Fleischmann, K.R. (2010) 'Investigating Multi-Label Classification for Human Values.' *ASIST 2010,* 47(5), October 22 – 27, Pittsburgh, PA.

James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R.* Springer Texts in Statistics. Casella, G., Fienberg, S., Olkin, I.(ed.) New York, 2013.

Jelinek, F (1991) 'Self-organized language modeling for speech recognition' In A. Waibel, K.F. Lee (ed.): *Readings in Speech Recognition*, Morgan Kaufman Publishers, 1991. pp. 450-506.

Jelinek, F. & Mercer, R. L. (1980). 'Interpolated estimation of Markov source parameters from sparse data.', In *Proceedings of the Workshop on Pattern Recognition in Practice*, North-Holland, Amsterdam, May, pp.381–397.

Jespersen, O. (1924). *Philosophy of Grammar*. London: George Allen & Unwin.

Jindal, N. and Liu, B. (2006a). 'Mining Comparative Sentences and Relations.', In *Proceedings of National Conference on Artificial Intelligence (AAAI-2006)*, Boston Massachusetts, July 16 - 20, pp. 1331-1336.

Jindal, N. and Liu, B. (2006b). 'Identifying Comparative Sentences in Text Documents.' In *Proceedings of ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR-2006),* Seattle, Washington, August 6-11, pp. 244-251.

Jindal, N. and Bing, L. (2007) 'Review Spam Detection'. In *Proceedings of the 16th International Conference on World Wide Web,* Banff, Alberta, May 8-12, pp. 1189-1190.

Jindal, N. and Bing L. (2008) 'Opinion spam and analysis.', In *Proceedings of the Conference on Web Search and Web Data Mining (WSDM-2008),* Palo-Alto, California, February 11-12, pp. 219-130.

Joachims, T. (1999) 'Making Large-Scale SVM Learning Practical', In Advances in Kernel Methods - Support Vector Learning, B. Schölkopf, C. Burges, and A. Smola, (ed.) MIT press.

Jung, Y., Park, H. & Myaeng, S. H. (2006). 'A hybrid mood classification Approach for Blog Text.', In Proceedings of the 9th Pacific Rim international conference on Artificial intelligence, Guilin, August 07-11, pp. 1099–1103.

Jurafsky, D. and Martin, J.H. (2009) *Speech Language Processing.* Oxford University Press, 2009, 2nd edition.

Jurka, T.P. (2012) 'Maxent: An R Package for Low-memory Multinomial Logistic Regression with Support for Semi-automated Text Classification.' *The R Journal*, Vol. 4/1, pp. 56-59.

Kahle, L. R., Poulos, B., and Sukhdial, A. (1988). 'Changes in social values in the United States during the past decade.' *Journal of Advertising Research*, 28, pp. 35-41.

Kantrowitz, M. (2000) *Method and Apparatus for Analyzing Affect and Emotion in Text.* U.S. Patent no. 6622140, 2003. Patent filed in November 2000.

Katz, S.M. (1987) 'Estimation of Probabilities from Sparse Data for the Language Model Component of a Speech Recognizer.', *IEEE Transactions on Acoustics, Speech and Signal Processing*, 35(3), pp. 400-401.

Kelly, J. (2016) 'Brexit: How much of a generation gap is there?' [Online]. Available at: http://www.bbc.co.uk/news/magazine-36619342 (Accessed: 15/12/2016).

Kim, S and Hovy, E. (2004). 'Determining the Sentiment of Opinions', In *Proceedings of International Conference on Computational linguistics (COLING-2004),* Geneva, August 23-27.

King, A.B. (2001) 'Affective dimensions of internet culture.', *Social Science Computer Review*, 19, pp. 414–430.

Kluckhohn, C. (1951). 'Values and value-orientations in the theory of action: An exploration in definition and classification', In T. Parsons & E. Shils (ed.), *Toward a general theory of action* (pp.388-433). Cambridge, MA: Harvard University Press.

Kneser, R. (1996) 'Statistical Language Modeling using a Variable Context Length', In *ICSLP-96,* Philadelphia, PA, October, Vol. 1, pp. 494-497.

Kneser, R. and Ney, H. (1995). 'Improved backing-off for M-gram Language Modeling', In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP-95,* May, Vol. 1, pp. 181-184.

Korphage, R. (1997). *Information Storage and Retrieval.* John Wiley & Sons, Inc., 1997.

Krippendorf, H.K. (2012) *Content Analysis: An Introduction to its Methodology.* 3rd edn. Sage: London.

Kubler, S., McDonald, R. & Nivre, J. (2009). *Dependency Parsing*. Morgan & Claypool Publisher.

Kukich, K. (1992). 'Techniques for Automatically Correcting Words in Text', *ACM Computing Surveys, 24*(4), pp. 377-439.

LDC (1993). *LDC Catalog: CSR-I (WSJ0) Complete.* University of Pennsylvania. Available at: www.ldc.upenn.edu/Catalog/LDC93S6A.html (Accessed: 03/12/2011).

Lev, I. (2005). *Syntax and Syntax-Semantics Interface.* [Lecture Ling221B] Spring, 2005.

Li, S., Lin, C., Song, Y. and Li, Z. (2010). 'Comparable Entity Mining from Comparative Questions', In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Uppsala, 11-16 July, pp. 650-658.

Lin, H.Y., Yang, C.H., and Chen, H.H. (2007). 'What Emotions Do News Articles Trigger in Their Readers?', In *Proceedings of 30th Annual International ACM SIGIR Conference on Research Development in Information Retrieval,* Amsterdam, 23-27 July, pp. 733-734.

Lin, H.Y. and Chen, H.H. (2008). 'Ranking Reader Emotions Using Pairwise Loss Minimization and Emotional Distribution Regression', In *Proceedings of Conference on Empirical Methods in Natural Language Processing*, Honolulu, 25-27 October, pp. 136-144.

Lin, W., Wilson, T., Wiebe, J., Hauptmann, A. (2006) 'Which Side are you on? Identifying Perspectives at the Document and Sentence Levels', In *Proceedings of the 10th Conference on Computational Natural Language Learning (CoNLL- 2006)*, New York, 8-9 June, pp. 109–116.

Lin, W. (2006). 'Identifying Perspectives at the Document and Sentence Levels using Statistical Models', In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Doctoral Consortium*, New York, 8-9 June, pp. 227–230.

Liu, B., Hu, M. and Cheng, J. (2005) 'Opinion observer: Analyzing and Comparing Opinions on the Web', In *Proceedings of International Conference on World Wide Web (WWW-2005)*, Chiba, 10-14 May, pp. 342-351.

Liu, B. (2011). *Web Data Mining: Exploiting Hyperlinks, Contents and Usage Data*. Springer.

Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan and Claypool.

Liu, H., Lieberman, H. and Selker, T. (2003). 'A model of textual affect sensing using Real-World Knowledge', In *Proceedings of the Eight International Conference on Intelligent User Interfaces*, Florida, 12-15 January, pp. 125–132.

Manjali, F.D. (1994), 'Dynamic Semiotics or the Case for Actantial Case', In *Semiotiques*, Paris, Vol. 5-6, pp. 85-97.

Manning, D.C., and Schutze, H. (1999). *Foundations of Statistical Natural Language Processing.* MIT Press.

March, S. T., and Smith, G. (1995) 'Design and Natural Science Research on Information Technology', *Decision Support Systems,* 15(4), pp. 251-266.

Markus, M. L., Majchrzak, A., and Gasser, L. (2002) 'A Design Theory for Systems that Support Emergent Knowledge Processes', *MIS Quarterly,* 26(3), pp. 179-212.

Maslow, A. H. (1965). *Eupsychian management*. Homewood, IL: Dorsey.

McCallum, A.K. (2002) 'MALLET: A Machine Learning for Language Toolkit.' Available at: http://mallet.cs.umass.edu (Accessed: 25/05/2012).

McDonald, P., and Gandz, J. (1991). 'Identification of Values Relevant to Business Research', *Human Resource Management*, 30(2), pp. 217-236.

McNair, D., Heuchert, J.P. and Shilony, E., (2003) *Profile of Mood States. Bibliography 1964-2002.* (Multi-Health Systems)

Melville, P., Gryc, W. and Lawrence, R.D. (2009). 'Sentiment Analysis of Blogs by Combining Lexical Knowledge with Text Classification', In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge Discovery and Data Mining*, Paris, June 28 - July 1, pp. 1275-1284.

Mejova, Y.A (2012) *Sentiment Analysis within and across Social Media Streams.* PhD thesis. University of Iowa [Online]. Available at: http://ir.uiowa.edu/etd/2943 (Accessed: 22/01/2017).

Metzler, D. and Croft, W.B. (2005). 'A Markov Random Field Model for Term Dependencies', In *Proceedings of the 28th ACM Conference on Research and Development in Information Retrieval (SIGIR)* pp. 472-479.

Metzler, D., Dumais S., and Meek, C. (2007) 'Similarity Measures for Short Segments of Text'. In *Proceedings of the 29th European conference on IR research*, Rome, 2-5 April, pp. 16-27.

Migration and Societal Integration Research Consortium, *Identity and Values,* Available at: http://www.migration.uni-jena.de/project4/about_the_project/index.php (Accessed: 23/05/2015)

Mihalcea, R. and Liu, H. (2006). 'A corpus-based approach to finding happiness', In *AAAI 2006 Symposium on Computational Approaches to Analysing Weblogs*, pp. 139–144.

Miller, G.A. (1995) 'WordNet: a lexical database for English', *Communications of the ACM,* 38(11), pp. 39-41.

Mohtarami, M., Lan, M. and Tan, C.L. (2013). 'From Semantic to Emotional Space in Probabilistic Sense Sentiment Analysis', In *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence,* Bellevue Washington, 14-18, July, pp. 711-717.

Morris, R. (1994). 'Computerized Content Analysis in Management Research: A Demonstration of Advantages & Limitations', *Journal of Management*. 20(4) pp. 903 - 931.

Muller, T. and Schutze, H. (2011). 'Improved Modeling of out-of-vocabulary words using Morphological Classes', In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, Portland, Oregon, 19-24, June, pp. 524–528

Murakami, A. and Raymond, R. (2010). 'Support or Oppose? Classifying Positions in Online Debates from Reply Activities and Opinion Expressions', In *International Conference on Computational Linguistics (ACL)*, pp. 869–875.

Musat C., Trausan-Matu S. (2010) 'The Impact of Valence Shifters on Mining Implicit Economic Opinions', In: Dicheva D., Dochev D. (ed.) *Artificial Intelligence: Methodology, Systems, and Application*, 6304, pp. 131-140.

Nasukawa, T. and Yi, J. (2003). 'Sentiment Analysis: Capturing favorability Using Natural Language Processing', In *Proceedings of the K-CAP-03, 2nd International Conference on Knowledge Capture*, Sanibel Island Florida, 23-25 October, pp. 70-77.

Neuendorf, K.A (2002) *The Content Analysis Guidebook*. Sage: London.

Neviarouskaya, A., Prendinger, H. and Ishizuka, M. (2009). 'Sentiful: Generating a Reliable Lexicon for Sentiment Analysis', In *Affective Computing and Intelligent Interaction and Workshops, (ACII 2009)*, pp. 1-6.

Ney, H. and Essen, U. (1991). 'On Smoothing Techniques for Bigram-based Natural Language Modeling', In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing '91*, 2, pp. 825–829.

Ney, H., Essen, U. and Kneser, R. (1994) 'On Structuring Probabilistic Dependences in Stochastic Language Modelling', *Computer Speech and Language*, 8, pp. 1-38.

Ney, H., Martin, S. and Wessel, F. (1997). 'Statistical language modeling using leaving-one-out', In *Corpus-Based Methods in Language and Speech Processing*, Kluwer, pp. 174–207.

Nigam, K., Lafferty, J., and McCallum, A. (1999) 'Using Maximum Entropy for Text Classification', In *IJCAI-99 Workshop on Machine Learning for Information Filtering*, pp. 61–67, Available at: http://www.kamalnigam.com/papers/maxent-ijcaiws99.pdf (Accessed on: 21/05/2014).

Nivre, J. (2005). *Dependency Grammar and Dependency Parsing*, Vaxjo University.

Norcross, J.C., Guadagnoli, E., and Prochaska, J.O. (2006). 'Factor structure of the Profile of Mood States (POMS): Two partial replications', *Journal of Clinical Psychology* 40, pp. 1270-1277.

Oepen, S., Kuhlmann, M., Miyao, Y., Zeman, D., Flickinger, D., Hajic, J., Ivanova, A. and Zhang, Y. (2014) 'SemEval 2014 Task 8: Broad-Coverage Semantic Dependency Parsing', In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, Dublin, August 23-24, pp. 63–72.

Onyimadu, O., Nakata, K., Wilson, T., Macken, D. and Liu, K. (2013). 'Towards Sentiment Analysis on Parliamentary Debates in Hansard', In: Kim W., Ding Y., Kim, H.G. (ed.) *Semantic Technology. JIST 2013. Lecture Notes in Computer Science*, 8388, pp. 48-50.

Onyimadu, O., Nakata, K., Wilson, T., Macken, D. and Liu, K. (2014). *Sentiment trends in Hansard.* Unpublished KTP Project between University of Reading and System Associates Ltd, Maidenhead.

Osgood, C.E., Suci, G.J. and Tannenbaum, P.H. (1957) *The Measurement of Meaning*. University of Illinois Press.

Pang, B., Lee, L. and Vaithyanathan, S. (2002). 'Thumbs up? Sentiment Classification Using Machine Learning Techniques', In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing (EMNLP),* pp. 79-86.

Peirce, C. S., Hartshorne, C., Weiss, P., and Burks, A. W. (1965). *Collected papers of Charles Sanders Peirce*. Cambridge, Mass: Belknap Press of Harvard University Press.

Perry, R.B. (1926) *General Theory of Value: Its Meaning and Basic Principles Construed in Terms of Interest.* New York: Longmans Green. Reprint, Cambridge, MA: Harvard University Press, 1954.

Polanyi, L & Zaenen, A. (2006). 'Contextual valence shifters', In *Computing Attitude and Affect in Text: Theory and Applications*, 20, pp. 1–10.

Polguere, A. and Mel'cuk, I.A. (2009). *Dependency in Linguistic Description*, University of Montreal, John Benjamins Publishing Company, Amsterdam Philadelphia.

Popescu, A. and Etzioni, O. (2005). 'Extracting Product Features and Opinions from Reviews.' In *Proceedings of the Human Language Technology Conference and the*

Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP), Vancouver, British Columbia, October 06-08, pp. 339-346.

Qiu, G., Bu, J., Liu, B. and Chen, C. (2009). 'Opinion Word Expansion and Target Extraction through Double Propagation.', In *Association for Computational Linguistics*, pp. 9-27.

Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. (1985) *A Comprehensive Grammar of the English language*. Longman, 1985.

Ratnaparkhi, A. (1996). 'A Maximum Entropy Part of Speech Tagger.' In *EMNLP 1996*, Philadelphia, PA. pp. 133-142.

Resnik, P. (1995). 'Using Information Content to Evaluate Semantic Similarity in a Taxonomy.', *International Joint Conference for Artificial Intelligence (IJCAI-95)*. Pp. 448-453.

Rescher, N. (1969). *Introduction to Value Theory*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Robinson, D.T and Smith-Lovin, L. (2006) 'Contemporary Social Psychological Theories, chapter Affect Control Theory.' *Stanford Social Sciences*, 2006.

Rohan, M. J. (2000). 'A rose by any name? The values construct.' *Personality and Social Psychology Review,* 4, pp. 255-277.

Rokeach, M. (1960). *The Open and Closed Mind.* New York: Basic Books.

Rokeach, M. (1973). *The Nature of Human Values.* New York: Free Press.

Roseman, I.J. & Smith, C.A. (2001). *Appraisal Theory: Overview, Assumptions, Varieties, Controversies.* New York, NY: Oxford University Press, USA.

Ruppenhofer, J. (2013). *FrameNet and Sentiment Analysis*, [Lecture at Universitat Hildesheim] November 14.

Ruppenhofer, J., Ellsworth, M., Petruck, M.R.L., Johnson, C.R., Baker, C.F., & Scheffczyk, J. (2016). *FrameNet II: Extended Theory and Practice*. Available at: https://framenet2.icsi.berkeley.edu/docs/r1.5/book.pdf. (Accessed: 17/03/2017).

Russell, S. and Norvig, P. (2002). *Artificial Intelligence: A Modern Approach.* 2^{nd} edn. Prentice Hall.

Scherer, K. R., Shorr, A., and Johnstone, T. (Ed.). (2001). *Appraisal Processes in Emotion: Theory, Methods, Research.* Canary, NC: Oxford University Press.

Schmidt, J. (2007). 'Blogging practices: An analytical framework.' *Journal of Computer-Mediated Communication*, 12(4). Available at: http://jcmc.indiana.edu/vol12/issue4/schmidt.html. (Accessed: 21/01/2011)

Schön, D. A. (1983) *The Reflective Practitioner: How Professionals Think in Action, Basic Books*, New York, 1983.

Schwartz, S.H. (1992). 'Universals in the content and structure of values: Theory and empirical tests in 20 countries.' In M. Zanna (Ed.), *Advances in experimental social psychology,* Vol. 25, pp. 1-65. New York: Academic Press.

Schwartz, S.H. (1994). 'Are there universal aspects in the structure and contents of human values?' *Journal of Social Issues*, 50 (4), pp. 19-45.

Schwartz, S.H. (1996). 'Value priorities and behavior: Applying a theory of integrated value systems In C. Seligman, J. M. Olson, & M. P. Zanna (ed.), The psychology of values: The Ontario Symposium, vol. 8, pp. 1-24. Hillsdale, NJ: Erlbaum.

Schwartz, S.H. (2004a). 'Basic human values: their content and structure across cultures.' In *Values and Behaviour in Organizations,* (ed.) A. Tamayo & J. B. Porto (ed.), pp. 21-55. Brazil: Vozes.

Schwartz, S.H. (2004b). 'Mapping and interpreting cultural differences around the world.' In *Comparing Cultures, Dimensions of Culture in a Comparative Perspective*, (ed.) Vinken, H., Soeters, J., Ester, P. Leiden, The Neth.: Brill.

Schwartz, S.H. (2006). 'Basic human values: Theory, measurement, and applications.' *Revue française de sociologie.*

Schwartz, S.H. (2007). 'Value orientations: measurement, antecedents and consequences across nations.' In R. Jowell, C. Roberts, R. Fitzgerald, & G. Eva (ed.), Measuring Attitudes Cross-Nationally: Lessons from the European Social Survey, pp. 169-203. London: Sage.

Schwartz, S.H. (2012). 'An Overview of the Schwartz Theory of Basic Values.' *Online Readings in Psychology and Culture,* 2(1). Available at: http://scholarworks.gvsu.edu/cgi/viewcontent.cgi?article=1116&context=orpc. (Accessed: 27/03/2013).

Schwartz, S.H. (n.da). Basic Human Values: An Overview, Basic Human Values: Theory, Methods, and Applications. Available at: http://segr-did2.fmag.unict.it/Allegati/convegno%207-8-10-05/Schwartzpaper.pdf. (Accessed: 14/02/2014).

Schwartz, S.H. (n.db). 'A Proposal for Measuring Value Orientations across Nations' Available at: http://www.europeansocialsurvey.org/docs/methodology/core_ess_questionnaire/ESS_core_questionnaire_human_values.pdf. (Accessed: 01/03/2014).

Schwartz, S.H. and Bilsky, W. (1987). 'Toward a Universal Psychological Structure of Human Values.', *Journal of Personality and Social Psychology*, 53, pp. 550-562.

Schwartz, S. H. and Boehnke, K. (2004). 'Evaluating the structure of human values with confirmatory factor analysis.', *Journal of Research in Personality*, 38, pp. 230-255.

Schwartz, S.H. and Rubel-Lifschitz, T. (2005) 'Sex Differences in Value Priorities: Cross-Cultural and Multi-Method Studies.' *Journal of Personality and Social Psychology,* 89(6), pp. 1010-1028.

Schwartz, S. H. and Rubel-Lifschitz, T. (2009). 'Cross-national variation in the size of sex differences in values: Effects of gender equality.', *Journal of Personality and Social Psychology*, 97, pp. 171-185.

Scott, W. A. (1965). *Values and Organizations: A Study of Fraternities and Sororities*. Chicago: Rand McNally.

Searle, J.R. (1970) *Speech Acts: An Essay in the Philosophy of Language*, Cambridge University Press.

Shannon, C.E. (1948). 'A Mathematical Theory of Communication', *Bell Syst. Tech. j.,* 27, pp. 398-403.

Shorthouse, R. and Kirby, D. (2015) 'A Balanced Centre-Right Agenda on Immigration - A Manifesto for Immigration', *Bright Blue Campaign*, 2015.

Simon, H. (1996). *The Science of the Artificial*, 3rd edn., Cambridge, MA, MIT Press.

Smith, M.B., Bruner, J.S. & White, R.W. (1956) *Opinions and Personality,* New York, Wiley.

Smith-Lovin, L. & Douglas, W. (1992) 'An affect control analysis of two religious subcultures.' *Social Perspective in Emotions*, 1, pp. 217–248.

Somasundaran, S. and Wiebe, J. (2009) 'Recognizing Stances in Online Debates.' In *Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP (ACL-IJCNLP-2009),* Suntec, August 2-7, pp. 226-234.

Somasundaran, S. and Wiebe, J. (2010). 'Recognizing stances in ideological on-line debates.', In *NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pp. 116–124.

Sparck, J.K. (1972). 'A Statistical Interpretation of Term Specificity and its Application in Retrieval.', *Journal of Documentation,* 28(1), pp. 11-21.

Sridhar, D., Getoor, L., Walker, M. (2014) 'Collective Stance Classification of Posts in Online Debate Forums.', In *Proceedings of the Joint Workshop on Social Dynamics and Personal Attributes in Social Media*, Baltimore, Maryland, June 27, ACL, pp. 109–117,

Stolcke, A. (2002) 'SRILM—an Extensible Language Modeling Toolkit', in J. H. L. Hansen and B. Pellom, (ed.), *Proc. ICSLP*, vol. 2, pp. 901– 904, Denver, September 2002.

Stolcke, A. and Shriberg, E. (1996). 'Statistical Language Modeling for Speech Disfluencies.', In *ICASSP-96,* Atlanta, GA. pp. 405-408.

Stolcke, A., Zheng, J., Wang, W., Abrash, V. (2011) 'SRILM at sixteen: Update and Outlook', In *Proceedings IEEE Automatic Speech Recognition and Understanding Workshop*.

Strapparava, C and Mihalcea, R. (2007). 'Semeval-2007 task 14: Affective text.', In *Proceedings of the 4th International Workshop on the Semantic Evaluations (SemEval 2007)*, Prague, June 23-24, pp. 70-74.

Subasic, P. and A. Huettner (2001). 'Affect Analysis of Text Using Fuzzy Semantic Typing.' *IEEE-FS*, 9, pp. 483–496.

Stamper, R.K. (1973). *Information in Business and Administrative Systems*. John Wiley and Sons, New York.

Stamper, R.K. (1992). 'Language and Computer in Organised Behaviour.' In Riet, R.P.v.d. and Meersman, R.A., (eds.), *Linguistic Instruments in Knowledge Engineering.* Elseiver Science, Amsterdam, pp. 143-163.

Stamper, R. K., Liu, K., Sun, L., Tan, S., Shah, H., Sharp, B, and Dong, D. (2004). 'Semiotic Methods for Enterprise Design and IT Applications.', In *Proceedings of the 7th International Workshop on Organisational Semiotics*, pp. 190-213.

Su, Q., Hu, X., Guo, H., Guo, X., Wu, X., Zhang, X., Swen, B. and Su, Z. (2008). 'Hidden Sentiment Association in Chinese Web Opinion Mining.', In *Proceedings of International Conference on World Wide Web (WWW-2008)*, pp. 959-968.

Takayama, Y., Tomiura, Y., Ishita, E., Oard, D., Fleischmann, K.R. and Cheng, A.S. (2014). 'A Word-Scale Probabilistic Latent Variable Model for Detecting Human Values.', In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management CIKM, 14,* November 3-7, Shanghai, pp. 1489-1498.

Tang, Y. and Chen, H. (2011) 'Emotion Modeling from Writer/Reader Perspectives Using a Microblog Dataset.', In *Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP), IJCNLP 2011,* Chiang Mai, Thailand, November 13, pp. 11–19.

Templeton, T.C., Fleischmann, K.R. and Boyd-Graber, J. (2011a). 'Comparing Values and Sentiment Using Mechanical Turk.' In *iConference 2011*, Seattle, W.A., February 8-11, pp. 783-784.

Templeton, T.C., Fleischmann, K.R. and Boyd-Graber, J. (2011b). 'Simulating Audiences: Automating Analysis of Values, Attitudes and Sentiment.', In *Proceeding of the Third IEEE International Conference on Social Computing,* Boston, October 9-11, pp. 734-737.

Tesniere, L. (1959), *Elemenets de Syntaxe Structurale*. Librairie C. Klincksieck, Paris.

Thomas, M., Pang, B. and L. Lee. (2006). 'Get out the vote: Determining support or opposition from Congressional floor-debate transcripts.', In *Conference on Empirical Methods in Natural Language Processing (EMNLP),* Sydney, July 22-23, pp. 327–335.

Titov, I. and McDonald, R. (2008). 'A Joint Model of Text and Aspect Ratings for Sentiment Summarization.', In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL-2008),* Colombus, Ohio, June, pp. 308-316.

Tong, R. (2001). 'An Operational System for Detecting and Tracking Opinions in online discussion', *SIGIR 2001 Workshop on Operational Text Classification*, pp. 1-6.

Toprak, C., Jakob, N. and Gurevych, I. (2010). 'Sentences and Expression Level Annotation of Opinions in User-Generated Discourse.', In *Proceedings of ACL-10, the 48th Annual Meeting of the Association for Computational Linguistics*, Uppsala, July 11-16, Association for Computational Linguistics, pp. 575-584.

Tsoumakas, G. and Katakis, I. (2007). 'Multi label classification: An overview.', *International Journal of Data Warehouse and Mining*, vol. 3(3), pp. 1–13.

Turney, P. (2002) 'Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews.', In *Proceedings of the Association for Computational Linguistics (ACL),* Philadelphia, Pennsylvania, July 07 - 12, pp. 417–424.

Valentin, J., de Rijke, M. and Weerkamp, w. (2010). 'Generating Focused Topic-Specific Sentiment Lexicons.', In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL-2010),* Uppsala, July 11-16, pp. 585-594.

Van Rijsbergen, C.J. (1975). *Information Retrieval.* Butter-worths.

Verplanken, B. and Holland, R.W (2002). 'Motivated decision making: Effects of Activation and Self-Centrality of Values and Choices and Behaviour.', *Journal of Personality and Social Psychology,* 82(3) pp. 434-447.

Walls, J. G., Widmeyer, G. R., and El Sawy, O. A. (1992) 'Building an Information System Design Theory for Vigilant EIS,' *Information Systems Research* (3:1), March, pp. 36-59.

Wiebe, J. (1990) 'Identifying Subjective Characters in Narrative', In *Proceedings of the International Conference on Computational Linguistics (COLING)*, Helsinki, August 20-25, pp. 401–408.

Wiebe, J (1994). 'Tracking point of view in narrative'. *Computational Linguistics*, 20(2) pp. 233–287.

Wiebe, J., Bruce, R. and O'Hara, T.P. (1999) 'Development and Use of a Gold Standard Data Set for Subjectivity Classifications.', In *Proceedings of the Association for Computational Linguistics (ACL)*, College Park, Maryland, June 2-26, pp. 246-253.

Wiebe, J., Bruce, R., Bell, M., Martin, M. and Wilson, T. (2001) 'A corpus study of evaluative and speculative language.', In *Proceedings of the 2nd ACL SIGdial Workshop on Discourse and Dialogue*, Aalborg, September 1-2, pp. 1-10.

Wiebe, J, Wilson, T and Cardie, C. (2005). 'Annotating Expressions of Opinions and Emotions in Language.*', Language Resources and Evaluation*, 39(2), pp. 165-210.

Wiebe, J. and Riloff, E. (2005). 'Creating Subjective and Objective Sentence Classifiers from Unannotated Texts.' In *Computational Linguistics and Intelligent Text Processing*, pp. 486-497.

Williams, R.M. (1979). 'Change and Stability in Values and Value Systems: A Sociological Perspective.' In Rokeach, M. (ed.), *Understanding Human Values Individual and Societal,* pp. 15-46, New York: Free Press.

Wilson, T., Wiebe, J. and Hoffmann, P. (2005). 'Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis.', In *Proceedings of HLT-EMNLP-2005*, pp. 1-8. Available at: https://people.cs.pitt.edu/~wiebe/pubs/papers/emnlp05polarity.pdf. (Accessed: 02/03/2012).

Wissler, L., Almashraee, M., Monett, D. and Panschke, A. (2014). 'The Gold Standard in Corpus Annotation'. *IEEE, GSC*.

Yang, C.H., Lin, H.Y., and Chen, H.H. (2007a). 'Building Emotion Lexicon from Weblog Corpora.', In *Proceedings of 45th Annual Meeting of Association for Computational Linguistics*, Prague, June, pp. 133-136.

Yang, C.H., Lin, H.Y., and Chen, H.H. (2007b). 'Emotion Classification Using Web Blog Corpora.', In *Proceedings of 2007 IEEE/WIC/ACM International Conference on Web Intelligence*, Washington DC, November 2-5, pp. 275- 278.

Yang, C.H., Lin, H.Y., and Chen, H.H. (2008).' Sentiment Analysis in Weblog Using Contextual Information: A Machine Learning Approach'. *International Journal of Computer Processing of Languages* 21(4) pp. 331–345.

Yang, Y. and Liu, X. (1999). 'A Re-Examination of Text Categorization Methods'. In M.A. Hearst, Gey, F. and Tong, R, (ed.), *Proceedings of SIGIR0-99, 22nd ACM International Conference on Machine Learning – ICML-97,* pp. 42-49.

Yessenalina, A., Yue, Y. and Cardie, C. (2010). 'Multi-level structured models for Document-Level Sentiment Classification'. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, MIT, Massachusetts, October 9-11, pp. 1046–1056.

Zhang, Y., Shen, D. and Baudin, C. (2011). *Sentiment Analysis in Practice*. EBay Research Labs.

Zwicky, A. M. (1985). 'Heads'. *Journal of Linguistics*, 21, pp. 1–29.

APPENDIX 1: Sentiment Analysis Types and Opinion Structures

A1.1 Direct and Indirect Opinions

Regular opinions are expressed at the sentence level and can be divided into direct and indirect opinions. Direct opinions are a type of regular opinions. They are mostly simple, short sentences where the sentiment of the subject, entity or aspect is expressed directly. On inspection, most direct opinions will feature a single predicate, which could be a verb, verb phrase or compound verb. For example, *"The new MacBook Pro has great picture quality."* The subject entity here is *"The new MacBook Pro"* and the sentiment expressed is the phrase *"great picture quality"*. In addition, the expressed sentiment could be an adverb, adjective string or phrase e.g. *The mobile phone melted, broke apart and fell into the flames"*. Conversely, indirect opinions are "expressed indirectly on an entity or aspect of an entity based on its effect on some other entities" (Liu, 2012) e.g. *"The loss of his dog led to his depressed state of mind"*. Here the phrase *"depressed state of mind"* is an indirect consequence of *"loss of his dog"*.

Direct/indirect sentiment expressions could be explicit or implicit (Balahur et al, 2011; Liu, 2011; Liu, 2012). In the explicit case, the subject of the opinion, which is an entity or aspect is clearly expressed and not subject to inferential interpretations - *"MacBook's have great battery life"*. Conversely in implicit sentiments, the aspects are not explicitly mentioned in the sentence but implied e.g. *"The new PC will not easily fit into any average sized pouch"*. Implicitly, the sentiment expressed is for the size of the PC.

A1.2 Comparative Opinions

A comparative opinion expresses a comparison of entities based on their shared aspects (Jindal and Liu, 2006a, 2006b; Zhang et al, 2011; Liu, 2012). People will sometimes express their sentiment towards a target by comparing similar entities. For instance, one might say *"I prefer PCs to Macs"*. In this example, the user expresses a preference for one product referencing another as a base line. For comparative opinions, the sextuplet representation fails. In Jindal and Liu (2006a, 2006b), "comparative opinions are said to express a relation of similarities or differences between two or more entities and or/a preference of the opinion holder based on some shared aspects of the entities". The expression of comparative opinions is signaled by the presence of adjectival superlatives e.g. *"Car A is better than car B"*. The superlative 'better' in the example sentence is indicative of a comparative opinion. Importantly this sort of features can be used in building syntactic pattern engines for identifying opinions in a corpus.

Comparative opinions can also be expressed subtly. Liu (2012) calls this type opinionated comparative sentences e.g. *"I prefer the smaller iPad to the older, bigger one"*. This example is opinionated because the speaker expresses a clear preference for one product in comparison to another. However, it is possible to express a comparison without explicitly expressing a sentiment. For instance, *"The older iPad is far bigger than the newer one"* provides a factual commentary of two products while expressing no explicit sentiments.

One significant difference between comparative opinions and regular opinions is in their syntactic structure. Strong patterns involving comparative keywords are key pointers to comparatives (Jindal and Liu, 2006b). Comparative opinions can be identified from their word order and the use of comparators (*more, suffix -er, less)* and the equative (Hu and Liu, 2004). Comparisons can be grouped into two main categories (Jindal and Liu, 2006a, 2006b) - Gradable comparatives and Non-gradable comparatives (see A1.3 and A1.4 for notes).

A1.3 Gradable Comparatives

Gradables express a measure of an individual's sentiment on a graded scale characterized by an expressive adjective. For example, an individual's use of the word "*tall*" translates to a measure of height, and "*heavy*" references a measure of weight. In linguistics, gradable comparatives fall into several different cases (Lev, 2005). In the predicative instance, the comparison is expressed through an adjectival predicate that compares two entities for instance *"Car A is faster than car B"*. Lev (2005) and Liu (2012) defines this predicative type non-equal gradable comparison because it emphasizes a form of ranking where the expression ranks one entity higher than the other, hence the non-equality. The attributive case bases the comparison around an attribute or attributes of the subject: *"Mr A read a longer book than he read a magazine"*.

Another case may involve an "explicit specification of the difference between the compared degree" (Lev, 2005). Comparatives in this instance are like the non-opinionated examples mentioned earlier. Lev (2005) also highlights examples where the comparison is direct but also a comparison to a reference point e.g. *"Mr A is taller than 7 feet"*. The comparison is about another comparative measure.

In addition to non-equal gradable comparison, two more types of gradable comparatives are emphasized in Liu (2012). They include equative comparison and superlative comparison. The former expresses a relation of the type *equal to* illustrating that two or more entities are equal based on some shared aspect. For instance, *"Sony TVs and Samsung TVs both sound the same"*. Superlative comparison expresses a relation of the type greater or less than all others. It is a form of ranking one entity over all the others e.g. *"Of all the tablet brands, iPad is the best"*.

A1.4 Non-Gradable Comparatives

Non-gradable comparatives do not express scalable sentiments and are often expressed using adjectives such as *different* or *similar*. For example, *"Mr A's salad taste differently to his son's"*. The phrase *"taste differently"* does not give any indication as to the level of difference in taste neither does it delineate the taste levels in question yet there is an obvious comparison expressed. As such, such opinions are called non-gradable comparatives. There are three known categories (Jindal and Liu, 2006b; Liu, 2012). The first expresses similarity or difference between two opinion targets based on one or more of their shared aspects e.g. *"Macs and PCs operate differently"*. The second type normally features two different entities with different aspects that cannot be substituted e.g. *"Radios have in built speakers while mp3 players require external speakers"*. The presence of *'while'* is normally a good indicator. In the third type, one entity has an aspect that the other does not have for instance, *"The PCs come with cd players while the new MacBooks do not"*.

Prominent features of comparatives are the presence of adjectives or adverbs. In fact, most opinionated sentences whether regular or comparative are marked by the presence of adjectives and adverbs. This is a vital feature for developing word patterns and rules for identifying sentiment sentences in research and industry.

Modeling comparatives involves two critical aspects: Given a set of sentences, first identify and classify the comparative sentences, secondly, model the comparative sentence by extracting comparative relations from the identified sentences. The second step involves the extraction of entities and the features to be compared (Jindal and Liu, 2006a, 2006b; Liu, 2012). Jindal and Liu (2006a) achieve this by first identifying the entities mentioned. Most times there will be two entities *(entity1)* and *(entity2)*, the feature or shared aspect for which the comparison is based *(a)* and the word used in relating the features *(w)* resulting in a quadruplet model *(entity1, entity2, a, w)*. For instance, in the sentence, *"Diesel cars perform better than petrol and electric cars"* the extracted relation will be: *({"Diesel cars"}, "{petrol cars, electric cars}", "performance", "better")*. In Li et al (2010), the model is expanded to a sextuplet featuring the objects *(entity1)* and *(entity2)*, their shared aspect *(a)* the preferred object set of the opinion holder *(p_0)* and the time *(t)*. This classification is more expansive and fine grained as it includes aspects of time (for observing changes in opinion holder sentiment) and most importantly the inclusion of the preferred object to provide a means by which polarities could be allocated to the entities. Note that this sextuplet classification is different from the structure expressed in Onyimadu et al (2013, 2014). Just like in the regular opinion model, this kind of opinion expression is a formal structure of opinion which can be cached and processed in a database or search engine to support complex information retrieval queries.

A recurring theme in comparatives is the presence of adjectival comparatives and superlatives in sentences. It has even been shown that almost every comparative opinion has a comparative keyword indicator (Kantrowitz, 2000; Lev, 2005), however not all sentences containing adjectives are comparatives for example, *"I will not go any further"*. This point is quite important because as mentioned earlier, comparative words (mostly adjectives and adverbs) form strong syntactic patterns which can be used as machine learning features in the identification of comparatives.

A1.5 Opinion Structure

Central to the concept of regular opinions and essential to differentiating opinion types is the structure of the opinion. Opinions are expressed by individuals, called opinion holders *(h)* (Kim and Hovy, 2004; Wiebe et al, 2005). The expressed opinion is targeted at a subject or entity *(s)* called opinion target. Likewise, the opinion itself will have an associated polarity *(p)*, which could be expressed categorically or numerically. Since opinions change, it is important to include the time the opinion was expressed *(t)*. Inclusion of time facilitates the observation of sentiment ebb and flow over a period. These variables *(h, s, t, p)* are called opinion quadruple (Liu, 2012).

The quadruple model is quite deficient, as it does not always capture the actual subject of the sentiment. For instance, in the sentence *"The sound quality of the new MacBook is fantastic"*, the sentiment is directed at *'the sound quality' (s)*. Representing the quadruple with this value of *s* does not actually tell the whole story because it is of little importance without any knowledge of the fact that it refers to *'the new MacBook'*.

Hu and Liu (2004), introduce the concept of entity decomposition wherein the target can be broken down into a structured hierarchy of sub-components and aspects to enable mining of opinions. Decomposing the previous example will result in, *'the new MacBook'* as the entity and *'the sound quality'* as an attribute. According to Liu (2012), entities have parts (aspects), resulting in a part-of relationship between entities and their parts. A '*MacBook*' (A type of computer) could have parts keyboard, speaker, microphone, screen etc. each possessing their own individual aspects or features - the screen for instance could be high resolution LCD or CRT. The final opinion model based on the decomposition of entities is an opinion quintuple *(h, s, t, p, a)* where (*a*) is an aspect of the entity *(s)* (Hu et al, 2004; Liu, 2012).

Onyimadu et al (2013), enhances the quintuplet to include the sentiment intensity *(i)*. The rationale behind this was that sentiments, whether positive or negative have different intensities. *"Very good"* and *"good"* while both positive connote different levels of positivity, same applies to *"unhappy"* and *"depressed"*, the intensity, stress or emotional level expressed in these examples can be ranked because of grammatical intensifiers such as *'very'* or by adjectival superlatives - e.g. *'good', 'better', 'best'*. This gives an opinion sextuplet - *(h, s, t, p, a, i)*.

Comprehensive as this model of regular opinion is, it still does not cover all situations. Liu (2012), suggests that the model fails to account for situations where the entity is referenced in terms of multiple features for example in the sentence, *"The view finder and lens are too close"*. Here opinion on two different parts is expressed. Liu (2012) indicates further that it also does not adequately cover the context of the opinion. The sentence *"The dress will not fit big people"* is not really referencing the *'dress'* but size (*"big people")*. Also, Liu (2012) suggests further that for the model to work all opinion components must be present and that the absence of any of the six components i.e. *(h, s, t, p, a, i)* will render it useless. For instance, the absence of time *(t)* means opinion transitions might be unobservable, the absence of opinion target *(s)* realistically makes the model useless because sentiments are often directed at a target. Onyimadu et al (2013) shows however that the model still works in the absence of some of the components and that opinion target *(s)* and polarity *(p)* are the only indispensable parts.

One benefit of this model is that it provides a framework for transforming the unstructured text to a structured data model so that it can be easily employed for computational purposes. Basically, the sextuplet represents a data structure that can be modelled as a database schema, object oriented data model or represented relationally. In Onyimadu et al (2013), components are stored and implemented in Mongo[94], followed by a similar JSON style representation in SOLR[95] to support information retrieval (Onyimadu et al, 2014). Based on these works, the absence of components such as time and opinion source will not necessarily nullify the model because queries such as *"All positive comments on Health where the source is unsubstantiated i.e (source is null)"* can be served.

A1.6 Analysis Levels

The level of analysis refers to the text span that must be captured to sufficiently analyze the sentiment. This section provides a review of the existing levels of analysis adopted in industry and research.

Analysis of sentiments can be carried out at various levels of granularity. By levels, reference is made to the object of analysis or the unit of analysis. In other words, given a sentence, what aspect or aspects can the sentiments be interpreted from? Identification of entities and their aspects discussed earlier is an analysis level that focuses on identifying the sentiment polarity of the entities and features mentioned in the sentence. In other words, the objective of the sentiment analysis task is to take as input a sentence and identify the polarity of entities and features mentioned. Two other levels exist in current research. They include document level and sentence level.

Sentence level opinion furnishes a lower analysis level of abstraction where the unit of analysis is the sentence. As compared to document level analysis, the objective here is to determine the polarity of each sentence. Much of the early work on sentence level analysis focused on the identification of subjective sentences and their classification into one of two classes - objective and subjective. Also of importance is the fact that some sentences may fall on the borderline between objective and subjective and this accounts for the use of objective and subjective ratings in human subjectivity annotation in Wiebe et al (1999).It is also important to distinguish both subjectivity and sentiment: According to Liu (2012) and Wiebe et al (1999), subjectivity is not equivalent to sentiment because many objective sentences can imply opinions for instance, *"My new iPad has suddenly stopped working!"*. In this example, the speaker expresses a fact, an objective utterance for which as humans some sort of sentiment can be associated. However, the sentence is not subjective. In addition, some subjective expressions do not express any sentiment for instance, *"I assume your bride will arrive today"*. It is conclusive to say that while subjectivity does not equate

[94] Mongo DB is a crossplaform NOSQL database.

[95] SOLR is an open source enterprise search platform

to sentiment, it can imply sentiment just as objectivity can imply sentiment but does not equate to sentiment. In sentence level opinion analysis, the common approach to detecting sentiments in a document is to filter non-subjective and non-opinionated sentences and subsequently determine the polarity of the entity.

At the document level, the problem is presented as a text classification problem involving the classification of an entire document as positive, negative or neutral (Pang et al, 2002; Turney, 2002; Lev, 2005; Valentin et al, 2010).

Central to document level opinion analysis is the assumption that the document is

- Written by a single person
- The expressed opinion is directed at a single entity making it quite difficult to carry out document level analysis on blogs or postings because they often express opinions on multiple subjects.

Realistically, document level analysis is too coarse for most opinion needs as individuals often express several opinions on diverse opinion targets. As a classification problem, document level analysis will in most cases require the collection of feature vectors like adjectives, term part of speech, word order and opinion words.

APPENDIX 2: Value Inventories (VI)

A2.1 Schwartz Value Inventory

Inventory	Motivation/Goal	Value Items
Self-Direction	Independent thought and action. Need for control and mastery (Bandura, 1977).	Choosing, creating, exploring, autonomy, independence, curious, self-respect, intelligent, privacy.
Stimulation	Need for variety and stimulation in order to maintain an optimal, positive rather than threatening level of activation (Berlyne, 1960)	Excitement, novelty, challenge, daring, a varied life, exciting life
Hedonism	Organismic needs and the pleasure associated with satisfying them (Schwartz, 2012).	Pleasure, self-indulgent, enjoying life
Achievement	"Personal success through demonstrating competence according to social standards" (Schwartz, 2012)	Successful, capable, ambitious, influential, intelligent, self respect
Power	Need for individual dominance and control. Also control or dominance over people and resources. (Schwartz, 2012)	Social power, authority, wealth, preserving my public image and social recognition
Security	Need for personal, group and national safety. Both physical and mental safety.	Clean, national security, social order, family security, reciprocation of favours, healthy, sense of belonging
Conformity	"The requirement that individuals inhibit inclinations that might disrupt and undermine smooth group interaction and functioning." (Schwartz, 2012)	Obedient, self-discipline, politeness, honour
Tradition	Acceptance and respect for culture, customs and ideas.	Devout, accepting portion in life, humble, moderate, respect for tradition, detachment

Benevolence	Derives from the need for affiliation (Maslow, 1965) and smooth group functioning (Kluckhorn, 1951)	Preserving and enhancing the welfare of others, kindness
Universalism	Concern and acceptance of others, nature and the world	Broadminded, social justice, equality, unity, harmony

A2.2 Rokeach Value Survey (RVS)

Terminal Values – An exciting life, pleasure, mature love, true friendship, inner harmony, social recognition, a sense of accomplishment, family security, national security, self-respect, health, a comfortable-life, freedom, salvation, equality, wisdom, a world at peace and a world of beauty.

Instrumental Values – Ambitious, broad-minded, capable, clean, cheerful, courageous, forgiving, helpful, honest, imaginative, independent, intellectual, logical, loving, obedient, polite, responsible and self-controlled.

A2.3 Personal Values Questionnaire

(England, 1967; Cheng and Fleischmann, 2010)

Group 1: Goal of Business Organizations – High productivity, industry leadership, employee welfare, organizational stability, profit maximization, organizational efficiency, social welfare, organizational growth.

Group 2: Personal Goals and Individuals – Leisure, dignity, achievement, autonomy, money, individuality, job satisfaction, influence, security, power, creativity, success, prestige.

Group 3: Groups of People – Employees, customers, my co-workers, craftsman, my boss, managers, owners, my subordinates, labourers, my company, blue collar workers, government, stockholders, technical employees, me, labour unions, white collar employees.

Group 4: Ideas associated with people – Ambition, ability, obedience, trust, aggressiveness, loyalty, prejudice, compassion, skill, cooperation, tolerance, conformity, honour.

Group 5: Ideas about general topics – Authority, caution, change, competition, compromise, conflict, conservatism, emotions, equality, force, liberalism, property, rational, religion, risk.

A2.4 Personal Value Scale (PVS)

PVS (Scott, 1965) was derived empirically for analysing individual concepts of ideal relations among people. A survey of college students inquiring about what traits they admired in others formed the basis of the survey questions. A multi-question instrument to measure the values acknowledged by the students followed resulting in twelve values. They include: Intellectualism, kindness, social skills, loyalty, academic achievement, physical development, status, honesty, religiousness, self-control, creativity and independence.

A2.5 Managerial Moral Standards

Managerial moral standards by Bird and Waters (1987) was designed for organizations, focusing on the moral standards held by managers in their work life. Moral issues arising in the daily life of managers were identified through interviews across a sample of managers. Identified commonalities across the respondents formed the managerial moral standards. The values include honesty in communication, fair treatment, special consideration, fair competition, organizational responsibility, corporate social responsibility and respect for the law. This value system is ideal for individuals in an organization and not applicable to a wide set of domains.

A2.6 Meta-Inventory of Human Values (MIHV)

Significant portions of existing inventories are designed for survey purposes and applicable only to certain domains or contexts. For instance, the SVI is best suited to general social contexts or situations, while the PVQ and LOV are particularly suited to organizations. Considering this, Cheng and Fleischmann (2010) developed a one-size fits all inventory from a synthesis of 12 VIs including the ones mentioned above. Their aim was the development of a wide ranging one-size fits all inventory that can be tailored by researchers to measure human values. The inventory consisted of 16 value categories aggregated from different domains that address "general individual values, work values, managerial values and values for technology design" (Cheng and Fleischmann, 2010). Compared to other value inventories the MIHV is more manageable and offers users flexibility in the choice of values to use in their research. Most importantly, it is prone to less ambiguity as it unifies related value items and concepts. For instance, concepts such as achievement and success may be ambiguous in the same inventory, however when synthesized, under the term accomplishment, the ambiguity is eliminated.

A2.7 Latent Variable Model (LVM)

LVM (Takayama et al, 2014) was developed for detecting values in text. It frames the value detection problem as a multi-category classification problem. The intuition behind LVM is based on some theoretical standpoints:

- Human values are latent components that influence behaviour and as such can be modelled as latent probabilistic variables (Verplanken and Holland, 2002).
- People will often use words and phrases that reflect their values and so for every sentence, each word used can be associated with a particular value.
- The syntactic context of content is a clue for detecting values expressed in snippets of content.

These assumptions served as the primary basis for applying supervised Probabilistic Latent Variable Model (PLVM) capturing the relationship between word, sentences and their values. Since words reflect values, they surmised that the sum total of values reflected by each word in a sentence would represent the values expressed in the sentence. In summary, the LVM's methodology involved first detecting word level human values corresponding to each word in a sentence as latent variables and then aggregating these variables logically to estimate the sentence level human values. Summation of word values equate to sentence level values.

Training data derived from net neutrality debates were annotated for values using the MIHV as its type classification (Cheng, 2010). Six value types from the MIHV were considered relevant to the domain. For instance, the sentence *"I am one of the network engineers involved for many years in designing, implementing and standardizing the software protocols that underpin the internet"* was assigned the value type *Honour* by coders and the sentence *"Congress enacts* ***safeguards*** *to preserve American consumers' longstanding* ***freedom*** *of internet content* ***choice****"* was annotated as *social order* because of the word '***safeguards***' and *freedom* from the words *'**freedom**'* and *'**choice**'*. Challenges faced in annotating passages include – annotated passages could be of any length, and annotated passages often overlapped indicating evidence of multiple values. To deal with this, they supported multiple value annotations per sentence where the unit of annotation was a single sentence.

A2.8 Heuristic Values Model

The work of Bengston et al (2004) is considered in automatically identifying forest values orientations in written text. Like in previously discussed approaches, the underlying principle is the assumption that when humans communicate, their words are structured to portray their values.

A set of value categories were first identified by manually reviewing literature on forest planning, management and policy. From this emerged an inventory consisting of three categories:

- Bio-centric values are values that assert the importance of non-human life.
- Anthropocentric values assert that humans are central and more important than other species.
- Moral/spiritual values promote non-instrumental values of forests such as moral values, spiritual and sacred values.

This inventory acted as an a priori for subsequent analysis. First, a corpus of forestry and environmental content was codified by domain experts to identify instances of the value types. As individual sentences were the unit of analysis, sentences with similar values were grouped into representative inventory types. From each group, a list of ideas, words and phrases relevant to some of the concepts mentioned were collated into a dictionary. Syntactic patterns specifying how pairs of ideas in the dictionary combine to give meanings were identified to form *Idea Transition Rules (ITR).* New sentences to be classified are subsequently checked against the syntactic patterns in the ITR for any clues reflecting its most likely value. ITRs represent manually composed rules that specify the presence or absence of certain words or patterns representative of distinct values.

To illustrate, one of the dimensions of anthropocentric values is *concern over loss of commodity related jobs* (Bengston et al, 2004). This is established by humans at the start of the annotation. *'Job loss'* thus becomes a concept of importance for people with anthropocentric values and entered in the dictionary. Syntactic patterns in the training set spanning the phrase *'job loss'* are used to form the ITR. For instance, in the sample sentence, "Last week the forest service and the Bureau of land management said that if the Thomas recommendation is adopted, timber harvests on North Western Federal Lands will be cut nearly in half over the next five years causing a net loss of about 13000 jobs in Oregon, Washington and Northern California (Sonner, 1990)". This example was marked up as anthropocentric value orientation, because in the dictionary, the word '*loss'* was one of many words phrases that connote the idea of decrease or cut backs which is a concept used by people with anthropocentric values.

Developing the dictionary and idea transition rules is a tedious iterative process. It is also not particularly exhaustive because words can have several meanings and so there is a constant need to develop more rules to satisfy new word semantics and relevance. In addition, it is also domain specific. The idea transition rules developed for one domain will not be applicable to another. Developing idea transition rules for every domain is tedious impractical plan. Finally, the approach is dependent on humans for the generation of the a priori inventory and the ITRs. In spite of these issues, this approach showed the highest values classification accuracy, correctly predicting 78.4% of anthropocentric values, 86.8% of bio-centric values and 93% of moral and aesthetic values.

APPENDIX 3: Trigram Language Models

A LM consists of the following,

- A finite set V of words
- A set of parameters $\gamma(r|p,q)$ for each trigram p,q,r such that $r \ \in V \ \cup \{STOP\}$, and $p,q \ \in V \ \cup \{START\}$

Given these definitions, a LM defines a distribution such that for any sentence, $x_1 \dots x_n$, where $x_i \ \in V$ for $i = 1 \dots (n-1)$, and x_n = STOP, the probability of the sentence under the trigram language model is

$$p(x_1 \dots x_n) = \prod_{i=1}^{n} \gamma(x_i|x_{i-2}, x_{i-1}),$$

$$\text{where } x_0 = x_{-1} = START.$$

The modelling task then requires estimating the parameters $\gamma(r|p,q)$. A natural estimate of γ, for the trigram x_{i-2}, x_{1-1}, x_i can be estimated by the Maximum Likelihood Estimate (MLE) expressed as the ratio of the bigram to the trigram

$$\gamma(x_i|x_{i-2}, x_{i-1}) = \frac{c(x_{i-2}, x_{i-1}, x_i)}{c(x_{i-2}, x_{i-1})}$$

Given a training sample of such sentences, the goal is to learn a distribution p that satisfies the following conditions. For all possible languages or sentences in V' sum to 1

$$\sum_{x \in V'} p(x) = 1$$

$$and$$

$$p(x) \geq 0, for \ all \ x \ \in \ V'$$

To this end p is a well-formed distribution of all sentences in the language from which probability estimates can be assigned to sentences.

APPENDIX 4: Absolute Discounting

The idea behind absolute discounting is to subtract a fixed discount d from seen ngrams and assign it to unseen ngram counts (Ney & Essen, 1991; Ney et al, 1994). The aim of subtracting small discounts from seen events with high counts is to adjust the unreliable probability estimate of low count unseen events. It is also expected that the discounted count from high count ngrams would not have a significant effect on their probability estimate. The interpolated absolute discount smoothing is expressed as

$$P_{absolute}\left(w_i \middle| w_{i-n+1}^{i-1}\right) = \frac{\max\{c\left(w_{i-n+1}^{i-1}\right) - d, 0\}}{\sum_{w_i} c(w_{i-n+1}^{i})} + (1 - \lambda(w_{i-n+1}^{i-1}))P_{absolute}(w_i|w_{i-n+2}^{i-1})$$

where, λ= interpolation weight and d is the small fixed discount. This equation interpolates a discounted higher order ngram $\frac{\max\{c\left(w_{i-n+1}^{i-1}\right)-d,0\}}{\sum_{w_i} c(w_{i-n+1}^{i})}$, with an interpolated weight $\lambda(w_{i-n+1}^{i-1})$ and a lower order ngram probability $P_{absolute}(w_i|w_{i-n+2}^{i-1})$ also called the back-off probability. In combining lower order ngrams with the discounted ngram, the resulting model combines the benefits of higher order ngrams which have more context and lower order ngrams which are unlikely to have zero counts. However, a downside to absolute discounting is the back-off probability which can be quite unreliable because of its high bias towards words with high frequency ngrams. Kneser-Neys smoothing addresses this issue.

APPENDIX 5: Interpolated Kneser-Ney Smoothing

Kneser-Neys smoothing has its origins in absolute discounting (see appendix 7) and it aims at combining information from lower order ngrams towards improving the estimate of higher order ngrams. The primary innovation in this algorithm is the enhancement of the lower order ngram back-off probability $P_{absolute}(w_i|w_{i-n+2}^{i-1})$[96] (see equation 8). The back-off probability is applied in the ngram model when the higher order ngram is unseen or has very few counts. However, this back-off probability can be quite unreliable because of its high bias towards high frequency words.

$$P_{absolute}\left(w_i\middle|w_{i-n+1}^{i-1}\right) = \frac{max\{c\left(w_{i-n+1}^{i-1}\right) - d, 0\}}{\sum_{w_i} c(w_{i-n+1}^{i})} + (1 - \lambda(w_{i-n+1}^{i-1}))P_{absolute}(w_i|w_{i-n+2}^{i-1})$$

To describe bias problem, we use the popular illustration expressed by Jurafsky and Martin (2009). Assuming we have the task of completing the sentence:

'I can't see without my reading _____'

'*Glasses*' appears to be the most likely suggestion instead of the word '*Francisco*'. However, what if '*Francisco*' appears more times than '*Glasses*' in the corpus. Because of this, assuming backing-off to a unigram model, the model would choose '*Francisco*' instead of '*Glasses*'. To address this, "we would like to capture the intuition that although Francisco is frequent, it is only frequent after the word 'San', that is the phrase 'San Francisco'" (Jurafsky and Martin, 2009). Kneser-Neys suggests a unique back-off estimate that is based on the 'number of different contexts that the word w_i has appeared in' (Jurafsky and Martin, 2009). The Kneser-Neys back-off probability is simply a count of the number of unique ngrams a word w_i completes divided by the total number of ngram counts. For instance, in the bigram case, the back-off probability is given as

$$P_{continuation}(w_i) = \frac{|\{w_{i-1} : c(w_{i-1}, w_i) > 0\}|}{\sum_{w_i} |\{w_{i-1}: c(w_{i-1}, w_i) > 0\}|}$$

Where, the numerator is the set of continuation words preceding w_i and the denominator is the total number of word bigram counts. $P_{continuation}(w_i)$ is substituted to the absolute discount smoothing equation so that the Kneser-Neys bigram formulation becomes

$$P_{KN}(w_i|w_{i-1}) = \frac{\max(c(w_{i-1}, w_i) - d, 0)}{c(w_{i-1})} + \lambda(w_i)\, P_{continuation}(w_i)$$

Where, $\lambda(w_i)$ is a weighted constant expressed as:

$$\lambda(w) = \frac{d}{c(w_{i-1})} |\{w : Count(w_{i-1}, w) > 0\}|$$

The general recursive formulation for Kneser-Neys algorithm thus becomes,

[96] This is called $P_{continuation}$ in Kneser-Neys smoothing

$$P_{KN}\left(w_i \middle| w_{i-n+1}^{i-1}\right) = \frac{\max(c\left(w_{i-n+1}^{i}\right) - d, 0)}{\sum_{w_i} c(w_{i-n+1}^{i-1})} + \frac{d}{\sum_{w_i} c(w_{i-n+1}^{i-1})} + \lambda\left(w_{i-n+1}^{i-1}\ .\right) P_{KN}(w_i | w_{i-n+2}^{i-1})$$

This research's LM implementation was carried out using SRILM toolkit (Stolcke, 2002; Stolcke et al, 2011) primarily because it supports a wide variety of LM implementations including interpolated KN smoothing.

APPENDIX 6: Perplexity and LM Evaluation

LM Evaluation can be intrinsic or extrinsic (In vivo). In extrinsic evaluation, the performance of the model is determined on real life data or scenario. It is therefore, more expensive but realistic. Intrinsic evaluation on the other hand is cheaper but less realistic and perplexity is a form of intrinsic evaluation. Perplexity (PP) is the most common LM evaluation metric and it is the probability that the implemented model assigns to a test data. Perplexity is an information theoretic approach for measuring the predictive power of LMs on test data (Stanley et al, n.d).

In the development of LMs, the data set is typically divided into training and test sets. The LM is built using the training set and the perplexity is estimated on the test set. The reasoning behind this is that given two probabilistic models, the better model is the one that has a tighter fit to the test data or predicts the details of the test data better (Jurafsky and Martin, 2009).

Given a test set T that is made up of test sentences $\{t_1 \dots t_n\}$, the probability of the test set $p(T)$ is computed as the product of all the sentences in the test set, normalized by the number of words. This is expressed as,

$$PP(T) = p(w_{1t} \dots w_{nt})^{-\frac{1}{N}}$$

Using the chain rule to expand the probability of T, the perplexity for the bigram and trigram case are expressed below:

$$PP(T)_{bigram} = \sqrt[N]{\prod_{i=1}^{N} \frac{1}{p(w_i|w_{i-1})}}$$

$$PP(T)_{trigram} = \sqrt[N]{\prod_{i=1}^{N} \frac{1}{p(w_i|w_{i-1}, w_{i-2})}}$$

APPENDIX 7: List of Sample Party Documents

Liberal Democrat Party	
Liberal Democrat Manifesto 2015, published and Promoted by Tim Gordon on behalf of the Liberal Democrats	17 pages
Liberal Democrats Manifesto 2015, Stronger Economy, Fairer Society, Opportunity for everyone.	158 pages
Liberal Democrat Manifesto for the 2014 European Parliament Elections	47 pages
Liberal Democrat Conference Report, Glasgow 4-8 October, 2014.	59 pages
Liberal Democrat, Federal Conference Report, 14-18 September, 2013	56 pages
Liberal Democrats Policy Consultation, Immigration, Asylum and Identity, Consultation paper 115, August 2013	24 pages
Labour Party	
Labour Party European Manifesto	27 pages
The Labour Party Manifesto 2015	87 pages
2014 Labour Conference Speech[97] (Ed Miliband)	28 pages
National Policy Forum Report 2014[98]	218 pages
Conservative Party	
Conservative Party European Election Manifesto 2014	74 pages
The Conservative Party Manifesto 2015	84 pages
A Balanced Centre-Right agenda on Immigration – A Manifesto for Immigration[99] (Shorthouse and Kirby, 2015)	31 pages
UKIP	
The UKIP Manifesto 2015	76 pages
UKIP Europe Manifesto 2014	8 pages

[97] http://www.labour.org.uk/blog/entry/2014-labour-conference-speech - Last accessed 29-02-2015
[98] http://www.policyforum.labour.org.uk/uploads/editor/files/NPF_Annual_Report_2014.pdf - Last accessed 29-02-2015
[99] https://www.barrowcadbury.org.uk/wp-content/uploads/2015/04/A-manifesto-for-immigration.pdf - Last accessed 11-04-2016

APPENDIX 8: Latent Dirichlet Allocation (LDA)

The goal of LDA is to infer thematically relevant topics for an unobserved document from words in observed documents and it accomplishes this through the fundamental idea that documents consists of a blend of topics. Each topic is defined to be a distribution over a fixed vocabulary of words. For instance, the distribution of words over a sample topic *'Football'* would span a range of high probability *'Football'* words thematically associated to football (call these football words) and irrelevant low probability words which have little or no association with football.

To elucidate, consider the following documents:
Document 1 – *'The English football team lost to Germany'*
Document 2 – *'The Germans have decided to leave the EU'*
Document 3 – *'We lead them until the last minute before they scored the equalizer'*
Document 4 – *'Who will lead us out of the EU?'*
Document 5 – *'We've left the EU but we can still beat the Germans at football'*
Applying LDA, commences with an assumption about the number of topics. Assume that for these documents there are 2 topics – *'Football'* and *'Europe'*. The classifier might output a classification result as seen in table 44 which suggests that 85% of the words used in document 1 relate to *'Football'* while 15% relate to *'Europe'*. So, a reasonable inference would be that document 1 is about *'Football'*. Therefore, based on the distribution of words in each document, reasonable thematic assignations can be made for each document by setting a threshold percentage limit.

Table 44: Sample Thematic Classification of 5 Documents (Adapted from Blei et al, 2003)

	Proportion of 'Football'	Proportion of 'Europe'	Assigned topic
Document 1	85%	15%	Football
Document 2	0%	100%	Europe
Document 3	100%	0%	Football
Document 4	0%	100%	Europe
Document 5	65%	35%	Football

LDA Algorithm

The LDA algorithm is described using the following pseudocode.

1. Assume topics $\beta_{1:k}, where\ k\ is\ the\ number\ of\ topics.$
2. Set k for the set of documents D.
3. For each document d_j in D

a. Randomly assign each word $w_{i,j}$ to a topic β_k - Output – A poor distribution of words for each topic and a poor distribution of topics for each document. That is each word receives a temporary topic assignment which will be updated in the next step

4. $For\ each\ document\ d_j\ in\ D$

   a. (Update each word's topic) For each word $w_{i,j}$ in d_j

      i. For each topic β_k, (Two parameters are estimated)

         1. How prevalent is the word across all topics, that is $p(w_i|\beta_k)$ where $p(w_i|\beta_k) = \frac{Total\ Count\ of\ w_{i,k}\ in\ D}{Total\ number\ of\ words\ assigned\ to\ \beta_k}$, where $w_{i,k}$ is the number of w_i assigned to topic β_k.

         2. How prevalent is the topic β_k across d_j or $p(\beta_k\ |d_j)$, or total number of assignations of β_k or total number of words in d_j assigned topic β_k divided by total number of words in d_j.

            $p(\beta_k\ |d_j) = \frac{\beta_{k,j}}{Count\ of\ words\ in\ d_j}$, where $\beta_{k,j}$ is the total count of topic β_k in d_j.

         3. Compute $p(w_i|\beta_k) *\ p(\beta_k\ |d_j)$

         4. Repeat the iterative process of topic assignment across all documents until convergence.

APPENDIX 9: Sample Perplexity Scores

Model Perplexity for Conservative-Immigration

Model	Perplexity
Good-Turing	135.83
Linear	161.01
Witten-Bell	144.13
Absolute	148.10
Kneser-Neys (Back-off)	136.11
Kneser-Neys - Interpolated	124.05

Model Perplexity for Labour Immigration

Model	Perplexity
Linear	292.07
Good-Turing	255.94
Witten-Bell	231.61
Absolute d=0.1	236.4344
Absolute d=0.2	234.4507
Absolute d = 0.3	233.1169
Absolute d = 0.4	232.1276
Absolute d = 0.5	231.38
Absolute d = 0.6	220.84
Absolute d = 0.7	220.48
Absolute d = 0.8	220.29
Absolute d = 0.9	220.27
Absolute d = 1.0	220.43
Absolute d = 1.1	220.43
Kneser-Neys (Back-off)	220.26
Kneser-Neys – Interpolated $\lambda 1 = 0.5$, $\lambda 2 = 0.5$	173.2671
Kneser-Neys – Interpolated $\lambda 1 = 0.0007$, $\lambda 2 = 0.6$, $\lambda 3 = 0.3993$	169.2671
Kneser-Neys – Interpolated $\lambda 1 = 0.0004$, $\lambda 2 = 0.7$, $\lambda 3 = 0.2996$	169.5733
Kneser-Neys – Interpolated $\lambda 1 = 0.00032$, $\lambda 2 = 0.77$, $\lambda 3 = 0.22968$	169.54
Kneser-Neys – Interpolated $\lambda 1 = 0.00032$, $\lambda 2 = 0.57$, $\lambda 3 = 0.42968$	160.1149
Kneser-Neys – Interpolated $\lambda 1 = 0.00022$, $\lambda 2 = 0.57$, $\lambda 3 = 0.42978$	160.1148
Kneser-Neys – Interpolated $\lambda 1 = 0.00002$, $\lambda 2 = 0.57$, $\lambda 3 = 0.42998$	**160.1133**

Model Perplexity for Labour EU

Model	Perplexity
Linear	298.12
Good-Turing	261.07
Witten-Bell	244.29
Absolute d=0.1	248.109
Absolute d=0.2	247.13
Absolute d = 0.3	245.98
Absolute d = 0.4	245.67
Absolute d = 0.5	244.22
Absolute d = 0.6	240.396
Absolute d = 0.7	238.118
Absolute d = 0.8	236.44
Absolute d = 0.9	236.35
Absolute d = 1.0	236.81
Absolute d = 1.1	236.813
Kneser-Neys (Back-off)	230.18
Kneser-Neys – Interpolated $\lambda 1 = 0.5$, $\lambda 2 = 0.5$	190.77
Kneser-Neys – Interpolated $\lambda 1 = 0.0007$, $\lambda 2 = 0.6$, $\lambda 3 = 0.3993$	181.39
Kneser-Neys – Interpolated $\lambda 1 = 0.0004$, $\lambda 2 = 0.7$, $\lambda 3 = 0.2996$	180.03
Kneser-Neys – Interpolated $\lambda 1 = 0.00032$, $\lambda 2 = 0.77$, $\lambda 3 = 0.22968$	180.03
Kneser-Neys – Interpolated $\lambda 1 = 0.00032$, $\lambda 2 = 0.57$, $\lambda 3 = 0.42968$	177.611
Kneser-Neys – Interpolated $\lambda 1 = 0.00022$, $\lambda 2 = 0.57$, $\lambda 3 = 0.42978$	177.602
Kneser-Neys – Interpolated $\lambda 1 = 0.00002$, $\lambda 2 = 0.57$, $\lambda 3 = 0.42998$	**177.6**

APPENDIX 10: Content Word Clustering for Dimension Reduction

The objective of the clustering approach is to reduce the horizontal dimension of a large matrix of feature vectors, so that it is computationally easier to construct multiple maxent models for each of the constituent matrices. For instance, the matrix $\begin{bmatrix} x_{1,1} & \cdots & x_{1,241} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \cdots & x_{n,241} \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$, where n is the number of the content word instances is split into a cluster or set of matrices:

$$\left\{ \begin{bmatrix} x_{1,1} & \cdots & x_{1,241} \\ \vdots & \ddots & \vdots \\ x_{k^1,1} & \cdots & x_{k^1,241} \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_{k^1} \end{bmatrix}, \begin{bmatrix} x_{1,1} & \cdots & x_{1,241} \\ \vdots & \ddots & \vdots \\ x_{l^2,1} & \cdots & x_{l^2,241} \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_{l^2} \end{bmatrix}, \dots, \begin{bmatrix} x_{1,1} & \cdots & x_{1,241} \\ \vdots & \ddots & \vdots \\ x_{z^p,1} & \cdots & x_{z^p,241} \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_{z^p} \end{bmatrix} \right\}$$

Where, in k^1, k represents the number of content words so that k^1 represents the number of content words in the first cluster. Similarly, l^2 represents the number of content words l in the second cluster. Different alphabets are used for the content word size because the clusters have different sizes.

Since the goal is to provide unique groups for content words in the vocabulary set, this approach involves observing quantitative properties of the content words, identifying similarities or patterns, subsequently use the observed patterns or similarities in grouping content words. Thus, for each content word vocabulary, words are arranged in ascending order of frequency. Figure 37 show a distribution of EU-Conservative and Immigration-Conservative content words grouped by frequency of occurrence. In the distribution, three distinct groups were observed as portrayed in figure 38. Group A in figure 38 consists of content words with very low frequencies of occurrence but high collective frequency. For instance, there are 7623 unique words that occur 4 times in the Conservative-EU training corpus. Group C consists of very high frequency words, while group B represent words that fall between very high and low frequency. This is consistent with Zipf's law. It is also observed that a significant portion of content words fall into this category (The distribution for Labour and Liberal Democrats is shown in figure 39 and 41. Clearly, the distributions take the same form).

To distinguish the groupings identified in figure 38, additional properties of the content words are explored by computing information content (IC). IC is a measure of specificity that quantifies the amount of information required to encode a piece of text (Shannon, 1948; Resnik, 1995). The information content of an event c is computed from its probability $p(c)$ using the formula $IC = -\sum_{i=1}^{n} p(c) \log_2 p(c)$. Based on this, words with the same frequency have the same IC and if they have the same IC it is assumed that they belong to the same group since they contain the same amount of information required to predict or generate them.

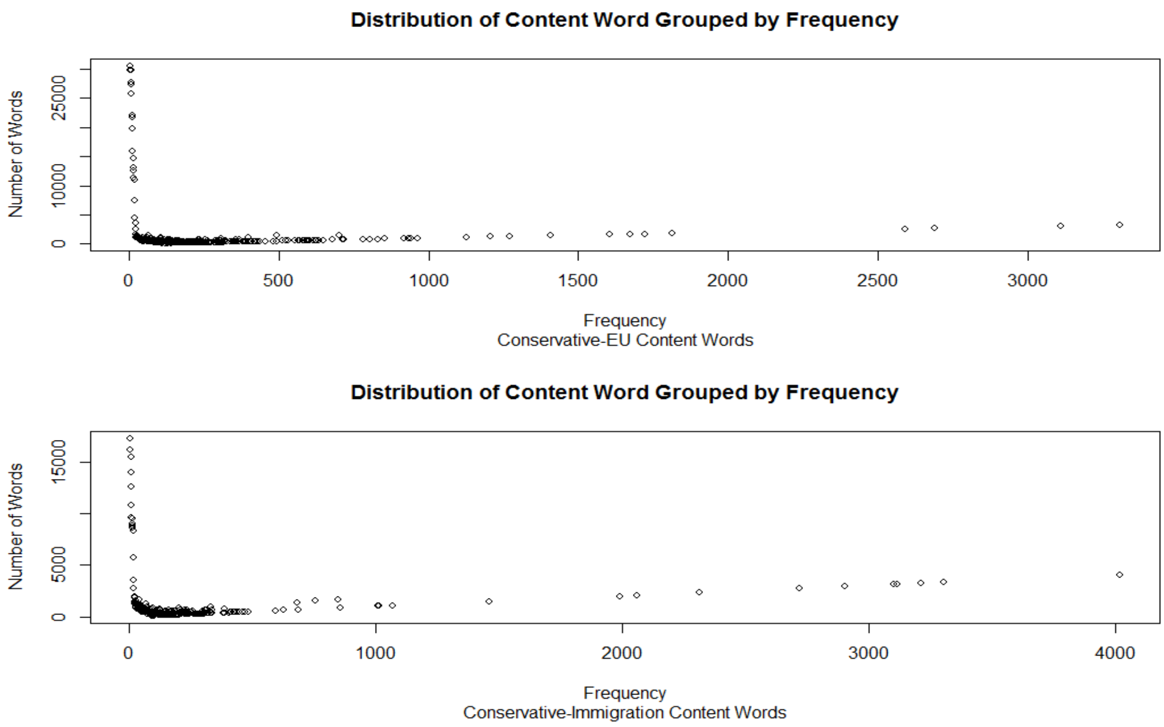


Figure 37: Distribution of content words grouped by Frequency (Conservatives)

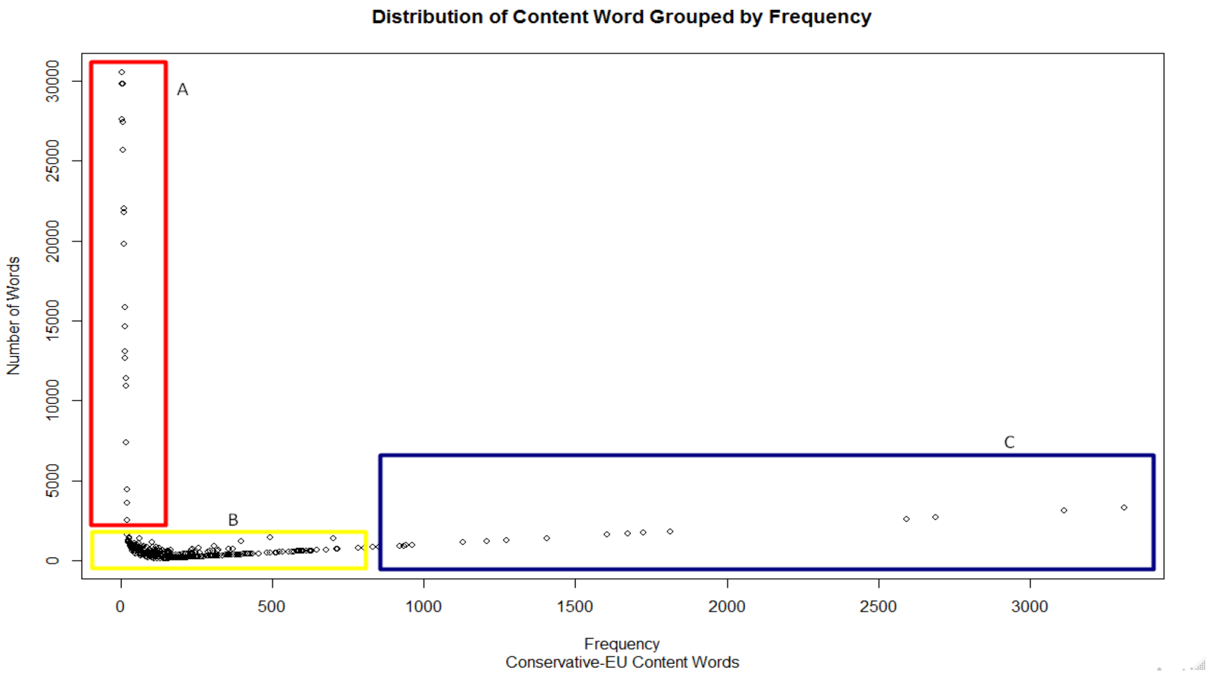


Figure 38: Illustrating groupings of content words distributions

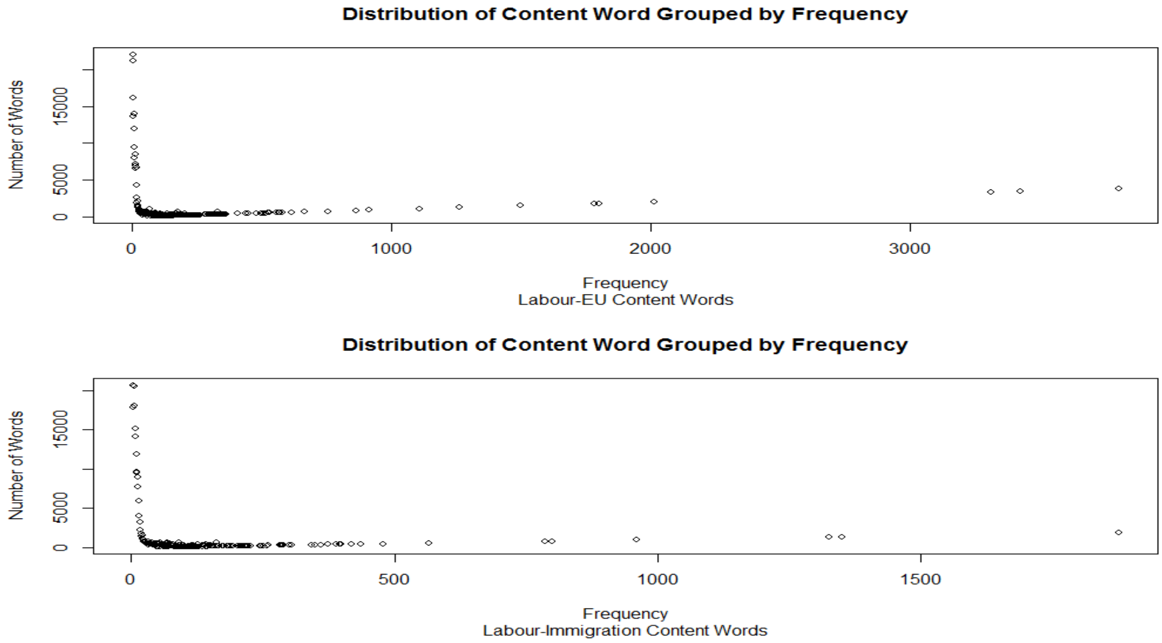


Figure 39: Distribution of content words grouped by Frequency (Labour)

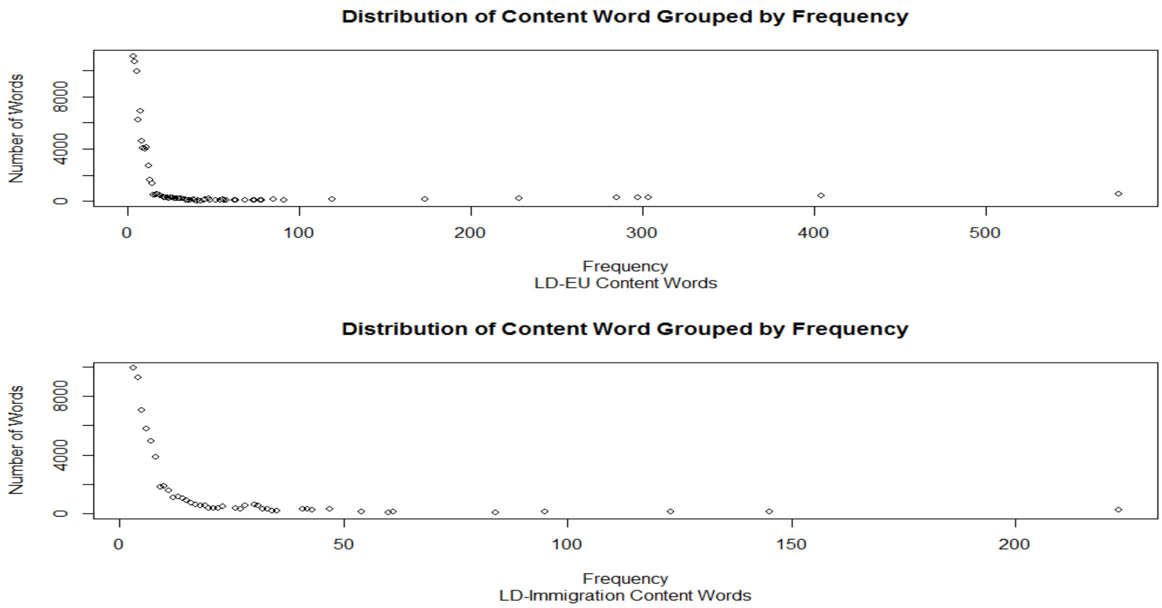


Figure 40: Distribution of content words grouped by Frequency (LD)

To explore this, the count of count of words (N_x) with the same information content and the IC of all content words with frequency f (where N_x is the count of count[100] of events seen x times) are computed. As an illustration, in the conservative EU data, it is observed that the word that occurs 3310 times in training has a count of count of 1 (that is it is the only word in the corpus to occur 3310 times. In training it is observed that words with very high frequencies tend to have a count of count of 1), while words which occur 45 times have

[100] The expressions count of count and frequency of frequency are used interchangeably

a count of count of 14. Events seen N_x times should have the same IC, so the IC is computed for each count of count as:

$$IC_x = -\sum_{i=1}^{n} p(c) \, logp(N_x), \text{ where } p(N_x) \text{ is } \frac{total\ number\ of\ events\ that\ occur\ x\ times}{total\ number\ of\ content\ words}$$

In this illustration, words with N_{45} and N_{3310} have IC values of 9.632521 and 7.239123 respectively[101]. With IC computed for all frequencies, a distribution of count of count against IC was plotted. This is illustrated in figures 41, 42 and 43 for each party-context pair. The distributions shown in figures 41, 42 and 43 reveal distinct data groupings (see figure 44). Each group represent a distinct distribution of words that lie within a range of IC values. The position of points in the graphs means that some groupings like the first three in figure 44 are fairly obvious while others are not. To this, each data point is selected as a unique group only if the number of content words is at least twice the number of features. If a single cluster point does not satisfy this criterion, it is merged with the point closest to it. This process of grouping is a manual exercise and it is not a problem because the number of points or the dimensionality of the problem has been reduced to a set of points that can be easily assigned to groups. With the manual selection of groupings completed, a maxent model for each group can be computed because of the reduced size of the vector Y.

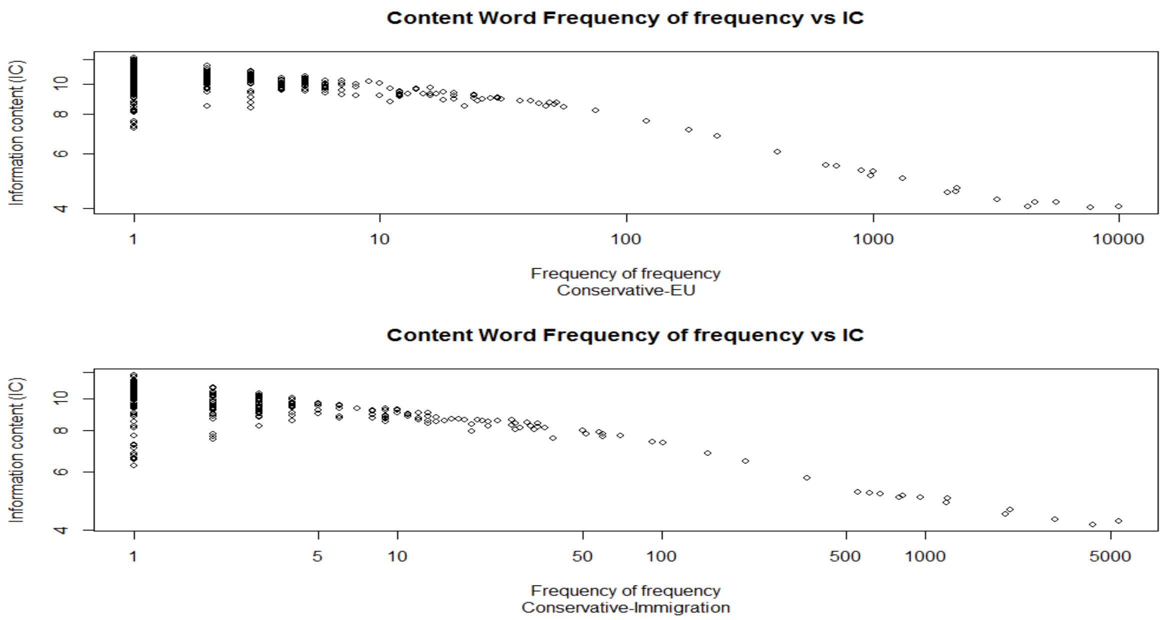


Figure 41: Log plot showing distribution of content words IC against Count of count (Conservatives)

[101] Note that we do not include words with very low frequencies. Words with counts of 1 and 2 are excluded. Thus, the total number of content words used in this computation (Conservative-EU) is 500059

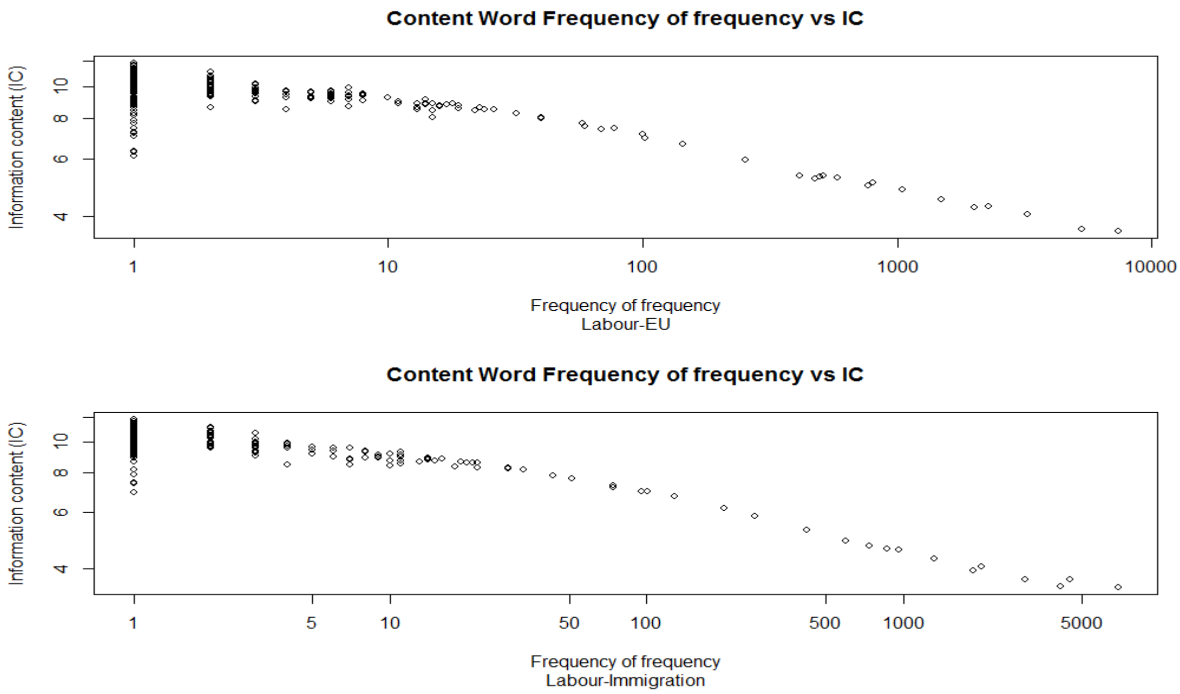


Figure 42: Log plot showing distribution of content words IC against Count of count (Labour)

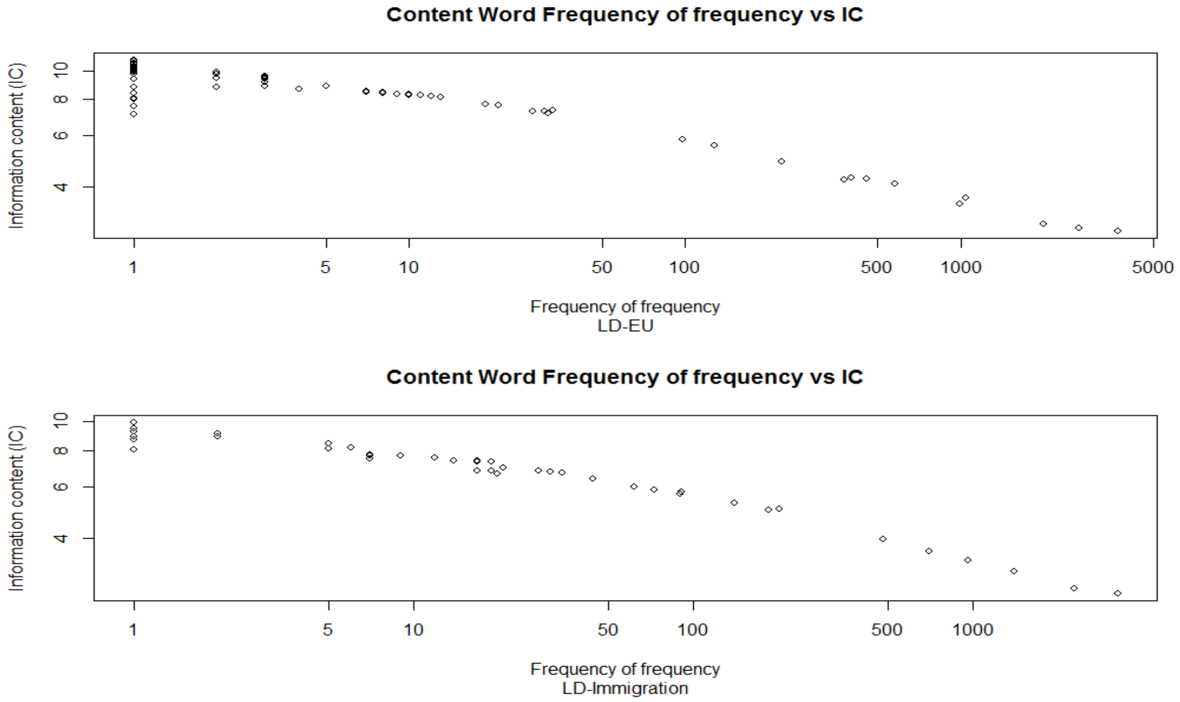


Figure 43: Log plot showing distribution of content words IC against Count of count (Liberal Democrats)

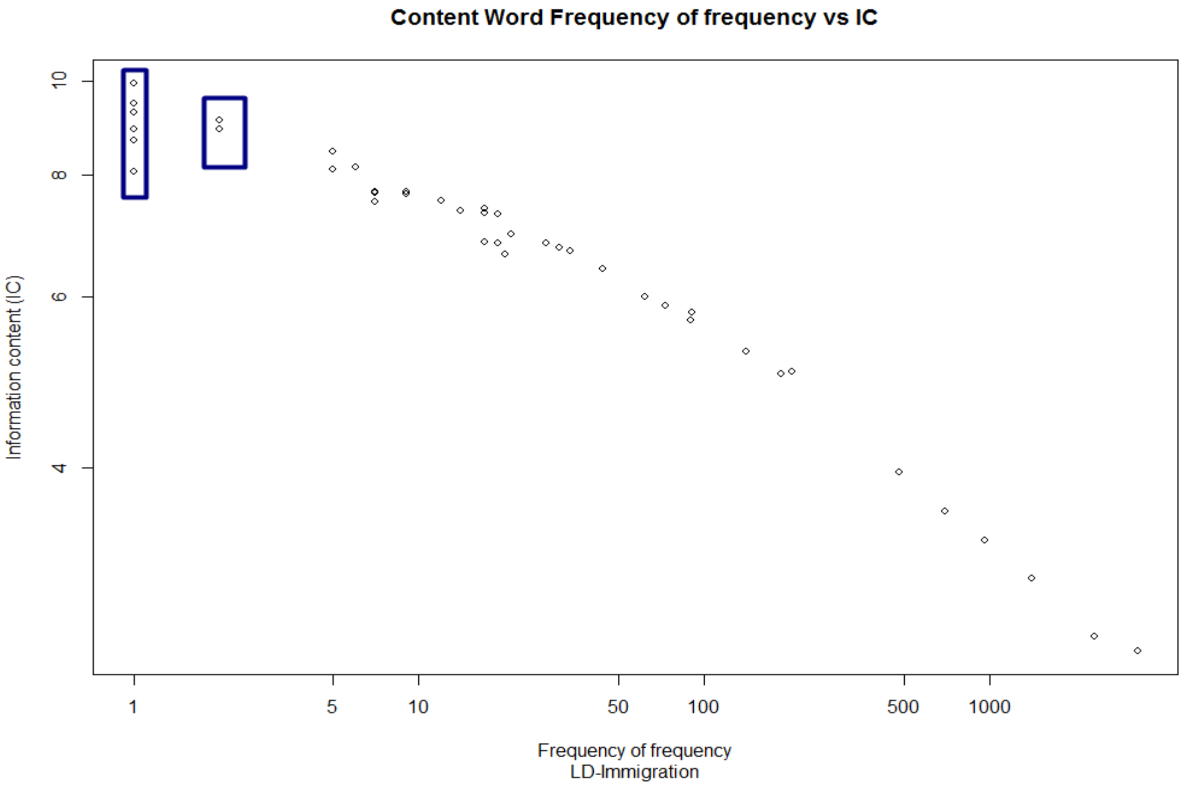


Figure 44: Log plot showing manual selection of word groupings(Labour)

APPENDIX 11: Modified VSEA Algorithm with Word Clusters

Algorithm 5 Modified Value Sentence Estimation Algorithm (VSEA) with Word Clusters

Inputs

(i) $V = \{v_1, v_2, v_3, \ldots, v_N\}$ where V is the vocabulary of content words and N is the size of the vocabulary.

(ii) The Interpolated KN Smoothed LM model LM_1

(iii) Maxent model clusters $M = \{m_1, m_2, \ldots, m_z\}$ where M is the set of maxent models and z is the number of clusters, so that m_z is the maxent model for cluster z. There is a 1:1 mapping from $C_i \rightarrow M_i$.

(iv) Sets $C_z = \{v_{z1}, v_{z2}, \ldots, v_{zn}\}$ where C is a set of content words in a cluster z and n is the number of words in the cluster.

Result: Product (P)
Initialize: $P = 1$
Process

(a) **for** each word, w_i, in $W = \langle w_1, w_2, \ldots, w_n \rangle$, where i is an integer **do**

i. **Identify** the word type of $w_i (i = 1 \ldots n)$, i.e. is it a content word or a function word

ii. **If** w_i is a function word estimate $p_f = p(w_i | w_{i-2}, w_{i-1})$ from LM_1, **Compute** $P := P \times p_f$

iii. **Else if** w_i is a content word

(i) **If** $w_i \in V$ **then iterate** through each set C to identify w_i's cluster C_i.
(ii) **Identify** M_i from C_i
(iii) **Build** feature vector X_i for w_i using features $f_1 \ldots f_{241}$
(iv) **Feed** vector X_i into M_i to estimate the probability p_{w_i} of w_i i.e $p(w_i | f_1 \ldots f_{241})$
(v) **Compute** $P := P \times p_{w_i}$

iv. **Else if** w_i is an unknown content word **do**

A. **Estimate** the probability p_{UNK} from the LM_1
B. **Compute** $P := P \times p_{UNK}$

(b) **Return** product P

APPENDIX 12: Sentiment Prediction Algorithm with Modified VSEA

Algorithm 6 Recipient Sentiment Prediction using Modified VSEA and Feature Switching

Inputs

(i) Sentence W

(ii) $V = \{v_1, v_2, v_3, \ldots, v_N\}$ where V is the vocabulary of content words and N is the size of the vocabulary.

(iii) Interpolated KN LM – LM_1

(iv) Maxent model clusters $M = \{m_1, m_2, \ldots, m_z\}$ where M is the set of maxent models and z is the number of clusters, so that m_z is the maxent model for cluster z. There is a 1:1 mapping from $C_i \rightarrow M_i$.

(v) Sets $C_z = \{v_{z1}, v_{z2}, \ldots, v_{zn}\}$ where C is a set of content words in a cluster z and n is the number of words in the cluster.

Process

(a) **Apply** VSEA on W and estimate a probability P.

(b) **Apply** VSEA on W' such that for all content words in W we assign opposite features and estimate P'.

(c) **Determine** $max(P, P')$

Output

If $max(P, P') = p$, **infer** sentiment Ψ of W (Ψ_W) is positive

Else, **infer** sentiment Ψ of W (Ψ_W) is negative

APPENDIX 13: Description of Value Holders/Judges

	1	2	3	4	5	6	7
Age	31	35	36	37	44	47	55
Qualification	MEng	Msc	MSc	MSc	MBA	BA	PhD
Nationality	British	British	British	British	British	British	British
Region	England	England	England	England	NI[102]	England	England
Party	LD	Lab	Con	LD	Con	Con	Lab
Married/Civil	Y[103]	Y	Y	Y	Y	Y	Y
Spouse Nationality	British	NEU[104]	British	E-NEU[105]	British	British	British
Race	White	Black	White	White	White	White	White
Occupation	Project Manager	Project Manager	Software Engineer	Software Engineer	Self Employed	Accountant	Teacher
Signed up to Party news letter	Y	Y	Y	Y	Y	Y	Y
Read or watch the news daily	Y	Y	Y	Y	Y	Y	Y
Follow, read, watch or listen to debates, discussions on the EU	Twice a week	Three times a week	Three times a week	Twice a week	Four times a week	Every day	Every day
Rank Regular Source of information on policies and	B - 5 NP - 4 SM -8	B - 7 NP - 5 SM -8	B - 6 NP - 6 SM -6	B - 6 NP - 7 SM -7	B - 6 NP - 8 SM -6	B -6 NP - 9 SM -7	B -10 NP - 10 SM -0

102 NI = Northern Ireland
103 Y = Yes
104 NEU = NonEU
105 E-NEU = European NonEU

facts[106] (0 (lowest)-10(highest)	PSM - 7 TV- 6 PC – 6 PN - 3 PP- 4	PSM - 6 TV- 7 PC – 6 PN - 3 PP- 6	PSM - 5 TV-8 PC – 4 PN - 5 PP- 7	PSM - 4 TV-8 PC – 4 PN - 5 PP- 7	PSM - 10 TV-9 PC –7 PN - 8 PP- 8	PSM - 7 TV-8 PC –6 PN - 6 PP- 8	PSM - 0 TV-10 PC –7 PN - 2 PP- 10
Regular Source of information on policies and facts during election	B - 6 NP - 8 SM -10 PSM -8 TV- 6 PC – 6 PN - 3 PP- 6	B - 8 NP - 6 SM -8 PSM - 7 TV- 10 PC – 3 PN - 3 PP- 10	B - 8 NP - 6 SM -7 PSM - 7 TV-10 PC – 4 PN - 5 PP- 10	B - 8 NP - 8 SM -8 PSM - 5 TV- 9 PC – 4 PN - 5 PP- 9	B - 8 NP - 8 SM -10 PSM - 10 TV-10 PC –8 PN - 10 PP- 8	B -7 NP - 9 SM -7 PSM - 8 TV-9 PC –6 PN - 9 PP- 9	B -10 NP - 10 SM -0 PSM - 0 TV-10 PC –8 PN - 4 PP- 10
Registered to a party	N	N	N	N	Y	Y	N
Ever Ran for public office	N	N	N	N	Y	N	N
Voted last two general elections	Y	Y	Y	Y	Y	Y	Y
2010 vote	LD	Lab	Con	LD	Con	Con	Lab
2015 vote	LD	Lab	Con	LD	Con	Con	Lab
Referendum vote	R[107]	R	R	R	L	L	L

[106] NP = Newspaper and Print media, SM = Political Social media, PSM = Party social media, TV = TV and radio News, PN = Party Newsletter including leaflets , PC = Word of mouth, personal communication, rumour, PP = Political Programmes, B = Blogs and opinion piece
[107] R = Remain, L = Leave

APPENDIX 14: Sample UKIP Immigration and EU Test Sentences

1. We can never control immigration while we continue to be members of the European Union.
2. Outside the EU we can manage our borders and decide who we want to come and live and work in the UK.
3. We will continue to honour our obligations to bona fide asylum seekers.
4. We can never control immigration while we continue to be members of the European Union.
5. Cutting and controlling immigration.
6. Allow British businesses to choose to employ British citizens first.
7. Restrict access to EURES the EU wide jobs portal that has become the go-to source for employers looking for cheap labour from overseas.
8. UKIP will do its utmost to deport foreign criminals and prevent those with criminal records from entering Britain when we leave the EU.
9. By leaving the EU and restricting immigration through the use of an Australian style points based system we will give back some hope to British workers for a brighter future.
10. UKIP will implement new border control technology solutions to ensure all passport and visa holders are counted in and out and to identify over-stayers including those on student visas.
11. Until we leave we are forced to abide by the EU's founding unshakable principle of the free movement of people meaning we cannot prevent the flow of citizens from all EU member states into Britain.
12. UKIP will increase the numbers of border agency staff by 20%.
13. We applaud the home secretary's responsible measures in the bill to control migration and I am sure that they will be widely supported throughout the country.
14. We do not have the ability to vote down any deals between the European Union and Turkey because of the deal we have struck.
15. We should not approve the EU deal with Turkey.
16. The deal with Turkey will give 5 million Turks visa-free unrestricted access to the Schengen areas from 2018.

17. The UK may not be part of Schengen but the deal with Turkey does affect us.

18. There will be no mechanism to log people coming into the Schengen area and none to log people out.

19. UKIP policies recognise the new openness in our world and the positive benefits controlled immigration has brought and can continue to bring to our nation.

20. The deal with Turkey can only add to the porousness of the EU frontiers which can only contribute to the increase in numbers of those camped outside Calais seeking entry into the UK.

21. The talks between the EU and Turkey mean that Turkish accession to the EU is back on the table.

22. We would not wish joining the EU on anyone certainly not a friend such as Turkey.

23. Within a short time those migrants assigned to Portugal will have every right to come and live in Peckham and those assigned to live in Italy will have every right to move to Ipswich.

24. We say no to ever closer union with the EU.

25. A few days ago, the EU announced what is in effect a four-part deal with Turkey.

26. We have signed away the right to reject a duff deal with Turkey made in our name the consequences of which will be with us for yours to come.

27. The deal with Turkey is not in our national interest.

28. We are sometimes made to deal with Turkey as an equal yet it does not have the belief in equality within turkey that we in the West of Europe and North America hold so dear.

29. The deal with Turkey has profound implications on us and we have no say over it.

30. We can expect many more thousands of migrants to find their way into this country as a direct consequence of this deal with Turkey.

31. Many voters out there will deeply resent the fact that they have simply not been asked about this deal with the EU and Turkey.

32. We should speed up the asylum process and seek to do so while tackling logjams in the system for those declined asylum status.

33. We value and want to encourage tourism however there are inequalities in the current system which treats some nationalities more favourably than others.

34. Our membership of the European Union and associated acceptance of the free movement of people principle means we are unable to prevent criminals arriving on our shores.

35. We must leave the EU to prevent those with criminal convictions coming here.

36. Our new immigration policies will begin when we confirm our intention to leave the EU with an out vote in a national referendum.

37. Any European Union citizen who is resident in the UK at the time of the referendum will be permitted to remain and work here.

38. Our key aim is to control immigration so we will abolish the EEA family permit scheme and reinstate the primary purpose rule.

39. Foreign nationals marrying British citizens will have to prove that the primary purpose of their marriage is not to obtain British residency.

40. We will also repeal Labour's human rights legislation.

41. UKIP immigration policy is built on fairness.

42. We would aim to reduce migration, guarantee border security accommodate sensible numbers of foreign students, protect asylum seekers and make sure new migrants do not place undue pressure on our NHS.

APPENDIX 15: Breakdown of User Ratings in Test Set Generation

EU-Conservative Agreement			
Test type	**Total agreement (595)**	**Percentage**	**# of o's**
-2-1012	186	31.26%	84
-2-1012 (without the o agreements)	$\frac{186-84}{595-84}=\frac{102}{511}$	19.9%	0
-101	297	49.9%	84
-101 (without the o agreements)	$\frac{213}{511}$	41.68%	0
-101 (with at least two people in agreement)	$\frac{213+101}{511}$	61.4%	0
Immigration- Conservative Agreement			
Test type	**Total agreement (575)**	**Percentage**	**# of o's**
-2-1012	151	26.2%	51
-2-1012 (without the disagreements)	$\frac{151-51}{575-51}=\frac{100}{524}$	19.08%	0
-101	339	58.9%	51
-101	$\frac{288}{524}$	54.96%	0
-101 (with at least two people in agreement)	$\frac{288+103}{524}$	74.61%	0
EU-Labour Agreement			
Test type	**Total agreement (595)**	**Percentage**	**# of o's**
-2-1012	320	53.78%	106
-2-1012 (without the disagreements)	$\frac{320-106}{595-106}=\frac{214}{489}$	43.76%	0
-101	373	62.68%	106
-101 (without the disagreements)	$\frac{267}{489}$	54.6%	0
Immigration- Labour Agreement			
Test type	**Total agreement (575)**	**Percentage**	**# of o's**
-2-1012	335	58.26%	182
-2-1012 (without the disagreements)	$\frac{335-182}{575-182}=\frac{153}{393}$	38.93%	0
-101	399	69.4%	182
-101	$\frac{217}{393}$	55.21%	0
EU- LD Agreement			
Test type	**Total agreement (595)**	**Percentage**	**# of o's**
-2-1012	314	52.77%	54
-2-1012 (without the disagreements)	$\frac{314-54}{595-54}=\frac{260}{541}$	48.05%	0
-101	441	74.11%	54
-101	$\frac{387}{541}$	71.5%	0
Immigration- LD Agreement			
Test type	**Total agreement (575)**	**Percentage**	**# of o's**

-2-1012	294	51.1%	38
-2-1012 (without the disagreements)	$\frac{294-38}{575-38} = \frac{281}{537}$	47.67%	0
-101	411	71.47%	38
-101	$\frac{373}{537}$	69.45%	0
EU-Conservative Agreement on UKIP data			
Test type	**Total agreement (95)**	**Percentage**	**# of 0's**
-2-1012	27	28.4%	11
-2-1012 (without the 0 agreements)	$\frac{27-11}{95-11} = \frac{16}{84}$	19.04%	0
-101	56	58.9%	11
-101 (without the 0 agreements)	$\frac{45}{84}$	53.57%	0
-101 (with at least two people in agreement)	$\frac{45+6}{84}$	60.71%	0
Immigration- Conservative Agreement on UKIP data			
Test type	**Total agreement (75)**	**Percentage**	**# of 0's**
-2-1012	21	28%	9
-2-1012 (without the disagreements)	$\frac{21-9}{75-9} = \frac{12}{66}$	18.18%	0
-101	47	62.6%	9
-101	$\frac{38}{66}$	57.57%	0
-101 (with at least two people in agreement)	$\frac{38+3}{66}$	71.2%	0
EU- Labour Agreement on UKIP data			
Test type	**Total agreement (95)**	**Percentage**	**# of 0's**
-2-1012	19	22.1%	7
-2-1012 (without the disagreements)	$\frac{19-7}{95-7} = \frac{12}{88}$	13.6%	0
-101	69	72.63%	7
-101 (without the disagreements)	$\frac{62}{88}$	70.45%	0
Immigration- Labour Agreement on UKIP data			
Test type	**Total agreement (75)**	**Percentage**	**# of 0's**
-2-1012	44	58.6%	2
-2-1012 (without the disagreements)	$\frac{44-2}{75-2} = \frac{42}{73}$	57.5%	0
-101	66	88%	2
-101	$\frac{64}{73}$	87.67%	0
EU- LD Agreement on UKIP data			
Test type	**Total agreement (95)**	**Percentage**	**# of 0's**
-2-1012	54	56.84%	2
-2-1012 (without the disagreements)	$\frac{54-2}{95-2} = \frac{52}{93}$	55.91%	0
-101	84	88.4%	2

-101	$\frac{82}{93}$	88.17%	0
Immigration- LD Agreement on UKIP data			
Test type	**Total agreement (75)**	**Percentage**	**# of 0's**
-2-1012	51	68%	3
-2-1012 (without the disagreements)	$\frac{51-3}{75-3}=\frac{48}{72}$	66.6%	0
-101	68	90.6%	3
-101	$\frac{65}{74}$	87.8%	0

APPENDIX 16: Focus Group Discussion Sentences

1. Immigration is not a concern for the UK.

2. There are too many migrants coming into this country.

3. Close British borders to people coming from outside the EU

4. The UK must not close her borders to people coming from outside the EU

5. The UK should act like Trump, by banning Muslims from coming into the UK.

6. Government must tackle the problem of migrants on benefits.

7. UK's position on EU migration should be to reduce it to a few thousand people, maybe 20000 people.

8. Migrants from Syria are not a danger to the UK

9. Child Migrants from Syria are not a danger to the UK

10. Migrants have the right to make visa appeal after appeal and this must be stopped

11. We need to change the direction of our visa and immigration system.

12. We should adopt a point based system for EU migrants

13. Cut non-EU migration to the lowest levels

14. Foreigners should have to pay for their NHS care

15. Government should make the visa application process tougher

16. Absolutely committed to tackling the exploitation of vulnerable migrants

17. Visitors applying for visas must show that they must not recourse to public funds

18. Foreign criminals should be able to prevent deportation simply by dragging out the appeals process

19. The jungle migrant camp should be closed because it is inhumane

20. We should be supporting vulnerable children coming in from Syria and Turkey

21. Should be supporting migrants coming in from Syria and Turkey

22 Immigration would add an extra 2 million people to the UK's population

23. Immigration would negatively affect our already strained NHS

24. Migrants trying to access the channel tunnel should be housed in a detention centre.

25. Detention centres for asylum seekers should be closed.

26. It is important to ensure that those who arrive in the European Union are properly fingerprinted and that we have the identification of those who come to our shores.

27. The police are also doing a great deal of very good work to tackle trafficking.

28. Many people who come to this country to study get a very good impression of it.

29. Many people who come to this country end up staying longer than they should.

30. Student visas should be cancelled for migrants who skip classes

31. We say no to the EU

32. We say no to the EU and yes to the Commonwealth

33. Introducing screening and monitoring of foreigners coming to the UK

34. Migrants on benefits will be made to pay their fair share towards the NHS

35. Opposed to Turkey's membership of the EU

36. Opposed to membership of the EU

37. Will never support the EU's migration policy

38. Should provide visas for exceptionally talented individuals.

APPENDIX 17: Focus Group Discussion Report

Sentence	Topic	CO	LA	LD	Comments and Observations
1. Immigration is not a concern for the UK.	I	n	n	y	Conservatives and Labour agreed that immigration is a major concern although all the participants agree that there is some media hype on the. They also agreed that the media hype has fuelled some of the public outcry. It is for this reason that both LD participants insisted that it wasn't a concern as significant as 'terrorism', and the 'welfare crisis' and that if the media stopped reporting on 'Immigration' then politicians would focus on other important issues.
2. There are too many migrants coming into this country.	I	2-n/1-y	n	n	Only one conservative participant assumed that to this statement. The remaining participants were of the view that migration into the UK is no different from any country in Western Europe
3. Close British borders to people coming from outside the EU	I, EU	n	n	n	All the participants agreed that this comment was quite extreme and are vehemently against it
4. The UK must not close her borders to people coming from outside the EU	I, EU	1 -y/2 -n	y	y	In this case one of the conservatives was of the view that there should be more controls. The others felt that non-EU migration is under control.
5. The UK should act like Trump, by banning Muslims from coming into the UK.	I	n	n	n	This question was initially posed in jest, and everyone was vehemently against it
6. Government must tackle the problem of migrants on benefits	I, EU	y	y	y	All the participants viewed this as a bigger problem than immigration numbers. Surprisingly the LDs and Labour were quite keen on this issue and feel it should be tackled.

7. UK's position on EU migration should be to reduce it to a few thousand people, maybe 20000 people.	I, EU	y	n	n	The LDs who are for open borders are against any restriction and I got the impression that Labour are against this because it is a conservative policy. Conservative participant also used the word 'restrict' and our LD participants were opposed to 'restricting' or 'reducing'
8. Migrants from Syria are not a danger to the UK	I	u	u	u	All the participants were unsure of this one. Most are of the view that there is some risk in accepting migrants because of the likelihood of infiltration from ISIS. However, when we asked about child migrants from Syria, they were all for it.
9. Child Migrants from Syria are not a danger to the UK	I	y	y	y	All participants were for accepting child migrants and felt that they do not pose much of a threat to the UK. However, both Labour and one conservative participant expressed some concern as to how many to accept at a time.
10. Migrants have the right to make visa appeal after appeal and this must be stopped	I	y	u	n	This was actually a UKIP election promise: To stop migrants who's visa have been refused from making appeal after appeal. The conservatives were for this. However, both LD and Labour were unsure primarily because they felt that by allowing appeal after appeal it boggled down the judicial system. And so they are unsure of its benefit for the migrant or the judicial system. All the LDs and Labour participants agreed however, that people should have their day in court. To this end, we have marked both Labour and LD as unsure.
11. We need to change the direction of our visa and immigration system	I	y	y	y	Everyone agreed with this. Each offering diverse reasons, but all in all there was universal agreement for this.
12. We should adopt a point based system for EU migrants	I, EU	y	u	n	The Labour participants on hearing some of the arguments for this were inclined to saying yes, even though Labour has no direct policy on this matter

13. Cut non-EU migration to the lowest levels	I, EU	y	n	n	This is a claim that has been made by the Conservatives, and the view we get is that Labour and LD only oppose this statement simply because it is something the conservatives say. The argument of Labour participants was that Gordon Brown's measures for curbing non-EU migration was already in place and working before the conservatives.[108]
14. Foreigners should have to pay for their NHS care	I	y	n	n	The argument here again also seemed to be one where the LAB and LD opposed the view because it was simply a Conservative policy. The arguments from Labour was that this would not fix the NHS and the argument from LDs was that British citizens are eligible to healthcare in the EU.
15. Government should make the visa application process tougher	I	Y	n	n	Again, to this LD and Labour participants responded that the visa application process was already hard enough. The impression we get here is that the Conservatives are in agreement because it is part of their policy. (This statement was taken out of the UKIP policy manual)
16. Absolutely committed to tackling the exploitation of vulnerable migrants	I	y	y	y	Everyone was in agreement
17. Visitors applying for visas must show that they must not recourse to public funds	I	y	y	y	
18. Foreign criminals should be able to prevent deportation simply by dragging out the appeals process	I	n	n	n	All disagreed with this, however we find that the Conservatives blamed Labour for this problem.

[108] This might just be another contextual issue

19. The jungle migrant camp should be closed because it is inhumane	I, EU	y	y	y	All agreed with this.
20. We should be supporting vulnerable children coming in from Syria and Turkey	I	y	y	y	All participants agreed with this, (see next question)
21. Should be supporting migrants coming in from Syria and Turkey	I	2- n/1-y	u	1-y 1-n	Two of the conservatives disagreed with this. The Labour participants were unsure stating that they would love to but wouldn't make a decision until they had a plan in place and the people had been consulted. One LD agreed stating 'As a rich nation, the UK should help the poor', the other simply disagreed siting concerns about the cost to the tax payer.
22 Immigration would add an extra 2 million people to the UK's population	I	2-n, 1-y	n	n	LD and Labour viewed this comment as bordering on scare mongering. Calling it 'daily mail style'. The Two of the conservatives agreed and 1 simply said it was a statement of fact and that it could have potential repercussions on welfare. (Again, this was a comment made by UKIP in their manifesto)
23. Immigration would negatively affect our already strained NHS	I	u	u	n	
24. Migrants trying to access the channel tunnel should be housed in a detention centre.	I	y	n	n	Labour and LD agreed with the concept of detention but disagreed in principle because the detention centres are not up to the standard of other countries
25. Detention centres should be closed down	I	n	n	n	All the groups agreed on this
26. It is important to ensure that those who arrive in the European Union are properly fingerprinted	I,EU	y	y	y	This was taken from Hansard and all parties agreed with this

and that we have the identification of those who come to our shores.					
27. The police are also doing a great deal of very good work to tackle trafficking.	I	y	y	y	All parties agreed with this, however, one of the Labour participants suggested that they could do more if the Conservatives in Government provide more funding.
28. Many people who come to this country to study get a very good impression of it.	I	y	y	y	
29. Many people who come to this country end up staying longer than they should	I	u	n	n	Most were unsure of this; however, Labour and LD chose to refute it because according to the Labour participant 'it's the kind of scare mongering comment you get from UKIP'
30. Student visas should be cancelled for migrants who skip classes	I	n	n	n	All participants felt this was too harsh including the Conservatives. However, this is a Conservative policy.
31. We say no to the EU	EU	n	n	n	All participants were pro EU
32. We say no to the EU and yes to the commonwealth	EU	n	n	n	This was a UKIP policy and all three parties disagreed with no to the EU. As for the commonwealth, they were of the view that there's more potential in the EU for business and growth
33. Introducing screening and monitoring of foreigners coming to the UK	I	y	y	y	
34. Migrants on benefits will be made to pay their fair share towards the NHS	I	y	n	n	
35. Opposed to Turkey's membership of the EU	EU	y	n	u	

36. Opposed to membership of the EU	EU	n	n	n	
37. Will never support the EU's migration policy	EU, I	1-y/2-n	u	n	The Labour participants are of the view that there should be some negotiation between the UK and the EU on migration. 2 Conservative MPs agreed with this.
38. Should provide visas for exceptionally talented individuals	I	y	y	y	All parties are of the view that exceptionally talented individuals should be granted visas

APPENDIX 18: Definition of Evaluation Metrics

The evaluation metrics applied in this research are as follows:

1. Accuracy (A) – This is the fraction of documents assigned to their correct classes by the classifier. $\frac{\#Correct\ classifications}{\#Total\ \ classifications}$ or $\frac{tp+tn}{tp+fp+tn+fn}$, where tp = true positive, tn = true negative, fp = false positive, fn = false negative

2. Precision (P)– The ratio or percentage of the number of correctly classified sentences to the total number of classified. It is expressed mathematically as $P = \frac{tp}{tp+fp}$

3. Recall (R) - This is the ratio of the number of correctly classified sentences to the total number of sentences that should have been labelled. It is formulated as $R = \frac{tp}{tp+fn}$

4. F-measure (Van Rijsbergen, 1975) combines the precision and recall scores into a single metric. F-Measure is given as $F_{\beta=1} = \frac{(\beta^2+1)PR}{\beta^2P+R}$, where β weighs the importance of precision and recall, such that when $\beta > 1$, recall is favoured while $\beta < 1$ favours precision. In this research, we make use of $\beta = 1$, hence the F_1 score formulation is expressed as $F_1 = \frac{2PR}{P+R}$. F-scores are based on computing F_1.

5. Misclassification error (M) – This is the fraction of misclassified documents and formulated as $M = \frac{fp+fn}{tp+tn+fp+fn}$

APPENDIX 19: Comparing Named Entities Samples from Immigration and EU Corpus

	EU	Immigration
Merkel	68	2
Blair	59	11
Putin	122	0
Miliband	43	5
TTIP	19	0
NHS	22	16
NATO	277	2
UKBA	235	264
FBI	2	0
HMRC	6	10
BRP	83	114
UKIS	252	306
European Army	313	165
Europol	24	6
Maastricht	82	0
Rome	43	0
Slough	1	27
Berlin	38	3
Colchester	1	4
Manchester	9	29
Swindon	10	3
Ukraine	625	11
Africa	49	27
Germany	137	29
Poland	61	21
Washington	21	1
Lewisham	3	4
London	115	76
Syria	28	14
Turkey	60	28
Russia	66	6