DOCTOR OF PHILOSOPHY

Mining Ethos In Parliamentary Debate

Duthie, Rory William

Award date:
2020

Mining Ethos In Parliamentary Debate

Rory William Duthie

Doctor of Philosophy

University of Dundee

Scotland

2020

Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank, first and foremost, my supervisor Dr. Katarzyna Budzynska. This research would not have been possible without her enthusiasm, wisdom, patience, dedication and overall engagement in my work.

I would also like to thank my second supervisor Prof. Chris Reed who has always been a soundboard for my ideas and a great source of knowledge.

Thanks also go to the whole of ARG-tech who have provided needed distractions, encouragement and motivation. I could not have done it without you all.

Finally, I would like to thank my family and friends who have been extremely supportive throughout this journey through the highs and the lows and have kept me sane.

The work was supported by EPSRC in the UK under grant EP/M506497/1.

Declarations

Candidate's Declaration

I, Rory William Duthie, hereby declare that I am the author of this thesis; that I have consulted all references cited; that I have done all the work recorded by this thesis; and that it has not been previously accepted for a degree.

[redacted]

Rory Duthie

Abstract

Argumentation has played a fundamental role in society for centuries from debate in the public sphere to everyday conversation. Most recently computational research into argumentation has focussed upon argument mining, the automatic extraction of reasoning structures from natural language. Although the content of argument (logos) is a fundamental part of persuasion, **ethos**, the character of the speaker, also plays a significant role in communication as one of Aristotles modes of persuasion. This is identified within both argumentation (e.g. argument from expert opinion) and within conversation and debate, as sometimes a stronger character outweighs logical reasoning. Despite the importance of ethos, it has not been considered computationally in a solitary sense rather than as a bi-product of argumentation schemes or when performing argument mining on debate.

This body of research aims to perform the novel annotation of ethos, as the sole phenomenon of focus, and develop a set of ethos technologies showcasing the importance of ethos in relation to political events through **ethos mining**. Due to its consistent debate structure and the large amount of available historical data, **Hansard** (the UK parliamentary debate record) is used as an exploratory domain. Four research questions were then formulated as the objectives of this PhD project: **(RQ1)** Can ethos be reliably annotated in natural language? **(RQ2)** Can ethos be reliably extracted automatically from natural language? **(RQ3)** Can fine-grained ethos types used by speakers be reliably annotated and automatically extracted? **(RQ4)** Can the analysis of appeals to ethos give an insight into

the dynamics of the political landscape through the interactions between politicians?

Following Aristotle, ethos is specified as a property of an identifiable individual or identifiable group of individuals. Two types of ethotic linguistic activities are considered. The individual or individuals can be **supported** by others, e.g. "Mr. John Moore said, *My hon. Friend is assiduously pursuing his constituents' interests*" (positive ethotic statement) or they can be **attacked** by others, e.g. "Mr. Bruce Grocott said, *Is it not the simple truth that the Government are making the country sick?*", (negative ethotic statement).

In order to answer research questions RQ1 and RQ2, two natural language processing (NLP) pipelines, a **rule-based** and a **deep learning** approach (RQ2), were designed to automatically mine ethos in a manually annotated corpus (RQ1). To answer RQ3 a further NLP pipeline, building upon the supports and attacks of ethos extracted for RQ1 and RQ2, was created with the goal of differentiating ethos types, the grounds for which a speaker attacks or supports ethos. Finally, to answer RQ4 the output from RQ1 and RQ2 of positive and negative ethotic statements were used to produce ethos analytics based upon counts of supports and attacks for individual politicians, comparisons with external publications and relationships between politicians.

In the first approach to answer RQ1 and RQ2, this body of research produced the first corpus of ethos supports and attacks grounded in the rhetoric theory and focussing solely on ethos using transcripts from Hansard for manual annotation. The corpus incorporated 60 transcripts overall (70,117 words in total) annotated at a sentence level to allow for the full context of an ethotic support or attack to be captured. An inter annotator agreement study gave a Cohen's kappa of κ =0.67 on the identification of ethos, when determining support or attack κ =0.95, κ =1 for determining the source speaker and κ =0.84 for determining the target speaker. This result shows the reliability of the annotation guidelines and thus provides a positive outcome in regard to RQ1. Following this was the first rule-based automatic extraction of ethos, trained and tested on data from the manually annotated corpus addressing RQ2. The pipeline of standard NLP techniques and domain specific rules developed specifically for ethos mining gave an F1-score of 0.70 (53% above baseline) when determining if a sentence contains ethos and a macro-averaged F1-score of 0.78 (16% above baseline) when determining the polarity (positive or negative). The evaluation of the rule-based approach indicates reliable extraction results delivering RQ2.

To aid in generalisation this process was repeated. For RQ1, a further thirty transcripts from Hansard were manually annotated for training data and the 60 transcripts from the first iteration were re-annotated to a new set of annotation guidelines, ensuring ethos is present on the surface of a sentence, resulting in a total of 90 transcripts. The improved guidelines gave a Cohen's kappa of κ =0.67 on the identification of ethos, when determining support or attack κ =1, κ =1 for determining the source speaker and κ =0.93 for determining the target speaker. This evaluation shows a reliable annotation in line with RQ1, in particular consistency when determining ethotic statements and improvements on identifying the polarity and target of each sentence. For RQ2, deep learning methods from image classification were developed for the novel application in text classification. A Deep Modular Recurrent Neural Network (DMRNN) was created to automatically identify ethotic statements and then determine their polarity making use of a NLP pipeline. The DMRNN and full pipeline gave an F1-score of 0.74 (21% above baseline and 6% when compared with the rule-based approach) for identifying ethotic statements and a macro-averaged F1-score of 0.84 (31% above baseline and 8% when compared with the first approach) for determining if an ethotic statement is positive or negative indicating that deep learning is reliable for the automatic extraction of ethos in the case of RQ2.

Addressing RQ3, all 90 transcripts were further annotated using the Aristotelian distinction of **elements of ethos**, Wisdom, Virtue and Goodwill (wisdom referring to practical experience, virtue to character traits and goodwill to aligning with the audience). An annotation evaluation on a 10% subset of the data gave an average Cohen's κ of 0.52. The results, defined as moderate agreement (Landis and Koch, 1977) show the reliability of annotation and yet the difficulty of the task at hand which can be improved through further annotation iterations. An automatic classification using pairwise classifiers and one versus all classification gave F1-scores averaging 0.62 showing room for improvement.

Three **applications of ethos** were identified in relation to RQ4: graph analytics, qualitative analytics and quantitative analytics. The output from RQ2, both for the rule-based and deep learning approaches, allowed for the **analysis of ethotic relations** between politicians and political groupings to give an insight into the political landscape. Firstly, this shows which politicians attacked or supported one another as a directed graph. Secondly, qualitative analytics were developed exploring time series data for supports and attacks on individual politicians enabling the investigation of correlations with political events such as individual party position appointments. Finally, quantitative analytics were developed exploring time series data for supports and attacks on political parties enabling the investigation of correlations to general election results. Each of these steps give a visual, qualitative and quantitative insight into the political landscape of the time addressing RQ4.

In summary, this research has described novel advances in the new sub-field of argument mining, ethos mining, contributing: (i) manual corpora of ethos supports and attacks; (ii) the automatic classification of ethos supports and attacks; (iii) the creation of deep learning methods (DMRNN) in text classification for extracting ethos; (iv) manually annotated corpora containing ethos types; (v) the development of an NLP pipeline for classifying ethos types; and (vi) a set of ethos analytics. These advances are widely applicable to various domains not only as a tool to gauge political opinion, but to extract public opinion from various sources of natural language. Social media discussions or public deliberation can be used to build profiles of individuals over large time periods from the natural language and identify individuals or groups at the centre of popular (or unpopular) opinion.

Chapter 1

Introduction

1.1 Motivation and Background

Argumentation is at the core of society from day to day conversation to the highest levels of democracy. A more recently established field within argumentation, argument mining, has the primary goal to automatically extract the reasoning structures used within natural language. Rather than just determining *what* opinion a speaker holds, as is the case in sentiment analysis and opinion mining, argument mining looks to determine *why* they hold that opinion using the structure of argument to extract the information. In recent years the automatic extraction of these structures has advanced in several domains, and yet the whole focus has been upon the content of what is said and not about who, in particular, said it.

For millennia in rhetoric, the character of the speaker has been recognised as just as important as what the speaker is trying to say. To this end, Aristotle (1991) defined three modes of persuasion in which a speaker can influence others through communication: *logos*, reasoning and the content of what is said; *ethos*, the character of the speaker; and *pathos*, appealing to the emotions of an audience. One domain for which the modes of persuasion are particularly prominent is parliamentary debate. Politicians need to argue effectively, whilst being seen to have a positive ethos in order to persuade an audience during a speech in parliament. In the Aristotelian sense, ethos refers to establishing one's own character, particularly making this positive character clear to an audience. In parliamentary debate, the dynamic of ethos shifts as politicians are in parliament through merit, and therefore a priori have already established ethos. Essentially over time what many politicians realise is that their ethos can far outweigh what they say in the eyes of voters, especially when this ethos is built or damaged by others. In this case and specifically defined for this study, a politician can utilise a strategy of supporting the ethos of other politicians, through positive sentiment, to solidify their statement (see example 3 where "every support" highlights the positive sentiment to the target "he" ,i.e., Mr Younger). At the same time a strategy of attacking another politician's ethos, through negative sentiment, can be utilised to undermine a statement (see examples 1, 2 where "no longer able" signals the negative sentiment to "he", i.e., Mr Younger and "bad position" to "hon. Gentelman", i.e., Mr Bruce). Thus the intuition behind ethos mining is to extract mentions of specific entities and their related sentiment from the linguistic surface.

(1) **Mr Malcolm Bruce said,** *Will the Secretary of State (Mr. Younger) acknowledge that there is real concern in Scotland among parents and teachers that he is no longer able to maintain an adequate education system?*

(2) **Mr George Younger said,** *If the hon. Gentleman (Mr. Bruce), as a Member of Parliament with every access to information, thinks that that is a proper description of the offer, we are indeed in a bad position.*

(3) **Mr Gordon Wilson said,** *I assure him that he (Mr. Younger) will have every support from my party on this matter.*

In some cases the notion of ethos defined above can be hard to distinguish against the concept of trust due to the fact they are complementary. Trust, the idea of an entity acting dependably, though is an all encompassing construct which relies heavily upon contextual information and is focussed upon making decisions based on this trust (*cf.*

(Castelfranchi and Falcone, 2010; Grandison and Sloman, 2000)). In the latter case ethos plays a role in the trust decision process, whereby the character of a speaker is one aspect to consider. Also similar to trust, as an output of ethos mining, is the traversing of networks to determine the relationship between two entities. In trust this is determined through networks, either transactional (Berners-Lee and Hendler, 2001) or social, which are largely unweighted (weight is used here in the sense that a relationship between two entities can be quantified) whereas through ethotic statements these relationships are created and a weight can be established through the quantity of such statements. Thus, ethos mining can be utilised for the exact purpose of automatically generating such networks to allow the production of insightful analytics based on political dynamics.

Whilst the importance of ethos is clear in the context of politics and parliamentary debate the focus of automatic extraction has been on logos in argument mining. The only exception to this was the study of stance classification of politicians (Hirst et al., 2014), reputation defence speech (Naderi and Hirst, 2017, 2018) and linking argument frames with pro and con arguments (Naderi and Hirst, 2015)[1]. Furthermore, the automatic extraction of arguments specifically in the form of argumentation schemes can be used for analysing debate dynamics (Lawrence and Reed, 2016; Toledo-Ronen et al., 2016). In the case of ethos, this has been extracted as a by-product of argumentation schemes. For instance in the argument schemes, position to know and expert opinion, ethos is not defined as a conclusion within the scheme that can be supported or attacked. In fact ethos in these cases is instead used within the wider argument structure as a premise for a particular argument and therefore the notion of schemes cannot be used to model ethos support or attacks. Ethos in this case and others is not then the main focus of extraction in argument mining. For example in (Hidey et al., 2017) only 3% of sentences in the corpus, equating to 30 sentences in total, were annotated as containing ethos. This is also the case in (Carlile et al., 2018) where only 25 claims are annotated as containing ethos. In (Habernal et al., 2018) the focus was shifted to ad hominem, (similar to that of attacking ethos shown

[1] Each of the works here are defined as argument mining but do not conduct the task of extracting an overall argument structure.

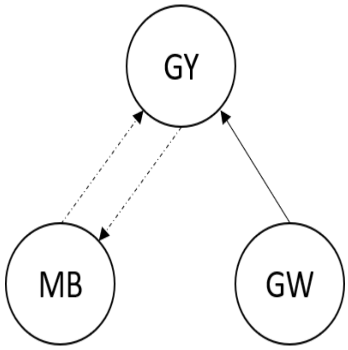

Figure 1: A basic network of ethotic statements from examples 1, 2 and 3. Dashed lines show an ethotic attack whereas a solid line shows support. Nodes show speakers Malcolm Bruce (MB), George Younger (GY) and Gordon Wilson (GW).

in examples 1 and 2) of which 1,267 instances were annotated and classified, although there was no counterpart to ethos support. Each of these works highlights the need for a comprehensive study solely focussed upon the annotation and extraction of both supports and attacks of ethos particularly in a political setting where this type of language use is prevalent.

Political science plays a major role in the analysis of the language used within debate (*cf.* (Laaksonen et al., 2017; Nooy and Kleinnijenhuis, 2013)) and in studying the shape of the current political landscape. In the former case, manual analysis is relied on heavily to determine political outcomes, whilst sentiment analysis plays a particularly import role in the latter case speculating on how members of the public will vote in parliamentary elections (*cf.* (Burnap et al., 2016; Franch, 2013; McGregor et al., 2017)). To the same extent, a politicians' stance can be utilised to gain insight from a debate as can a full argument structure. What is not clear is the role ethos can play for analysing the same data specifically when taking into account supports and attacks. Thus a set of analytics to determine the power of ethos within the political sphere in comparison to the media of the time can be utilised to determine the power in ethos.

1.2 Goals

The main goals of this thesis are the manual annotation of ethos to create a publicly available resource and the automatic extraction of ethos supports and attacks within parliamentary debate. From a manual annotation perspective, no other work has tried to build an annotated corpus solely focussed upon ethos supports and ethos attacks. As such their is little research on empirically grounded ethos supports and attacks which this corpus enables.

In the case of automatic extraction no other research has been conducted in the focussed area of ethos mining, which tackles the task of building text classifiers to extract ethos supports and attacks. Instead (see also section 1.1) the main area of research in text extraction in argumentation has been upon argument mining where any ethos extraction was achieved on small datasets or through ad hominem (AH) related to ethos attacks. By building ethos mining pipelines created using both tried and tested and novel methods of extraction this will then allow the independent study of ethos to determine what role it plays within parliamentary debate.

Looking to logos as a motivator for further research on ethos can also be useful. In the case of argument extraction the classification can be further continued using argumentation schemes. These schemes can help to extend automatic classification using the scheme structure to extract premises or to develop specific large scale analysis. To this extent ethos schemes or types for more fine grained analysis can be developed in the same vein as argumentation schemes. There are two goals in this case: the annotation of a fine-grained set of ethos schemes or types; and, the automatic classification of these same schemes or types. The rationale being that in the future these types can be used to improve the classification of ethos supports and attacks or for further strategic analysis. This point then leads to a fifth goal of this thesis, as to what role ethos supports and attacks play in a wider political context.

This goal is motivated by the output of ethos mining, as illustrated in figure 1, the extraction of ethos allows the creation of ethos networks which show the speaker dynamics within parliamentary debates. Until now, no research has been conducted on how these dynamics are reflected in the public sphere. For example, are all the important interactions between two politicians highlighted by the media? Or, does the media have a more specific focus on politicians deemed to be important? In each of these cases supports and attacks on ethos, constructed by ethos mining can be utilised to investigate particular political traits. Provided below are the overall set of goals of this research:

- The manual annotation of ethos to create a publicly available resource.
- The automatic extraction of ethos supports and attacks within parliamentary debate.
- The annotation of a fine-grained set of ethos schemes or types.
- The automatic classification of these same schemes or types.
- Determining what role ethos supports and attacks play in a wider political context.

1.3 Research Questions

Following on from the motivations and goals described above there are four main research questions which are addressed in this thesis.

- **(RQ1)**: Can ethos be reliably[2] annotated independently of logos in natural language?
- **(RQ2)**: Can ethos be reliably[3] extracted automatically from natural language?
- **(RQ3)**: Can fine-grained ethos types used by speakers be reliably annotated and automatically extracted?
- **(RQ4)**: Can the analysis of ethos give an insight into the dynamics of the political landscape through the interactions between politicians?

[2]Reliability in the case of manual annotation can be determined through the scale provided in (Landis and Koch, 1977).

[3]Reliability in the case of automatic extraction is determined through scores above baseline metrics.

To answer these research questions several steps were taken in this thesis. In the first step, to answer RQ1, a working definition of ethos was specified taking inspiration from Aristotle. Rather than ethos only relating to one's own ethos, ethos is defined here as: a property of an individual or a group which can be supported or attacked. This is further specified through the terms: ethotic sentiment expression (ESE), a sentence which contains ethos referred to through sentiment; ¬ethotic sentiment expression (¬ESE), a sentence deemed to hold no ethos; a positive ethotic sentiment expression (+ESE); and, negative ethotic sentiment expression (-ESE).

As a next step transcripts of parliamentary debate, from the UK House of Commons, were chosen as a relevant and promising domain for ethos classification, due to the volatile language. From these transcripts ethos supports and attacks were annotated using the Online Visualisation of Argument tool (OVA) (Janier et al., 2014). OVA was specifically chosen for this task due to the ease of annotation as it allows for the incorporation of the original text transcript from which the annotation of ethos can be directly made.

To answer RQ2, as a first step, a rule-based ethos mining approach was taken. This involved using the manually annotated corpus to construct domain specific rules for the identification of ESEs. As part of the pipeline sentiment classification, using an ethotic domain specific lexicon and a generic sentiment word lexicon, was utilised to determine +/- ESEs. This rule-based approach was deployed specifically to determine how effective rule-based methods were for the task of ethos mining against standard feature representation methods which may not capture the complex phenomena. Lexicons were used for sentiment classification to ensure reliable results, hence the domain specific lexicon, as a standard bag-of-words approach would not perform optimally due to the relative size of the manually annotated dataset.

As a second step in the research conducted for this thesis, RQ1 and RQ2 were revisited to improve and extend the manual annotation and to employ deep learning approaches to ethos mining. The corpus was revisited to extend its overall size which provides the advantage of more data for training in machine learning and a confirmation of annotation consistency. Deep learning approaches were deployed due to the widely known improvements they have made in similar tasks such as argument mining and sentiment analysis. The deep learning approach also has the advantage of being more generalisable when compared with domain specific rules.

In the case of fine-grained ethos types in RQ3, Aristotle's elements of ethos were used: wisdom referring to practical experience; virtue to character traits; and goodwill to aligning with the audience. These ethos types identify the language used by politicians to support and attack. After a step of annotation guideline construction, ethos elements were annotated on top of the existing +/-ESE annotation. As a next step an NLP pipeline for ethos elements using domain specific and general features (such as proper and plural noun identification and principal component analysis) was created. These features alongside pairwise classifiers and one versus all classification then determined which element of ethos should be assigned to a +/-ESE. In this case the pairwise and one versus all classifications provide binary classifications ensuring that only one label can be assigned to each sentence both of which are essential for Support Vector Machines and Logistic Regression classifiers. The same methodology is undertaken for all classifiers to ensure comparability.

Finally to answer RQ4, +/-ESEs are used to define three applications of ethos in the context of a wider political analysis. Both the rule-based and deep learning approach to ethos mining allow for the creation of ethos analytics and visualisations of ethotic interactions between politicians and political groupings to give an insight into the political landscape. Graph analytics, are used to visualise the supports and attacks of ethos between politicians and groups. Qualitative analytics are then deployed to determine correlations between political events and the quantity of ethos supports and attacks on an individual. Quantitative analytics are then used in a final step to analyse time series data over a political event. This allows +/-ESEs to be tracked over this event where the change in the values can indicate the outcome of said event.

1.4 Technical Contributions of Ethos Mining

The overall goal of this thesis is to, for the first time, develop techniques solely for ethos mining in parliamentary debate. In order to achieve this goal several technical contributions were made.

The first corpus containing solely ethos supports and attacks was created. The corpus incorporated 60 transcripts overall (70,117 words in total) annotated at a sentence level to allow for the full context of an ethotic support or attack to be captured. An evaluation annotation by a second annotator, comprising of a 10% subset of the data, was used to determine inter-annotator agreement. This gave a Cohens kappa of κ=0.67 on the identification of ethos, when determining support or attack κ=0.95,κ=1 for determining the source speaker and κ=0.84 for determining the target speaker all of which are relative to the identification of ethos.

Following this annotation the first ethos mining pipeline was created using rule-based domain specific features and standard NLP techniques such as anaphora resolution, part-of-speech tagging, named entity recognition and sentiment classification. This pipeline gave an F1-score of 0.70 (53% above baseline) when determining if a sentence contains ethos and a macro-averaged F1-score of 0.78 (16% above baseline) when determining the polarity (positive or negative).

In order to improve the rule-based ethos mining pipeline a further 30 transcripts were annotated and the previous 60 re-annotated to an improved set of annotation guidelines to ensure reliability of annotation. The improved guidelines gave a Cohens kappa of κ=0.67 on the identification of ethos, when determining support or attack κ=1, κ=1 for determining the source speaker and κ=0.93 for determining the target speaker. A novel Deep Modular Recurrent Neural Network (DMRNN) was then created to automatically identify ethotic statements and then determine their polarity making use of a NLP pipeline. This pipeline used similar techniques as the rule-based pipeline but was extended to use external information from Wikipedia for anaphora resolution and the removal of domain specific text from sentences such as a question format. The DMRNN gave an F1-score of 0.74 (21% above baseline and 6% when compared with the rule-based approach) for identifying ethotic statements. The full pipeline gave a macro-averaged F1-score of 0.84 (31% above baseline and 8% when compared with the first approach) for determining if an ethotic statement is positive or negative.

Following the annotation of ethos supports and attack, the first corpus of ethos types was created. Building upon the support and attack annotation all 90 transcripts were annotated with wisdom, virtue and goodwill. An annotation evaluation on a 10% subset of the data gave an average Cohens κ of 0.52. Following this annotation an automatic classification using pairwise classifiers and one versus all classification, with features such as proper noun detection and principal component analysis, gave F1-scores averaging 0.62.

Finally, three novel applications of ethos mining were created forming a set of ethos analytics, graph analytics, qualitative analytics and quantitative analytics. The output from the rule-based and deep learning approaches, allowed for the analysis of ethotic relations between politicians and political groupings to give an insight into the political landscape. Firstly, this shows which politicians attacked or supported one another as a directed graph. Secondly, qualitative analytics were developed exploring time series data for supports and attacks on individual politicians enabling the investigation of correlations with political events such as individual party position appointments. Finally, quantitative analytics were developed exploring time series data for supports and attacks on political parties enabling the investigation of correlations to general election results. As a result ethos mining pipelines have been applied to large amounts of data, determining the relationships between politicians not normally seen by the general public and providing empirically grounded insights between these relationships and political positions previously unidentified.

The technical contributions of this thesis can then be summarised as several firsts for ethos mining: the largest corpus of both ethos supports and attacks; an ethos mining pipeline to automatically extract ethos supports and attacks; an ethos focussed sentiment lexicon; a deep modular recurrent neural network for text classification and ethos supports and attacks; a corpus of ethos types (elements of ethos); A natural language processing pipeline for automatically classifying ethos types; and a set of ethos analytics.

1.5 Thesis Structure

The chapters of this thesis build upon several research publications which have been further extended. In order to address the research questions these papers form the basis of the chapters and were reported in the following way: chapters 2 and 3 give the background to ethos mining; chapters 4, 5, 6, 7 and 8 provide the main contributions of this thesis to the state of the art in the ethos mining; and chapter 9 concludes this research and provides possible future extensions.

In chapters 2 and 3 an overview is given of the theoretical background to ethos, focussing upon rhetoric and argumentation theory. Also outlined are the computational fields and methods closest to that of ethos mining. Although not directly possible a comparison is made to the literature that performs aspects of ethos mining.

Chapter 4 then outlines the domain of parliamentary debates and the base annotation guidelines for ethos supports and attacks. Specifically addressed is the political landscape of the time period as well as the formalities of the UK parliament. Also outlined is the core data used for ethos mining, how it is subsequently stored, the infrastructure used to do this and a description of the base annotation undertaken as part of this research.

Chapter 5 describes the initial manually annotated corpus of ethos supports and attacks and the first rule-based ethos mining pipeline. This chapter extended the research undertaken in (Duthie et al., 2016a)[4].

Chapter 6 then describes an extended ethos mining pipeline making use of new methods in text classification, a Deep Modular Recurrent Neural Network built upon an extended corpus of ethos supports and attacks the largest publicly available. This chapter extended

[4] Duthie, R., Budzynska, K., & Reed, C. (2016). Mining Ethos in Political Debate. In Proc. of the Int Conference on Computational Models of Argument (COMMA, 2016) (pp. 299-310) [**Won Best Student Paper Award**]

the research undertaken in (Duthie and Budzynska, 2018b)[5].

Chapter 7, then describes the process of using ethos supports and attacks to annotate and extract elements of ethos (ethos types). The ethos types are annotated on top of the ethos support and attack relations to build the first corpus of ethos elements. The first natural language processing pipeline for ethos types was then built. This chapter extended the research undertaken in (Duthie and Budzynska, 2018a)[6].

In Chapter 8 a description is given of possible applications of ethos supports and attacks. The output from the ethos mining pipelines described in Chapters 5 and 6 is used to construct graph, qualitative and quantitative analytics for individual politicians and political groups. Again this chapter extended the research in (Duthie et al., 2016a) and (Duthie and Budzynska, 2018b).

Finally, Chapter 9 concludes this thesis whilst providing an outline of the possible future work avenues.

Several other works are also mentioned in this thesis:

- Visser, J., Duthie, R., Lawrence, J., & Reed, C. (2018). Intertextual correspondence for integrating corpora. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018).[7]

- Lawrence, J., Duthie, R., Budzynska, K., & Reed, C. (2016). Argument Analytics. In Proc. of the Int Conference on Computational Models of Argument (COMMA) (pp. 371-378).[8]

- Duthie, R., Lawrence, J., Budzynska, K., & Reed, C. (2016). The CASS technique for evaluating the performance of argument mining. In Proceedings of the Third Workshop on Argument Mining (pp. 40-49).[9]

[5] Duthie, R., & Budzynska, K. (2018). A Deep Modular RNN Approach for Ethos Mining. In Proc. of the International Joint Conference on Artificial Intelligence (IJCAI 2018) (pp. 4041-4047)

[6] Duthie, R., & Budzynska, K. (2018). Classifying types of ethos support and attack. In Proc. of the Int Conference on Computational Models of Argument (COMMA 2018)

[7] http://www.aclweb.org/anthology/L18-1554

[8] http://ebooks.iospress.nl/volumearticle/45276

[9] http://www.aclweb.org/anthology/W16-2805

- Visser, J., Konat, B., Duthie, R., Koszowy, M., Budzynska, K., & Reed, C. (2019). Argumentation in the 2016 US presidential elections: Annotated corpora of television debates and social media reactions. In Language Resources and Evaluation.[10]

[10] https://link.springer.com/article/10.1007/s10579-019-09446-8

Chapter 2

Theoretical Foundations

This chapter describes the philosophy behind and applications of rhetoric and argumentation related to the notion of ethos, the character of the speaker. These foundations are crucial in the effective automation of this phenomenon, ethos mining, in that a sound theoretical base is needed when attempting to perform any kind of natural language processing.

Rhetoric describes ethos, alongside logos and pathos, as a mode of persuasion which has been considered since Greek antiquity. Over time the definition of ethos has evolved to consider discourse in general and not just in the legal or political domain and how it is structured within natural language.

Argumentation theory defines ethos in discourse through well known appeals such as ad hominem or argumentation schemes such as expert opinion. Despite this continued focus on ethos, neither rhetoric or argumentation theory clearly define ethos in such a way so that it can be automatically extracted. In rhetoric, rather than a phenomenon which can be annotated reliably through the definitions, ethos is instead defined through the topics of interest at the time (for example, in politics or law). In argumentation theory, on the other hand, the focus has mainly been upon logos and the art of reasonable and logical argument. This chapter then looks to describe the foundations of ethos and the steps towards forming a definition which can be utilised for automatic extraction methods.

2.1 Ethos in Rhetoric

In order to effectively persuade an audience, various techniques can be used. The use of sound logic (logos) can persuade some, whilst appealing to the emotions of the audience (pathos) can be more effective for others. Also appearing to be a credible source (ethos), through past experience and or character virtues can effectively persuade.

Each of these techniques have been studied extensively as rhetoric, the art of persuasion through words and actions. Rhetoric has been studied since antiquity and has taken many different forms, however, these all have the commonalities of reasoning, emotion and credibility.

2.1.1 Logos, Ethos and Pathos

Rhetoric can be defined in several ways and has been over time, but, it was not always defined specifically as rhetoric. Initially the Sophists, experts in public speaking, taught the skills needed to address an audience, for compensation, which coincided with the need for public speaking in Greece due to public forums (Taylor and Lee, 2016) (see figure 2 for a timeline of the period). The Sophists looked to teach the skills needed for persuasion to allow the student to gain an advantage in the political sphere using any possible means motivated by the idea that "knowledge and truth are illusory" (Johnson, 1998). Particularly, they believed that persuasion was subjective and that in order to convince others you must appeal to an individual through arguments that target them specifically through emotion.

A distinction can be made between Sophists and the later works in rhetoric in that these later works looked to create manuals on rhetoric which can be followed to communicate effectively rather than rhetoricians travelling to teach for a fee, as was the case for Sophists. A distinction can also be made in the reasoning behind the teaching of rhetoric - rather than as a tool to manipulate - rhetoric was taught as a set of ideals and rules that should be followed (a normative approach). These rules then allowed a speaker to be perceived as acceptable and therefore the assertions they make are listened to.

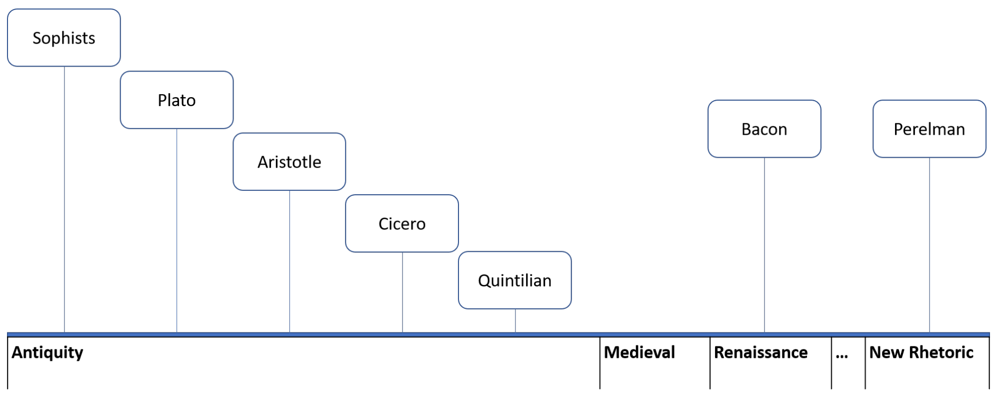

Figure 2: Timeline of rhetoric from antiquity to, medieval, renaissance, and new rhetoric, starting with the Sophists and ending with Perelman.

Plato defined rhetoric as a “rhetorical art” and the “power to persuade” (Gagarin, 2001) and taught rhetoric through a set of dialogues involving another philosopher - Socrates. In these dialogues Socrates makes several points on rhetoric and how to effectively persuade (Brownstein, 1965; Kraut, 2017). Firstly, a speaker must know about the topic at hand in order to communicate the truth effectively, contrary to the motivations of the Sophists. Secondly, a speech must follow the specific structure of the subject in order to persuade. Finally, Socrates stipulates that the speaker must know the audience that they are addressing, so as to effectively communicate with them.

Aristotle saw rhetoric as “an ability to see the means of persuasion” (Gagarin, 2001). Three elements were defined in a speech: the subject; the speaker; and the audience. From this, three modes of persuasion were also identified: logos, the use of logical reasoning to persuade an audience on a particular subject; ethos, the use of the speakers character and credibility to persuade the audience; and pathos, the use of emotion to persuade the audience (Aristotle, 1991). In each case a set of rules are defined, so as to determine how one should persuade and how all the means of persuasion should be utilised. In the case of logos, arguments must be structured in a particular way with any conclusion made preceded or proceeded by a premise (this is argument or reasoning in its simplest form). In the case of ethos a speaker must have a particular set of attributes, i.e., practical wisdom (practical experience), moral virtue (good moral values) and goodwill (alignment with the audience to ensure that the speech is understood). For pathos then the speaker must use the emotions of the audience based on previous experiences, perhaps also experienced by the audience, to persuade them. Each of the means of persuasion identified by Aristotle clearly spell out the already identified ways to persuade by Plato. Aristotle, however, defined them in such a way that they can be followed effectively as a guide to effective persuasion, something which is not so easy from Plato's account.

Roman rhetoric further utilised Aristotle's means of persuasion. Cicero re-purposed the means of persuasion using them in speeches when representing another in-front of a jury. Particularly pathos which could be used to "rouse emotion in an audience" (McCormack, 2014), a very effective strategy in the court. Cicero further specified rhetoric through five cannons: invention, refining an argument; arrangement, organising your argument; style, how to present the argument; memory, memorising the speech; and delivery, how the speech should be given (May, 2006). These cannons use a base of Aristotle's means of persuasion and then define how they should be used in a larger structure. Quintilian further specified rhetoric through a series of textbooks forming a curriculum which should be followed in order to become an effective orator. He specifically defined rhetoric as "the art of speaking well" although mainly redefined the notions of rhetoric that came before (Kennedy, 2008).

Following roman rhetoric, in the medieval era the art of rhetoric declined significantly. The need for sound public speaking was reduced and rhetoric was mainly utilised in letters and law (McKeon, 1942). Instead classical rhetoric was mainly taught due the prestige that it held in this era and was then re-purposed as a tool to address the public in court or public assembly, and became prevalent in public sermons (Purcell and Benson, 1996). In each of these use cases, the classical definitions of rhetoric were used, mainly that of Roman rhetoric due to the influence of the Roman empire.

In the Renaissance era, the classical forms of rhetoric, studied in Greek antiquity, were revived following its decline in the medieval era. Rather than being a subject of prestige, rhetoric was taught more generally and covered all areas of the humanities, including politics and history, as well as generally in science (Plett, 2004). This era drove the production of many handbooks on rhetoric in multiple languages with a focus upon the ancient language (Mack, 2011). Towards the end of this era, however, there was a shift away from the art of rhetoric to a focus mainly upon logic (Skalnik, 2002). This was driven by Peter Ramus, who considered rhetoric mainly as an art of style, when and how something should be said, rather than an art of the content of what is said (Skalnik, 2002). This move pushed for the considerations of logical reasoning only which tend to discount ethos and pathos as they are seen as more subjective. Francis Bacon, deemed rhetoric as a form of psychology as it can be seen as more personal and therefore viewed differently in the eyes of an individual. This view came about as it was clear that logic defined reasoning, but rhetoric was needed to communicate that reasoning (Miller, 1997).

After a long period of rhetoric being neglected the "New Rhetoric", by Perelman, re-ignited the definitions and explanations of Aristotle in Greek antiquity flanked by more modern areas of study, linguistics and psychology. Instead of being limited to specific areas of speech, rhetoric was instead applied more generally (Hochmuth, 1952). Perelman drove the use of Aristotelian rhetoric in argumentation encouraging rational reasoning, not necessarily based on logic (Perelman and Olbrechts-Tyteca, 1973):

> "*[New rhetoric] then, is more of a renewed rhetoric, aimed at demonstrating the great value that can be attained through reintroducing Aristotelian rhetoric and dialectic into humanist discussion in general and philosophical discussion in particular.*" (Frogel, 2005, p. 35)

Overall rhetoric has been defined in several ways in history, yet there are common trends through the applications of ethos, logos and pathos (named differently in some cases). The descriptions of the means to persuade within rhetoric show that they play a crucial role, and therefore Aristotle's definitions can be deemed as the most promising first steps for further investigation. Aristotle defined the means of persuasion, not only as something which should be used in a particular form of dialogue (like the court), but more generally for use in all discussion and topics. This is supported through the continued use and resurgence of the Aristotelian ideals throughout the history of rhetoric, in particular by Bacon and Perelman. The rest of this chapter then further explores ethos, taking Aristotle's description as the core definition.

2.1.2 The Nature of Ethos

In order to form meaningful conversation and debate, each participant in a conversation must have a level of character or credibility that allows them to speak and be trusted in what they say. Ethos is defined as the character of a speaker (Aristotle, 1991) and, as mentioned in the previous section, is one of three modes of persuasion conceived by Aristotle along with pathos, persuading an audience through their emotions, and logos, persuading an audience by the use of reasoning.

Taking Aristotle's definition, ethos, as considered in this thesis, is then developed only by the speaker during a speech:

> "*reinforced by Aristotle's conceiving of Ethos exclusively in terms of what is done within the speech (rather than as a matter of prior reputation).*" (Brinton, 1986)

This allows any analysis of ethos within dialogue to have no predefined state. That is, no predefined level of character or credibility is necessary when analysis of ethos occurs, allowing the ethos of a person to be defined at a speech or conversation level. In this case for a speaker to conduct a speech in the public domain, they need no prior reputation as long as they build ethos extensively throughout a speech. This same fact is present in (Braet, 1992) where two methods of eliciting ethos are highlighted either through the conclusions of the audience and or the second instance through self-defined ethos in-keeping with the environment in which Aristotle defined ethos. In ancient Greek culture, citizens brought in front of a judge were expected to represent themselves, unlike today where barristers are used. It was then important that a citizen could build their own character from within a speech, especially when the judge had no prior knowledge of this person other than the crime they were accused of committing. It was then extended further to any public forum and speaking, where it was important to build ones character in order to be perceived as credible.

In Roman rhetoric, a court format more similar to that of present day was established where a citizen could be represented by a defendant to speak on their behalf (May, 1988). This defendant was more likely to come from noble birth meaning a prior sense of their ethos played an important role. In public speaking at this time it was desirable to align oneself with the audience to build up the sense of ethos the audience holds for the speaker (Herrick, 2015).

The same elements of ethos that appear in ancient Greece and Rome can also be observed in contemporary rhetoric where ethos can be appealed to through works of an expert which are well known (Fahnestock and Secor, 2003). Two types of ethos are then described, intrinsic, meaning the sense of ethos gathered from what a speaker has said, and extrinsic, the speakers prior reputation or the ethos they bring to a debate or speech. Again these elements are highlighted in (Crowley and Hawhee, 2004), where modern examples of ethos are given such as a list of books of an author or the actors present in a film advertisement. Also covered is the importance of ethos in presidential elections, where all aspects of a candidate's life are put under scrutiny to ensure they have good character.

As Brinton reveals, modern rhetoric defines ethos in one common way which is the use of the informal fallacies ad vericundiam and ad hominem (Brinton, 1986) (see section 2.2 for a more in depth description). Each of these two methods of defining ethos tend to be based on Aristotle's view of rhetoric (Aristotle, 1991), but are defined as fallacies rather than effective moves of persuasion as Aristotle found them to be. In (Garver, 1994) this point is emphasised, whereby if sound, reasonable and logical advice is given, only reasons given to conclude a person is not sound are effective in persuasion. This case then shows the need to consider some fallacies, which use personal attack, as an effective but not rationally justified means to persuade or even to discount an argument.

In (Braet, 1992), the use of a personal attack on ethos is considered as “inevitable”. This is further explained through Aristotle, where judgement is not considered to be wholly rational and therefore cannot rely solely on logos. Instead logos only would be used in an ideal situation, but in real practice ethos and pathos are more heavily relied upon. In this case then a balance between all modes of persuasion is needed.

2.1.3 Ethos Elements: Wisdom, Virtue and Goodwill

Aristotle further defined ethos through three elements (ethos types): to display some form of ethos they must display practical wisdom, moral virtue and goodwill (Aristotle, 1991). These elements help to further specify ethos making it less ambiguous and more easily explainable. Aristotle defines these elements through negative examples:

> “*through lack of practical sense they do not form opinions rightly [practical wisdom], though forming opinions rightly they do not say what they think because of bad character [moral virtue], and they are prudent and fair-minded but lack good will, so that it is possible for people not to give the best advice although they know what it is*” (Aristotle, 1991)

In the first case this refers to practical wisdom which a speaker must possess in order to show that they have positive ethos. A speaker must have experience and must be knowledgeable about the topic of conversation otherwise their reasoning may be incorrect. The second case refers to moral virtue, character traits of a speaker which in this case mean that the speaker may not be honest because of bad character. Finally, the third case refers to goodwill where a speaker must be tell the truth to their audience, providing the best advice possible when they know it.

Fahnestock and Secor (2003) give more practical examples of ethos elements which provide further details of the concepts within each element. In the case of practical wisdom one must “command the material” of a speech or essay to show traits of knowledge. For moral virtue an open-minded and consistent nature must be shown. Finally for goodwill, the audience must be treated as an equal where the speech or text must be understandable to all. In (Murphy et al., 2013), wisdom and virtue alone are further outlined. In the case of wisdom a speaker must perform their function well, contribute effectively and have moral excellence, whilst for virtue a speaker must show they are just, courageous, selfless, noble and can exercise self-control. Goodwill is further described in (Garver, 1994), where a speaker must produce emotions in an audience by sharing pains and pleasures and at the same time align with the audience and their values.

2.2 Ethos in Argumentation Theory

Whilst ethos in rhetoric enables a speaker to persuade effectively, the ancient literature does not define ethos as a structure within language. In this case, by looking at the specified nature of ethos, argumentation can be utilised as the means to identify these ethotic structures within language, in some cases alongside logos and pathos.

As mentioned in Section 2.1, ethos has largely been considered as an accompaniment with logos and pathos and in argumentation theory has largely been considered as a by-product of an argument used as a premise in a structure of logos, i.e., in argumentation schemes (Walton et al., 2008) which define common structures of argument to be identified independent of the domain. In many of these argumentation schemes, such as an argument from position to know (defined in this section), no reference to the nature of ethos is specified, rather ethos is defined in a single premise to be used as part of the overall argument structure. Explicit references to conclusions of ethos are instead implicitly determined from the critical questions for these schemes. Walton does, however, identify a set of ethotic arguments that have a similar structure to that of ad hominem which will also be explored in this section.

2.2.1 Argumentation Schemes

Argumentation schemes are built upon a premise and conclusion structure, where the conclusion of the argument is supported by one or several premises. This conclusion and premise set are then tested through critical questions which can be utilised to determine the validity of the scheme. The critical questions in this case can be formed of explicit references to premises, or they may refer to implicit assumptions associated with the credibility and reliability of the speaker.

Critical questions "represent additional factors that might cause an argument to default" (Walton and Reed, 2003). By asking a critical question associated with an argument scheme it shifts the burden of proof to the proponent of the argument, as in order for the argument to be accepted as part of the scheme they must prove that the argument satisfies the critical question asked. Thus critical questions are utilised to determine the validity of premises within a scheme, yet the answers of those questions are not explicitly shown within the scheme itself.

Argumentation schemes are not a modern concept determined only very recently in history, their origins lie in Greek antiquity and have been re-defined, elaborated on, and extended throughout history (see figure 3 for a timeline of the origins of argumentation schemes). Argumentation schemes originate, as was the case for ethos, from Aristotle and the topics that he defined know as *topoi* (Walton et al., 2008). The *topoi* are a set of rules or circumstances that can be followed to create arguments for a particular topic or situation making use of a set of premises that have been agreed by the many as acceptable.

Continuing from Aristotle's topics, Cicero defined *loci*, a reduced set of twenty of the topics broken down into intrinsic and extrinsic. The intrinsic set of arguments are further split into two categories which Quintilian later defined as topics within and related to the discussion at hand. Extrinsic, on the other hand, only contained a single topic that of an argument from authority.

As was in the case in section 2.1, the medieval era continued to use the definitions of the topics defined in antiquity with Boethius re-organising Cicero's topics using two

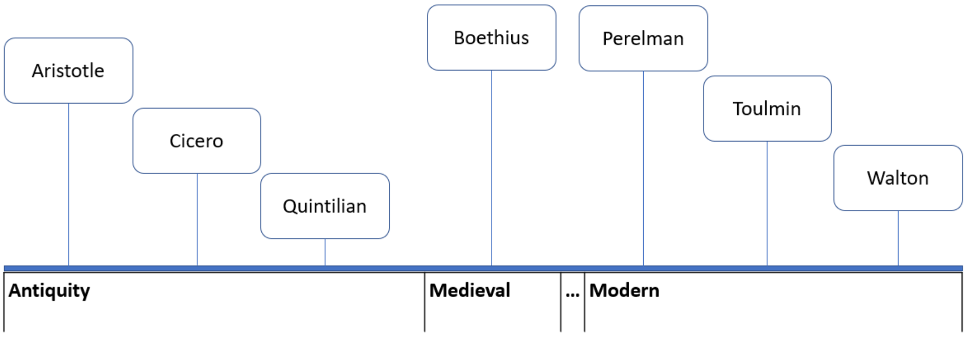

Figure 3: Timeline of the origins of argumentation schemes from antiquity to, the modern era.

distinctions: the distinction between the necessary and plausible; and, the distinction between the dialectical and rhetorical which combines two of Aristotle's distinct works.

In the modern era Perelman, defined a set of thirteen argumentation schemes which offered a different perspective to Aristotle's topics as they were seen as dependent on the culture and society of the time rather than a set of general arguments that can be utilised across topics. Perelman splits these schemes into two parts: association and dissociation, association covering the relation between arguments, logical arguments and arguments that show the structure of reality and disassociation as a class of its own which looks to demerit opposing arguments.

Following Perelman, Toulmin defined a means to apply argumentation through notation which allowed the graphical representation of arguments using schemes to show the relationship between conclusion and premise (Toulmin, 1958) (see figure 4). The model of argumentation developed by Toulmin clearly shows the claim or conclusion preceded by: a qualifier eluding to the fact that the claim may not always hold true; warrant, providing the implicit or general knowledge required for the argument; and the data or grounds which provides the premise for the conclusion. Additional backing can also be provided to the warrant or the overall argument can be rebutted through the qualifier in the case of the moment where the claim may not hold true. These together can be considered as an argumentation scheme.

Finally, and the main focus within this thesis, Walton et al. (2008) defined a set of

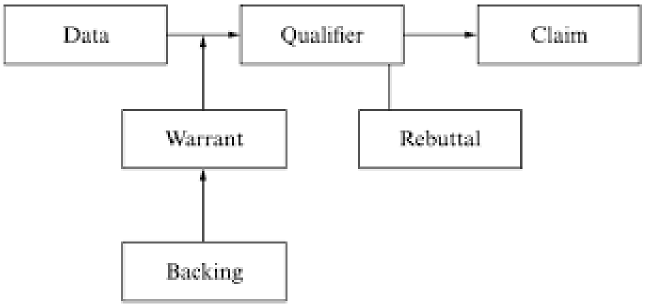

Figure 4: Toulmin's model of argumentation (Toulmin, 1958).

argumentation schemes that can be generally applied across domains and focus on many different ways to argue including those schemes which have a focus on authority and more closely ethos. These schemes are the most up-to-date and widely applied set within the argumentation community and have been shown to have applicability within computational argumentation (see section 3.3.1.6).

2.2.1.1 Argument from a Position To Know

One example of ethos in argumentation is an argument from a position to know (Walton et al., 2008) in which we reason that what is said is true from the fact that a person is in a position to know about a subject. This argument has the following scheme where it can be noted that there is no explicit conclusion of ethos, rather ethos is explored as a premise:

ARGUMENTATION SCHEME FOR AN ARGUMENT FROM A POSITION TO KNOW

a is in a position to know whether A is true or false.

a asserts that A is true (false).

Therefore, A may plausibly be taken to be true (false).

(CQ1) Is a in a position to know whether A is true or false?

(CQ2) Is a an honesty, trustworthy or reliable source?

(CQ3) Did a assert that A is true (false)?

Three critical questions are used for this argument that directly affect the ethos of the proposer of the argument (CQ1, CQ2), whilst one critical question (CQ3) concerns itself with the speech acts associated with the scheme. If any of the critical questions cannot be answered with assurance then the argument is not valid. This is true for all critical questions which cannot be answered in any argumentation scheme.

Explicitly in the structure of this scheme, no reference to ethos is made instead a premise is used to identify that a has ethos. References to ethos can be added as an external addition to these schemes, supporting that a is in a position know or that A is true. In this case ethos here can be defined as either positive or negative, this can be achieved by supporting a or attacking a. In turn then, through these supports or attacks, ethos can be utilised to validate or invalidate either of these premises. For example the premise "a is in a position to know whether A is true or false" can be supported by a's ethos, this in turn can be supported by an ethotic statement proclaiming why a has positive ethos. Thus an extended structure is created, "a asserts that A is true", "a has ethos" and "b says that a's ethos is positive". Here then the critical questions act as a prompt to explicitly define ethos within the structure of the scheme.

2.2.1.2 Argument From Expert Opinion

A further example of ethos in argumentation schemes is that of, an argument from expert opinion. Again, like in the argument from a position to know where the argument is valid because a is trustworthy, the fact that there is an expert is enough to accept the argument being made (Walton et al., 2008).

ARGUMENTATION SCHEME FOR ARGUMENT FROM EXPERT OPINION

Source E is an expert in subject domain S containing proposition A.

E asserts that proposition A (in domain S) is true (false).

Therefore, A may plausibly be taken to be true (false).

(CQ1) How credible is E as an expert source?

(CQ2) Is E an expert in the field that A is in?

(CQ3) What did E assert that implies A?

(CQ4) Is E personally reliable as a source?

(CQ5) Is A consistent with what other experts assert?

(CQ6) Is Es assertion based on evidence?

An expert can be defined as a person with considerable knowledge and credibility within a field of study ((Walton and Koszowy, 2014) also explores the extensions of this scheme to authority). This notion of credibility, is only defined within the critical questions associated with the scheme which look to identify a prior sense of ethos, however, the notion of ethos is not bound by this predefined credibility (as explained in 2.1).

Argumentation schemes, such as the argument from expert opinion, also focus on how ethos or credibility affect the logos side of persuasion and ignore more general ethotic manoeuvring. This is to say that an attack or support of ethos does not necessarily always have links with an overall argument structure. This is achieved through personal supports or attacks without a prior debate. Thus, ethos must first be defined as a separate entity from logos and then investigate the links between both logos and ethos as defined by Aristotle.

Furthermore, ethos in this case is implicitly added into a premise to support an overall conclusion that A is true. The scheme can be further expanded to show an authoritative figure and how this authority is used as a premise to support a conclusion. By addressing CQ1 and CQ4 the scheme can also be further expanded where, like in the argument from position to know, ethos can be utilised to support or attack the premise “Source E is an expert in subject domain S containing proposition A.”.

2.2.1.3 Ad Hominem

The ad hominem (AH) argument considers that certain types of argument are unacceptable as they do not utilise logical reasoning, and instead tend to use the character of the speaker to dismiss claims made. In this sense AH is used to attack the speaker which in some areas of research, like pragma-dialectics (see section 2.2.2) is considered inappropriate. Walton et al. (2008) considers several types of AH scheme, e.g. abusive AH, circumstantial AH, bias AH, and poisoning the well AH (Krabbe and Walton, 1993), which are largely considered as inappropriate to use within a debate. The critical questions within the argument from expert opinion and position to know, however, show that character attacks are appropriate in certain situations and therefore it is important to define an argumentation scheme to show when AH is valid (Budzynska and Reed, 2012). The generic AH scheme shows the main structure for AH schemes:

ARGUMENTATION SCHEME FOR GENERIC AH

a is a bad person.

Therefore, a's argument A should not be accepted.

(CQ1) Is the premise true (or well supported) that i is a bad person?

(CQ2) Is the allegation that i is a bad person relevant to judging i's argument α?

(CQ3) Is the conclusion of the argument α should be (absolutely) rejected even if other evidence to support α has been presented, or is the conclusion merely (the relative claim) that α should be assigned a reduced weight of credibility, relative to the total body of evidence available?

Each of the critical questions (CQ1, CQ2 and CQ3) are associated with an AH argument and are used to check the criteria of the argument. If an AH argument is found to violate any of these questions then it is considered to be a fallacious argument.

When looking at the AH argument scheme, an AH attack can be interpreted as an attack purely on logos, the reasoning in the argument rather than on ethos, on a person or an author of the argument. Also worth noting is that the attack in this case is instead modelled as inference for a non-truth premise instead of as an attack on the speaker directly. An ethos attack then is considered by redefining the AH scheme removing inference to be replaced by conflict. In addition to this, in order for a speaker to make an assertion they need to have some form of ethos so as there assertion is accepted (see again the critical questions in expert opinion and position to know).

2.2.2 The Structure of Ethos

As alluded to in section 2.2.1, ethos is prevalent in argumentation schemes yet is rarely defined explicitly as a conclusion in the overall argumentation scheme structure. Instead ethos tends to be expressed as a premise within the schemes which does not accurately explain the polarity of ethos which is seen in references to this premise of ethos. To better understand ethos and the role it plays within argumentation as a whole it must be represented as a conclusion which allows references of support and attack to this conclusion. Thus to use an ethotic support the theory on inferential argumentation schemes must be re-defined. In the case of ethos attacks this change, from ad hominem, is only in the polarity of the scheme moving from inference to conflict.

Ethos in argumentation theory tends to follow two main paths. That is treating ethos as a valid form of argument or treating ethos as a fallacy. In the latter case this is motivated by an ideal world where only logic and reasoning would be used to persuade. In reality, however, this ideal situation rarely occurs and therefore ethos can be legitimately used to persuade. This is highlighted in (Van Eemeren and Grootendorst, 1987) where the pragma-dialectical approach to argument is one of which logos is seen to be the only acceptable way to argue. This is disputed by Brinton.

> "*It has become the custom in modern philosophy to regard appeals to or*

attacks upon character in argument as logically irrelevant. This is reflected in the tendency to think of them in terms of the informal fallacies ad vericundiam and ad hominem. It was not always so." (Brinton, 1986)

An ethotic argument constitutes any argument which looks to add or remove credibility from the conclusions drawn (Brinton, 1986). This, can be achieved in three ways:

- Citation - referring to another persons work in text to persuade.
- Exemplar - referring to examples in history to persuade.
- Spectator - referring to spectators in order to create witnesses to persuade.

The use of citation, referring to the work of others is an effective way to persuade. In this case the ethos of another is effectively used as the method of persuasion assuming that this other person is considered as having a positive ethos. In exemplar historical references to people are made to show the way in which a person should act. These people, again assuming they are held in a positive light, effectively persuade an audience to follow in their footsteps. In the case of a spectator a person is used as an example to highlight how one should act in their presence.

In (Leff, 2009), a comparison is made between the ethotic argument present in rhetoric and the ad hominem argument in a dialectical approach. In the first case the dialectical approach considers the ideals of argument and defines how a person should argue. In the case of ad hominem this would then be considered as a case where a speaker should not use personal attacks. In rhetoric the opposite view is held, rather than restricting the means to persuade, ethos in this case can be used effectively.

In a second instance ad hominem in the dialectical approach is only considered within that particular set of utterances and not, as is the case in rhetoric, to the wider social impact of a personal attack. In the third case credibility and authority in the dialectical approach are used to assess the strength of an argument. Whilst in rhetoric, ethos is considered as a trait which can be built or damaged through speech. In the final case the dialectical

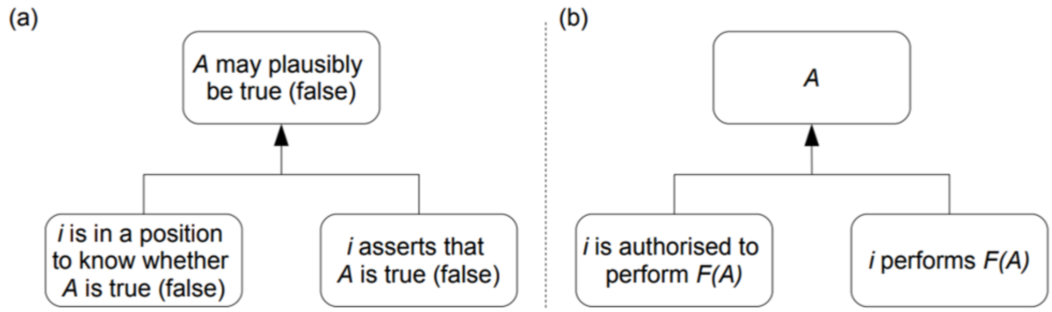

Figure 5: Ethotic structures with support relation in: (a) Argument from Position to Know (Walton et al., 2008); (b) a model with a speech act $F(A)$ (Budzynska, 2010).

approach is used to determine an outcome for a debate. In the case of ad hominem this means splitting each case into various types which then need to be assessed in different ways depending on a dialogue type. In the case of ethos in rhetoric, types can also be developed but instead to understand how a speaker argues rather than to assess.

As alluded to in section 2.2.1, each of the argumentation schemes treat ethos as a premise in a wider argument structure rather than as a conclusion which can be supported or attacked. Instead ethos must be pulled out separately to show the effect it has on these schemes.

2.2.2.1 Argument from Position to Know and Ethos

In section 2.2.1.1 the argument scheme from a position to know defined ethos within the premise of the scheme but did not show references can be made to this ethos through support or attack. Figure 5 shows the standard argumentation scheme representation and an adapted argument from position to know showing the addition of authorisation to perform a speech act relating to the conclusion. This addition, although specific to the argument from position to know and expert opinion, provides a step towards the ideal structure of having ethos as a conclusion which can be supported or attacked.

This new diagram shows that instead of the premise supporting the conclusion with this argument scheme it can instead be shown to have a deeper analysis. The argument from a position to know instead supports the ethos of a speaker on the right to perform a speech act. This extension can be bundled into the premise i is in a position to know where A is true but this does not show a concrete conclusion of ethos, which is a vital part of the scheme. Thus, the addition of the authorisation to perform a speech act is a step closer to allowing the support or attack of ethos desired for the research conducted in this thesis.

As explained in section 2.2.1.2 an argument from expert opinion has a similar structure to that of an argument from position to know. In this case then figure 5 can be adapted for an expert rather than position to know. This extension was defined in (Budzynska, 2010) where the argument from expert opinion was adapted to show an authority as a premise.

2.2.2.2 Ad Hominem and Ethos

In (Budzynska, 2010) the generic ad hominem scheme is further investigated. In this scheme the standard premise of "i is a bad person" is used to support a conclusion that "i's argument A should not be accepted". The idea that ad hominem is used to support rather than as an attack on ethos is counterproductive as demonstrated in section 2.2.1.3 and does not allow for attacks upon the ethos of speaker and as a consequence does not fit the goals of the research conducted in this thesis. The structure of ad hominem is further investigated in (Budzynska and Reed, 2012) in a rhetorical sense rather than a dialectical one. That is to say only the structure of the ethotic attack is investigated and not the outcome of using such an attack. In the case of ad hominem, (Budzynska and Reed, 2012) add an ethos proposition which supports the illocutionary connection of "asserting". This means that ethos allows a speaker to perform an assertion and therefore make an argument A. In the case of ad hominem the proposition "i is a bad person" thus attacks the ethos proposition. This then provides a more logical format of conflict for ad hominem or ethotic attack instead of inference (see figure 6).

This new construct of ad hominem takes into consideration the changes to an argument from position to know or expert opinion, that of authorisation to perform a speech act (see section 2.2.2.1). The ad hominem argument then is used to determine that this authorisation should not be given as the speaker is a bad person. This adaption of the standard scheme

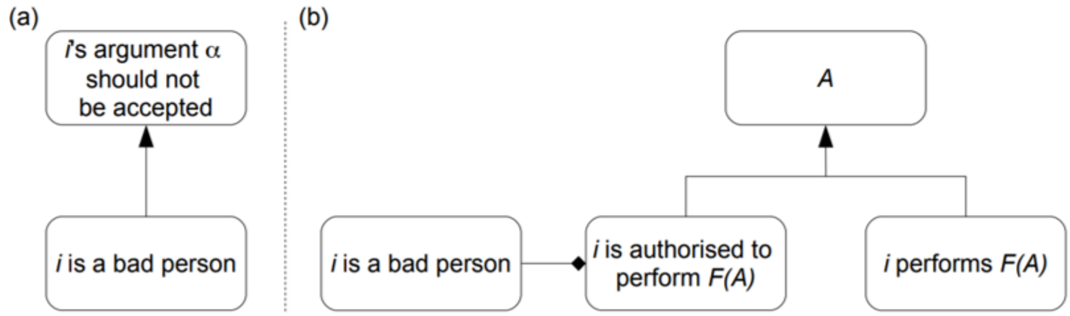

Figure 6: Ethotic structures with attack relation in: (a) Generic Ad Hominem, AH (Walton et al., 2008); (b) AH model with a speech act $F(A)$ and attack relation (graphically represented differently than inference relation) (Budzynska, 2010)

looks at ad hominem in two ways: a valid form of argument; and as an attack rather than inference. These changes again are a step closer to the attacks of ethos which are desired for the research in this thesis, however, the adaption is specific for both an argument from position to know and expert opinion. Thus, the structure defined must be generalised.

2.2.3 Types of Ad Hominem: Prudence and Morals

Ad Hominem can be further broken down into several ethotic argument scheme types (Walton, 1999). The types identified in this section directly relate to the elements of ethos, defined in section 2.1.3, when considering the basic definitions of these elements: practical wisdom referring to active judgement and decision making; virtue to the morals shown within an argument; and, goodwill to the audience. That is argumentation schemes have been created that overlap with the elements of ethos but have not been explicitly linked to those defined by Aristotle. Furthermore, whilst there are more explicit counterparts to wisdom and virtue in AH from prudence and AH from morals no such scheme exists for goodwill. The schemes here again show a relation of inference, as is seen in the standard AH scheme, which does not necessarily reflect the polarity of the point being made and thus must be adapted in order to allow attacks of ethos elements.

An ethotic argument from prudence can be associated with the Aristotelian idea of practical wisdom. In the scheme (see below) which can be interpreted as a lack of practical wisdom, a conclusion of a's argument should not be accepted is supported by the fact that they do not show good judgement. Here the "bad character" is again used as a premise within an argument, showing this as an attack on the authorisation to perform a speech act or as a further step as an attack on ethos in general.

NEGATIVE ETHOTIC AD HOMINEM ARGUMENT FROM PRUDENCE

a has a bad character for prudent judgement.

Therefore, a's argument α should not be accepted.

A negative ethotic ad hominem argument from morals relates to the element of ethos moral virtue. In this argument scheme (see below) a lack of moral standards in a person can be utilised as a premise for not accepting an argument. Again this shows the relation between "bad character" and the argument through inference rather than an attack on a conclusion of ethos.

NEGATIVE ETHOTIC AD HOMINEM ARGUMENT FROM MORALS

a has a bad character for personal moral standards.

Therefore, a's argument α should not be accepted.

Both argumentation schemes above only refer to the negative aspect of morals and prudence which in reality could be mentioned in a positive way. In this instance the positive counter-part of these schemes does not exist which is needed in the case of supports and attacks of ethos. Thus a different structure of ethos must be defined which allows this flexibility.

2.3 Ethos in Inference Anchoring Theory

To determine the structure of ethos required in this thesis, a theory is needed which allows for conclusions of ethos and supports and attacks of ethos to be specified. This theory must also allow the connection to a larger argument structure to ensure the maintainability and extension of this research. In this research this will be explored through Inference Anchoring Theory (IAT). Although other argument theories are available the advantage of IAT lies in the interaction between argument and dialogue an important element of this thesis. IAT provides the flexibility to add conclusions of ethos for individual speakers which can be directly attached to the logos side of argument, thus creating the ability to directly link logos and ethos showing how they interact, but also how these conclusions are directly linked to the dialogue.

IAT is used to represent inferential structures alongside dialogical structures (Budzynska and Reed, 2011; Visser et al., 2018a). This allows the structure of a conversation (dialogical structure) to be represented and the argumentation structure (inferential structure) for the same conversation to also be represented showing the relationship between them. This is particularly important when it comes to ethos, as any conclusion or premise relating to ethos then relates to the person who said it and the person who the conclusion or premise is about. The only way this can be shown is through the dialogical and propositional structures, and thus as eluded to above separate ethotic conclusions or propositions must be included to show this.

2.3.1 IAT Structure

The example below (example 4) gives a basic sample of how the dialogical and inferential structures interact. Within the text alone it is relatively clear what argument is being made "q therefore p". What is not clear, however, is how that argument came to be or the reason why a premise is needed. In this case the question "why p?" forces the premise. Example 5 is then a real world example from Hansard showing an argument.

(4) a. A: p

b. B: *why* p*?*

c. A: q

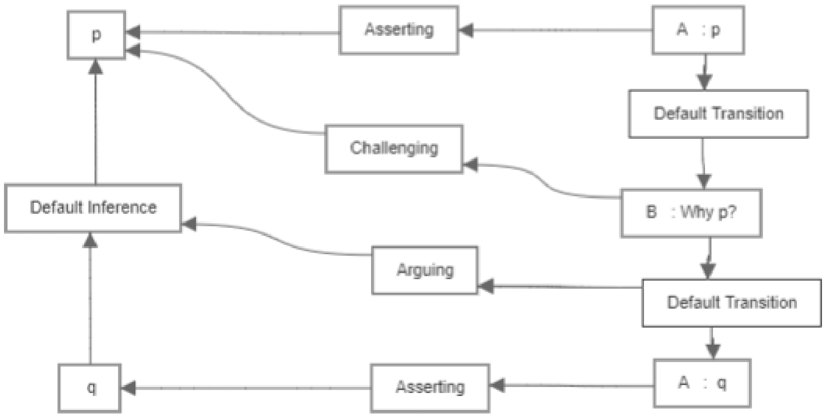

Figure 7: IAT diagram of an assertion with challenging from example 4.

(5) a. **Mr. Jessel** said, *The library must be kept together*

b. **Mr. Jessel** said, *because its 300,000 volumes comprise a resource that commands respect and admiration throughout the world.*

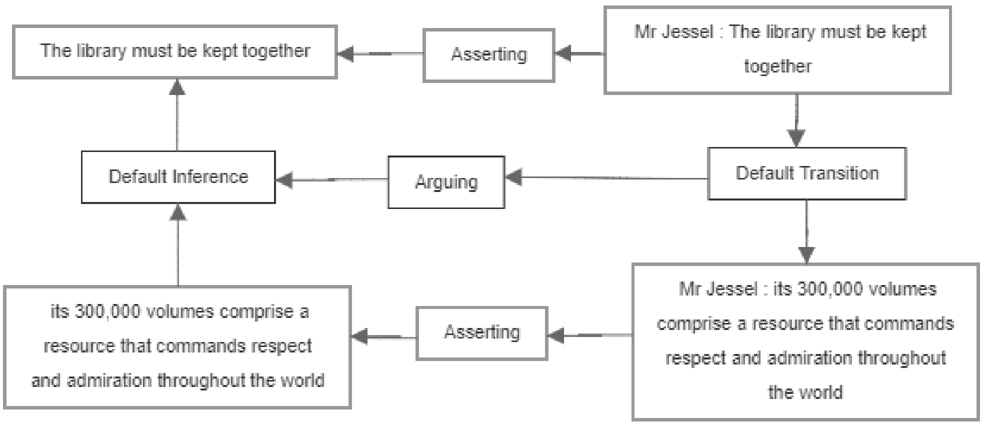

Figure 8: IAT diagram of an assertion in example 5 from Hansard.

In figure 7, on the left side of the diagram, there is a relationship between the propositions p and q which show the resolved content of the spoken text. On the right hand side, the dialogue of a conversation is represented through locutions with the names of participants labelled within the locution. Transitions between locutions show the steps the dialogue takes, i.e., a transition appears when there is a step from one part of the dialogue to another and these parts have a relation to one another due to the path the conversation takes. On the left hand side, the propositional content of locutions is represented along with the argument structure. Inference or conflict can be displayed between nodes of propositional content with inference showing that nodes of propositional content support each other, and conflict representing a conflict in views between the propositions. In example 4, between p and q there is inference as the statement q supports p.

In between the propositional content and dialogical actions of an IAT structure are the illocutionary connections which represent the connection between a locution and its propositional content through illocutionary forces (Austin, 1962; Searle, 1969) which show the communicative intentions of a speaker. In this example (example 4) “asserting” is present between the locution (4-a) and the propositional content p. As speaker A makes a point within the conversation there is an implied illocutionary force in which they are “asserting. In this case the illocutionary connection is anchored in the locution, however, illocutionary connections can also be anchored in transitions. This is clear for the illocutionary force, arguing which is the main illocutionary force used when inference is produced within the argument. The illocutionary connection is anchored in the transition as the argument is not implied from a locution but instead the transition between the previous locution and the present locution.

Figure 8 is a further example of IAT showing the illocutionary connections of “asserting” and “arguing”. In this instance the word “because” is used to indicate an inferential relation and as a consequence is not shown within the locutions of the structure and is instead represented by a default transition linking the conclusion and premise.

The advantage of IAT lies in its adaptability. That is to say IAT is flexible to any set of argument theories meaning that they can be plugged directly into the IAT structure. This is important for argumentation schemes where the set created in (Walton et al., 2008) can be applied in IAT and at the same time this set can be adapted and still represented.

In the context of this thesis IAT is an important tool that provides the flexibility to further explore ethos structures independently of logos and yet at the same time affords the opportunity to re-connect ethos to a larger argument structure.

2.3.2 Ethotic Structure in IAT

Ethotic structures present within IAT were first defined in (Budzynska, 2012) which is the main focus of this section. In this instance ethotic structures were created through the pretence of ethotic circularity.

Circularity by definition is when a premise is used to prove a conclusion and that same conclusion is used to prove a premise. Ethotic circularity can then be described in a similar way, i.e., when the credibility of a speaker is put into question and the speaker asserts that they are credible. The audience, to accept this, has to then accept prior to this that the speaker is credible, this then generates an ethotic cycle.

There are three main types of cycle in ethotic structures:

- Self-referential - when the speaker refers to their own credibility.
- Embedded testimony - when one speaker evokes an others testimony to confirm their own credibility.
- Ethotic begging the question - when a testimony is supported by another speakers credibility and that credibility is supported by the first speakers credibility.

An example of self-referential circularity (figure 9) is as follows, the speaker asserts that they are credible which creates an illocutionary connection of "asserting" with the propositional content of the speaker being credible. As a consequence of the illocutionary connection, credibility is already present from this force and therefore a cycle is created. In this instance the ethos of a speaker is used to support their right to perform a speech act.

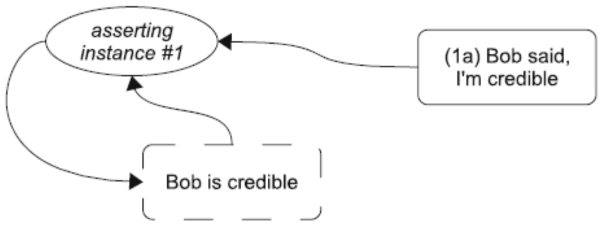

Figure 9: IAT diagram of self-referential circularity (Budzynska, 2012).

A similar example can be shown for embedded testimony (figure 10). Embedded testimony is shown when a speaker asserts that a second speaker says they are credible. This first implies that the speaker is credible as an audience has to trust that what they say, another speaker has said, is true. Secondly if the audience believes this then they also have to assume that the second speaker is credible as the first speaker implies this with their assertion, thus again completing a cycle.

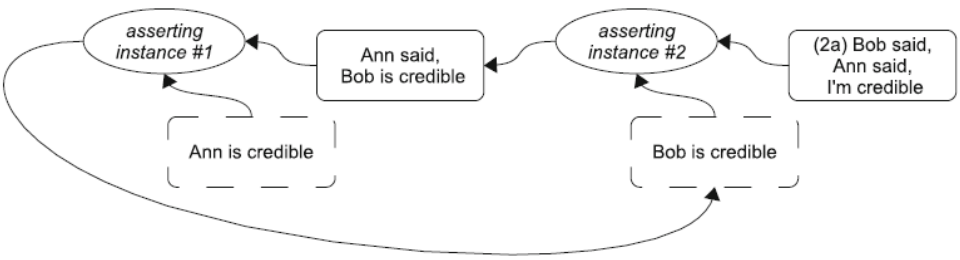

Figure 10: IAT diagram of embedded testimony circularity (Budzynska, 2012).

Each of the ethotic cycles above present the need to model ethos within IAT showing that ethos is a product of an assertion and can be modelled as such. The examples above, however, are specific to a larger argument structure assuming that ethos is only present within this structure. To model supports and attacks of ethos this structure again has to be adapted slightly as supports or attacks upon ethos do not necessarily have to be towards a speakers assertion. A speakers ethos may be attacked without first having performed any kind of speech act or before putting forward any kind of argument. Thus attacks or supports of ethos are not necessarily reactionary.

2.3.3 Ethos Support and Attack

In this thesis the aim of ethos mining is to identify supports and attacks of ethos. Thus, only a fragment of an IAT structure with ethos will be considered such as the structures shown in figures 11 and 12. As is the case when modelling arguments a locution with the speaker is on the right side of the structure, an illocutionary connection is then made to a propositional content. After this support and attack relations are used (in this case default inference represents support and default conflict, attack) to connect the proposition to a conclusion that a speaker has ethos. As alluded to in section 2.3.2 the conclusion of a speaker having ethos has been pulled away from a larger argument structure meaning that this does not have to be connected to an illocutionary connection.

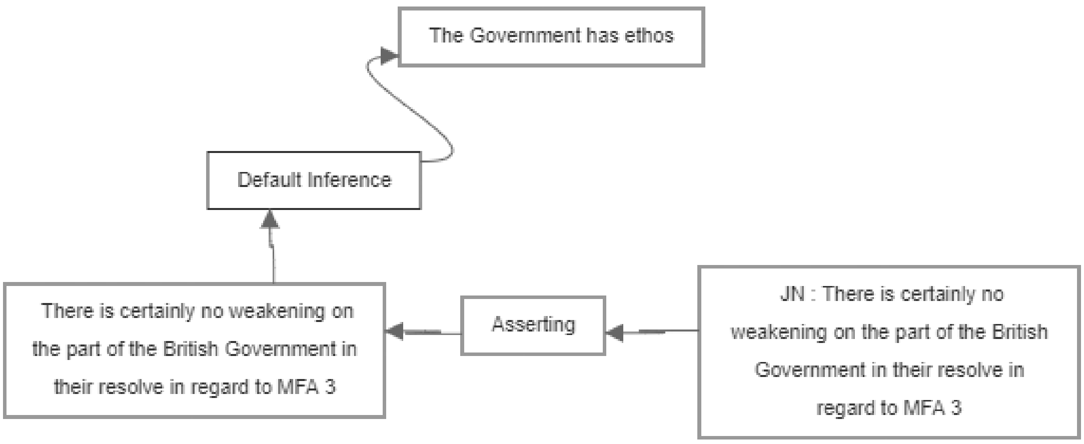

Figure 11: IAT diagram of an ethotic support taken from Hansard the UK parliamentary debate record (see section 4 for a description of Hansard).

In each of these diagrams no assumption is made about the overall ethos of the speaker as this, in some cases, could be subjective and alone is too little information to make any claims about the overall ethos. In each of these cases the polarity also reflects the feeling towards the entity. A positive ethotic statement is labelled with support and a negative ethotic statement with an attack, rather than the case of argumentation schemes where the ethos relations are bundled in with logos on inference.

Figures 11 and 12 then provide the needed constructs for ethos supports and attacks

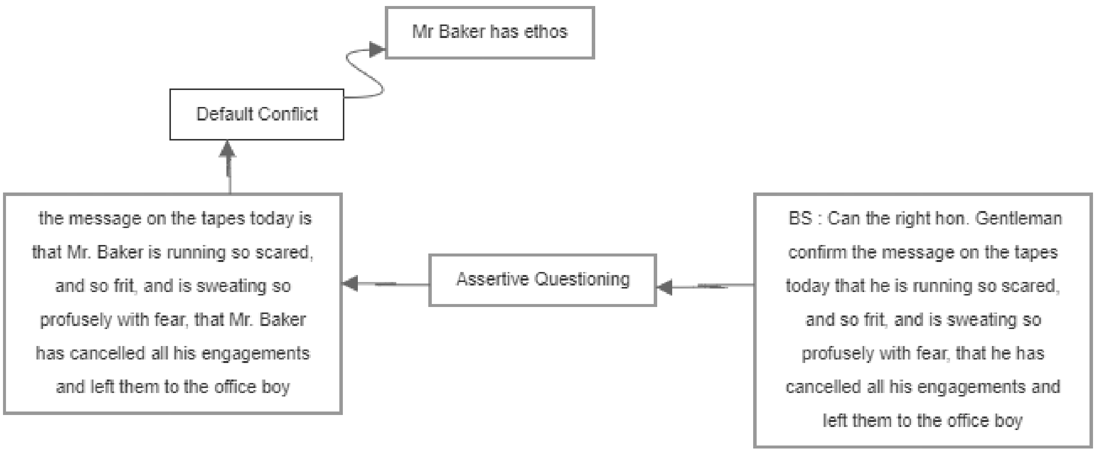

Figure 12: IAT diagram of an ethotic attack taken from Hansard the UK parliamentary debate record (see section 4 for a description of Hansard).

where a conclusion of ethos is specified and can be supported or attacked through the propositional content of a locution. The crucial aspects in this research then are the declaration of ethos and the locution presenting the original text fragment that gives the opinion towards this ethos. The use of IAT means that the proposed structures are then extendable for a wider argument annotation which may use this conclusion of ethos as evidence to support an argument.

2.4 Discussion

Overall, ethos in rhetoric can be defined as a speaker building their own character through speech. This notion is then extended to a prior sense of ethos which comes from a reputation through scholarly work or through appearing to an audience to have a positive (or negative) ethos. The notion of ethos can then be extended to elements of ethos, wisdom, virtue and goodwill, which are all necessary to have in order to have ethos. In the case of ethotic argument, this is most commonly used to attack the ethos of another speaker. Much of argumentation theory defines these cases as a fallacy and yet when used to support an argument they are considered as a premise within argumentation schemes. In this thesis ethos is considered in parliamentary debate, where a prior sense of ethos for each politician is present. This means that rather than building one's own ethos the objective is to build or remove the ethos of others.

This chapter also highlights the structures of ethos within argumentation theory. In the first instance two main approaches can be taken to ethos in argumentation. A rhetorical approach where ethos is investigated to determine how it can be effectively utilised to persuade and a dialectical approach where the strength of an argument is evaluated considering ethos attacks as fallacies. In argumentation schemes ethos is only considered as a premise within a wider argument structure and not as a standalone conclusion which can be supported or attacked. Extensions to these schemes were then built to show authoritative figures and how they are effectively used as premises. Ad hominem fallacies are also described, particularly how they can be used as effective means of persuasion or defended against. Ad hominem as an argumentation scheme, however, is given as inference rather than modelled as an attack and was thus adapted to allow such manoeuvring.

The research in this section highlights an area not investigated so far within argumentation theory, that of ethos supports and attacks, treating ethos as a conclusion in the larger structure. As highlighted in section 2.1, a comprehensive study of solely ethos must be undertaken to understand its prevalence. The polarity of ethos should also be apparent within any research on ethos and therefore ethos must be examined as a conclusion alone within argumentation. This is clear in the argumentation schemes for ad hominem where the "bad character" of a speaker has the negative ethotic sentiment in the premise and then this premise is used to determine that an argument in not valid through inference. Instead, and in this thesis, ethos is pulled away from this premise to show the relation and reason why a speaker may have "bad character" but the ethos in this case does not come with a conclusion that the argument is not valid.

Ethotic supports and attacks are considered as the counterparts of argumentation schemes and ad hominem. In the case of an ethotic support, the positive argumentation schemes do not directly reflect an ethotic support and thus are re-defined with extensions of ethos. These extensions, relating to the argument schemes from position to know and expert opinion, are then too specific for the goals of this thesis and as such need to be expanded through the use of IAT. In the case of ethotic attack, ad hominem is more directly applicable and yet must also be re-defined to accurately show the attacking nature rather than through inference as is the current state in argumentation schemes.

Chapter 3

Literature Review

Mining ethos in parliamentary debate requires the creation of novel techniques for the extraction of ethos which span several areas of active research (see figure 13). The definition of ethos, that of the character of the speaker, can bring associations to trust in computer science, however, this area of research has focussed upon multi-agent systems where these systems use trust for decision making which is not directly comparable to ethos (see section 3.1).

Closer to the research conducted in this thesis, specifically the application of ethos mining in ethos analytics, is the use of ethos in political science which applies both manual and automatic techniques to evaluate wider public and political opinion to gain an understanding of the political events of the time. Whilst this research does not explicitly use ethos, the applications of ethos to this area are clear, especially in the ability to gauge public opinion of political figures (see section 3.2).

Finally, the section on ethos in text mining explores the areas of computer science research that are most closely related to this study, in particular the areas of sentiment analysis and argument mining. In the former this is due to the polarity of ethos which when automatically classified will need sentiment analysis. In order to perform this automatic polarity detection, domain specific data-sets will need to be built and proven methods of classification explored. Argument mining can be intuitively seen as related to ethos mining with argument mining focussing on the logos side of Aristotle's modes of persuasion. Thus the research in this area may provide techniques which can be directly applied or extended for the purpose of ethos mining. The combination of both argument mining and sentiment analysis is used specifically to study expert opinion, persuasion on social platforms and through reputation, which all provide the research most closely related to the area of study in this thesis (see section 3.3).

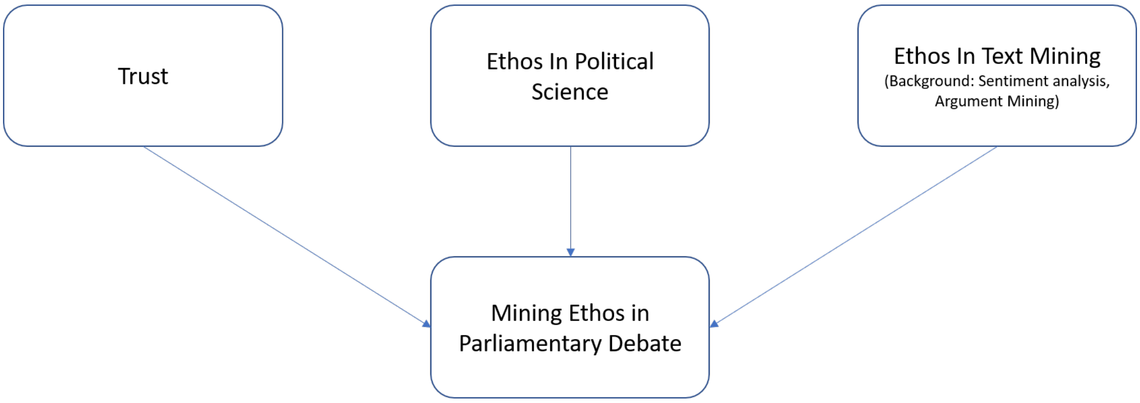

Figure 13: Roadmap of the work related to ethos mining with trust, ethos in political science, and ethos in text mining.

3.1 Trust and Ethos

Trust is an integral part of any process that requires a decision to be made which relies upon another entity. This is particularly evident in multi-agent systems (MAS), any form of computational transaction or, at the most basic level, trust placed in people. Despite the wide spread applications of trust within computer science, there is not a uniform definition of trust used by all of the systems developed, instead there are multiple definitions that encompass different aspects of trust.

3.1.1 The Concept of Trust

Much of the literature engaging in trust specify it's definition in a slightly different way. In this section several of the most influential papers have been highlighted that look to review trust generally each time specifying their own definition of what trust encompasses. This means that comparing ethos and trust is generally difficult with some definitions being closer to ethos than others.

Grandison and Sloman (2000) define trust as "the firm belief in the competence of an entity to act dependably, securely and reliably in a specific context". Trust in this setting is considered as a relationship where a trustor trusts a trustee to perform a specific action in a specific context, although this is not a bi-directional relationship in that an entity may trust another but not vice versa.

In (Mui et al., 2002), trust is a subjective behaviour only which considers the previous encounters between entities and is based upon an expectation about future behaviour. The previous encounters of an entity are defined as reputation which is based upon the social network of the entity which is evaluating trust.

In (Sabater and Sierra, 2005), trust and reputation are said to come from a cognitive perception where trust is made up of "underlying beliefs", whilst from a game-theoretical perspective trust is a subjective property where an individual expects another to perform an action they are dependent upon making trust a relationship of dependency between two entities. Sabater and Sierra also identify several information sources for which knowledge of who to trust can be gained. Trust is established through the direct experiences of entities, whether this is through a direct interaction with another entity or through an observed interaction, for example, in a multi-party dialogue. Witness information considers the word-of-mouth experience of an entity. Instead of being party to an interaction, an entity can gain information through the indirect story of interaction, however, this raises more questions of trust in the witness. Sociological information is also considered in a trust decision. That is, the social relation between entities which considers any current positions of trust they hold, such as high profile important roles. Finally, Sabater and Sierra define the visibility of trust in an individual as two kinds of property: global and subjective. Trust as a global property considers past opinions in interactions between entities that are discoverable in publicly available data, e.g. reviews. Trust as a subjective property is an individual trust decision based upon personal history between entities, known trustworthy relations and witness information.

Artz and Gil (2007) consider trust in four major research areas: policy based trust, reputation based trust, general models of trust, and information sources of trust, whilst defining trust in the same way as Grandison and Sloman (2000) and Mui et al. (2002). Policy based trust is the credentials or access properties of an entity using a third party to determine trust much like witness information in (Sabater and Sierra, 2005), but with more of a focus on transactional web based trust. Reputation based trust is defined as the past performance and past behaviour of entity from both first hand experiences (subjective trust in (Sabater and Sierra, 2005)) or based upon others experience (global trust in (Sabater and Sierra, 2005)), which can be evaluated through social networks. General models of trust take into account the factors that play a role in a trust decision such as beliefs, risks, importance and utility. Finally, information sources for trust consider the provenance of information such as the relation between people and information to support a trust decision, the propagation of trust through links and reviews of information or data to determine trust.

In (Castelfranchi and Falcone, 2010), trust is defined as a mental attitude towards another, a decision to rely upon another or as a behaviour. In (Falcone and Castelfranchi, 2004), they also emphasise that reputation (the past experience of an interaction with an entity) or trust is not useful without context. For example, the interactions of a user on an auction website is not useful in determining if they are a trustworthy accountant. Castelfranchi and Falcone (2000) also recognise that trust should be reason based, i.e., determined from well motivated evidence and good inference from credible sources of information. This is supported in (Paglieri and Castelfranchi, 2014) and in (Parsons et al., 2014) with the latter extending this idea to consider that argumentation (constructing arguments or reasons to enhance belief) is an appropriate avenue for this reasoning.

Finally in (Golbeck et al., 2008), trust is defined as a concept that helps entities to make a subjective decision about conduct, services or people only when there is uncertainty or a belief and willingness to take some action based on that belief (Golbeck, 2008).

Golbeck also recognises trust in context through information sources, but adds trust in web services, such as peer-to-peer systems or banking transactions, and trust in people taking into account feedback through websites or through friend of a friend (FOAF) structures in social networks (this can be considered as witness information in (Sabater and Sierra, 2005)) where a decision to trust an entity can be obtained through a friend's decision. There are, however, problems with methods that rely upon the decisions made by others. That is, a FOAF system can be susceptible to "gaming". One party can use these systems for their own gain, creating a false sense of trust in another so as to deceive the system. In these cases ethos can play a role, particularly when extracted from natural language text.

Each of the works above show that ethos, as defined in rhetoric and argumentation and adapted in this thesis, can then be seen as related to trust in that knowing the ethos of a person or system can aid in making the decision on whether to trust someone or not. While a general sense of a persons ethos can be used to make the decision of trust it is advisable to use the context of that particular situation. Trust then is considered as a wider concept which requires multiple inputs for implementation.

3.1.2 Implementation of Trust

Outlined in this section are some of the most well known works on computational models of trust and those that are the closest in relation to the research undertaken in this thesis. In general, trust in computation and AI has focused upon multi-agent systems, secure and trustworthy transactions as well as trust between agents and to a greater extent people in a social network situation. This is particularly evident in the semantic web with the principle being that any person or agent can provide information in a uniform way which allows automatic scraping of information and connection of multiple sources (Berners-Lee and Hendler, 2001).

In (Richardson et al., 2003), it was identified that although there is a uniform basis for information in the semantic web, it is not clear which information sources should be trusted. In order to discover trust, users of the semantic web would be asked to identify a small number of others that they trust which then would allow for the propagation of trust through a social network of users. This propagation can then be used to identify trust for a new information source. The degree of trust in an information source, combined with the source's belief in a statement, can then be used to determine an overall belief value in a statement.

In (Guha et al., 2004), a framework to propagate trust and distrust is developed containing four parts: propagation of distrust, iterative propagation, rounding, and atomic propagation. *Propagation of distrust* forms three parts in the framework: the propagation of trust scores only which ignores distrust; one step distrust which only considers individual distrust of a source discounting everything from this source, but not propagating this distrust; and propagated distrust where both trust and distrust scores are propagated. *Iterative propagation* involves inferring a final trust score using eigenvalue propagation and weighted linear combinations. *Rounding* looks to interpret the final trust scores from the iterative propagation as either trust or distrust achieved through global, local and majority rounding. Global rounding takes the trust or distrust decisions of all agents into account to determine a fraction based threshold of trust and distrust which are then met by creating a ratio between the trust and distrust for a particular agent against all others. Local rounding instead involves thresholds determined only by the interactions of a single agent or entity which is then compared against the trust fraction of the pair of agents or entities. Majority rounding takes into account the trust decision made by all other agents around the agent making the trust decision where the majority trust or distrust decision is then accepted. Finally, *atomic propagation* consists of direct propagation where trust is inferred (if i trusts j and j trusts k then i should trust k), co-citation where agents with similar trust feelings are determined to have the same trust for a new agent (i trusts j, k trusts j and k trusts l because i and k both trust j, i and k should also both trust l).

In (Huynh et al., 2006), a trust model incorporates interaction based trust (direct interactions between agents), role based trust (which takes into account the roles of each other), witness reputation and certified reputation (third party reputation references).

Witness and certified reputation both are susceptible to malicious or wrong data provided by a third party therefore the framework makes two assumptions that agents want to share the information they hold and that agents are honest. This makes for an easier implementation of the trust framework, but as more recent research has shown (see below) malicious data or agent identification is a problem.

An attempt to combat malicious agents is made in (Burnett et al., 2011) where delegation controls are used to mitigate any risk in an initial agent interaction. An *explicit incentive control* is used to agree compensation (agreement to utilise that agent more) if the agent performs to a particular standard. *Monitoring* is then used to check on the behavioural choices of an agent which can then allow for a trust decision to be made. *Reputation incentives* are then used to provide positive or negative feedback to an agent community where a trustworthy service by an agent leads to good feedback for that agent being distributed amongst the network. The final goal is to then delegate tasks to an agent while mitigating the initial interaction risk through five methods. *Simple delegation* uses an initial trust evaluation of an agent and trusts the agent to do the task delegated. *Delegation with monitoring* sets the effort that an agent should adhere to and then monitors the agent to observe the choices that agent has made based upon effort. *Delegation without monitoring* again sets the effort level, but does not monitor the decisions made by the agent. *Delegation with reputational incentive* involves setting different effort levels for an agent where it will gain or lose reputation depending on the effort level it chooses. Finally, *abstaining from delegation* occurs when there are no agents where the benefits of delegating a task outperform the risks.

In (Teacy et al., 2012), a hierarchical and Bayesian inferred trust (HABIT) model was created. The model consists of two tiers where opinions and reputation sources are modelled in the bottom layer and the correlation between the opinions and actual behaviour is modelled in the top layer. This is applied with individual agent behaviours modelled in the bottom layer and the connections between the behaviours of agents in the top layer. HABIT focuses on the need for no common ground, a normal requirement for trust models.

That is, in most systems communicating agents have the same goal, are performing the same tasks under the same parameters or can only share trust information on parameters that they have in common. In HABIT this is not compulsory and therefore any trust information can be shared. HABIT also attempts to extract information from sources which are known to be misleading. Once an agent is determined as being misleading the information it provides is discounted and the opposite of this information is said to be true. Finally, predictions about the behaviour of an agent can be made through statistics. Correlations between agents are made through the Bayesian analysis of known agents and beliefs which can allow for the prediction of a behaviour especially for agents with no prior interactions.

In (Hunter and Booth, 2015), trust is used to determine what new information provided by a source should be integrated into a set of already known beliefs. A trust decision is made about a source as a precursor to any addition of the information into the set of beliefs so that only genuine, trusted information is passed on and not malicious or wrong information. To do this, a set of domain specific rules are developed to determine which agents have trust in which domains. An agent is given a trust state for a particular domain, so that a decision can be made.

In (Xiang et al., 2017), a mixture of Gaussian processes model was created to asses the reliability of crowd sourced data. Crowd sourced data can be manipulated by faulty sensors in mobile devices (noisy data and outliers in data), inappropriate measurements by participants or dishonest workers. The framework updates the trustworthiness value of sensors as well as of workers after each task. Dishonest workers are then less likely to be selected for crowd sourcing tasks.

In this section the creation of networks and trust relationships are similar to the goals of this thesis in extracting ethos supports and attacks to build networks. The difference though lies in the ultimate goal of these works where trust networks look to infer relations of trust quantified through completion of tasks or verification, whereas, ethos mining looks to build these networks from scratch using an already known phenomenon of ethos as the base. In this case, and as described in the previous section, ethos mining is a step before a full trust network although some works in trust have attempted similar classifications.

3.1.3 Trust Extraction using NLP

Although the focus of computational trust has been mainly in the areas described above, some works focus on automatic trust extraction through NLP which is more closely related to the research in this thesis. In (O'Donovan et al., 2007), trust and distrust classifications were made from free text feedback comments in Ebay using a set of predefined nouns and domain specific features. Each comment was stemmed and then classified for binary sentiment polarity, positive denoting trust and negative denoting distrust. Evaluation of a set of machine learning algorithms was conducted against a manually generated dataset where participants were asked to rate comments on a likert scale and the average rating was used to determine the binary classification. The AuctionRules classifier outperformed all others with an accuracy of 97.5%. This work, however, did not take into account any neutral comments made which were instead classified as trust or distrust and no evaluation of agreement between participants in this task was conducted due to the averaging of ratings.

In (Rubin, 2009), an initial study of linguistic features for automatically extracting trust, distrust and trust rhetoric is conducted. In this framework trust can be extracted by looking for statements of praise, recommendation and acting upon advice. Distrust is recognised by identifying WordNet synsets of known distrust words, anger or blame. Trust rhetoric can be extracted by qualifying appeals to trust, promises or loyalty pledges.

In (Agarwal and Zhou, 2014), a trust model was created to identify malicious tweets. Known malicious Twitter users were utilised to develop the trust network where tweets, users and topics are nodes in a directed graph which in turn allows for the back propagation of trust scores. A dataset from Twitter containing 10,000 users, 20,000 tweets and 3,000 topics was created by asking 12 annotators to label each entity (in this case entity refers to tweets, users and topics) as either malicious (given a score of 0), borderline (0.5) or legitimate (1). The overall score for a label was then given by averaging the scores from all annotators to decide a final label. Due to this averaging no annotator agreement was provided which could remove unreliable annotators. The trust network was then used to classify each entity which gave an overall F1-score of 0.958, although no comparison was made with a baseline.

In (Ceolin and Potenza, 2017), graph centrality is used to determine the trust of users in the social web. All users are connected via their interactions where users with a large centrality score are determined to be more likely to trust other users due to the large amount of connections they have. The approach was then evaluated by investigating the trust in the social media application and trust in other users computed from user activity achieving an accuracy between 43% and 99%.

In each of the works described above there is no direct comparison to ethos mining, instead there has been a focus on what could be considered as statements about the trustworthiness of a source of information rather than explicitly about their character. In essence the extraction of trust is closer to inferring the provenance of a data source than it is to ethos mining.

3.1.4 Provenance and Trust Management

Provenance, studying the trustworthiness or reliability of data (Artz and Gil, 2007), and trust management, managing the already known trust relations (Richardson et al., 2003), relate to further applications of trust. When applying trust to any system, there must be a consideration of the provenance of the data source and the provenance of the information the data source provides. Once this information is gained there needs to be a way of managing this with other known trust relations, for which provenance can be used for this knowledge.

In both trust management and provenance, there is no direct comparison to the work undertaken in this thesis, or for that matter with the work of ethos or ethos mining in general. To reiterate, ethos is a step in the direction for the ability to make a trust decision.

In that, the information garnered from ethos mining (relations between entities with a value relating to a single support or attack of the ethos of an entity) can aid in the trust decision process, but it is not the only piece of information that should be considered. For example, knowing that a particular politician does or does not have an adequate knowledge to rebut statements on the economy, does not mean that they should or should not be trusted in regard to their ability to rebut a statement on agriculture (unless of course the lack of adequate knowledge is down to their ability as a politician in general). In relation to provenance, ethos can be utilised as a component of the provenance of a source or the provenance of information. In relation to trust management, ethos can be managed for multiple sources either through provenance or by ignoring provenance and applying the ethos data through to other entities through inference. Therefore, provenance and trust management can be seen as a next step in the applications after ethos mining has been conducted.

3.2 Ethos in Political Science

Much of the political science literature has relied upon the manual analysis of large amounts of data to determine various hypotheses which are then validated through the use of statistics. Hypotheses range from the types of interaction there are between politicians, whether attacks or supports of policy are more likely, the strategies of the media within elections or the prediction of election outcomes. Ethos can then become the subject of political science applications for instance: analysing ethotic supports and attacks between politicians; analysing ethotic supports and attacks on political entities from the media; or, using ethotic supports and attacks for the purposes of election prediction (an assumption here would be that politicians alter their behaviour based on the pressure they receive from members of the public in their constituency and therefore an election would validate this). This analysis aligns with the fourth research question of this thesis, that being to design a set of ethos analytics based on the output of ethos supports and ethos attacks between politicians from ethos mining. Due to this motivation, the tools and applications of computer science methods must be explored alongside the particular political data analysed in political science.

3.2.1 Tools for Political Science

Rather than focusing on political analysis specifically, much of the computational community instead develop methods to aid in political analysis of large amounts of data. Some of these approaches focus on topic modelling, where documents are automatically scraped or extracted and clustered into topic areas to ease analysis (Hillard et al., 2008; Karan et al., 2016). Other approaches instead focus on the classification of various political aspects such as political stance and party affiliation for social media users, or the opinions in tweets with a focus on politics (Maynard and Funk, 2011; Sarmento et al., 2009; Yu et al., 2008; Zhou et al., 2011) (see also section 3.3.1). Whilst these methods of extraction do not make any suggestions of election outcomes or generalisations about politics they provide interesting data that can be utilised in political science studies. For example, knowing the political stance of a social media user can in turn make an election outcome more predictable assuming they are eligible to vote.

In (Grimmer and Stewart, 2013), a survey was conducted of the various computational methods at the disposal of political scientists to aid in textual analysis. As described above these methods range from topic modelling to classification as well as word score calculations which determine where in the political sphere a word sits. The need for the validation of automatic methods is highlighted whether that is focused on manual annotation within supervised learning methods or reviewing data in the case of unsupervised methods. Grimmer and Stewart also stress the need for new methods of textual analysis, either supervised or unsupervised, to identify commonalities in large amounts of data which can aid political scientists and at the same time emphasise the pitfalls of these methods in that they cannot be generally applied without some thought as to what the extracted data may achieve. In the latter step either hypotheses must be constructed in order to determine what to look for in the data, or more advanced computational methods must be used which automatically determine trends and explain them. For example, automatically determining that a particular sentiment towards a known political party is statistically significant could aid in predicting the outcomes for the party at an election.

In the case of ethos mining, analytics can be built using the assumption that a support or an attack on ethos holds a degree of weight with the positions of a person in a political party. For example, a politician that comes across as strong and agreeable to all sides on a political debate may be more likely to achieve higher office. Hence, analytics looking at specific roles within parliament can be created.

3.2.2 Media as a source of interaction

Much of the focus of the political science field has been on social media. This is due to the large number of people who use and express political opinions on different web platforms whether that is politicians, the media or the general public. Many of these social media analyses still heavily rely upon manual annotation and analysis of subsets of the data before applying a statistical analysis.

In (Nooy and Kleinnijenhuis, 2013), analysis of negative campaigning in the media by political actors (politicians or parties) was conducted. The analysis took place three months prior to the Dutch elections of 2006, where supports and attacks of political actors were identified to accept or reject fourteen hypotheses. The manual analysis contained 4,280 statements from media sources which encompass 160 political actors. Due to the large volume and the need for manual analysis this was limited to 27 political actors leaving 1114 statements to be analysed and of this attacks and supports of people or policy were also limited to one per day. Of the 14 hypotheses 9 were identified as significant. It was determined that: the media are more likely to report attacks on political actors than supports, using direct quotes; ideological difference between parties encourage attacks; members of the same political party are more likely to support one another; the media are more likely to attack large parties; members of large parties are more likely to be attacked;

the media are more likely to attack incumbent parties; previous supports or attacks of an another actor increase the likelihood of more of the same; if support or attack increase polarisation then they are likely to be used more; and attacks are more likely to be used. The rejected hypotheses also allowed other conclusions to be formed such as, attacks on party leaders are more likely, this can be attributed to the fact they are a figurehead for the party. The effect of agreement between political actors is stronger than that of polarisation perhaps due to the rarity of such cases of agreement. Party size did not make a party more or less likely to attack or support others meaning that small parties still engage in negative campaigning. Finally, incumbent parties are determined to not be more or less likely to attack or support, perhaps casting doubt on the idea that incumbent parties need to defend their position. Although the methods in (Nooy and Kleinnijenhuis, 2013) provide a large amount of coverage and conclusions, the analysis conducted is resource intensive and could benefit from applications in machine learning or data science in general.

More recently in the investigation of political interactions, Facebook comments from Finish election candidates were automatically scraped (Laaksonen et al., 2017). Analysis of the data relied upon two separate methods: field notes and big data solutions. The field notes found that candidates were mainly negative when mentioning another candidate from an opposing party, and actively pursued these interactions in an attempt to pick up on mistakes. The data science approach involved sentiment analysis on all the Facebook comments extracted (137,000) showing 359 candidate to candidate interactions. This analysis found that the majority of the interactions were actually neutral with more positive interactions than negative, although the negative comments obtained more attention. Laaksonen *et al.* suggest a combined ethnography and data science approach using both field notes and data trends. Despite the advantage of this approach, in that further more detailed insight will be obtained, the authors do concede that any study would still be very resource intensive. There are a number of solutions the authors do not consider such as automating as much of the manual process as possible to reduce the field notes needed for analysis. This can be done through many fine grained classifications, to first remove neutral data then further classify the remaining data. Although initial observations or tasks for the data will be needed this can reduce the overall cost of analysis.

In the case of ethos analytics the approach is focussed on a mainly automatic solution and their interaction with media of the time. This uses the same combined approach of field notes and automatic solutions but does not rely so heavily on the former. There are two reasons for this: manually analysis is a timely process; and manual analysis is more prone to errors in tasks which require analysing large volumes of data as long as the automatic methods for extraction are deemed as reliable.

3.2.3 Election Prediction

One of the main topics in political science is to use analysis of social media to predict election outcomes. In (Franch, 2013), data from Facebook, Twitter, Google and YouTube was used to predict the outcome of the 2010 UK general election through mentions of Prime Minister candidates. This analysis was only provided for the leaders of the three main political parties in the UK (Conservatives, Labour and Liberal Democrat) and relied upon mentions of the political leader by name. A set of independent variables were used to test significance ranging from popular Facebook group comments to Twitter sentiment predictions. Although much of the analysis was collected manually this allowed for the pinpointing of contentious moments in the campaign such as spikes in mentions of Gordon Brown on Facebook pages opposing him when he insulted a voter. Overall nine independent variables were significant in predicting the vote share in the election (5 for the Conservatives, 3 for the Liberal Democrats and 1 for Labour). The vote percentage prediction matched the outcome of the election almost perfectly and was more accurate in comparison than any of the prediction polls. The sustainability of this method is uncertain however due to the focus on political leaders and vote share. Although political leaders attract a high proportion of voters, within the UK election system voters are instead voting for a party or politician in their area which can influence the outcome. Vote share although being a good indicator for election outcomes, does not necessarily show the overall outcome of the election due to the size of some larger constituencies which only equate to one seat in the election. The reliance on manual methods of analysis also makes this an expensive task.

In (Burnap et al., 2016), a different approach was taken to predict the 2015 general election prior to knowing the outcome rather than a retrospective prediction. Tweets (13,899,073) containing sentiment, adjusted for Twitter population bias, seat share and power distributions were extracted using political party and political leaders. Any tweet containing multiple entities were removed, so the target of each tweet was easily identified. Leader and party data were then combined to give an overall sentiment value for a political party, which was then used to calculate the vote share for the 2015 election constituency by constituency across the UK utilising the vote share from the 2010 election. Overall the prediction was that the Labour party would win the most number of seats but not a majority. This proved to be wrong with the Conservative party winning the election although the vote share prediction for the election was accurate. Several problems were identified that may have contributed towards the incorrect prediction. There was an inability to geo-locate tweets meaning tweets from outwith the UK and in different constituencies could influence results. Tweets were also attributed to the wrong source, validated from a manual annotation of 1000 tweets, such as the Green party tweets mainly involving the Australian Green party. There were issues around attributing tweets to political leaders who despite being figure heads are not always the reasons for voting a particular way. Manipulation of data can also occur through trolls or bots and the data is not always representative of the population as a whole. Finally in (McGregor et al., 2017), Twitter discourse is extracted to determine if the outcome of a US senate election can be predicted. Political races of 70 candidates were followed encompassing 35 seats in the 2014 election with incumbency, money spent and competitiveness used as control variables. A list of political Twitter handles was manually compiled encompassing politicians, and media personal. Tweets (3,131,721) were then extracted for three categories, political elites, the media and the general public. Overall the results found that rather than the volume of mentions on Twitter being appropriate to determine the election outcome, Twitter is instead used as a tool for political communication. The study also found that incumbents had no advantage over other candidates and that attention was focused on the candidate that spent the most money. This study again focused on the use of manual annotation to determine any possible significant trends in the data and had a large focus on manually retrieving Twitter handle names.

What each of these studies highlight is a lack of prediction from the content of what is said in parliament. As was suggested at the beginning of this section there may be a link between what is said in parliament and the feeling the general public have towards politicians. Pressure on a politician from outside should be reflected inside the parliament whether that comes from the media or public, therefore, extracting supports or attacks on ethos could provide a reliable analysis.

3.3 Ethos in Text Mining

Ethos in text mining considers the automatic extraction of arguments or ethos in the most comparable way to that of ethos mining in this thesis. While many of the tasks in this section do not concern themselves with the exact identification of ethos, they do include those which can be improved by extracting ethos supports and ethos attacks, or that extract a particular argument component which could be considered as similar to ethos. For example, an argument from expert opinion includes mentions of ethos within the scheme, such as appeals to a higher authority in a particular domain as a premise (see work by (Walton et al., 2008) and (Budzynska, 2010)).

In this section specifically, the background for ethos mining, opinion mining and sentiment analysis, and argument mining, are explored. These fields constitute the research most closely related to ethos mining in which the methods and tasks of classification can be used as inspiration. The research explored in this section also has a relation to ethos mining through the methods of extraction, using text based features like part of speech tags and dependency trees or a bag of words approach, with machine learning.

Table 1 outlines the remainder of the section where specific works in both of these fields have been further described. The works in section 3.3.2 describe expert opinion classification either through the Walton et al. (2008) schemes or variants which explore ethos and authority. Section 3.3.3 describes persuasion in social media specifically investigating ethos in Reddit and extracting ad hominem. Finally, section 3.3.4 explores one area of research made up of three papers which look to Canadian Hansard to automatically extract question and answer pairs related to reputation.

Section	**Paper**	**Relation to Ethos**	**Domain**	**Dataset**
3.3.2	(Lawrence and Reed, 2015)	Expert Opinion	Annotated Arguments	AIFdb
3.3.2	(Toledo-Ronen et al., 2016)	Expert Stance	Encyclopedia	Wikipedia
3.3.2	(Carlile et al., 2018) & (Ke et al., 2018)	Authority	Persuasive Essays	Student Essay Corpus
3.3.3	(Hidey et al., 2017)	Classical Ethos	Social Media	Reddit Change My View
3.3.3	(Habernal et al., 2018)	Ad Hominem	Social Media	Reddit Change My View
3.3.4	(Hirst et al., 2014) & (Naderi and Hirst, 2017) & (Naderi and Hirst, 2018)	Reputation	Parliamentary Debates	Canadian Hansard

Table 1: An outline of the work most related to ethos mining. Specified is the natural language phenomenon of interest including the exploratory domain and dataset.

3.3.1 Background

This section outlines the background research which contributes towards ethos mining namely opinion mining and sentiment analysis and argument mining.

Opinion mining and sentiment analysis are techniques used within Natural Language Processing (NLP) to gauge the feelings or opinions a person expresses in a sentence, dialogue turn or document. The feelings or opinions in this case can be towards an entity (person, location, or company) or a product and can be expressed for any topic. Sentiment analysis and opinion mining (also known as opinion extraction, sentiment mining and subjectivity analysis) have applications in ethos mining where the sentiment expressed in a sentence can be used to determine the feelings an entity expresses towards another, while ethos mining can be considered as a more specific opinion mining task. Specifically, ethos mining relies upon the connection between entities with an expression of positivity or negativity while opinion mining and sentiment analysis[1] are more broad, extracting less relevant information. Despite this, both techniques can be utilised either as a step in an ethos mining pipeline or to provide techniques to extract ethos.

Argument mining (also called argumentation mining, see e.g. (Budzynska and Villata, 2016; Moens, 2013; Peldszus and Stede, 2013; Schneider, 2014; Wyner et al., 2010) for an overview) is the automatic extraction of argument from text over many different domains[2]. Overall the motivation behind argument mining is to find the reasons for a particular opinion. Sentiment analysis and opinion mining can find the sentiment and opinion a person has towards something, but what they cannot do and what argument mining strives to do is automatically extract the whole reasoning structure behind that opinion.

Furthermore the methods used to perform sentiment analysis can be generalised for other text mining tasks. Finally, the area of research most closely related to ethos mining, sentiment analysis of political discourse is also explored to determine if the methods used can be utilised in this research.

3.3.1.1 Defining Sentiment Analysis

Sentiment analysis encompasses several tasks depending on the goal of the classification. These tasks vary from classifying sentences, larger paragraphs, or even whole documents as positive and negative, to extracting opinions and sentiments towards a specific product or event in reviews.

In (Jurafsky, 2015), sentiment analysis is defined in three stages. A simple task, a more complex task and an advanced task. A simple task asks the question: ‘Is this text negative or positive?’. This shows sentiment analysis in its simplest form where blocks of natural language can be automatically traversed to find positive or negative sentiments about an object. A more complex task is defined as ranking the attitude of the binary sentiment task on scale. This involves taking negative or positive sentiments and then

[1] From this point both sentiment analysis and opinion mining will be referred to as sentiment analysis.
[2] In the Aristotelian sense argument would be classed as logos.

deciding what numerical scale should be applied to them. Finally, an advanced task involves detecting a specific target for a review or detecting a more complex sentiment within the text. Detecting a specific target involves searching for certain attributes of an object and attaching a sentiment to that attribute whist detecting more complex sentiments considers the discourse structure within a document.

In (Liu, 2010), sentiment analysis is considered in four separate categories. The first is sentiment and subjectivity classification encompassing classifying sentences as positive and negative and as subjective or objective. Classifying sentences as positive and negative allows the propagation of this sentiment to the whole document, whilst classifying sentences as subjective or objective can allow specific features of an object to be classified as negative or positive. The second category is feature-based sentiment analysis which involves defining categories for a document or piece of natural language and then deciding whether these categories are positive or negative. The third category defined is sentiment analysis of comparative sentences which can be achieved in two ways. The first is providing positive or negative sentiments for an object and that object alone. The second is to have direct comparison with other products. This means that positive or negative sentiments are decided purely on the comparison of an object with another object of similar stature. The final category defined is opinion search and retrieval which is more general and instead of finding negative and positive sentiment from an individual document, it involves finding positive and negative opinions on a selected subject.

Within this thesis the final category of opinion search and retrieval is used for ethos mining, however, the first category of classifying sentences as positive and negative are also utilised.

3.3.1.2 Methods for Sentiment Analysis

In order to perform sentiment analysis several methods can be deployed. These methods range from rule-based systems which count the occurrences of sentiment holding words to deep learning based classification methods which use large amounts of data to determine trends. The methods used for classification are not unique in all cases to sentiment analysis. This is particularly the case in the machine learning based classification which can be used for any text classification task. In the case of a lexicon based classification, this is more common for sentiment classification, although could be utilised in other areas.

In the case of rule-based methods, these normally operate on a sentence level for classification. The reason being that the presence of sentiment holding words are used to calculate the overall sentiment value for the sentence. The rules then need to stipulate what to do in difficult classification situations. For example, when a positive word is preceded by negation or when there is the presence of a word like “but” which switches the sentiment of a sentence.

In (Hu and Liu, 2004), rule-based methods are used to determine the polarity of sentences in product reviews. Three main rules are developed to detect this polarity. In the first instance if there are a majority of either positive or negative sentiment holding adjectives then this is the basis of the classification. In a second case where there is an equal number of adjectives with both polarities then past sentiment values for a product are used to generate an average sentiment. In all other cases, such as negation or the use of words like “but”, the sentiment of the sentence is switched. In order to gain sentiment values WordNet is used, with a set of key sentiment holding words, to generate a lexicon. This lexicon is commonly used for sentiment classification due to its ability to generalise reasonably well to other domains.

In (Pang et al., 2002), a different approach was taken. A machine learning approach using three classifiers, Naive Bayes, Maximum Entropy and Support Vector Machines (SVM), was used. Sentiment classification using an SVM outperformed the other classifiers in (Pang et al., 2002) when performed on large feature vector sets using only unigrams as features. This is though, dependent on the size of the training data used, where an increase in data would mean the training time for the SVM grows, whilst also becoming less generalisable to other domains.

Sentiment classification has also been improved through the use of deep learning

(LeCun et al., 2015). In (Zhang et al., 2018), several works using deep learning techniques for sentiment analysis are described, specifically tasks such as document, sentence and aspect based classification are investigated. In the case of sentence classification, the most closely related task for ethos mining, various deep learning techniques have been used. In these cases different architectures are used for classification which tend to have word embeddings as an input into the neural network. Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) layers are combined in (Wang et al., 2016b) and CNN and Recurrent Neural Network (RNN) layers are combined in (Wang et al., 2016a). The research in the latter paper makes use of the fine-grained aspect of a CNN whilst at the same time making use of the sequential aspect of an RNN layer. Particularly the RNN layer will allow for the classification of longer dependencies in the text which would not be picked up by the CNN. In (Qian et al., 2017) linguistic features of the sentiment task are incorporated through an LSTM. In this case rather than learning the features of the text from scratch, the sentiment features can aid in classification. Overall these works improve sentiment classification through deep learning. The improvements, although not particularly large over lexicon based methods, show the applicability of various CNN and RNN models. In these cases large training datasets are used which means the applicability to ethos mining may prove difficult. In the case of ethos mining though, inspiration can be taken from these methods.

3.3.1.3 Opinion and Sentiment in Political Discourse

The most closely related area of sentiment analysis to the research conducted in this thesis is that of political discourse. These works tend to focus on determining political opinion either in a parliamentary setting or through social media data. As the target domain of ethos mining is parliamentary data, the works explored here focus on this aspect.

Parliamentary data does pose a challenge for sentiment analysis due to several reasons. In the first instance the language used to express sentiment in politics is different to the more colloquial language used within reviews or in social media. Furthermore, there are less works which explore parliamentary data because of the difficulty of such language which and therefore there is less data to create more advanced methods.

In (Salah, 2014), NLP techniques such as part-of-speech (POS) tagging were applied to parliamentary debates to obtain features for machine learning classifiers. These were then compared to two lexicon based sentiment approaches, an off-the-shelf lexicon approach, SentiWordNet 3.0 and a domain specific lexicon approach. When compared, the machine learning classifiers out performed the lexicon based sentiment approaches.

In (Thomas et al., 2006), a corpus of U.S. floor debates from the House of Representatives was compiled with added voting records, to train and test an SVM classifier to determine if a politician supports or opposes a piece of legislation. Although this does not encompass relationships between politicians, it is highlighted in (Thomas et al., 2006) that this is an intriguing path to explore.

In (Rheault et al., 2016), a word embedding approach is used to generate a domain specific lexicon in Hansard. To generate the lexicon a set of seed words (100 in total) were manually crafted for both the positive and negative class. Following this step synonyms were found using WordNet to generate more seed values. Three rules were then applied in order for a seed work to be used in the final lexicon creation. The words must be commonly used in english, ruling out unique words. The words cannot be associated with an entity, for example a name or institution. The words may not have varying polarity across synonyms. Word embeddings are then generated from the text and cosine similarity used to determine how similar a word is to the set of seeds. Following this step the overall set is reduced to 4200 words in the lexicon overall. This approach does generate a large domain specific lexicon, however, there is still the chance that the context a word may appear in will not hold a sentiment value. Therefore, false positives will occur.

3.3.1.4 Argument Mining

There are several subtasks involved with argument mining, ranging from component identification (whether a piece of text is part of the argument structure) and relation identification (depending on the theory used if the relationship between components is that of inference, conflict or a rephrase of what has already been said) to argument scheme classification (classifying the inference or conflict into different types). While the focus of argument mining does not necessarily overlap with ethos mining, there is a clear relation between the two tasks as defined by Aristotle in his modes of persuasion (see section 2.1) with logos directly related to argument mining. In the case of ethos mining then, argument mining can be used as an inspiration for possible methods of extraction or scheme identification, as well as an avenue for possible future work (see section 9.2). Moreover, the difficulty of both argument mining and ethos mining makes for a good comparison in relation to both the manual annotation of arguments and the overall classification results.

3.3.1.5 Tasks and Methods in Argument Mining

The task of argument mining can be broken down into a number of sub-tasks which incorporate varying methods for automatic classification. Initially, natural language text must be broken down into argumentative segments, also known as argumentative discourse units (ADUs), the role of the segment must then be identified, and then the relations between the segments identified (Peldszus and Stede, 2013).

Segmentation and Component Identification

In the first instance identifying ADUs involves breaking a larger section of text down into smaller components that could be considered as arguments. These smaller components vary in size, in some cases consisting of multiple sentences, in others a full sentence, and sometimes below a sentence level.

In (Persing and Ng, 2016), a set of manually defined rules are used for the task of segmentation specifically investigating punctuation marks which start and end a segment and characteristics of syntactic parse trees such as a shallow subordinate clause. While in (Lawrence et al., 2014), a classification approach was taken using two Naïve Bayes classifiers one to identify the start of a segment and one to identify the end of a segment. Both use the same feature set made of individual tokens, incorporating the current word, preceding word, proceeding word (including punctuation as a word), word length and POS tag. In (Stab, 2017), the segmentation problem is considered as a sequence labelling task operating on each token to define the segment. In this case a conditional random field (CRF) classifier was used with structural text features (location of a token within the whole text), syntactic features (POS tags), a combination of lexical and syntactic features and, the probability of a token being the beginning of a segment. Most recently, (Eger et al., 2017) have used deep learning techniques for argument mining considering the task as multi-task problem, which as a by-product includes segmentation. In this case an LSTM (see section 3.3.1.2) model using dependency parse features and built for resolving entities and relations was adapted for the argument mining task, meaning that the task again could be considered as a sequence labelling problem.

In many works, partly due to the number of sub-tasks within argument mining, segments are given as a prior input to a classifier, simplifying the overall task and ensuring that no errors are passed from the difficult segmentation task. This then allows for a different task to be conducted, that of argument component detection which involves classifying segments as either containing an argument or not or as a first step classifying if sentences contain an argument or not. This is the case in (Moens et al., 2007; Palau and Moens, 2009; Rooney et al., 2012; Stab and Gurevych, 2014) where sentences are first classified as containing an argumentative component or not. Each of these works use various textual features such as uni-grams, bi-grams and trig-rams as well as syntactic (POS tags), discourse indicators and structural features (dependency tags) as inputs to machine learning classifiers (max entropy, naïve bayes, decision trees and SVM).

While the issue of segmentation is important for argument mining, the research on ethos mining described in this thesis instead focuses on a sentence level segmentation. There are two reasons for this. Firstly the role ethos plays in the segmentation is not yet clear in comparison to arguments. Secondly, any further segmentation would therefore have to be conducted with a full argument analysis of the text of which the context would sometimes be considered as an argument and in other cases not.

Role and Relation Identification

Following the identification of segments or sentences comprising argumentative components the roles they play must be identified. This depends upon the theoretical foundations used to define an argument structure and the dataset it is applied to. What must be determined is the premise conclusion structure, whether or not that is made up of one main conclusion and a set of claims and premises (as in (Cabrio and Villata, 2012; Eger et al., 2017; Palau and Moens, 2009; Rooney et al., 2012; Stab and Gurevych, 2014; Stab, 2017)) or multiple conclusions and premises (see (Lawrence and Reed, 2015, 2017; Lawrence et al., 2014; Park and Cardie, 2014; Villalba and Saint-Dizier, 2012)). Coupled with this task is that of relation identification, ideally if the premise and conclusion structure is known then the relation can be inferred particularly if this relation is one of inference or conflict.

The task of automatically identifying the role of an argument is normally conducted through the use of statistical machine learning approaches using a combination of features which are also used for segment identification. However, earlier approaches also made use of a linguistic analysis to determine rules of extraction (Villalba and Saint-Dizier, 2012). Linguistic analysis of text involves investigating the structure of dialogue and extracting argument through the use of discourse indicators.

There tends to be a defined structure within the text with an evaluation followed by an attribute. Adverbs are normally found at the start of the text which can describe the attribute, the adverb is the evaluation. Discourse relations in argument support are used to try and persuade a reader to believe an argument through the use of discourse structures, such as illustrations, justification and elaborations.

In (Villalba and Saint-Dizier, 2012) each discourse relation is then implemented using

TextCoop (Saint-Dizier, 2012). TextCoop allows discourse analysis of text using linguistic techniques to identify the discourse relations, through a set of rules, for identification within text. The rules make use of the punctuation in text, keywords which tend to appear in certain discourse relations and the expressions which are generally expected in the different discourse relations.

In (Lawrence and Reed, 2015), different argument mining techniques are combined to extract the overall structure including the role of components and the relation between them. This involved combining structural features of schemes (using arguments from position to know and arguments from expert opinion) and general text features: component similarities from (Lawrence et al., 2014) and discourse indicators. In order to improve the classification in (Lawrence and Reed, 2017) high precision techniques of similarity and discourse indicators were used to extract argument component pairs from the web which can then be utilised as training data. This extra data can then be combined with topical similarity thresholds.

In (Stab and Gurevych, 2014), a corpus of persuasive essays was created and used for the argument mining task using text classification techniques to determine argument component roles and relations between them. The dataset used a main claim, claim and premise structure, where there is one main claim in a text, followed by subsequent claims either support or attacking the main claim and then premises supporting the claims. For the combined tasks an SVM classifier, using a combination of lexical, syntactic, discourse indicators and structural features, was used.

Finally in (Eger et al., 2017), the same dataset of persuasive essays as in (Stab and Gurevych, 2014) was used to extract arguments. The authors evaluate several deep neural network models on their ability to identify different essay components, relations and a combination of the tasks with an LSTM which uses the dependency tree structure. The use of the LSTM model allowed for a multi-tasking classification with a treeLSTM from a dependency parse used as the input.

Overall the tasks and methods described for argument mining are not directly comparable to that of ethos mining described in this thesis but they do hold some similarities which can be used as an inspiration and rough comparison when conducting an automatic classification. The task of classifying argument components as argumentative or not has a strong relation to that of classifying a sentence as ethotic or not although with varying phenomena. This does mean that the techniques used for this particular task can be utilised for ethos mining. The tasks of segmentation and relation identification, however, are not so related. In argument mining the relations can be of support and attack (similar to ethos mining), however, in terms of ethos mining the language used to support or attack varies to that of argument mining where an attack relation can be a contradicting statement rather than one that holds a negative sentiment. The task of segmentation, as highlighted above, would require a further investigation for ethos mining particularly when both the argumentative and ethotic structures are considered together.

3.3.1.6 Argument Scheme Classification

A further step in relation identification in argument mining is that of argument scheme classification (see section 2.2.1 for argument schemes). This involves automatically labelling the premise conclusion structures in an argument structure with a specific scheme type. The methods then used for argument scheme classification can be utilised for ethos type (elements of ethos) classification as the task can be considered as similar despite the varying language and structure (see section 3.3.2 for a more specific comparison of argument schemes and ethos mining).

In (Feng and Hirst, 2011), five common argumentation schemes are identified (Argument from Example, Argument from cause to effect, practical reasoning, argument from consequence and argument from verbal classification) and classified using a One Vs All classification. Data from the Aracaria database was used to train a decision tree on each class using a total of 393 arguments. The highest accuracies were achieved for argument from example (90.6%) and practical reasoning (90.8%) while consequences, cause to effect and classification obtained 62.9%, 70.4% and 63.2%. Following the One vs All classification a pairwise classification was undertaken. In this case the pairwise classification is used to resolve any classification discrepancies found in the One Vs All classification such as a particular relation classified as belonging to more than one scheme. Overall this classification step produced accuracies between 64.2% and 97.9% with the large gap due to the varying language used to express each scheme meaning that schemes which are semantically closer will have a lower classification accuracy.

In (Lawrence and Reed, 2016) four schemes were classified (argument from analogy,cause to effect,practical reasoning and verbal classification) using One Vs All classification and 226 arguments from AIFdb. Overall a Naive Bayes classifier gave the best overall performance for each scheme with an average (combined premise and conclusion classification) F1-score of 0.65 for Analogy, 0.70 for cause to effect, 0.74 for practical reasoning and 0.75 for verbal classification. The same second step as in (Feng and Hirst, 2011) was taken, that of a pairwise classification to resolve relations given the same scheme. The resulting classification, again using a Naive Bayes classifier, gave F1-scores ranging from 0.60 to 0.88 with the same gaps due to the varying language used to express the different schemes meaning distinction can be made more easily for semantically different schemes.

In (Musi et al., 2016), a new set of argument schemes were proposed for three levels: the top level, intrinsic, extrinsic and complex; the middle level, causal, definitional, mereological, analogy, opposition, practical evaluation and alternatives, authority; and the low level, formal clause, final clause, material clause and efficient clause. The proposed new schemes act as an alternative to the Walton argumentation schemes which can be difficult to classify due to the vast number of available schemes and in some cases the similarity between the various schemes [3]. Using untrained annotators for manual scheme classification gave a $\kappa = 0.1$ for the middle level, while the top level was not significantly higher. After training the annotators agreement increased to $\kappa = 0.31$ while this is a significant improvement over the previous results it is unclear as to whether the new

[3]Take for example an argument from position to know and an argument from expert opinion, where an expert would be considered as someone providing an argument from a position to know.

scheme set proposed would perform any better than annotating using the Walton schemes.

In (Visser et al., 2018b), the Walton schemes are revisited as the same problem, of identifying which schemes are appropriate for annotation from the large number available, was found. A second set of argument schemes, Wagemans' periodic table of arguments (Wagemans, 2016), were also chosen for the purposes of comparison with the Walton schemes on the grounds that a scheme set with a more clearly defined selection criteria may provide a more reliable annotation. Due to the large number of Walton schemes a decision tree was also created to aid in the choice of argument scheme. In both evaluations a 10% subset of the data was taken for IAA which gave a κ of 0.69 for the Wagemans schemes and 0.72 for the Walton schemes. Despite the closeness of the IAA results the Walton schemes still outperformed the annotation of Wagemans schemes this could be explained by the creation of a decision tree making the scheme selection a simpler task. Despite this, the Wagemans schemes could play a further role in automatic classification. The periodic table of arguments is broken down into simple classification tasks, first-order or second-order, subject or predicate and fact, value or policy. Once these simple classifications are made the choice of scheme is presented.

Whilst the task of argumentation scheme classification both in an annotation sense and automatic classification sense is not directly comparable to the task of ethos mining or the automatic classification of ethos elements, there is a twofold comparison that can be made. In the first instance the need for revised annotation guidelines is comparable as is the complexity and difficulty of the task (section 7). In the second instance for automatic classification, there is an overlap between the task being performed. That is to say both scheme classification and ethos element classification operate on a single relation it is only the content of the linked propositions which differs and therefore the classification task differs overall.

3.3.1.7 Argument Mining in Various Domains

The process to extract arguments automatically remain the same no matter the domain of application, however, some techniques will perform better than others. Domain specific language can mean that the same techniques may not be generalisable across domains. For example, the language used in legal texts will be very different to those used in social media data, but, the language used in parliamentary debates may have similarities with legal text. Whilst argument mining can be applied to any domain which consists of textual or spoken data, this section focuses on those closest to the domain of parliamentary discourse and ethos mining in general which is more broad than just political discourse due to the relatively young field of argument mining.

Political Texts

Political texts provide a rich resource of data for the problems of argument mining and ethos mining. In particular parliamentary debate in many countries is freely available, updated daily and transcribed manually creating a very accurate data source. This section aims to explore argument mining in the political domain providing a brief outline of the possible avenues of extraction, as well as, the methods used.

In (Naderi and Hirst, 2015), a corpus of 274 argument frames (see example 6) from Gay Marriage political debates in the existing ComArg corpus (as a training set) were connected to Canadian parliamentary debates on same sex marriage made up of sentences pro or con towards the topic. The data was annotated by three coders with an inter-annotator agreement using weighted kappa Cohen (1960) of 0.54 for stance (a users stance on a sentence) and 0.46 for connecting pre-existing frames to stances (pre-existing arguments which highlight an aspect of an argument), with 90% agreement on statements between at least two of three annotators.

(6) **Frame:** *"There was no need to change the definition of marriage in order for gays*

and lesbians to establish meaningful, long term relationships that are recognised in law."

The statements which had agreement between two annotators were chosen leaving a total of 121. An SVM classifier was trained using a bag of words approach, distributed word representations of stance and frames with similarity calculations and the stance of each statement, either pro- or con-, as a feature. Frames were then identified in political speeches connected to sentences on gay marriage with an overall accuracy of 73.5%.

In (Lippi and Torroni, 2016), a combination of speech and text features are used to classify claims of three candidates from live TV debates during the UK general election of 2015. Initially a corpus of claims and evidence (where claims are considered conclusions and evidence premises) was created by two expert annotators for the three leaders in the TV debate (Cameron, Clegg and Milliband). Overall the annotation task gave a Cohen's κ of 0.53 for statements after which an overall agreed annotation was reached.

The automatic extraction process involved a pipeline containing a speech recognition tool, to extract both text and acoustic features from the recorded TV debates. The text and acoustic features of the audio were then passed to a feature extraction function to allow classification using an SVM classifier. This gave $F1$-scores of 0.53, 0.53 and 0.29 for the three candidates over random baselines of $F1$-score 0.47, 0.44 and 0.30.

Overall the works explored here have no direct relation to the task of ethos mining, however, the domain explored does provide similarities. The tasks explored can provide possible future avenues for the extension of methods to extract ethos, particularly in the use of datasets outwith the remit of parliamentary debate or extracting ethos directly from parliamentary speeches which are continuously televised each day.

Legal Texts

The automatic classification of legal texts holds a two-fold relation to the task explored in this thesis, that of ethos mining. The first relates to the language used with the legal domain and its similarities to that of parliamentary discourse. Within parliamentary discourse the use of a legal style of language is common due to the topic of debate mainly pertaining to the implementation or non-implementation of legislation. The second similarity is that the domain of legal texts was one of the first for argument mining, meaning that the tasks and methods explored are at a similar level as to expect from ethos mining.

In the pioneering work on argument mining, (Moens et al., 2007), arguments were explored in a number of domains (discussion fora, newspapers, parliamentary records and weekly magazines) but with a target of legal texts. Data was taken from the Araucaria database of arguments (Reed and Rowe, 2004), totalling 1899 sentences which contain an argument and 827 sentences which contain no argument, this was also supplemented with 1072 sentences which contain no argument but were extracted from the same sources of the Araucaria database. This gave a balanced dataset to use as training data for automatic classification. A number of text features were tested with the best set containing: text statistics (sentence length, word length and punctuation marks); verbs; and word couples (the combination of all the possible word pairs in a sentence) overall this gave a classification accuracy of 68% on legal texts ("He is aware of the risks involved, and he should bear the risks." an argumentative sentence from legal text) and a highest accuracy on newspaper articles (73%).

Whilst again not directly comparable to the task of ethos mining, the methods are comparable. There is no problem of segmentation in both tasks and they both operate on a sentence level. This means the classification methods used in (Moens et al., 2007) can be utilised for ethos mining.

The research on arguments in legal text was continued in (Palau and Moens, 2009). Rather than classifying sentences as containing an argument or not, they were classified as premises, conclusions or containing no argument. In this work the datasets were not balanced this is to keep the classification task as close to the real natural language text problem and were increased using data from the European Court of Human Rights a total of 2516 sentences. The same text features were used for argument and non-argument detection with additional features then used to classify the arguments as conclusions or premises. The additional features used were: sentence length; verb tense; history (previous and next sentences); rhetorical patterns (support, against, conclusion, other, none in sentences); discourse indicators; and part of speech features (main verb, articles). Overall classifying conclusions gave an F1-score of 0.74 and classifying premises gave an F1-score of 0.68.

Again the argument mining problem investigated in this task is not directly comparable to ethos mining, yet the text feature used for classification could provide an avenue to improve classification methods. Whilst the use of a two step classification, from argumentative to non-argumentative and then to premise and conclusion structures, provides a good first step in argument mining more recent methods using neural networks and combining all the argument mining tasks through multi-task learning aim to avoid carrying errors. The two works in this section, (Moens et al., 2007; Palau and Moens, 2009), are the pioneering works of argument mining specifically addressing legal texts. While there are further works in automatically processing legal texts, those defined here are the closest to the problem of ethos mining in that they display the first automatic classifications of argument.

3.3.2 Expert Opinion

In conversation, experts are continually used to support a conclusion or claim made. In these cases, the weight of the expert and the notion that they will be knowledgeable about the given topic is used as a premise along with a supporting statement the expert has made on the topic of the conclusion (see section 2.2.1.2 for a description of the expert opinion scheme). While this does not strictly relate to the notion of ethos used in this thesis (that of a reference of support or attack on the character of the speaker), it does involve the topic of using authority to support a claim and, when taking into account the argument scheme of expert opinion and the critical questions associated with that scheme, relates to questioning whether a given expert really is an expert in this case. In Table 1, three works including a total of four papers associated with experts have been identified encompassing several domains and corpora with varying annotation of experts.

In (Lawrence and Reed, 2015), the problem of extracting arguments using techniques such as discourse indicators, similarity and argumentation scheme structures including argument from expert opinion was investigated. In the case of automatically classifying expert opinion, the individual components (premises and conclusions) of the scheme (see Example 7) were identified using a range of features, for example, punctuation, similarity and unigrams and bigrams. In turn, by extracting the individual components of the scheme, such as premises or conclusions, the relations between them can be inferred using the typical argumentation scheme structure. Just for the case of expert opinion, and for all the scheme components an $F1$-score of 0.79 was achieved using a Naive Bayes classifier.

While the extraction of expert opinion is not strictly comparable to the problem of ethos mining, it provides an indication of the acceptable performance level of any ethos mining system although for a more specific task. Furthermore, ethos mining operates only on one of the premises of the expert opinion scheme where this premise would instead be considered as a conclusion of ethos which can be supported or attacked. In the case of example 7 from (Lawrence and Reed, 2015) ethos does not operate on (7-a) or (7-c) but instead on (7-b). This means that the expert, in this case Sir Simon Wessely, can either be supported or attacked through (7-b) which could play a role in the overall validity of the argument (this is not a point addressed in this thesis but is instead addressed as a question for future work).

(7) a. **Conclusion:** *"An explosion of charities offering different and sometimes unproved treatments to veterans with mental illness could be harming rather than helping."*
b. **Premise:** *"Sir Simon Wessely, an expert in the field"*
c. **Premise:** *"said there was a lack of regulation in tackling post-traumatic stress disorder"*

In (Toledo-Ronen et al., 2016), the focus shifts from using argument schemes to identify arguments, to instead using the stance an expert has to identify if an unseen expert is for or against a controversial issue. Essentially this allows automatic stance classification through the use of expert entity recognition. Once the expert is identified their previous statements can be manually searched for references to particular topics and from this classified as pro or con based on their underlying beliefs. This approach can then allow for the increased volume in datasets which can be used for argument mining, perhaps by leveraging the category classification of an expert with the statements for that given category. To create a dataset for automation, Wikipedia was searched for categories of discussion which are then utilised to identify experts on these areas. After the identification of experts a set of six manual annotators were used to determine the experts stance on the category (pro, con or none) which achieved a Cohen's κ of 0.92. A next automatic step was then taken to identify pro or con experts for a category. A rule based classifier implemented using a development set gave a precision of 0.94 and recall of 0.72 on determining the stance of a category and 0.94 and 0.60 when determining the stance of the expert.

This work represents a step in the direction for obtaining contextual information in argument mining through the use of expert profiling. Whilst this is not directly related to ethos mining, it does show a possible application of ethos mining to argument mining in that the stance of a person towards another can be utilised to increase the size of datasets (although with some added noise). For example, if a speaker holds a negative opinion of another speaker in a majority of cases (e.g. the relationship between them is mainly ethos attacks) but data is only available for a specific topic, such that a natural language processing model would have a limited vocabulary, then inference can be made for other topics assuming that any reference between these two speakers will be negative.

In (Carlile et al., 2018; Ke et al., 2018), argument persuasiveness was annotated on top of 102 student essays from the Argument Annotated Essays corpus (Stab and Gurevych, 2014). For the essays corpus, students were prompted to answer an essay question on a topic, they were encouraged to provide both sides of the argument and provide reasons for their claims. The argument analysis consisted of major claims, which refer to the main conclusion of the essay; claims, which are used to support or attack the major claims; and premises, which are considered as the reasons for a claim. In this paper Carlile et al. label major claims, claims and premises with three attributes with diffent scales: eloquence (on a scale between 1 and 5); specificity (between 1 and 5); and evidence (between 1 and 6). Eloquence relates to the style and presentation of the text specificity relates to how detailed the text is and evidence if the statement supports a parent statement. A value of one differs in meaning for the various attributes from a non appearance of an argument to errors in the argument, while a value of five or six also differs in meaning from a strong argument persuading most readers, to an argument free of errors. Persuasive strategies (logos, ethos, pathos) are also labelled on the major claim and claims. This labelling takes into account all of a major claim's child nodes and constitutes a binary value, either containing the persuasive strategy or not. It should be noted, however, that the definition of ethos here is different to the definition proposed in section 2.3.3 and is closer to trust described in section 3.1 or expert opinion in section 2.2.1. In this case ethos is considered as an appeal to an authority, "trust in a higher authority" which is closer to that of an argument from expert opinion (Walton et al., 2008), while this thesis is concerned with the character of a speaker specifically character mentions. The nature of the dataset and the prompt to the students, who in some cases will not be fully knowledgeable on a particular topic, does limit the application of ethos to persuade. Inherently, it becomes more difficult to quote experts on a topic with no previous knowledge of a field and becomes nearly impossible to openly criticize an experts character.

In (Ke et al., 2018), this work is continued employing neural networks to classify the overall persuasiveness of an argument and each of the attributes and strategies. Two approaches were considered, a pipeline approach where predictions are made for each attribute and passed to the next classification task, and a joint approach where persuasiveness and attributes are both considered as inputs and the persuasiveness score and attribute classifications are given as outputs. Focusing on the classification of attributes, as this involves ethos, the joint model and pipeline models gave similar results with the lowest $F1$-scores of 0.045 and the largest 0.967. While the results for ethos classification varied widely in the two classifications it should be noted that the overall frequency of these values in the datasets, 25 instances overall split between training, development and testing, reduces the likelihood of a generalisable ethos classification in this case especially to a wider ranging ethos definition.

Despite this research extracting a notion of ethos, the results are not directly comparable to the research in this thesis as the definitions of ethos are different. There is also a mismatch in the volume of ethos present, as will be shown in chapters 5 and 6, where more than 25 instances of ethos are manually annotated and used for automatic extraction.

3.3.3 Persuasion in Social Media

Aristotle recognised that there were a number of ways to persuade and therefore defined his three modes of persuasion (logos, ethos and pathos), however, when exploring persuasion it can be difficult to determine if a particular statement is persuasive or at which point a change of attitude occurs. A potential solution to this problem is possible by investigating social media posts which include information pertaining exactly to this. In table 1, two papers are identified which investigate Change My View (CMV) on reddit. This data encompasses a statement made by an original poster with the task for other users being to change the original view expressed in the statement. When the original poster's view has been changed they provide an indicator of this, called a Δ point, and the rationale behind the indicator.

Annotation of ethos along with logos and pathos was conducted in (Hidey et al., 2017), on CMV posts, involving eight student expert annotators identifying claims and premises and crowd-sourced annotation of the semantic types of claims and premises. In the case of the semantic types the annotators were provided with six possible labels for claims and three labels for premises. The annotation of ethos was limited to premises only and was annotated when a premise contained a reference to expertise or a reference to a title or reputation. Identifying premises involved a binary decision on all thee labels (logos - reasoning, ethos - expertise and reputation and pathos - emotions) where the majority vote of the crowd-sourced data was taken and compared against that of an expert annotator giving an IAA value of 0.73 (see example 8 for premise types). Although this research did not stretch as far as the automatic extraction of ethos, it was noted that ethotic statements made up only 3% of the overall corpus.

The annotation of claim types (see example 9 for claims) was limited to: interpretation (prediction or explanation); evaluation (positive or negative judgment); evaluation-rational (an evaluation based on a rational reason, evidence, or source credibility); evaluation-emotional (opinion based on emotional reasoning); agreement; or, disagreement. Again for the IAA of claim types, the majority vote was compared to one expert annotator providing an IAA value of 0.46. In this case it was noted that the segments used for annotation give rise to errors due to multiple propositions or claims being present in one segment. Furthermore the labels defined for premises only, overlap with those for claims. That is to say the labels given to claims overlap with those given to premises which in any case should not be restricted to premises only. To rephrase, logos, ethos and pathos are not limited to a premise of a conclusion (or in this case a claim) and can instead be found within a conclusion or premise or in the relation between them. It is also worth noting that the definition of ethos used in this work is closer to that of trust and reputation defined in section 3.1 and differs from that of ethos supports and attacks used in this thesis.

(8) Premise types from (Hidey et al., 2017)

a. **Logos -** *"He will probably win the election. He is the favorite according to the polls"*

b. **Ethos -** *"I trust his predictions about climate change. They say he is a very sincere person"*

c. **Pathos -** *"Doctors should stop prescribing antibiotics at a large scale. The spread of antibiotics will be a threat for the next generation"*

(9) Claim types from (Hidey et al., 2017)

a. **Interpretation -** *"I think he will win the election."*

b. **Evaluation-Rational -** *"He is a very smart student"*

c. **Evaluation-Emotional -** *"Going to the gym is an unpleasant activity"*

As mentioned earlier in this section ethotic statements made up only 3% of the overall CMV corpus, although statements referring to ethos are more rare than logos, the domain used plays an important role in this. CMV on Reddit restricts the use of direct attacks and is more focused on formal argumentation where any type of personal attack, such as ad hominem (AH), is against the rules.

The annotation of AH was undertaken in (Habernal et al., 2018), using a similar corpus although with the addition of comments previously deleted by the moderators of Reddit. Habernal et al. (2018) annotated 200 AH instances for types (abusive, tu quoque, circumstantial, bias and guilt by association) using 16 mechanical turk workers. Three main research questions were investigated: What qualitative and quantitative properties do AH arguments have and is this reflected in the theory?; Is there a need for context when annotating instances of AH?; and What are the triggers for AH arguments?. To verify the annotation, through an IAA task, 50 random AH instances were sampled with the 50 other arguments also sampled as negative cases with no context given. A Cohen's κ of 0.79 comparing six workers was achieved when selecting if a component, in comparison to semantically similar arguments, was AH or not. A convolutional neural network was then used to automatically extract these cases of AH from CMV threads. 7,242 instances of AH were extracted for training and testing purposes using cross fold validation giving an accuracy of 0.81 on the identification of AH.

The AH types were then annotated on top of the AH base to determine if AH types are "empirically relevant". This annotation step achieved a low percentage agreement (score was referred to as low but not reported) where the authors highlighted that the AH arguments from within CMV did not fall under distinct classes identified within the theory, because many of these cases were multifaceted. A list of empirically driven AH categories were then compiled, moving away from the theoretical background on fallacies. No further

IAA study was taken to verify the new empirically driven categories, meaning that it is unclear if these categories provide any further clarification from an annotation perspective.

Many of the proposed new categories defined also fall under the parent categories defined within the theory. For example, “vulgar insult”, “illiteracy insult” and “idiot insults” all fall under the standard understanding of abusive AH, while “accusations of ignorance” could be defined as an Ad Ignorantiam argument. While it is the case that many theoretical works do not always fit the data in annotation tasks, the theory must be considered and from this a concrete set of annotation guidelines created which do fit the data to be annotated and can be verified using IAA studies. CMV data also has the potential to restrict the use of AH, rather than having a comment deleted posters could, in theory, choose not to use AH arguments on this basis.

In the case of ethos mining in this thesis the research in (Habernal et al., 2018) is similar to that of ethos attacks and ethos types (described in section 2.1.3). For ethos attacks, the annotation of statements as either being AH or not is similar to the steps taken for ethos mining. In the case of ethos types the annotation undertaken in AH types is similar, the results of which act as an indicator of how to conduct the study, essentially building a comprehensive set of guidelines through iterations from the core theory. For this thesis, the research in this work cannot be used directly as the motivations do differ on the goals of extraction and the possible extensions to analytics.

3.3.4 Reputation in Question Answer Pairs

In political discourse it is common for politicians to support or attack one another on the basis of their actions (or lack of action) or their statements within parliament. In table 1, three papers are given which all relate to the use of attack and defence speech within the Canadian Hansard. In particular the notion of reputation within question-answer pairs is investigated using the oral answers to questions period of the parliamentary record which is emulated across a number of countries within the Commonwealth.

In (Hirst et al., 2014), automatic tasks were undertaken to determine if the use of

SVM classifiers in classifying a politicians party were sensitive to attack and defence speech. Data was taken from the Canadian Hansard to train the classifiers and tested on a separate parliamentary term to verify the disintegration of the bag-of-words SVM approach. Through this work it was highlighted that the majority of the oral question period, in Hansard, consists of hostile speech between politicians. This avenue of work was continued in (Naderi and Hirst, 2017) where an annotation task was conducted on reputation defence strategies (denial, excuse, justification, concession and no strategy) in the Canadian Hansard (see table 2 for an indication of when the labels apply). This featured, questions from the Opposition to the Government only, and incorporated question-answer pairs which contain multiple sentences (see example 10 for a question-answer pair). These pairs were provided to multiple annotators where they had to agree for a pair to be included in the dataset (493 pairs overall from 1500). The reputation strategies denial, excuse, justification, concession and no strategy were classified using both mutliclass classification and pairwise classification. An SVM trained using five-fold cross validation gave an overall F1 score of 0.57 (range from 0.1 to 0.65).

(10) a. **Question:** Gerard Deltell said, *Madam Speaker, this question period is very informative. Earlier, when asked a question by the member for Carleton, the Minister of Environment and Climate Change finally acknowledged that large emitters will not pay 100% of the tax because that could result in job losses. The Liberal carbon tax could affect jobs. The question for the minister is very simple, why is there a double standard? Why will small businesses pay 100%, while large emitters get a 90% writeoff?*

b. **Answer:** Catherine McKenna said, *Madam speaker, I am really surprised by this Quebec member. Every political party in Quebec, both federal and provincial, supports carbon pricing. Why does the federal Conservative Party and the member from Quebec not support a price on pollution? We know that we must tackle climate change and that there is a cost to pollution.*

I hope that the member will listen to Quebeckers, who want us to address climate change, want a price on pollution and want a clean economy.

Questioner	Question Type	Government Answer
Opposition	Threat	Reputation Defence
Government Backbencher	Non-Threat	None

Table 2: Annotation scheme for reputation in question answer pairs showing the respective position of the source and their role in the question answer pair (Naderi and Hirst, 2017).

In (Naderi and Hirst, 2018), this classification was continued instead using an automatically generated corpus of the first question answer pairs in a topic discussion. The question answer pairs were classified as reputation threats, if the question came from an opponent to the government, and friendly non-threats, if the question came from a "back-bencher" of the governing party (see table 2 as an indicator of where the labels apply). Example 10 shows a question-answer pair which would be annotated with the following: threat, and justification. Two neural network architectures were then used to classify question-answer pairs as threats or not. While the results of this classification show an outstanding accuracy and $F1$-score (highest value of 98%), a number of issues arise in the dataset. Annotating and classifying on large question answer pairs brings some level of noise into the dataset. That is to say, parts of the question answer pair will not relate to the threat or non-threat nature of the text. This is a common problem in other fields like sentiment analysis, where whole texts can be classified as positive or negative. This annotation could be more fine-grained through the use of ethos, pinpointing exactly which parts of the text relate to "threats". A second problem with the dataset relates to the automatic creation of the dataset. There is a direct relation between threat and non-threat question-answer pairs and opposition or not question-answer pairs. Ultimately, this means that it is hard to draw conclusions from the classification of threat or non-threat pairs, as in this case, 100% accuracy would be achieved by extracting the names of politicians (which are provided for all turns in the transcripts) and from that determining their political party from readily available information in sources like Wikipedia or even the parliamentary websites.

Although the topic of interest is very similar to that of ethos mining explored in this thesis, the units of text used, full paragraph question answer-pairs rather than fine grained sentences, differ. Example (10-b) illustrates this where only two sentences out of five relate to the justification of the stance and the rest of the answer challenges the view given in the question and refers to the questioners ethos, specifically at their inability to listen to their constituents. While the results are promising it is apparent that extracting full question-answer pairs can add sentences which do not necessarily aid in the classification task, thus adding noise. For this thesis, the research described here, again, is not directly comparable due to the difference in goals of extraction, however, there are similarities between both works on the domain chosen for extraction and in some of the answers extracted and defined as reputation defence. Reputation defence extraction in essence is close to the task of ethos mining the difference though lies in the use of a wider argument frame incorporating many sentences which contribute reasons for defence in the case of (Hirst et al., 2014; Naderi and Hirst, 2017, 2018), whereas ethos mining in this thesis looks to extract individual sentences of supports or attacks to another speaker.

3.4 Discussion

Ethos mining covers a large area of the literature within computer science as a whole. As a base ethos mining can use techniques from sentiment analysis and argument mining, whilst having similarities to trust and having applications in political science. Despite this there is still very little research directly comparable to that of ethos mining with the research most closely related defined in sections 3.3.2, 3.3.3 and 3.3.4. Referring back to the research questions specified in chapter 1 it is clear that no direct comparison can be made with the current research in the fields of trust, political science, or argument mining, however, many of these works provide inspiration for automatic extraction methods, comprehensive analytics and sound research methods.

In the case of trust described in section 3.1, the work undertaken in this thesis is again not directly comparable. That being said, ethos plays an important role in trust overall especially in decision making on trust, as statements made referring to another persons ethos (positive or negative references to character) can be utilised to determine if they should be trusted. Whilst using ethos only to determine trust is possible it may not include the overall context which is needed to make a trust decision, and therefore ethos would instead be considered as an element along with context and any other factors in determining trust. The literature on trust has a more transactional focus, that is an algorithm can be trusted if it performs pre-defined tasks to expectation. Within argumentation trust is considered more as a value which can be propagated through an argument structure to determine the validity of arguments. Applications in natural language are less common but any research conducted on ethos mining may improve this area of research.

Section 3.2 described applications in political science that closely involve ethos. These mainly focus upon the use of sentiment analysis. Although social media and sentiment analysis can play a role in predicting vote shares and overall outcomes of an election, the methods to do this rely upon large amounts of manual annotation, manual dataset curation and largely remain inconsistent. The use of social media, although encompassing a large number of people, does not always represent the full population due to the lack of social media use by some. With the increased engagement of politicians in social media a method of scaling up which addresses political speeches could instead be utilised, with the assumption that politicians will adapt their behaviour to online criticism whether through holding other politicians to account for their actions or to appease the wants of the public. This solution would then have two advantages, it will encompass, in part, some of the social media criticism and will encompass criticism between politicians within parliament. Ethos mining then can be utilised in political science as a means to examine politicians, particularly in their effectiveness around election periods.

Sentiment analysis is closely related to the task of ethos mining, particularly as ethos is defined as holding sentiment to a target in speech. The research explored in section 3.3.1

though, is not directly comparable to ethos mining. In most cases the datasets available for sentiment classification are larger than that of ethos mining and thus the methods used for classification can be more complex. The tasks in sentiment analysis also vary from full document classification to sentence level and normally investigate any sentiment attribution rather than to a specific entity.

What section 3.3.1 does show is that the methods used for classification are applicable to other areas in text classification. More specific tasks such as argument mining, more closely related to ethos mining in the sense of data and task, can utilise these methods for classification and generate more specific opinions and sentiments. Particularly important are the reasons why these opinions or sentiments are held.

The tasks within argument mining (section 3.3.1.4) relate to those within ethos mining. In particular, classifying sentences as containing arguments or not relates to the same task in ethos mining (determining if a sentence contains ethos or not), whilst automatically classifying argument schemes relates to the same task of classifying ethos elements [4]. There are also similarities in the domains explored and the methods used to classify sentences within these domains.

Taking inspiration from argument mining can allow ethos mining to follow a similar trajectory in classification improvement, making use of novel methods identified within argument mining. These methods will not necessarily be explored within this thesis but it does allow an extensive area of future work to be defined meaning that ethos mining could have as much as an impact as argument mining currently is.

Although the work in sections 3.3.2, 3.3.3 and 3.3.4 closely relate to that of ethos mining it is important to note that they are not directly comparable. Each of the tasks presented differ in either domain, the target of any automatic classification or the text units used (e.g. segments vs sentences vs paragraphs). In the case of expert opinion ethos mining works as a further premise or rebuttal in the scheme validating the claim that a person is an expert. The case of AH looks at one distinct part of ethos mining, that of

[4] It should be noted that these tasks are not directly comparable as they are performed on different training and test data sets with a different output.

an attack on ethos while reputation analysis takes a more global perspective than ethos mining, operating on paragraphs and question-answer pairs rather than more fine grained sentences referring to ethos.

This section has highlighted how difficult and nuanced the tasks are in argument mining. At the first stage the manual annotation tasks require long processes of annotation guideline construction and fine tuning after which the results, while promising and reliable, are still not perfect. At the stage of automatic classification the task is equally, if not more, difficult with any classification system needing to rely upon contextual and consistent information. This is particularly difficult in the case of natural language with a number of words, such as “because”, “and” and “so”, playing different roles for cases of argument mining and the need for world knowledge in the cases where a particular word normally positive can be seen as negative in a context. Ethos annotation and classification is directly related to this, where the need for world knowledge can hamper the automatic classification task which in some cases is obvious to human annotators.

Chapter 4

Data for Parliamentary Debates

This chapter outlines the domain and data source for annotating ethos. In order to provide the largest dataset for annotating ethos supports and attacks (see section 2.3.3), several decisions had to be made around the exact domain of debates; the structure of those debates; the time period; and the complexity of the language or the language structures for future automation. The choice and rationale in this instance are crucial due to the need of large annotated datasets for building any kind of automatic system using natural language. The data must also be generalisable enough to ensure that the results obtained for an automatic system are not too domain specific[1].

This chapter then describes the rationale behind the choices made for each of the points above: the domain of Hansard, the UK parliamentary debate record (see section 4.1); the structure of the oral answers to questions sessions in Hansard and the complexities of the language (see section 4.1); and, the time period 1979 to 1990 (see section 4.2). Finally the chapter outlines, how the data was obtained and stored (see sections 4.3 and 4.4) and the base set of annotation tags which are used throughout this thesis (see section 4.5).

[1]Domain specificity is a problem in any natural language processing system and therefore will need to be adapted for the particular domain.

4.1 Hansard, the UK Parliamentary Record

The domain of choice to conduct annotation and automatic extraction of ethotic supports and attacks is that of political discourse, specifically parliamentary debates. Parliamentary discourse provides a rich dialogue between several political entities (parties, politicians, government, agencies etc.) at one time where character appeals are consistently made to add support or attack what has been said. In the 21st century, the resources available for political discourse have increased rapidly through the mass use of social media, which provides a platform for discussion in terms of political events important to the public and as a means for politicians to interact with their constituents.

The rise of artificial intelligence and data mining techniques has also forced the need to publish parliamentary proceedings in a digital format as a step towards transparency and accountability. This process tends to produce a more structured output than what can be gained in social media discussions which are riddled with spelling and grammar mistakes, reactionary posts and unstructured formatting. On the other hand, parliamentary records provide a structured text, a structured format for the text including information on the topic of debate, and very few transcription errors.

The UK parliamentary debate record, Hansard (`http://hansard.millbanksystems.com/`), provides a rich resource for both manual annotation tasks and automatic classification tasks where the debate transcripts date back to, in the first instance, 1802 and continue to the present day with updates appearing daily. The data is also relatively well structured in that over the various forms of debate which occur within the UK parliament (oral answers to questions, committee meetings etc.) there is an agreed format, making the automatic scraping of the transcripts more simple (see figure 14).

In figure 14 the transcripts are split between the different houses of the UK political system, the House of Commons and the House of Lords. The house of interest in this study is the House of Commons as it is the main house in the UK which has the ability to create legislation. Each house is then split into several sessions for the day with the date specified. These sessions are then split into various topics ranging from internal issues such as farming to foreign affairs where topic consists of a transcript ranging in size from two-hundred words to more than a thousand. These transcripts are not considered as verbatim, although they are as close as possible to a verbatim report with only obvious errors fixed which do not add to the argument being put across, thus repetitions and stutters are removed unless they play a role in the overall debate.

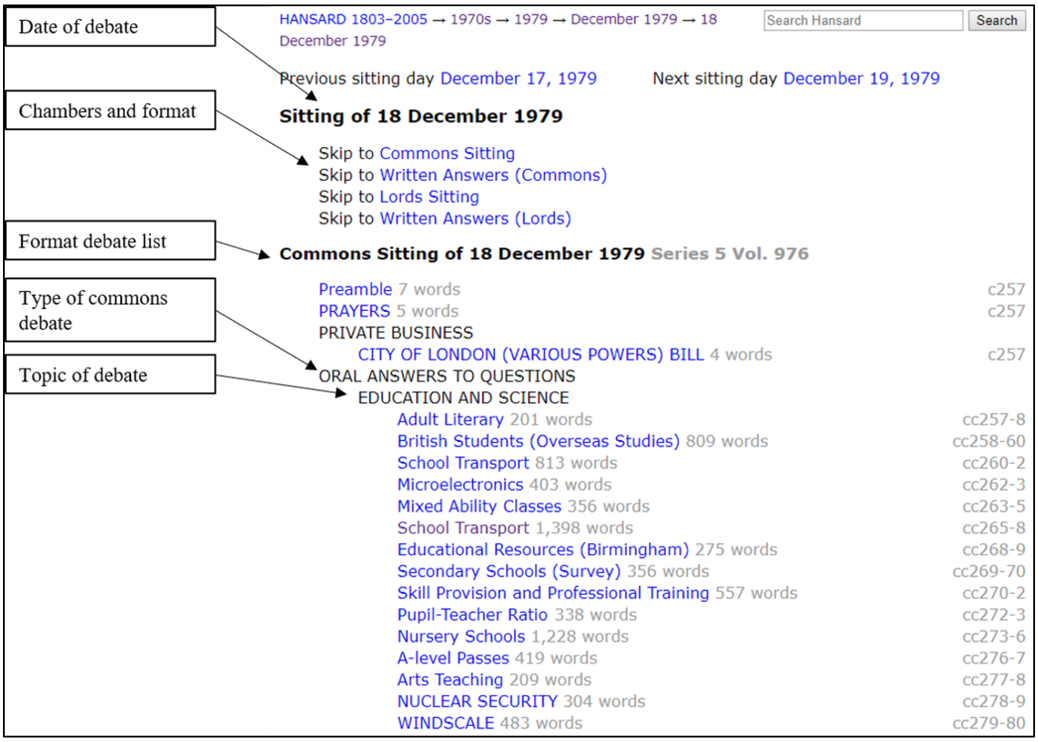

Figure 14: Structure of the Hansard millbank systems webpage containing parliamentary debates.

Throughout the record there is also a consistent speaker structure, illustrated in figure 15, attributing each utterance (sentence, paragraph or speech) to an individual normally identified by name and the constituency (the region of the country which has elected them to parliament) for which they are a member of parliament (MP) or through their position within the parliament (secretary of state etc.). If this is the first time the speaker is mentioned, then brackets with the speakers name are given. The speaker is then shown at the start of each utterance they make with replies beneath.

In this study, only one section of the UK parliamentary debates is utilised, that of oral

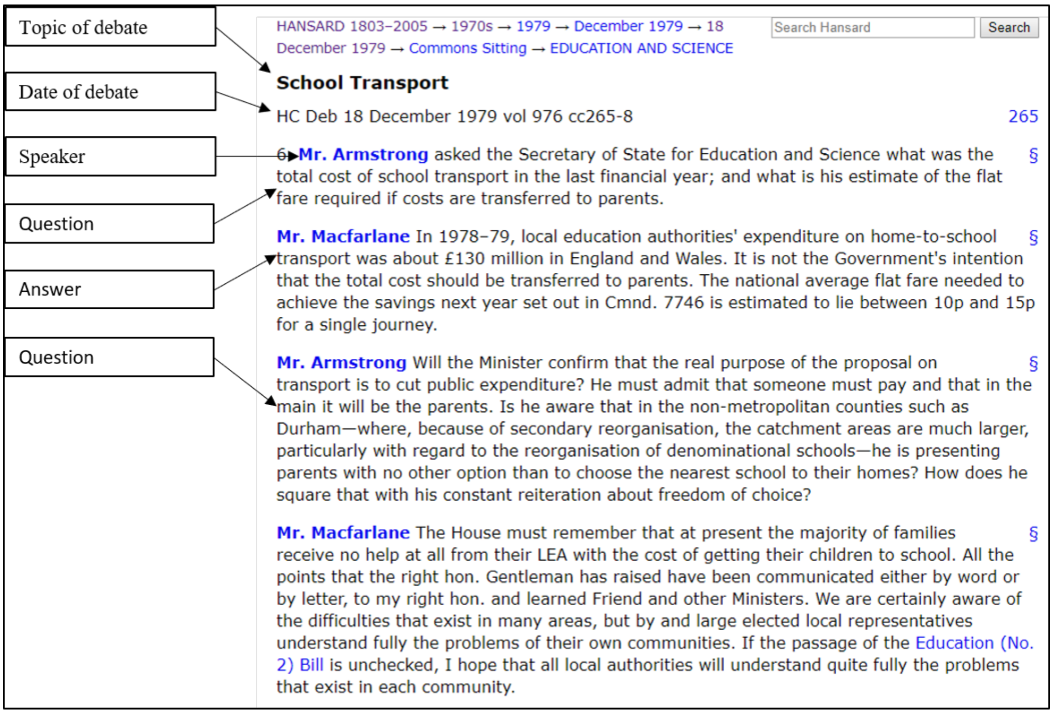

Figure 15: Structure of the Hansard millbank systems webpage containing a sample transcript from the oral answers to questions period.

answers to questions which constitutes members of parliament questioning government ministers on a wide range of topics from agriculture to foreign affairs. The oral answers session (which also includes the Prime Ministers question time - where the Prime Minister is questioned on government policies and issues) is the closest to an unconstrained dialogue, although some information is still corrected in the transcription process, for example, constituency names of the mentioned members of parliament especially in situations where the constituency could not be recalled.

The UK parliament rules for oral answers stipulate that the question must be addressed to another politician or the speaker. This enforces a common structure throughout the session of a question to a Government official seeking information, followed by a pre-determined statement, followed by a question normally asserting a point (figure 15 shows this structure of question-answer-question). Therefore there are many cases where an entity is mentioned yet the content will be neutral overall as it is purely information seeking (this is shown in the first question in figure 15). Making the problem of identifying ethos even more difficult is that the assertions may contain sentiment holding words. So a sentence can contain a mention of an entity, framed as a question, and surrounded by sentiment holding words and yet will not be ethotic.

The language of Hansard does create some issues when used in a natural language processing context. Particularly, the formality of the language means that any automatic extraction techniques, which focus on a bag of words, will not be directly generalisable, although this is an issue in all tasks of natural language processing. For example, traditions in the UK parliament are still up-kept meaning that instead of referring to another member of parliament by their name, the honourable (Hon.) member must be used instead (variations also include: hon. Lady, hon. Gentleman, Minister, Prime Minister, Chancellor, Opposition, Government, Lord and Dr.). There is also a reliance within the parliament on a legal style of language, that is the complex issues debated require a knowledge of various pieces of legislation which enforces a legal style when referencing this legislation. This mixes diplomatic language (rarely politicians will say exactly what they mean) and legal language (as this is the purpose of parliamentary debate to create legislation). A list of banned terms (e.g. "coward") within the parliament means that many politicians will reuse terms which do not have a negative connotation as having one. Despite these issues, making any pipeline built on this data generalisable to other domains would mean several out-of-the-box steps can be taken such as using Named Entity Recognition (NER) and anaphora resolution tools, and using a more specific training dataset.

4.2 Hansard Time Period

The chosen time period for the annotation and classification of ethos in this thesis is that of Margaret Thatcher's period as Prime Minister in the parliament, the longest serving Prime Minister in the 20th century and the 21st century, so far. This period of time in the UK parliament is considered to be volatile due to moving away from an industrial country, the "troubles" in Northern Ireland and the Falklands war. Thus the period, 1979 to 1990, was considered a good candidate for a source of ethotic appeals both positive and negative due to these events.

The constructs of the UK parliament mean that the party with the overall majority of seats form the government with the party with the second largest majority forming the official opposition. The role of the opposition is to create shadow government roles (appoint MPs to directly oppose Government positions), oppose and debate government legislation, and create alternative legislation than that offered by the Government. In the event of no party holding a majority, deals can be struck between political parties on the basis of a coalition Government, a supply and demand agreement (the agreed parties must vote with the Government on crucial legislation), or the largest party in the parliament may form Government with no majority or agreements in place, although this is rare. These constructs are incredibly important when tracking speakers, as they are sometimes referred to only as their position within the parliament. The person holding this position can change frequently.

During the period of interest of this research, 1979 to 1990, the Labour and Conservative parties consistently won the most number of parliamentary seats over the three general elections, followed by the Liberal coalition. The Conservative party won the 1979 general election taking power from the Labour party with a total of 359 seats (see table 3). In the 1983 election, the Conservative party increased their majority by 38 seats overall, producing 397 seats and the Labour party lost 52 seats leaving 209 seats. In the last election of this period, the Conservative party lost 21 seats with the Labour party, gaining 20 producing 376 and 229 seats respectively.

4.3 Obtaining Hansard Data

The data was obtained by performing divisions in Hansard (freely available at `http://hansard.millbanksystems.com/`) over the chosen time period of 1979 to 1990. Transcripts were initially scraped for the two end points of the period and then the midpoint.

Party	Seats Per Year		
	1979	1983	1987
Conservative	359	397	376
Labour	261	209	229
Liberal	9	23	22
SNP	2	2	3
Plaid Cymru	2	2	3
Other	17	17	17

Table 3: Number of seats won at the studied period between 1979 and 1990 in general elections in the UK.

Chapter	Corpus Name	Sessions	Words	Speakers
chapter 5	EtHan_Thatcher_3	60	70,117	253
chapter 6	Ethos_Hansard1	90	90,991	198

Table 4: Total volume of sessions, words and speakers for each of the two main datasets in chapters 5 and 6.

The midpoint was taken from each remaining portion, this continued until a set of 60 transcripts were extracted. This set was then extended by a further 30 transcripts using the same dates, but, with different topics of debate (see table 4 for the datasets used in chapters 5 and 6). A transcript was only used if it had greater than 500 words, this was to ensure a question and reply structure which is not always present in smaller transcripts. The data was selected by taking a random sub-sample of Hansard according to the following rubric: select the first two House of Commons debates over 500 words in length from the day closest to the date(s) at the midpoint(s) of the largest uninterrupted date range(s) (initially the midpoint in the range 4th May 1979 and 22nd November 1990 - viz., 11th February 1985; then at the midpoints between 4th May 1979 and 11th February 1985, and between 11th February 1985 and 22nd November 1990, etc.).

The data was split over the full time period to ensure a range of politicians were involved in the data, thus making the outcomes more generalisable for other speakers. The split over the full period can also be used to confirm that ethotic statements were not limited to a certain period of parliamentary discourse. Finally, the full period means that a range of topics can be covered which is crucial for any machine learning applications. In any bag of words approach, there may be appeals to ethos in a particular context, such as farming or war, taking the same volume of transcripts over consecutive days would not provide a diverse range of topics, which splitting by time period allows.

For chapter 8, the datasets were again further extended multiple times through automatic annotation where dates can be specified to scrape a full set of Hansard sessions and where no word minimum limit was applied. This means a much larger volume of transcripts can be extracted from the process of ethos mining allowing large scale data analysis.

4.4 Storing Hansard Data

To ensure that the data pulled from Hansard remains open to the public, a set of tools for annotation, storage and analysis are required as well as a theory to annotate ethos supports and attacks. To adhere to each of these steps, Inference Anchoring Theory (IAT) was used to annotate supports and attacks of ethos in Hansard (see section 2.3). The advantage of IAT is its flexibility as a theory, allowing the creation of ethos nodes which can be supported or attacked which can then be later adapted to include full argument structures. As a consequence of this choice the full suite of tools associated with the Argument Interchange Format (AIF) (Rahwan et al., 2007), which supports IAT, can be used. The goal behind AIF is to develop a theory free language which contains an ontology to express argumentation. IAT then builds upon AIF using this ontology to develop the theory using many of the node types which are defined in table 5. In the case of annotating ethos supports and attacks, the argument schemes RA and CA, for denoting a relationship between two I-nodes, are used. In the case of the conclusion of ethos, e.g. “speaker has ethos” defined in chapter 2 an I-node is used.

The data can then be analysed using the OVA+ annotation tool (Janier et al., 2014) (freely available at `http://ova.arg-tech.org`) which takes the original text of a transcript and allows the creation of each of the AIF node types with the content of the text (see figure 16). The text on the left side of the transcript is highlighted and a

Node	Full name	Category	Node	Full name
I-node	Information (propositional contents)		I-node which is not L-node	contents of locutions
			L-node	locutions
S-node	Schemes (relations between contents)	Argument schemes	RA	support (inference)
			CA	conflict
			MA	rephrase
		Illocutionary schemes	YA anchored in L	illocutionary connections associated with locutions
			YA anchored in TA	illocutionary connections associated with transitions
		Dialogue schemes	TA	transitions

Table 5: Types and sub-types of nodes (vertices) in graphs represented according to the Argument Interchange Format standard; their full names; and the categories of schemes.

node produced on the main canvas following a click with the option to create an IAT structure instantly. Following an annotation session the underlying IAT structure can be stored in the AIFdb database (Lawrence et al., 2015) (`http://aifdb.org`) which stores all annotation in a unified way creating a graph structure. This means that a single argument annotated multiple times will not have repeated entries in the database and will instead be uniform allowing a linkage of the arguments of many people. As an interface into AIFdb, AIF corpora allows the creation of publicly accessible datasets which can be downloaded in several data formats for multiple use cases such as ethos or argument mining (`http://corpora.aifdb.org`). Figure 17 shows a corpus name, a description of the contents of the corpus and the volume of annotation maps which are contained within the corpus.

Whilst other tools and theories would allow the annotation of ethos and arguments as a whole, none have such an array of tools backing the theory and therefore would not be applicable to this research.

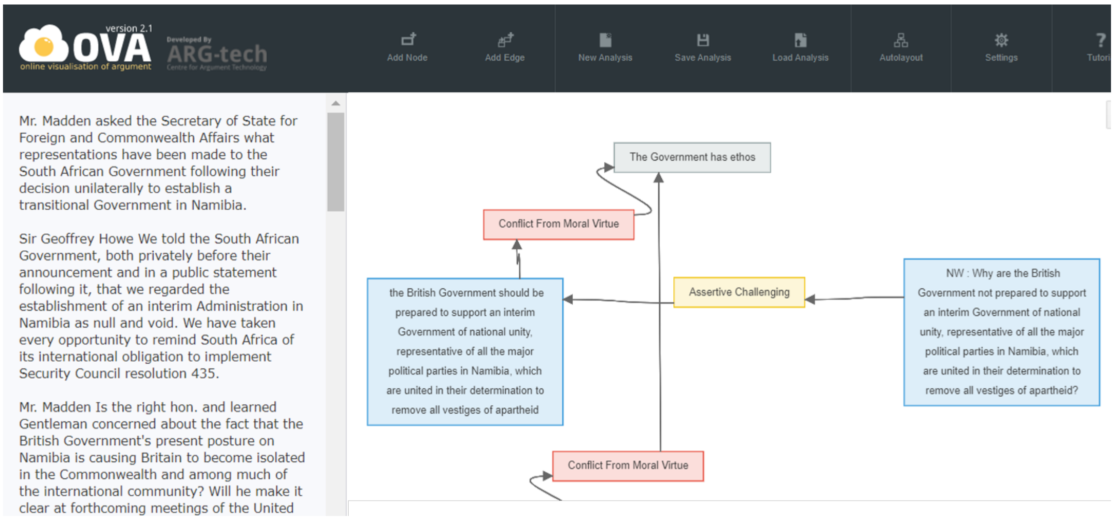

Figure 16: Screenshot of OVA+ with the original text from a transcript on the left side and the annotation on the right.

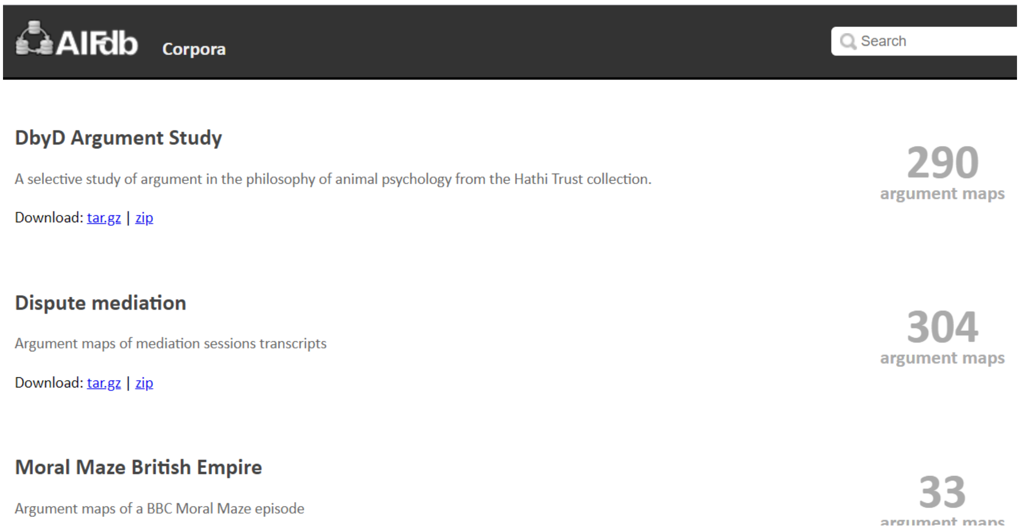

Figure 17: Screenshot of AIF corpora an interface to AIFdb.

4.5 Annotating Hansard Data

Throughout this research, a base set of annotated data was used to perform ethos mining which took only the locutions within IAT for automation, for two reasons. Firstly, so that the task in this thesis focussed solely on the automatic extraction of ethos and therefore tasks such as automatic illocutionary connection detection and propositional content reconstruction did not need to occur. Secondly, this meant no text segmentation need be undertaken and instead the annotated text in a locution used. This makes a distinction between ethos mining conducted in this research and some of the issues shown in argument mining. Manual or automatic segmentation errors can be passed through any kind of automatic pipeline making the problem of extraction more difficult. As this is the pioneering work in ethos mining, like in the pioneering work of argument mining (Moens et al., 2007), the task of segmentation has been removed.

In order to annotate ethos from natural language text, a set of base tags were created which form the minimum annotation for all the tasks conducted in this research. Sentences comprise the smallest units where the task then becomes identifying the source speaker of each sentence, the target speaker of each, whether or not this contains a reference to ethos and if that reference is positive or negative. All other sentences are then considered as not containing ethos for training and testing purposes.

The annotation on each corpus (see table 4) then applied four tags (specific guidelines are given in chapter 5 and 6):

Source person is used to mark a person who utters the statement.

Target person is a person (or entity) who is described by the statement.

Ethos support should be identified when the statement makes explicit mentions of a person, organisation or other entity in a positive frame (see figure 11 for an ethos support).

Ethos attack should be identified when the statement makes explicit mentions of a person, organisation or other entity in a negative frame (see figure 12 for an ethos attack).

The statements in which speakers support or attack others are called **Ethotic Sentiment Expressions, ESEs** and the statements which do not contain reference to others are denoted as ¬**Ethotic Sentiment Expressions,** ¬**ESEs**. The polarity of these statements is then expressed by the use of abbreviation **+ESE** for positive sentiment (ethos support) and **-ESE** for negative sentiment (ethos attack) or +/-ESE to describe both [2].

Overall the annotation of ethos is difficult, as is described in this chapter and will be shown in chapters 5, 6 and 7. In some instances the construct of the transcript makes the annotation task very simple, for example determining the source of an utterance. On the other hand, determining the target of an utterance can be very difficult, because of the use of pronouns or general terms of reference. This process is sometimes made easier as] when a speaker is first mentioned so is their constituency and therefore this can be used for identification purposes. Many references to ethos also rely upon some implicit information (see appendices .A and .B for annotation examples) referring to past legislation or to current and past events that hold a specific connotation and therefore some real world knowledge may be required to annotate ethos.

4.6 Discussion

The domain of parliamentary debate was specifically chosen as a source of data due to the large volume available in a format which can be easily scraped.

By using Hansard for the task of ethos mining, the ethotic relationships can be summarised through extraction particularly in historical data where the outcome can be verified against the media and outcomes of the time. Hansard in particular is more reliable than other sources of political debate, for example television debates. This can be attributed to the rules which are in place in parliament allowing speakers the time to speak over a more extended period of time, not in-front of larger audiences and not in a setting where they

[2]The annotation is visualised as directed graphs where support is marked as Default Inference, attack as Default Conflict, source-person is in the node with the statement, and target-person – in the node which refers to ethos.

are likely to be interrupted without a moderator (the speaker in Hansard) intervening. Also given is an extended period of time to prepare for debates as questions are sent prior to the debates occurring (although this only holds for the initial question in a debate).

The data does come with challenges, however, with speakers not identified using names, the language being more formal than is usual for debate, and the rules which define the discourse preventing specific terms. Although these challenges remain, the advantage of the data lies in its consistency, particularly with the time period chosen of 1979 to 1990. This is due to the same Prime Minister being present over the whole period and a relatively constant set of politicians contributing. This ensures that the language is persistent despite the long period of time and that there is some continuity in the topics discussed. This time period then allowed for the selection of two datasets for the purposes of ethos mining, annotated using the suite of tools associated with the AIF which ensure publicly available resources in an easy to view format. A further advantage of parliamentary debate, and particularly Hansard, is its continually expanding set of data with new transcripts being added every day that the parliament sits. This means that any automated solutions can continue to run providing new summary information.

In summary this chapter has contributed: a description of parliamentary debate, the domain in use for this thesis, Hansard, and the particular parliamentary session, oral answers to questions; a description of the complexities of the language used in parliamentary debate; a rationale for the chosen time period for the creation of datasets; details of how the data was obtained; an explanation of the tools and ontology used for storing the datasets; and finally defined are the base tags used for annotation ethos supports and ethos attacks from this data.

Chapter 5

Ethos Mining: Domain Specific Rules

This chapter defines the first steps in the newly established area of ethos mining and describes the first corpus focussed entirely on ethotic structures in line with RQ1 and RQ2. [1]. More specifically, the corpus creation task applies the base tags established in chapter 4 and an adapted definition of Aristotelian ethos in parliamentary debate (see also chapter 4 for a description of the domain).

Ethos is defined as a support of or attack on an entity which is the participant of communication (see section 2.3.3). In this context, the identification of ethos will aid in the understanding of how it is used to influence in a persuasive sense in interactions between multiple entities. Particularly important for ethos within politics is knowing who supports whom (see example 11); who attacks whom (see example 12); and which political party the person represents (e.g. supports or attacks from parties on different ends of the political scale could indicate an interesting dynamic, such as politicians who cross party lines).

(11) **Mr. Chris Patten said,** *The hon. Member for Falkirk, East (Mr. Ewing) in his admirable speech, put the position much more clearly than I could.*

(12) **Mr. Giles Radice said,** *In doing so he (Mr. Pawsey) failed to face up to his*

[1]The work in this chapter extends a published work (Duthie et al., 2016a) where ethos was annotated in sentences from the UK parliamentary record and then automatically extracted. Some passages are used verbatim from the source. This work was in collaboration with Duthie's PhD supervisors, Katarzyna Budzynska and Chris Reed who contributed extensive feedback and general ideas for implementation.

responsibility both to the House and to the schools of England, Scotland and Wales.

To mine ethos, a pipeline of natural language processing techniques is used to extract information from the linguistic surface. For example, the phrase, “admirable speech” in example 11 can be used to support Mr. Ewing’s ethos, while “failed (...) to his responsibility” can be used as a cue to determine that Mr. Pawsey was attacked.

This task, however, is particularly difficult due to parliamentary language (see chapter 4) where the list of banned terms within the parliament (e.g. “liar”) forces the re-purposing of words (e.g. “terminological inexactitude”) meaning that domain specific lexicons have to be created especially in the case of attempting to determine the polarity of a sentence. Several other challenges also have to be addressed. For example, the dialogical context encourages the use of pronouns (see “he” in example 12) so anaphora resolution has to be performed. Reported speech (see example 13) includes references to other people which are ethotically neutral and therefore such references should not be considered for classification. Also some phrases, which seem positive such as “honorable”, are in fact a part of political etiquette and thus sentiment analysis methods for determining polarity cannot make use of this data.

(13) **Mr. Giles Radice said,** *The hon. Member for Rugby and Kenilworth (Mr. Pawsey) said that in the United States and Australia this was a local decision.*

Also present within these sentences are sentiment holding words which are not targeted at a speakers ethos. This makes identification of sentences which hold ethos even more difficult and therefore basic sentiment classification or opinion mining techniques alone will not perform well enough on this task (see section 5.4).

This chapter then encompasses descriptions of: the first annotation and corpus of ethos supports and attacks (section 5.2) and its evaluation (section 5.2.1); the first rule-based NLP pipeline for ethos mining and a comprehensive evaluation of this (section 5.3 and 5.4); and a discussion of the potential improvements, or changes which can be made to both the manual annotation and the automatic classification (section 5.5).

5.1 Annotation

In order to annotate ethos from natural language text a set of annotation guidelines were created which build upon the tags specified in chapter 4. The annotation task then becomes identifying the source of each sentence, the target of each, whether or not this contains a reference to ethos and if that reference is positive or negative.

The annotation was performed by applying each of the four tags specified in chapter 4 according to the following guidelines for ethos supports and ethos attacks:

Source person is used to mark a person who utters the statement.

Target person is a person who is described by the statement.

Ethos support should be identified when: (*a*) the statement makes explicit mentions of a person, organisation or other entity (excluding groups and assemblages) except when this is reported speech; and (*b*) it takes the form of supporting a person's credibility or looking to put them in a positive frame through character supports or supports of work; and (*c*) a support to a person's own ethos should not be analysed as this is deemed to be a fallacy (Budzynska, 2012). See example 11 for an ethos support.

Ethos attack should be identified when: (*a*) the statement makes explicit mentions of a person, organisation or other entity (excluding groups and assemblages) except when this is reported speech; and (*b*) it takes the form of attacking a person's credibility or looking to put them into a negative frame; or (*c*) it may take the form of trying to unbalance authority on a subject giving the attacker more of a right to talk about the subject. An attack of one's own ethos would be a rare occurrence in any speech especially in a political context and is therefore not considered. See example 12 for an ethos attack.

Corpus	Sessions	Words	Segments	+ESE	-ESE	Speakers	Location
Train	30	40,939	387	96	291	127	http://arg.tech/Ethan3Train
Test	30	29,178	352	80	272	126	http://arg.tech/Ethan3Test
TOTAL	60	70,117	739	176	563	253	

Table 6: Summary of the language resources in the EtHan_Thatcher_3 corpus for mining ethos in Hansard.

5.2 Corpus

The domain of choice for ethos mining is Hansard, the UK parliamentary debate record, spanning the time period 1979 to 1990, in which Margaret Thatcher was Prime Minister in the UK (see chapter 4 for a description of the rationale). Following the annotation tags, specified in chapter 4 and the annotation guidelines above, a corpus, EtHan_Thatcher_3[2] (see table 6), of manually annotated ethos supports and attacks was created. A selection rubric (described in section 4.3) yielded 60 transcripts, the data in each of which was then cleaned such that any titles, section markers and unwanted information was removed to leave only the speakers, organisations or other entities and the statements they had made. In addition these 60 transcripts were then split evenly to give a training set of 30 transcripts and testing set of 30 transcripts. The training set formed the training data for the sentiment polarity classifier and was used as the basis for developing domain specific rules for recognising ethotic sentiment expressions.

The training and test sets combined then gave a total of 739 ethos supports (24%) and attacks (76%), with 1,194 unique tags used overall (see table 7 for their frequency) showing a balance in the number of unique speakers and unique targets for each statement. Despite this volume of unique speakers there is still an overlap in speakers between the train and test sets although the speakers are not used in any classification meaning that this overlap does not play a role. In the case of ethos supports and attacks the data is unbalanced, although this is to be expected as the official opposition's job is to hold the Government to account. Despite a support being less frequent in the data they may hold

[2]The corpus is named as so due to the annotation of session transcripts at different time periods. EtHan_Thatcher_1 containing an original 30 sessions which was extended to EtHan_Thatcher_3 and EtHan_Thatcher_2 containing a subset of EtHan_Thatcher_3 for agreement calculations.

Tag	Source	Target	Ethotic Support (+ESE)	Ethotic Attack (-ESE)	Total
#	243	212	179	560	1,194

Table 7: Occurrences of tags in EtHan_Thatcher_3.

more weight than attack when evaluating the outcome of a speakers ethos in debate. The task of evaluating the effectiveness a support or attack on ethos, however, is reserved for future work as it is out of the scope of this thesis. Overall the train and test sets produce imbalanced data with the volume of ESEs 387 and 352 respectively in comparison with 493 ¬ESEs in the former case and 852 ¬ESEs in the latter.

The number of ethotic statements, 739 (26%), is also imbalanced when compared with ¬ethotic statements, 2,085 (74%). As a consequence any proposed machine learning system will need to address this class imbalance which can be achieved by reducing ¬ethotic statements or artificially increasing the number of ethotic statements as is shown in 5.4 [3]. In the case of a purely rule based system, however, this same problem does not occur as the volumes of data do not play a role in the classification as is the case in a standard bag-of-words (BOW) model.

5.2.1 Inter-Annotator Agreement

In order to evaluate the annotation process, a subset of data used in the EtHan_Thatcher_3 corpus was selected. The selection followed the same method as applied to the whole dataset. The total size of this subset comprises 10% of the EtHan_Thatcher_3 corpus with 6 sessions containing 7,267 words, 91 segments and 30 speakers.

The inter annotator agreement (IAA) was calculated for two coders (see table 8 and figure 18 for a confusion matrix) where Cohen's kappa for recognising whether the statement is ESE or ¬ESE gave the value of $\kappa = 0.67$. For annotating ethotic statements as a support or attack, $\kappa = 0.95$[4]; for the source-person who utters an ethotic statement,

[3] It should also be noted that other methods can be utilised to combat imbalance such as regularization, error penalties, and adding class weights.

[4] All support and attacks comparisons are relative to the identification of ethos meaning that only when the annotators agree on the ethotic statement was the support or attack value compared.

<table>
<tr><th>Tag</th><th>Source</th><th>Target</th><th>Ethotic Support (+ESE)</th><th>Ethotic Attack (-ESE)</th></tr>
<tr><td rowspan="2">Kappa (κ)</td><td rowspan="2">1</td><td rowspan="2">0.84</td><td>0.95</td><td>0.95</td></tr>
<tr><td colspan="2">0.67</td></tr>
</table>

Table 8: Cohen's κ in EtHan_Thatcher_3.

$\kappa = 1$; and, for the target-person it was $\kappa = 0.84$. The two coders were experienced in argument annotation, with one a native speaker of English. For the purposes of this task a single session was used for training after-which the second coder annotated the six sessions. Cohen's kappa was specifically chosen for the IAA task due to the difficulty of annotating ethos in natural language where agreements only by chance are more likely, especially for a binary annotation. Percentage agreement, although more favourable in the case of the score, does not take into account any chance agreement and therefore may over exaggerate any agreement between the annotators.

The kappa score, when determining if a sentence is ethotic or not shows that the annotation is reliable overall (defined as substantial by Landis and Koch (1977)). Many errors in the case of this decision pertain to the framing of the statement. In nearly all error cases the opening sentence of a statement is a question, 269 of the total of 739 ESEs. The questions in these cases make it difficult to ascertain the assertion being made, for example, the use of "Is he/she aware", "Will he/she" at the start of the statement mean that unless there is an explicit second mention of another entity (e.g. "he" or "she" is mentioned later in the sentence) then it can be hard to determine if the statement is ethotic or not. Each of these starting points of questions can be rephrased to "he/she is aware" or "he/she is not aware" and "he/she will" or "he/she will not". On the other hand, each of these statements can easily be interpreted as information seeking and therefore do not need to be rephrased thus making this distinction difficult which leads to errors in the annotation.

In the case of determining the source and target of an ESE the kappa value is perfect for the former, as expected, and near perfect for the latter. Errors in this case pertain to multiple entities being present within a sentence thus creating a decision for the annotator as to who the target should be. This is a common occurrence with a reference made to

Ann 2

Ann 1

	ESE	¬ESE
ESE	55	36
¬ESE	4	281

Figure 18: Confusion matrix for the ESE/¬ESE annotation.

one speaker, but then an attack or support is produced for another speaker. Although one potential solution to this problem would be smaller units of segmentation this would require a further task in the annotation stage, of which the relative errors produced may outweigh those in the case of target identification (see chapter 3.3.1.4 for a description of segmentation and chapter 4 for a rationale behind the use of sentences in this thesis). In other cases, the entity mentioned is rather a group of unidentifiable entities. These mentions tend to use language such as "we" and "hon. Members". In both of these cases it is almost impossible to ascertain which group is being addressed and who belongs to it and therefore in these cases ethos should not be annotated.

5.3 System Architecture

To automatically extract ethos in natural language, an ethos mining pipeline was constructed which makes use of existing NLP methods and novel modules to address the complexities of the language in the UK parliament which are described in chapter 4. As defined in the introduction to this chapter, the ethos mining process requires: anaphora resolution, reported speech resolution and the identification of domain specific rules for the parliamentary language. The NLP pipeline combines a rule-based approach and a machine learning approach in the case of sentiment analysis.

In the former case a simple feature representation approach is not considered effective due to the specific nature of the language which can only be discovered through a manual rule-based approach which allows the nuances of the language to be discovered. Whilst rule-based approaches are less common within text classification tasks, they provide a basis for any improvements using machine learning, although the BOW approach is effective in showing the difficulty of a task like ethos mining.

The architecture of the software system for mining ethos consists of three stages, five layers and eight components (see figure 19). The three stages consist of the ESE / ¬ESE stage, the +/- ESE stage and the network stage. The ESE / ¬ESE stage takes an input of cleaned text transcripts from the EtHan_Thatcher_3 corpus and classifies each segment as either an ESE or ¬ESE. The +/- ESE stage then gives the polarity of ESEs, ESEs with positive sentiment (corresponding to ethos support), and ESEs with negative sentiment (corresponding to ethos attack). Finally, the network stage provides a visualisation of all ESEs as edges between each participant, represented as nodes, in the debate (the network stage is described in chapter 8 as it is rather considered as an application of ethos mining).

In the ESE / ¬ESE stage, there are three layers consisting of five components. The parsing layer uses plain text from the EtHan_Thatcher_3 test sub-corpus and applies three different methods to it: Named Entity Recognition (NER), Part-Of-Speech (POS) tagging and a set of domain specific rules. The output is Agent Reference Expressions (AREs) which are any statements referring to another person, organisation or agentive entity. Given the dialogical nature of the material, many statements do not refer to the target-person by their name explicitly, but by a pronoun (see "he" in example 11) or by a region the MP represents (see "The hon. Member for Falkirk, East" in example 12). Thus, AREs are then passed to the anaphora layer where both source-person and target-person of the statement are retrieved from the original text. The next challenge is that reports of what has been previously said can be ethotically neutral, especially when an MP refers to a statement which was proposed earlier in the debate (see example 13). Therefore, full AREs are passed to the reported speech layer where an ARE is removed if it is not an ethotic expression, but a reported speech.

In the +/- ESE stage, there is one layer, the sentiment layer, containing three components: the sentiment classifier and two word lexicons. The sentiment classifier and word

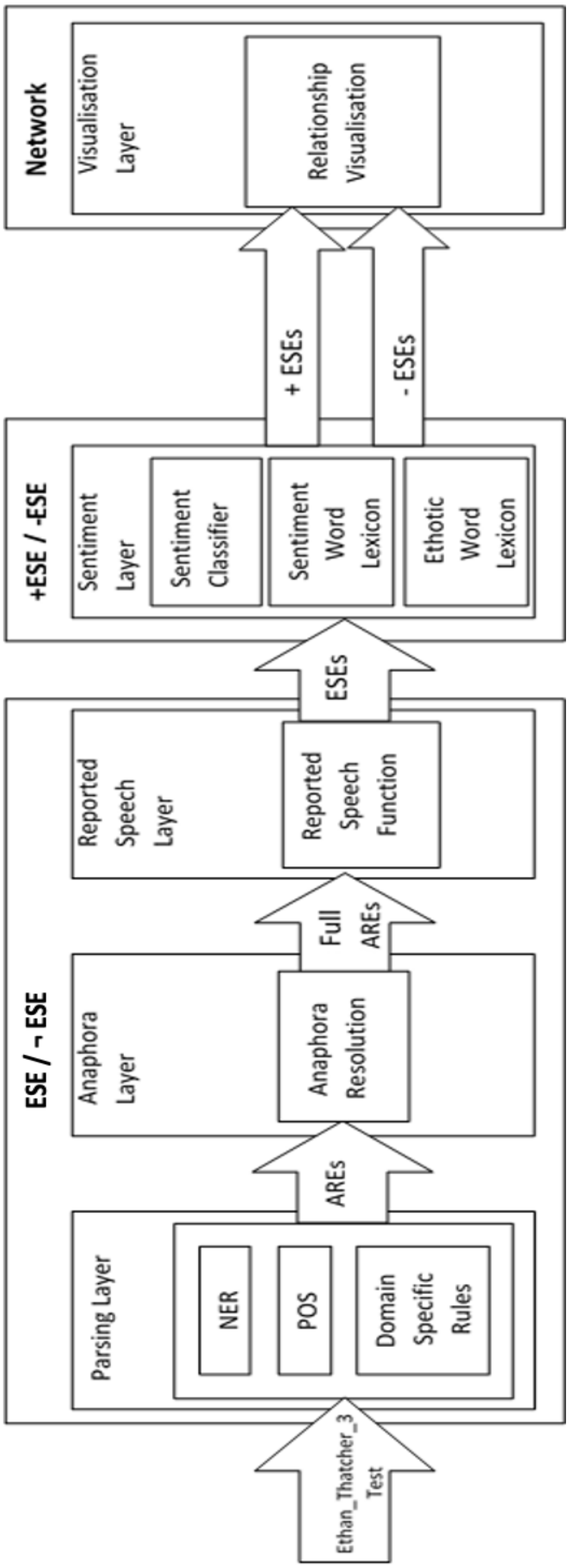

Figure 19: A text analysis pipeline for ethos mining: the extraction, polarisation and networking of ESEs from Hansard sessions in plain text transcripts.

lexicon components combine to classify ESEs as positive and negative. These two sets are then passed to the Network stage where the visualisation layer displays relationships between people, organisations and other entities.

Not shown in figure 19 is the training sub-corpus which is used just for defining domain specific rules in the ESE/¬ESE stage and the lexicon in the +/- ESE stage for the sentiment classifier. The techniques of domain specific rules, anaphora resolution, reported speech function and relationship visualisation were developed specifically for the tasks of ethos mining in parliamentary debate, and the method of sentiment classification was extended with the development of a lexicon to account for the characteristics of the domain. The ethos mining pipeline also applies existing NLP methods such as Part-of-Speech tagging, Named Entity Recognition and an SVM-based sentiment classifier with an existing sentiment word lexicon.

The remainder of this section is broken down into existing methods, those which have been used without any alteration, adapted methods, those which have been adapted for use in ethos mining, and novel methods, those which have been developed specifically for ethos mining.

5.3.1 Existing Methods

Existing NLP methods used in the pipeline consist of Named Entity Recognition and Part-Of-Speech tagging.

5.3.1.1 Named Entity Recognition (NER)

NER, uses the Stanford Named Entity Recognizer (Stanford NER) (Finkel et al., 2005). Its goal is to extract statements which contain names, organisations and locations from the plain text. This is applied to the original text from EtHan_Thatcher_3 corpus and produces a set of AREs on the assumption that any specific statement made to a named entity can in fact be a form of ethotic statement. This step aids in the reduction of the total number of sentences which contain no references to entities.

The Stanford NER module was specifically used due to its flexibility (the Stanford NLP tools can be imported into a number of programming languages as libraries or interfaced easily in languages where there is no library available) and the performance on the benchmark datasets. Although the datasets on which the NER module is trained vary from parliamentary data, an assumption can be made that in general entities of interest will be referred to in the same way.

5.3.1.2 Part-of-Speech (POS) Tagging

POS tagging, uses the Stanford POS Tagger (Toutanova et al., 2003). It is applied to extract statements which contain pronouns to account for situations such as in example 12. It uses a word based approach where a bidirectional dependency network is created and lexical cues are used in order to define the POS features. The POS Tagger then tags every word with its part-of-speech allowing for specific values to be searched for which can show names, organisations or other entities. Proper nouns were used as indicators of this, allowing the full segment to be extracted.

This was applied to the EtHan_Thatcher_3 test corpus and then run against the list of already extracted AREs from the NER to account for any duplicate segments extending the list of AREs. Again the Stanford POS tagger was used due to reasons of flexibility and the overall performance on the benchmark datasets, and although these datasets vary from parliamentary debate the assumption made is that sentences should be constructed in a similar fashion and therefore performance will be accurate.

5.3.2 Adapted Methods

Methods which have been adapted specifically for the ethos mining pipeline consist of sentiment classification.

5.3.2.1 Sentiment Classifier (SVM, NB, ME)

To perform sentiment analysis three machine learning algorithms were considered: Support Vector Machines (SVM), Naïve Bayes (NB) and Maximum Entropy (ME). A C-Support Vector Classification (C-SVC) algorithm (Chang and Lin, 2011) from the LIBSVM library was used to classify ESEs into two sets: positive and negative. To perform Naïve Bayes and Maximum entropy, the Stanford classifier library (Manning and Klein, 2003) was used [5]. In selecting these methods, we followed the conclusion formulated in (Onyimadu et al., 2014) that the discourse approach in sentiment analysis is not satisfactory and that supervised learning techniques are needed (which is demonstrated in (Pang et al., 2002)). Each algorithm was also chosen as linearly classifying the data is sufficient for sentiment analysis. Manually defined rules were not considered due to the advances in sentiment classification through machine learning methods with lexicons.

The lexicons (defined in section 5.3.3.4) were passed to the Stanford CoreNLP (Manning et al., 2014) library in order to perform lemmatization, allowing the frequency of words in the lexicon to be more accurately calculated using the morpohological base of lexemes (lemmas) by removing morphological inflection. POS tagging was again used in order to remove words representing any names, locations or entities. Names of MPs are stated throughout the manually tagged sentences, if an MP were to be tagged in a higher frequency of negative segments than positive segments this could influence the overall classification of a sentence. The bag-of-words approach was used to format training data with unigrams, bigrams and trigrams extracted from the manually tagged sentences.

The relationship between source and target were mapped through ESEs which allowed the frequency of attacks and supports of ethos to be modelled. The information of each segment, source, target for the statement and the ESE, including classification of positive and negative and the frequency of attack and support of each speaker, were added to a JSON file for use in the Relationship Visualisation stage (this is described in the applications of

[5] See also `https://nlp.stanford.edu/wiki/Software/Classifier` for the implementation of both classifiers.

ethos, chapter 8).

5.3.3 Novel Methods

Novel methods for the ethos mining pipeline consist of domain specific rules, anaphora resolution, a reported speech function, and sentiment lexicons, which have been developed specifically for this research.

5.3.3.1 Domain Specific Rules (DSR)

Rule-based expression recognition was developed for ethos mining to account for the specific language of the domain. In the House of Commons, the speaker and MPs are not allowed to refer to any other MP by name, but by phrases such as “Honourable Gentleman” or “Honourable Lady” or by using the constituency name of an MP such as “The hon. Member for Falkirk, East”. Organisations can also be mentioned under a different name, e.g. “the Government” will refer to the party in charge of the government at that time, and “the Opposition” – to the current official opposition. Smaller political parties will also not appear in training datasets for the NER module, for example “Plaid Cymru” or “SNP” which are specific to the UK. In these cases the only appropriate avenue for identification is the use of a rule-based approach, as the nuances of the data may not be identified by an automatic systems unless a comprehensive training dataset is created.

These rules are then extended with the creation of a list of ethotic words to determine if ethos is held in a particular ARE. A list of 326 ethotic statements were compiled from the EtHan_Thatcher_3 training corpus, containing some words not normally used in day-to-day conversation such as “penny-pinching” and “gerrymandering”, common with ethotic attacks. Again the new AREs produced from this component are checked against the list of already extracted AREs to remove duplicates. The assumption being that in some cases the NER module will miss the entities refereed to due to the peculiar language within the parliament.

5.3.3.2 Anaphora Resolution (AnaR)

This module was developed using manually defined rules to reconstruct all sources and targets in each AREs. For the source-person, the reconstruction is needed, when a sentence is not the first one in a turn in the dialogue (a turn corresponds to a paragraph in the transcript). In such cases, first the system associates a sentence with a paragraph. Since paragraphs are assigned a source-person, thus this person becomes a source for the sentence. For a target-person, there are several considerations that have to be made due to the complexity of the task, especially the mixing of pronouns with domain specific entities and those which are identified through NER. In (Mitkov, 2002) it is suggested that three steps must be taken to perform anaphora resolution, all candidates must be identified, impossible candidates must be removed through rules and a heuristic value applied to candidates. Within Hansard this task is made easier due to mentions of MP's, constituencies and the need to address an MP in a specific way removing the need for heuristic values. In the case of AREs produced from NER this is a simple task. The NER produces names of the target of the ethotic statement. This is applicable in cases such as "MP_1 said *MP_2 did this and he did that*" where MP_2 is identified through NER. A similar situation was produced in the case where the domain specific rules identify "Honourable Members" mentioned with a constituency location. The location can be extracted by performing NER on the segment and from this an MP's name can be applied to the given location. In the case of AREs produced by POS tagging and other domain specific rules further anaphora resolution is required.

In these two cases segment ID's related to each ARE produced to find each segment location within the document as a whole. The segments were traversed backwards to find earlier segments where NER was applied to get names. Previous segments can be utilised for anaphora resolution due to the structure in which POS tagged AREs and Domain Specific Rule AREs are applied. When pronouns are used to refer to a person, it implies earlier in the text the name of the person mentioned must have been said. This does not apply to all areas of debate, as hand gestures could be a reference to a person, but due to the rules of the House of Commons gesturing is not allowed. In cases such as "MP_1 said *he did this*" the method of backward traversing is applied. The same can be said for identifying the target from a Domain Specific Rule ARE. Again NER is applied to the previous segments to determine a speaker name if no name is found then the segment before is used.

5.3.3.3 Reported Speech Function (RSF)

A reported speech filter was developed which aims to remove segments containing neutral reports of what previously has been said by other speakers (thus no ethotic sentiment). The technique uses lexical cues such as "says", "you say" and "told me", and any segment containing these words is removed from the list of AREs. Any ethotic structure would instead be present in segments following reported speech where an organisation, person or other entity may be commented upon. The reported speech function then produces a list of ESEs which are passed to the sentiment classifier.

5.3.3.4 Lexicons (SWL, EWL)

To provide a lexicon for sentiment analysis one existing lexicon was used, the sentiment word lexicon (SWL) (Hu and Liu, 2004), and one lexicon was created, an ethotic word lexicon (EWL). The SWL, contains 2,006 words tagged as positive and 4,738 words tagged as negative. The EWL is a set of keywords developed using the EtHan_Thatcher_3 train sub-corpus containing 381 tagged sentences with 96 positive and 285 negative from which unigrams, bigrams and trigrams were extracted. Despite the relatively small volume of this set, its advantage lies in its adaptation to sentiment related specifically to ethos in parliamentary debate. The removal of non-sentiment bearing words, named entities, and the use of n-grams (uni-grams, bi-grams and tri-grams) in a BOW approach gave 32,858 features to be used as training data for machine learning.

5.4 Results and Evaluation

Results are reported for the anaphora layer, and the two stages of ethos mining shown in figure 19, the ESE / $\neg$ESE stage and the +/- ESE stage. A result is also given for the combination of these stages.

5.4.1 Anaphora Resolution

On the task of anaphora resolution, results are only provided for the task of determining the target of each ethotic statement as the source was provided in the transcripts, and methods used to tie sentences to full turns, meaning that the accuracy is 100%. On determining the target of an ethotic statement, the anaphora layer gave an accuracy of 71%, tested against a single target for each of the 352 statements in the test set. This compares with a Cohen's κ of 0.84 for the manual annotation, showing that the task of anaphora resolution is intuitive for human annotators, as it requires knowledge of English language terms and context which are not easily gained by machines. 71% accuracy in this instance shows results are well above random and considering the difficulty of the task are sufficient.

In this case errors pertain to the difficult task of determining if a pronoun, "he" or "she", is in reference to someone who has spoken before or are mentioned in the same sentence creating errors as the pronoun mentioned does not necessarily relate to the entity detected by the use of NER. Other errors pertain to multiple entities used within a sentence, which may be easy for a human annotator to ascertain but is difficult for the manual rules.

5.4.2 Recognition of Ethos

Table 9 gives the results of precision, recall and $F1$-score for the classification of sentences as an ESE or $\neg$ESE. Three groupings of results are provided in table 9, a baseline classifier which predicts the target class (ESE), three common machine learning algorithms (ME, NB and SVM) and the ESE / $\neg$ESE stage of our system, containing NER and with NER removed.

ESE/¬ESE	Precision	Recall	$F1$-score
Baseline	0.29	1	0.45
SVM BOW	0.30	0.30	0.30
NB BOW	0.20	0.94	0.32
ME BOW	0.46	0.27	0.34
Extraction Pipeline: NER, POS, DSR, AnaR, RSF	0.62	0.77	0.69*
Extraction Pipeline: POS, DSR, AnaR, RSF	0.64	0.76	0.70*

Table 9: Results of automatic extraction of ESEs from EtHan_Thatcher_3 Test corpus. Reported are precision, recall and $F1$-score for classifying sentences as ESE and ¬ESE. The star symbol (*) denotes the classifier above the baseline $F1$-score.

The baseline has a precision of 0.29 and recall of 1 due to the nature of the target class (a sentence containing ethos) meaning all sentences in the test set are classified as containing ethos, hence the low precision. In this instance either precision or the combination of both precision and recall ($F1$) are important to compare the performance of each algorithm. The machine learning algorithms perform so poorly on this test set due to the complexities of the language used and the imbalance of the datasets. In the case of the SVM classifier it performs worse than the baseline, this can be attributed to the bag-of-words approach used which is largely ineffective for ethos mining without some initial pre-processing in the same way as the manual pipeline constructed. In the case of the Naive Bayes approach it is clear that although the text features of an entity (e.g. "hon. Member") was discovered through the BOW approach (shown in the high recall value), this created many false positives. In the case of the Maximum Entropy classifier, there is an increase in precision, yet the lowest recall due to only a small number of ESEs being classified correctly but still a large number of false positives.

Of these algorithms both of the rule-based systems perform above the baseline $F1$-score by 53% and 56%. This occurs due to the inclusion of domain specificity around names used, common words of attack or support and less of a reliance on the specific vocabulary. To identify people within Hansard, a logical step would be to perform NER to extract names from text. Although this would be true for most cases of dialogue, due to UK parliamentary rules, the number of instances where names are used explicitly are few.

This can cause the problem of many false positives being extracted by the ESE / ¬ESE stage. With NER removed from the system, there is an observed increase in precision on the ethos mining system with only a slight drop in recall.

A confusion matrix has been constructed to identify where the implemented pipeline can be improved and to confirm where this focus should be. The confusion matrix in figure 20 shows the relative classifications of the pipeline against the manual gold standard test corpus. The results of this pipeline and this confusion matrix show that false positives (148) and false negatives (84) have contributed towards the overall $F1$-score and therefore both precision and recall can be improved.

Actual Values

Predicted Values		ESE	¬ESE
	ESE	268	148
	¬ESE	84	704

Figure 20: Confusion matrix for the ESE/¬ESE classification using domain specific rules without NER.

The nature of political discourse within the House of Commons adds to the complex nature of ethos. As explained earlier in this chapter automatically determining whether or not a question bears ethotic expression is a difficult task, quite apart from the many different forms of insult in the House of Commons. Together questions and hostile language make for a lower overall $F1$-score for the ESE / ¬ESE stage in table 9 with precision directly affected by these instances. The lower precision overall confirms this where sentences are falsely determined as ethotic by the ethos mining pipeline.

5.4.3 Recognition of Ethos Supports and Attacks

In table 10, the results of +/- ESE classification are reported with comparison of common machine learning techniques to a baseline classifier with a macro-averaged precision, recall and $F1$-score of the majority class (negative) and the minority class (positive)[6]. Comparison is made between the machine learning algorithms on two different lexicons, SWL and EWL in section 5.3.3.4. The results indicate that known ethotic words which were developed for the EWL are crucial in obtaining a high $F1$-score on sentiment classification of ESEs. Using the same set of features an SVM classifier outperforms both a Naïve Bayes Classifier and a Maximum Entropy Classifier with an overall $F1$-score 16% above the baseline. The use of the domain specific lexicon improves results over the standard sentiment lexicon as would be expected. The same can be said for the SVM classifier over the Naïve Bayes Classifier and Maximum Entropy classifiers as SVMs have been shown previously (described in section 3.3.1) to outperform other classifiers on the problem of sentiment analysis.

+/- ESE	**Precision**	**Recall**	**$F1$-score**
Baseline	0.50	1	0.67
NB, SWL	0.58	0.57	0.57
ME, SWL	0.6	0.65	0.62
SVM, SWL	0.64	0.59	0.62
NB, SWL, EWL	0.74	0.67	0.71*
ME, SWL, EWL	0.71	0.73	0.72*
SVM, SWL, EWL	0.78	0.78	0.78*

Table 10: Results for the sentiment classifier based on a macro-average of results of both positive and negative classifications. Reported are precision, recall and $F1$-score for a baseline classifier and machine learning classifiers two categories: (1.) Containing Sentiment Word Lexicon (SWL) (2.) Containing Ethotic Word Lexicon (EWL). The star symbol (*) denotes the classifier above the baseline $F1$-score.

In the case of +/- ESE a confusion matrix has been constructed to identify where the sentiment classification can be improved and to confirm where this focus should be. The confusion matrix in figure 21 shows the relative classifications of the pipeline against the manual gold standard test corpus. The results of this pipeline and this confusion matrix show that errors mainly pertain to the +ESE classification with false negatives. Errors in this classification also pertain to the size and scope of the sentiment lexicon and ethos

[6]This baseline is essentially two separate classifiers, one classifying all sentences as the positive class and one classifying all as the negative class, with the results averaged producing a recall of 1.

lexicon. Ideally a larger lexicon would be used meaning that there are less instances of unknown words to be classified in the test set. In the case of the ethos lexicon the words defined have the advantage of domain relevance. This suggests, along with the results in table 10, that any additional lexicon should be in a very similar domain. Although this will impact how generalisable the methods used are to other domains, it will improve classification.

Actual Values

Predicted Values	+ESE	-ESE
+ESE	51	25
-ESE	29	247

Figure 21: Confusion matrix for the +/- ESE classification using a SVM with EWL and SWL.

5.4.4 Combined Results

Table 11 and figure 22, give the results and a confusion matrix of the combination of the ESE/¬ESE stage and the +/- ESE stage. A true value is only given when the system correctly identifies an ESE and gives the correct sentiment polarity, when compared to manual analysis[7]. One of the drawbacks in a pipeline based system is that errors are passed between each stage meaning a lower overall accuracy at the end of the process. A drop in overall $F1$-score from table 9 is observed due to the error margin, reported in table 10. However, when calculating the baseline for the full system this gives $F1$-score 0.25, putting the full system, containing the ESE / ¬ESE stage and SVM +/- ESE stage, 140% above the baseline. As shown by the baseline values a precision of 0.55 shows a strong classification result, with a smaller number of false positives although this has meant a trade-off with the recall value.

[7] The combined scores are created by running the full pipeline over the test set and comparing the results against the manually annotated data

ESE/¬ESE & +/- ESE	Precision	Recall	$F1$-score
Baseline	0.14	1	0.25
Full System	0.55	0.65	0.60

Table 11: Results are provided for the combination of the ESE / ¬ESE stage and the +/- ESE stage.

Actual Values

Predicted Values		+ESE	-ESE	¬ESE
	+ESE	35	12	19
	-ESE	17	147	47
	¬ESE	28	113	786

Figure 22: Confusion matrix for the combined ESE / ¬ESE stage and the +/- ESE stage.

5.5 Discussion

To summarise, this chapter has described the pioneering research in the new sub-field of argument mining - ethos mining. Outlined is the first ethos support and attack corpus focussed entirely upon ethos, and the first ethos support and attack automatic extraction also focussed entirely on ethos. The evaluations of both the manual annotation (section 5.2.1) and the ethos mining pipeline (section 5.4), show the reliability of the methods answering RQ1 and RQ2. This line of research also highlights the importance of identifying ethos as the perception or character of a speaker is just as important as the content of what is said as will be shown in chapter 8.

The results of the manual annotation process, building the first corpus of ethos supports and attacks and evaluating the annotation guidelines, show that ethos can be reliably annotated independently of logos in natural language answering RQ1[8] with substantial agreement between annotators. The results of the ethos mining process, building a novel ethos mining pipeline, showed that ethos can be reliably identified and extracted in natural language text answering RQ2[9] as the results are above the baseline classifiers.

[8](RQ1): Can ethos be reliably annotated independently of logos in natural language?
[9](RQ2): Can ethos be reliably extracted automatically from natural language?

These results show that the application of ethos mining to natural language text is achievable, though there is still room for improvement in both the annotation of ethos and the automatic classification.

5.5.1 Annotation Improvements

In the case of ethos annotation, the guidelines can be improved in order to gain further reliability in the corpus which in turn is likely to improve automatic classification performance. This can be done in several ways. Firstly the guidelines surrounding questions can be updated. This would involve stipulating when a question is seeking further information or when it is framed as a question merely to assert. In the cases of "Is he/she aware", "Will he/she" outlined in the error analysis of section 5.2.1 the guideline must rely upon a second improvement. The annotation guidelines must stipulate that a second entity mention should be present within a sentence when the first entity mention is framed as a question. This second improvement coupled with the first guideline making annotators aware of the difficulty of annotating questions should enhance the annotation especially when determining if a sentence contains ethos or not or who the target of each sentence is.

A third improvement to the annotation guidelines can be made by stipulating which parts of a sentence can be ignored as they only implicitly refer to ethos. Overall this would mean that some sentences which are currently annotated as ethos would no longer be, as the annotator has relied too much on contextual world knowledge. This process would help alleviate cases where the sentence polarity is unclear due to the implicit information given. Any re-annotation in this respect would also aid in using machine learning. Whilst simple BOW feature approaches would still be ineffective, the enhanced annotation would aid in creating more complex features [10] assuming more data is annotated. This improvement would also mean less reliance of knowledge of the UK parliament as a whole meaning that annotation could, in the future, be crowd sourced.

[10] Specific linguistic features can be created making use of syntax parsing techniques such as dependency trees.

A final step to improve the annotation, would be extending the size of the corpus. This has a two fold advantage. In the first instance increasing the corpus size will provide more training data for automatically extracting ethos. The increased training dataset size will allow for further entity references to be discovered by domain specific rules. If the improved dataset also incorporates all of the guidelines outlined above then there will be an increase in the negative instances of sentences. Whilst providing a further imbalance to the data, it will increase the chances of identifying when a sentence is not ethotic. This coupled with further improvements for the detection of ethos should improve the classification results. In the second instance, increasing the corpus size will provide a larger set of data for annotation evaluation. The increased size should determine if the annotation guidelines are reliable and consistent. If the Cohen's kappa scores remain constant or improve the annotation guidelines can be determined as reliable. On the other hand, a reduction in kappa score would indicate that the guidelines are not consistent when scaled up.

5.5.2 Automatic System Improvements

In the case of the automatic extraction of ethos, through the rule-based ethos mining pipeline, improvements can be made in anaphora resolution, in the use of the parsing and reported speech layer and in sentiment classification for polarity identification. In anaphora resolution results can be enhanced through the removal of NER in the module. Despite the usefulness of NER in NLP tasks, for UK parliamentary debates the unusual language causes problems. In normal cases the NER module would determine the target in a sentence when they are mentioned but in the case of the transcripts used for ethos mining this instead creates false positives. A possible solution is only making use of the domain specific rules module which will be improved through the increased dataset size. The anaphora module can also be improved through the use of external data. Wikipedia provides a comprehensive list of all politicians within the UK parliament including their constituencies. This data can be utilised alongside domain specific rules to determine exactly who a politician is even if they share their name with another.

In the case of the parsing and reported speech layers, improvements can be made by first removing an entity mention when it is not related to any positive or negative sentiment and also removing keywords for reported speech identification. In the first instance an entity referred to in the opening of a question should be ignored if it has no association with a sentiment value or is not used again within the sentence. This step would require first creating a comprehensive list of questions referring to an entity, for removal and then making use of more comprehensive parsing methods. The list of questions can be obtained through the increased dataset, although this may not be comprehensive it will provide a starting point. The use of sentiment values would need to coincide with moving the sentiment layer in the ethos mining pipeline. Sentences can first be classified as positive or negative with key sentiment holding words identified this can then be combined with a full dependency parse. This will give a better understanding of the sentence structure, in particular, the relation between an entity and the related sentiment words. For the reported speech layer, keyword instances may tag ethotic sentences as ¬ethotic due to the complex nature of referring back to earlier speeches by the targeted politician. Whilst instances of reported speech will not be so easily identified, the sentiment and dependency features should alleviate the need for this function.

In the case of sentiment classification for the polarity of ethotic statements, results can be improved by using a more comprehensive set of lexicons. Whilst the results in the case of the sentiment layer are positive a larger lexicon in a similar, or the same, domain as parliamentary debates may be useful. This is particularly the case for the UK parliament with many banned words and also an archaic style of language used. Although the ethos lexicon developed does aid in this it is only as comprehensive as the training data it uses. New techniques making use of word embeddings and seed sentiment holding words can be utilised to build huge comprehensive sentiment lexicons. The use of similar domains also improves the classification whilst ensuring it is generalisable. Moving the sentiment layer to earlier in the extraction pipeline will also prove advantageous for detecting ethos especially when coupled with larger lexicons of both positive and negative holding words.

A final step can be made to improve the ethos mining pipeline developed. In much of the pipeline there is a reliance on manually defined rule-based modules. Whilst this ensures domain specificity it does provide a burden on the creator of such rules as the construction of these are expensive. The rule-based system also perhaps lacks portability to other domains. The use of machine learning techniques in some aspects of the pipeline may help to reduce the cost of ethos mining pipeline construction whilst at the same time improving generalisability. Providing a comprehensive set of features to any classifier some of which may be domain specific rules could improve the reliability of the results identifying features not realised in the manual rules.

Chapter 6

Ethos Mining: Deep Learning

This chapter contains annotation and automatic identification of supports and attacks on ethos[1] which looks to extend both the annotation and automatic extraction in the previous chapter by making improvements to the problems outlined in the discussion and answering RQ1 and RQ2.

The rule-based ethos mining pipeline showed the reliability of automatic ethos extraction, however, it also highlighted a problem with applying simple generic machine learning methods to a domain of complex language. The benefits of machine learning systems are clear, in that with the right training data they can provide a generalisable method of automatic classification. In most cases, the most generalisable machine learning systems tend to not provide the optimum performance on a specific problem. Thus, any application of machine learning for ethos mining will need to balance generalisability with domain specificity. The first step in creating this classifier is in ensuring a reliable and consistent training dataset.

Following the adaption of Aristotelian ethos and the tags specified in chapter 4 and structures used in chapter 5, ethos is specified as: properties of the individual or the group of agents which can be supported (Example 14, hereafter used as a running example)

[1]The work in this chapter relates to one published work (Duthie and Budzynska, 2018b) where ethos supports and attacks were annotated in sentences from the UK parliamentary record and then automatically extracted. Some passages are used verbatim from the source. This work was in collaboration with Katarzyna Budzynska who contributed extensive feedback and general ideas for implementation.

or attacked (Example 15) in order to influence the audience through communication. Extending and improving the corpus described in chapter 5 the aim remains the same to identify the relationships between politicians or between a politician and a party through ethotic sentiment expressions, and then to classify these relationships as having positive or negative sentiment. Additionally any reference made to ethos must be expressed on the linguistic surface, and have an explicit sentiment value rather than relying on contextual or domain knowledge. The intuition intended to be modelled is that the linguistic structure encodes both the target entity of the ethotic statement (in Example 14, “My hon. Friend,” and in Example 15, “the Government”) as well as the polarity of the ethotic statement (in Example 14, positive sentiment is signalled by “assiduously”, “pursuing” and “interests” and in Example 15, negative sentiment is signalled by “sick”). As a result, only part of each sentence provides the needed data for identification meaning the ordering present in the ethos mining pipeline has to be fundamentally changed as the sentiment classification is integral for classifying ethos at an early stage.

(14) **Mr. John Moore said,** *My hon. Friend is <u>assiduously</u> <u>pursuing</u> his constituents' <u>interests</u>.*

(15) **Mr. Bruce Grocott said,** *Is it not the simple truth that the Government are making the country <u>sick</u>?*

In order to automatically extract and classify ESEs, the corpus developed in chapter 5 was extended by an additional 30 transcripts from Hansard. This increase provides extra data for training any machine learning algorithm and to verify that both the sentiment and entity references appear in each ethotic sentence. A new ethos mining pipeline was then developed featuring a deep modular recurrent neural network, **DMRNN**, approach applied for the first time to the NLP task of text classification. More specifically, experiments are run with proven machine learning methods that boost accuracy (e.g. ensemble classification (Larkey and Croft, 1996)) and advances in neural networks through deep learning (e.g. recurrent neural networks, RNNs (Hochreiter and Schmidhuber, 1997)). The DMRNN also makes use of known classification methods for images, where image and metadata are combined for classification (Ma et al., 2016). The DMRNN approach is then created which provides several textual based inputs to word embedding layers (one-hot encoded vocabulary with a dense layer from the Keras library), passed into dropout layers to reduce over-fitting, then to RNN models, some featuring long short-term memory (LSTM) layers, and combines them into one model for binary classification of ESEs. The DMRNN allows for an increase in the available data that can be used for classification (due to the allowance of multiple separate modules) and for the application of multiple simple models with unique features from the inputs which, when combined, do not suffer from over-fitting. Each of the unique features already identified mean that the DMRNN model instead attempts to find trends over those features rather than searching for these trends in the word embedding model only. Whilst this does mean the DMRNN is stopped from identifying interesting features of the plain text it does reduce the chance of over-fitting to the text only. The improvements of the approach in this chapter are: the size of the dataset (30% more tokens); the annotation scheme; and the NLP method used to extract ethotic relations (a general purpose DMRNN not tailored to specific domain rules and the re-ordering of the pipeline).

The contribution of this chapter is as follows: (*1*) the largest publicly available corpus (90,991 tokens) and evaluation of ethotic relations between politicians or between politicians and political groups (see section 6.1 and 6.1.3); (*2*) an ethos mining pipeline with two modules employing existing NLP techniques, two modules extending standard text classification methods, and four original modules (see section 6.2); (*3*) a DMRNN approach which is new to the area of text classification (see section 6.2); (*4*) a comparison between the previous approach implemented in chapter 5 and the DMRNN approach (see section 6.3).

6.1 Annotation

The data obtained in this section extends that of the previous chapter, chapter 5, utilising the same techniques for transcript identification from the UK parliamentary record with the aim being to increase the size of the original training data, through additional negative examples (¬ESEs) and more specific criteria for detecting ESEs. Each of the additional transcripts holds the same structure as previously, part of a day parliament sitting containing an initial question asked by a member of parliament (MP) identified by their name and constituency, to which a reply is given by a government minister and further continued with subsequent turns in the debate.

To re-annotate the original corpus of ethos supports and attacks and to annotate the further 30 transcripts obtained, a comprehensive set of ethos annotation guidelines were created. The goal being to improve the reliability of ethos annotation. This can be achieved by keeping consistent or improving the Cohen's kappa scores for identifying sentences which contain ethos, identifying the polarity of those sentences, or by identifying the target of each sentence.

For the additional transcripts, ethotic statements are still annotated on a sentence level. ESEs contain a parliamentary entity (e.g. "hon. Member", "Conservative party") as a target of either an ethotic support or attack. The annotation scheme keeps the four types of tag defined in chapter 4: (*1*) **speaker**, the author of an ESE; (*2*) **target**, the referent of an ESE; (*3*) **ethotic support**, **+ESE**, when: (*a*) a single target entity, individual or group, can be established; and (*b*) it puts in a positive frame the target's character or achievements, or supports their credibility explicitly; and (*c*) it is not self-referential; (*4*) **ethotic attack**, **-ESE**, when: (*a*) a single target entity, individual or group, can be established; and (*b*) it puts in a negative frame the target's character, attacking their credibility or associating them with events of a negative connotation explicitly. For instance, Example 14 is tagged, speaker: Mr. Moore, target: Mr. Meyor (referred to as "hon. Friend"), and +ESE. All other sentences (a total of 3,007, 80% of the corpus) not labelled using these tags are considered as ¬ethotic sentiment expressions (¬**ESEs**).

6.1.1 Guidelines

To be annotated as ethotic a sentence must then contain (see appendices .A and .B for annotation examples):

A **target speaker** which the sentence refers to. This can be a person or an identifiable group and must be a known political entity within the parliament. For example "Prime Minister", "hon. Member for Bedford" or "hon. Lady" and not "the president of France", while this is a person it is outwith the context of the UK Parliament. The groups identified must also be political entities. For example "the Labour Party", "Conservative Members" or "the Government" and not "hon. Members", "British Gas" or "the German Government". While referring to "Conservative Members" is a generalisation this is on a small, identifiable scale, "hon. Members" is on large scale where it is impossible to determine which members. Companies and other countries again fall out of the remit of the UK parliament and while they have influence are not the subject of this annotation.

A **source speaker** from which the statement has come from. This must be a person or an identifiable group. This must have the same restrictions as the target speaker.

An **ethos support** of another entity but not self referential. It puts in a positive frame the target's character or achievements, or supports their credibility explicitly. For example, "the hon. Member is right", "the hon. Member is correct" and not "I congratulate the hon. Member on their appointement". While congratulations can show support in some cases, in this context it is merely politeness which is expected in the UK parliament. All supports must be explicitly stated on the linguistic surface of a statement. Or an **ethos attack** of another entity, but not self referential. It puts in a negative frame the target's character, attacking their credibility or associating them with events of a negative connotation explicitly. For example, "the hon. Member is wrong", "that was a terrible speech by the hon. Member" and not "the Scottish area of British Rail is one of the least reliable and punctual of all the areas in the British Rail federation". Although this is an issue for parliament to discuss British Rail is not a political entity and therefore should not be considered. All attacks must be explicitly stated on the linguistic surface of a sentence.

Table 12 explicitly states the changes made between the annotation conducted in this chapter and that in chapter 5. For both target and source speakers no changes have been made to the core guidelines, the only additions were that of further examples so that the annotators could be sure that the entities in the sentence are relevant to the UK parliament. In the case of identifying ethotic supports and ethotic attacks changes were made on the basis of identifying explicit words related to the sentiment. This means that a sentence must have an entity mentioned and there must be a word on the linguistic surface that explicitly shows the sentiment felt towards the entity.

Label	**Annotation Changes in Chapter 6**
Source Speaker	No Change (Further examples specified for annotators)
Target Speaker	No Change (Further examples specified for annotators)
Ethotic Sentiment Expression	Identifiable word or phrase must be in sentence that is related to sentiment of character. Must be on linguistic surface.

Table 12: Annotation guideline changes in chapter 6 from the base annotation specified in chapter 4 and used in chapter 5.

6.1.2 Corpus

Following the improved annotation guidelines, a new publicly available corpus was constructed, **Ethos_Hansard1** (`http://arg.tech/Ethos_Hansard1`, see table 13),

Corpus	**Sessions**	**Words**	**Segments (ESEs)**	**+ESE**	**-ESE**	**Speakers**
Train	60	61,813	395	106	289	116
Test	30	29,178	243	64	179	82
TOTAL	90	90,991	638	170	468	198

Table 13: Summary of the language resources in the Ethos_Hansard1 corpus for mining ethos in the UK Hansard.

Tag	Source	Target	Ethotic Support (+ESE)	Ethotic Attack (-ESE)	Total
#	149	188	169	469	975

Table 14: Frequency of tags in the Ethos_Hansard1 corpus.

by extracting an additional 30 transcripts leaving a total of 90 transcripts from Margaret Thatcher's period as Prime Minister, dating from 1979 to 1990. Additional criteria for identifying transcripts in Hansard was set, with the overall word count reduced to 500 words giving more scope for extracting transcripts. The data was then split into 60 training transcripts, of which 10% was used as validation data for machine learning, and the same 30 test transcripts re-annotated from the previous corpus to give a wide range of test cases.

The statistics given in table 13 show a reduction in the number of ethotic sentences and speakers overall, yet an increase in transcripts and words when compared with the corpus created in chapter 5 - 90 transcripts rather than 60 and 90,991 words rather than 70,117. This is expected due to the change of annotation guidelines which make the annotation of ethos more specific. Overall the train and test sets produce imbalanced data with the volume of ESEs 395 and 243 respectively in comparison with 2045 ¬ESEs in the former case and 962 ¬ESEs in the latter. In table 14 there are less tags shown overall in comparison to the previous corpus, 975 compared to 1,194, but not shown is the increase in the number of ¬ethotic sentences, 2085 to 3007. Table 14 also shows that the ratio of supports and attacks on ethos has stayed consistent with attacks being the majority class.

6.1.3 Inter-Annotator Agreement

To evaluate the extended corpus, a second annotator analysed a 10% subset of Ethos_Hansard1 (see Table 15 and figure 23 for a confusion matrix). This gave a Cohen's κ = 0.67 for distinguishing between ESE and ¬ESE (normalised for word count); κ = 1 for the polarity classification of ESEs (+/- ESEs) when both annotators have already agreed on an ESE; κ = 1 for source speaker tags; and κ = 0.93 for the target speaker tags. The two coders were experienced in argument annotation, with one a native speaker of English. For the

<table>
<tr><th>Tag</th><th>Source</th><th>Target</th><th>Ethotic Support (+ESE)</th><th>Ethotic Attack (-ESE)</th></tr>
<tr><td rowspan="2">κ</td><td rowspan="2">1</td><td rowspan="2">0.93</td><td>1</td><td>1</td></tr>
<tr><td colspan="2">0.67</td></tr>
</table>

Table 15: Cohen's κ for the Ethos_Hansard1 corpus.

purposes of this task a single session was used for training after-which the second coder annotated the six sessions.

Ann 2

Ann 1

	ESE	¬ESE
ESE	37	18
¬ESE	12	304

Figure 23: Confusion matrix for the ESE/¬ESE annotation.

When comparing the evaluation of the new extended corpus with that of the corpus constructed in the previous chapter, the kappa score for annotating ESE and ¬ESE remains consistent. This can be determined as a positive result. Although the guidelines have been improved and altered to only annotate ethotic statements which are signalled on the linguistic surface, the fact that the kappa score remains constant shows that the guidelines are robust and have not been affected negatively by the changes. The guideline examples also highlight the sometimes hostile language within the parliament and explain the consistent score rather than an improvement. This value and the examples given in the guidelines show that determining if a sentence is ethotic or not is extremely difficult, precisely due to this hostile language (see appendix .A for such examples) which makes determining the difference between sentiment holding and ethotic sentences hard. This same discovery was made in (Hirst et al., 2014) when investigating the Canadian Hansard transcripts.

The kappa score for polarity classification increased from 0.95 in the previous chapter to 1 where the small rise can be attributed to the updated guidelines. The fact that any entity mention must contain sentiment on the linguistic surface suggests that there will always be an identifiable entity and sentiment relation. The polarity identification increase coincides with that of target identification.

The use of running examples throughout the annotation guidelines also aids in the increased agreement between the annotators. The examples show the nature of parliamentary debates and indicate reasons as to why each sentence should be classified as ESE or ¬ESE, whilst at the same time show why an entity mention has no ethotic sentiment relation. The latter point can explain the increase in target identification, where questions referring to an entity are disregarded more often.

6.2 System Architecture

In order to mine ethos from parliamentary transcripts, an improved ethos mining pipeline was created with the motivation to enhance results, the identification of ethos, whilst combating some of the issues outlined in the the previous chapter. The new pipeline aims to make improvements in anaphora resolution, the parsing of sentences to aid in ethos identification and in sentiment analysis and also aims to be generalisable only requiring a small number of steps to move the pipeline to a new dataset or domain.

The pipeline (see figure 24) uses an input of raw natural language text and an output of +/-ESEs. The pipeline consists of components which are either employing existing techniques; or components which extend such methods for the purpose of ethos mining; or components which contain original techniques developed specifically for ethos mining.

The raw text is passed to five areas of the pipeline (see arrows coming top-down from raw text in figure 24): (*1*) directly to the DMRNN component; (*2*) the POS tagger; (*3*) the universal dependency (UD) tagger; (*4*) the sentiment classifier; and (*5*) the anaphora resolution component. Components (*3*) and (*4*) are involved in complex processes. The UD tags are passed to the entity extraction (EXT) module (which removes entity references not relevant for ethos mining) and then to the sentiment presence module (which determines whether a sentence contains a sentiment). The output from the sentiment classifier is passed

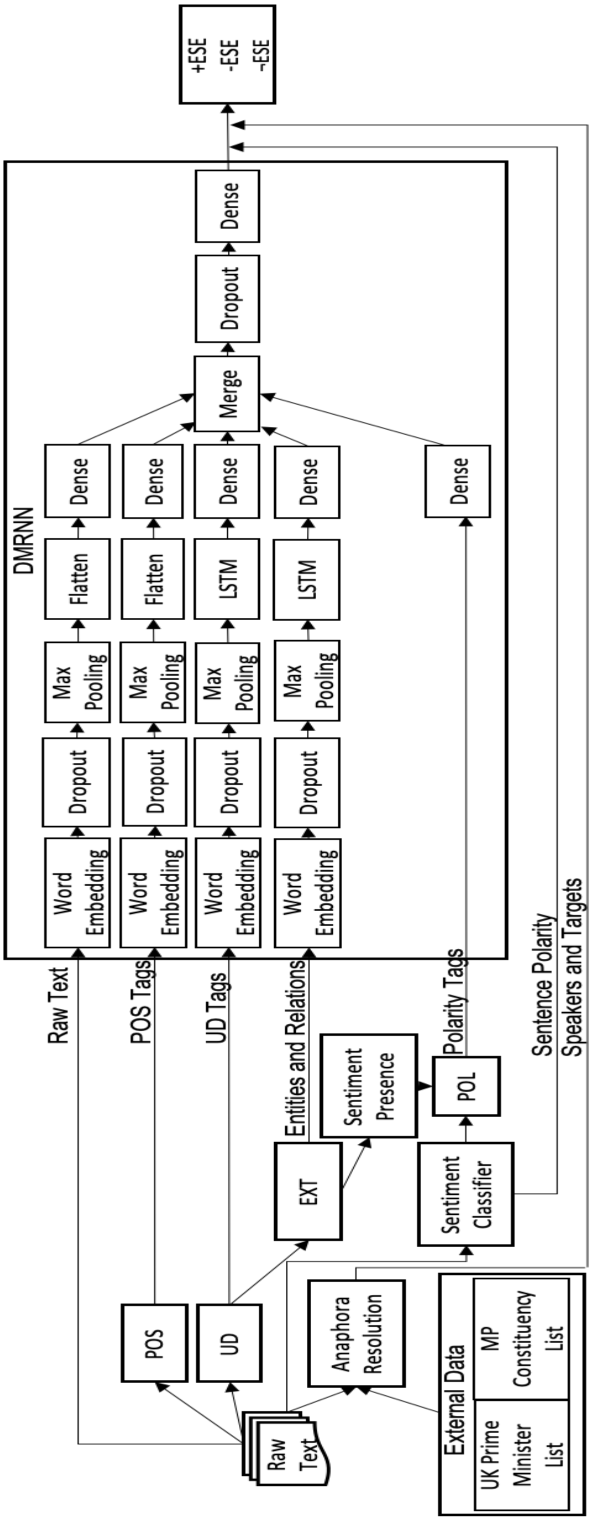

Figure 24: Pipeline for ethos mining featuring raw text; parsing (POS and UD); anaphora resolution with external data from Wikipedia; entity extraction (EXT); sentiment classification; sentiment presence; polarity combination (POL); and ESE/¬ESE classification performed by a DMRNN. The output of the pipeline is processed by ethos analytics.

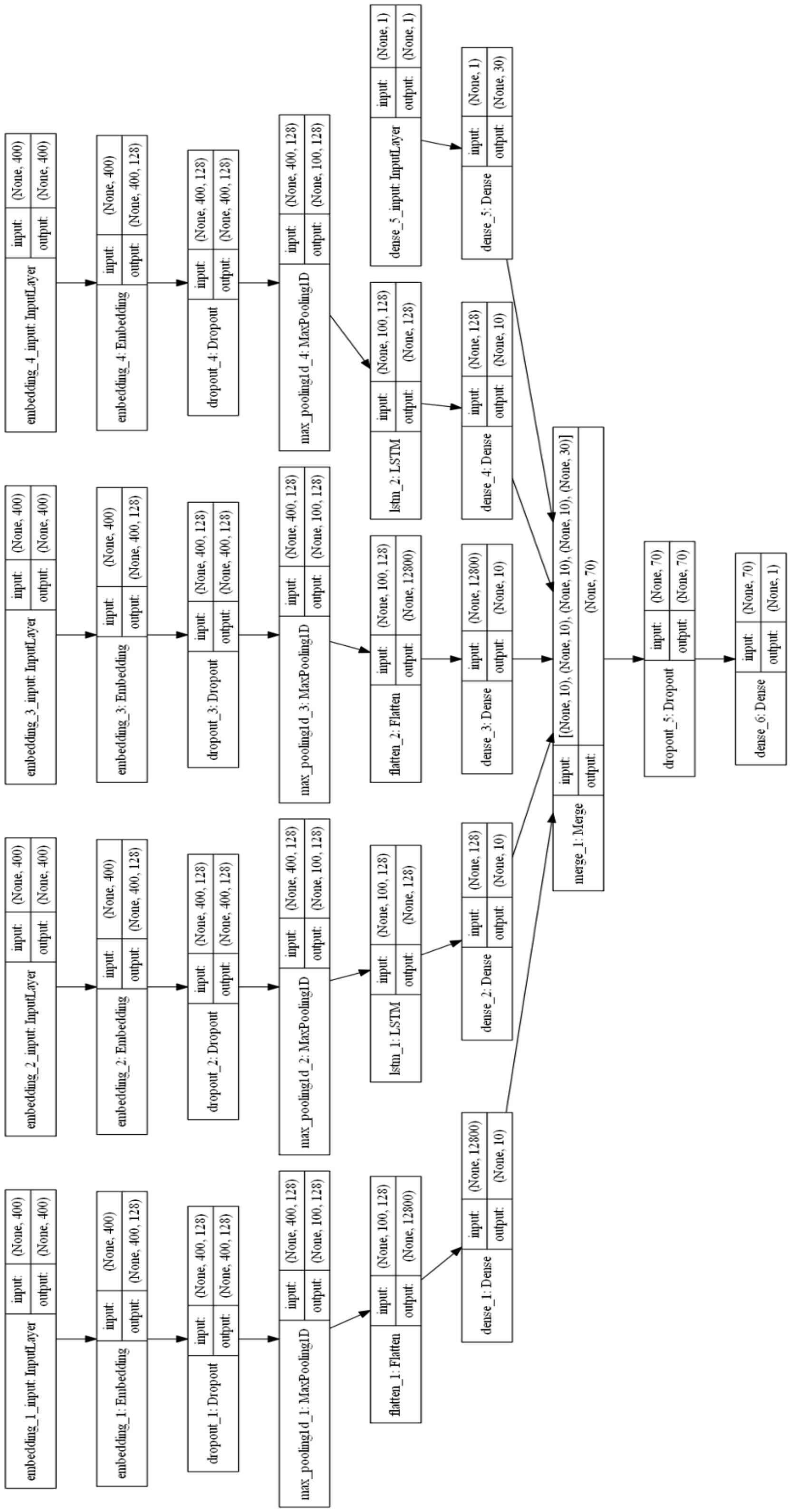

Figure 25: A closer insight into the model parameters of the DMRNN showing input and output vector sizes. A value of None specifies the batch size which is variable.

to the polarity (POL) module which combines the output from the sentiment presence module with the sentiment classifications. Next, the raw text, POS tags, UD tags, EXT output and POL tags are passed as separate inputs into the DMRNN (see figure 25 for vector sizes), returning ESEs/¬ESEs. The output of sentiment classification determines +/-ESEs with the anaphora resolution component tagging each +/-ESE with a source and target. Each element of the pipeline is further described.

6.2.1 Existing Methods

Two components in the pipeline consist of existing methods within NLP: part-of-speech tagging and universal dependency tagging.

6.2.1.1 Part-of-speech (POS) Tagging

POS tagging is applied to the raw corpus text, using the already existing Stanford parser (Toutanova et al., 2003). These are passed as an input to the DMRNN with the intuition that the syntax of each sentence plays a role in ethos classification.

6.2.1.2 Universal Dependencies (UD) Tagging

UD tags are obtained for the the raw corpus text, again, using the already existing Stanford parser (Schuster and Manning, 2016). The subjects of each sentence are tagged (e.g. "nsubj" in figure 26) allowing the extraction of subject entities later in the EXT module. The UD tags are also passed as an input directly to the DMRNN with the intuition that the syntax and relations in each sentence will play a role in classification.

6.2.2 Adapted Methods

Methods which have been adapted specifically for the ethos mining pipeline consist of anaphora resolution and sentiment classification.

6.2.2.1 Anaphora Resolution

The anaphora resolution component uses manual rules to determine the source and target entities of each sentence by extending the existing techniques described in chapter 5. The source is explicitly identified at the beginning of each of their turns. For the target of each sentence, manual domain rules relating to specific entity mentions are used. For entities where no unique identity details are given (e.g. "he", "hon. Lady"), the system tracks back over a sentence and a speaker's turn to determine if there are any entity mentions which do reveal unique values (e.g. "hon. Member for Bath") or unique roles (e.g. "Prime Minister"), but excluding entities without human features (e.g. "the Government").

In these cases, external data is used containing MP details and a UK Prime Minister list scraped from Wikipedia using the month and year of each transcript in order to pinpoint the target. If there are no entity mentions present, then the system sets the target as the speaker of the previous turn. In this chapter the use of NER has been removed. Following on from the error analysis and discussion in chapter 5 the use of NER provided too many false negatives in determining the target instead the domain specific rules (DSR) created and described in chapter 5 are more useful.

6.2.2.2 Sentiment Classifier

The sentiment classifier extends existing methods considering three machine learning classifiers: a linear SVM classifier, a Naïve Bayes (NB) classifier and a Logistic Regression (LR) classifier, all within scikit-learn (Pedregosa et al., 2011). They are used to classify the polarity of sentences into positive and negative, performing a binary sentiment analysis. A TF-IDF approach is also considered to create vectors from a sentiment word lexicon (LIU) (Liu, 2010) and extend the approach with an ethotic word lexicon (ETH) (Duthie et al., 2016a) as well as negative words from a word embedding generated Hansard specific lexicon (HAN) (Rheault et al., 2016). The addition of ETH and HAN provides a relevant domain classification which is needed for words that play a different sentiment role in day-to-day parliamentary speech. For example, "rich" has a positive meaning of wealthiness, but much of the time in the UK parliament it is used to describe someone as a hypocrite. This addition will increase the accuracy of classification although the methods used for generating the lexicon will produce some classification errors. The word embedding model, although useful for this kind of lexicon generation, will tag some words falsely as positive and negative.

6.2.2.3 Sentiment Presence

The sentiment presence component extends standard techniques with an input of entities and related words from the entity extraction component described in section 6.2.3.1. Using sentiment lexicons (LIU, ETH and HAN), sentences are tagged as holding sentiment or not. The aim of this module is not to determine the sentiment of a particular sentence rather the presence of sentiment only. This is in line with the hypothesis that each entity mention has a related sentiment value in the case of ethotic sentences.

6.2.3 Novel Methods

Novel methods for the pipeline consist of an entity extraction component, a polarity component and a Deep Modular Recurrent Neural Network which have been developed specifically for this research.

6.2.3.1 Entity Extraction (EXT)

The EXT component is a novel manual rule-based system to remove entity mentions which are not relevant for ethos mining. The dialogue protocol in the transcripts dictates that in order to elicit a response from a Government representative, MPs have to frame their statements explicitly as questions which introduces entities who are addressees of the questions. Thus any initial framing pertaining to entities is ignored (e.g. “Is he aware...”), unless the entity is mentioned again within the sentence.

The UD tags are used to extract all entity mentions in a sentence and any words related to these entities, as long as the tag between them is a nominal subject, clause or modifier (see figure 26). In example 14 the following entities and related words are extracted “interests”, “pursuing”, “assiduously”, “constituents”, “Friend” and “his”. They highlight the need for ethotic sentences to have an entity and sentiment relation which the dependency parse gives.

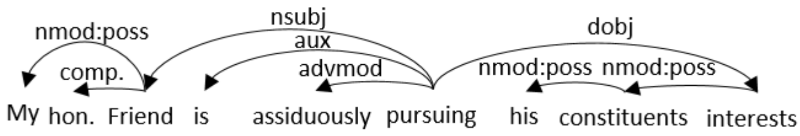

Figure 26: Dependency parse tree structure for Example 14.

The EXT module also allows for a large reduction of words which can be used for classification. Like stopwords, those words with less important relations are removed. This means that in most cases, as explained above, questions addressing an entity are removed and so is the context of the wider sentence which may mislead classification.

6.2.3.2 Polarity (POL)

The POL component is original and involves combining any entity relations tagged as containing sentiment with the corresponding sentiment classification. Sentences are grouped into positive, negative and non-sentiment bearing (example 14 is marked as positive by combining the sentiment presence output and the sentiment classifier output). This feature is particularly useful in the case of ESE and ¬ESE classification as all non-sentiment bearing utterances can be discarded as ¬ethotic.

6.2.3.3 Deep Modular Recurrent Neural Network (DMRNN)

A DMRNN is developed to classify sentences as ESEs or ¬ESEs. The keras deep learning library (`https://github.com/fchollet/keras`) is used to construct a novel DMRNN where the network hyper-parameters are specified using the validation set and early stopping (when the validation $F1$-score no longer increases) to determine the final model for testing. The DMRNN involves an ensemble of models as separate inputs (raw text, POS, UD, EXT and POL) which are then combined through concatenation in a final dense hidden layer with sigmoid activation. All of the models except polarity use the keras embedding layer with the output passed to a dropout layer to reduce over-fitting (Zaremba et al., 2014), and a max pooling layer to reduce the feature set (see table 16 for parameters).

More specifically, the text and POS inputs are flattened to reduce the dimensionality of the vectors and then passed to a dense hidden layer containing 10 units and sigmoid activation. The UD and entity relation inputs are passed to an LSTM layer, ensuring that long distance relations are taken into account. Its output is then passed to a final dense hidden layer containing 10 units and sigmoid activation. The POL input is passed only to a dense hidden layer with 30 units and linear activation due to the numerical values of the data where context is not needed. A final model is created by concatenating the individual models and it contains a dropout layer to combat over-fitting; a final output layer with one neuron and sigmoid activation which uses Adam optimization (Kingma and Ba, 2014); and a binary cross-entropy loss function for the binary classification to ESE and ¬ESE. The sentiment classifier is then used to determine the polarity and create +/- ESEs.

A large number of dropout layers have to be utilised in the case of ethos mining to combat over-fitting on the relatively small dataset. Each of the feature inputs then increases the data which can be used for classification with each feature determining its own classification with a final merge. Although some of the features operate using the same base (for example the UD features and entities and relations), the intuition is that any relations missed in the EXT module will be caught in the UD module.

Parameter	**Value**
Embedding dimension	input length: 400, output: 128
Hidden layer dimension	10
POL hidden layer	30
Dropout rate	0.20
LSTM layer	128
Max pooling size	4
Adam	α: 0.001, β_1: 0.9, β_2: 0.999
1D convolution	filter: 32, kernal: 5

Table 16: Hyper-parameter values for all models including the DMRNN.

6.3 Results and Evaluation

For the evaluation of the ethos mining pipeline, results are provided for anaphora resolution, the classification of +/-ESEs (as this is now an input into the later classification of ESE/¬ESE), the classification of ESE/¬ESE, and the combination of the highest performing classifiers for +/-ESEs and ESE/¬ESE. An error analysis is then provided for the +/-ESE and ESE/¬ESE classification.

6.3.1 Anaphora Resolution

The results for anaphora resolution are based upon the correct tagging of the target entity in each ethotic statement. The source is not considered as it will always have 100% accuracy due to the structure of Hansard transcripts. Target entities were tagged with an accuracy of 0.76% using the anaphora resolution function with errors mainly produced due to multiple different entities appearing in a sentence.

By removing the NER step which was used in the previous approach, for anaphora resolution, the overall performance has improved by 7%. The re-annotation preformed may aid in the increase in performance with entities more clearly defined in each ethotic sentence. Despite the increase there is still room for improvement, although these coincide with other possible changes which are discussed in the next section.

6.3.2 Recognition of Ethos Supports and Attacks

Table 17 provides +/-ESE results using machine learning classifiers, TF-IDF vectorization and a combination of lexicons. The results are compared against a baseline classifier classifying on the training set class distributions and against the +/-ESE classification from the previous work on ethos mining in chapter 5. All classifiers outperform the baseline (macro-averaged $F1$-score of 0.64) with the highest performing classifier, 31.3% above the baseline, consisting of LR and the LIU, ETH and HAN (Neg) lexicons (m-$F1$-score 0.84). The results can be attributed to the domain specific lexicons (ETH and HAN), in particular the HAN (Neg) lexicon which extends the set of negative domain specific words whereas, the full HAN lexicon does not perform any better due to the use of word embeddings in which many non-sentiment holding words (e.g. "point") are considered as positive, because of their context, skewing the classification. Whilst the use of word embeddings has allowed an increase in the size of datasets through unsupervised methods, they do still have issues in comparison to manual based dataset construction. Despite the best performance, LR with both the ETH and HAN (Neg) lexicons only showed a small increase in macro-averaged $F1$-score over Naive Bayes with only the ETH lexicon. This result shows that the ethotic lexicon is crucial for this domain, however, without the added Han (Neg) lexicon it will have a limited vocabulary thus making the classifier less general.

+/-ESE	**Precision**	**Recall**	**m-F1-score**
Baseline	0.63	0.65	0.64
(Duthie et al., 2016a)	0.78	0.78	0.78
SVM + LIU + ETH	0.81	0.81	0.81
NB + LIU + ETH	0.84	0.83	0.83
LR + LIU + ETH	0.80	0.80	0.80
SVM + LIU + ETH + HAN (Full)	0.78	0.79	0.78
NB + LIU + ETH + HAN (Full)	0.80	0.80	0.80
LR + LIU + ETH + HAN (Full)	0.79	0.80	0.79
SVM + LIU + ETH + HAN (Neg)	0.80	0.77	0.78
NB + LIU + ETH + HAN (Neg)	0.82	0.82	0.82
LR + LIU + ETH + HAN (Neg)	0.84	0.84	0.84*

Table 17: Classification of ESEs into positive and negative. Macro-averaged precision, recall and $F1$-score are reported for classification using machine learning classifiers and training lexicons, ethotic (ETH), Liu (LIU) and Hansard (HAN) compared against a baseline classifying on the training set distributions and against the previous work in ethos mining. (*) denotes classifier with highest $F1$-score.

In the case of +/-ESE a confusion matrix has been constructed to identify where the sentiment classification can be improved. The confusion matrix in figure 27 shows the relative classifications of the pipeline against the manual gold standard test corpus. The results of this pipeline and this confusion matrix show that errors mainly pertain to the +ESE classification with false negatives amounting to 34% of +ESEs overall. These errors pertain mainly to a combination of positive and negative words in each statement which would call for a conflicting classification or through a lack of domain knowledge meaning the positive aspects are not identified by the classifier.

Predicted Values \ Actual Values	+ESE	-ESE
+ESE	42	17
-ESE	22	162

Figure 27: Confusion matrix for the +/-ESE classification using LR.

(16) *"I bow to my hon. Friend's distinguished past and detailed knowledge of these matters."*

(17) *"He is right but the longer I allowed the situation to continue the worse might have become the gap between the overspenders and the underspenders."*

(18) *"I should stress to the hon. Lady that if one considers the different offices and the differential way in which the rules on community care grants in particular are applied one realises that serious application by social workers by local authorities and by diligent Members of Parliament I know that the hon. Lady is such a Member will ensure that community care grants are effectively spent."*

Example 16 shows that a lack of domain context means there is no positive classification in this case as "distinguished" is not present within the lexicon. While a domain lexicon has been utilised, further common knowledge is needed to determine that "detailed knowledge" is in fact positive. Example 17 shows how the lack of a context of the relevant parts of a sentence can create an error. While saying an entity "is right" shows support, words like "worse" and "gap" can shift the classification to negative. Example 18 continues this theme where the more complex structure of the statement paired with the many positive and negative terms creates an error. Homographs play a role in this error where "stress" can be seen as negative in a certain context, and "serious" is also determined as negative.

Overall this misleads the classification.

6.3.3 Recognition of Ethos

Table 18 shows the results of the ESE/¬ESE classification with a combination of standard machine learning algorithms and RNN and CNN models[2] compared against a baseline classifier using the training set distributions and against the ESE/¬ESE classification from chapter 5. All of the models outperform the baseline $F1$-score of 0.61 with the DMRNN model giving the highest $F1$-score 0.74 (21.3% above baseline) and macro-averaged $F1$-score 0.83 (13.7% above baseline). As opposed to chapter 5 macro-averaged $F1$-score is also provided to show the error margin of the classification[3]. The DMRNN model involves an ensemble method which does not suffer from the use of voting methods used in standard ensemble classifiers which explains the improvement of performance in the ethos mining pipeline. The addition of POL to any of the models provides an increase in $F1$-score, highlighting the role sentiment plays in ethotic statements, but alone it is not enough to classify ESEs ($F1$-score 0.66). CNN models have provided a boost in $F1$-score in areas like sentiment analysis, but this was not the case for ESEs. The relatively small dataset, in comparison to datasets in sentiment analysis, can explain this as a complex CNN model may be prone to over-fitting on smaller datasets[4]. Comparing the macro-averaged $F1$-scores of the previous ethos mining pipeline and the one described in this chapter, for the ESE/¬ESE classification, there is a 6% increase in performance. What this value shows is the relative error rate for each classification, 0.22 in the first pipeline and 0.17 for this pipeline. It should also be noted, however, that the results are not directly comparable as the training and test data are different due to adjusted annotation guidelines. The recall of the DMRNN also shows that any improvement has to be made in determining when a sentence

[2]All models used the same set of hyper-parameters for testing. The CNN layers also used relu activation as it is less expensive to train.

[3]This error rate, the inverse of 0.83 in this example, can be considered as the most important factor in classifier selection as it shows the practical application of the classifier which must weigh-up correct classification of the target class and incorrect classification.

[4]CNN models in text classification are prone to “memorising” the training data where the dataset is small, this can be evidenced through a high training accuracy and relatively low test accuracy.

is not an ethotic one as this will improve precision. Precision can also be improved by using a threshold on the output of the DMRNN. By replacing the Sigmoid layer, which has a binary output, with a softmax function, which provides an output between 0 and 1, a threshold can be set that optimises precision or recall. In this instance though, optimising precision could return too few true positives and on the other hand optimising recall could create too many false positives. As stated above the classifier, the DMRNN, was chosen as the macro-averaged F1-score provided the lowest error rate of classification.

ESE/¬ESE	**Precision**	**Recall**	**F1-score**	**m-F1-score**
Baseline	0.50	0.78	0.61	0.73
(Duthie et al., 2016a)	0.64	0.76	0.70	0.78
LR + Text	0.53	0.98	0.69	0.78
LR + Text + UD + POS + EXT	0.51	0.98	0.67	0.77
SVM + Text + UD + POS + EXT	0.52	0.98	0.68	0.77
POL	0.66	0.67	0.66	0.79
RNN + Text	0.52	0.98	0.68	0.77
CNN + Text	0.54	0.87	0.67	0.77
CNN + LSTM + Text	0.57	0.85	0.68	0.79
CNN + Text + UD LSTM	0.54	0.93	0.68	0.79
RNN + Text + UD LSTM	0.55	0.91	0.69	0.79
RNN + Text + UD LSTM + POL	0.58	0.85	0.69	0.80
RNN + Text + UD LSTM + POS + POL	0.58	0.88	0.70	0.80
RNN + Text + UD LSTM + EXT LSTM + POL	0.62	0.85	0.71	0.81
DMRNN	0.60	0.95	0.74*	0.83*

Table 18: Precision (P), recall (R) and $F1$-score are reported for the classification of ESEs and the macro-averaged $F1$-score (m-$F1$) is reported for the classification of ESEs and¬ESEs. Results from the previous work in ethos mining, standard machine learning classifiers, experimental classifications using different CNN and RNN modular combinations and our final DMRNN are compared to a baseline which classifies on the training set distributions. (*) denotes classifier with the highest $F1$-scores.

In the case of ESE/¬ESE a confusion matrix has been constructed to identify where the implemented pipeline can be improved. The confusion matrix in figure 28 shows the relative classifications of the pipeline against the manual gold standard test corpus. The results of this pipeline and this confusion matrix show that false positives are the main error. For the identification of ESEs only 6% are classified incorrectly indicated in the high recall, 19% of the ¬ESEs are classified as ESE which indicates where the main focus of this error analysis should be for overall improvement.

Actual Values

Predicted Values		ESE	¬ESE
	ESE	230	153
	¬ESE	13	809

Figure 28: Confusion matrix for the ESE/¬ESE classification using the DMRNN.

False positive errors mainly pertain to a failure in feature extraction, the identification of entities not relevant to the parliament, errors in features created from the sentiment lexicons, and politeness misconstrued as positivity. In the EXT module some entities and relations extracted are not relevant for the particular statement with these errors coinciding with the overall target of the sentence not being about an entity and rather about other parts of the statement which is relevant for those sentences which are mainly self referential. The identification of entities not relevant to the UK parliament or a lack of tense realisation alongside the wider context of a statement also cause problems. Errors from sentiment feature extraction relate to noise introduced by the lexicon and statements showing politeness can be misconstrued as support of ethos, which is common. Each of these errors listed above are illustrated in the examples 19, 20, 21 and 22.

(19) *"No doubt my hon. Friend will understand that those who resent the present system would be just as concerned about any substitute system in many areas."*

(20) *"On the first part of my hon. Friend's question I should be in even greater difficulty if I were to involve myself in legal arguments."*

(21) *"Will the Government now press South Africa to allow Mrs. Winnie Mandela to leave her own country next month to attend the women's conference in Nairobi?"*

(22) *"I too congratulate the Minister on his promotion within the Treasury and I wish him good fortune especially in solving the growing new problem of balance of*

payments deficits."

In example 19 there is no explicit attack on ethos, where the target of the sentence is "those who resent" rather than "hon. Friend". The EXT module has extracted "Friend" and the related words leading to the classification error. Example 20 is mainly a self-referential statement. The EXT module extraction fails to distinguish between the relations of the relevant entity "hon. Friend" and those referred to in the self referential part of the statement. Example 21 shows an instance where an entity not relevant to the UK parliament is identified, in this case "her" refers to "Mrs. Mandela" and should therefore not be extracted. Example 22 shows a different problem where the general politeness shown in the UK parliament can cause issues in the classification of ESEs, a Minister is being congratulated on promotion, but not praised in a sense that would be supporting ethos as this acknowledgement is down to parliamentary etiquette. The entity extraction would need knowledge of the wider context of politeness in order to distinguish this from praise.

From a qualitative analysis it is clear that general knowledge would increase the accuracy of the +/-ESE classification while increasing the positive word lists in each lexicon would also aid in this. Improvements on entity extraction would improve both classifications overall. This can be achieved through the use of a wider domain context or knowledge, or by distinguishing the main topic in a statement, although this will add noise to any extraction.

6.3.4 Combined Results

Each of the +/-ESE and ESE/¬ESE classifiers are then combined where a correct ESE classification is taken and then a +/-ESE classification is made. If this is correct then the classification is deemed as correct (see table 19 and figure 29 for results and a confusion matrix). The baseline trained on the class distributions gives the $F1$-score 0.39, and the previous work, in chapter 5, gives $F1$-score 0.60. The highest performing classifier in this approach, the DMRNN, gives an $F1$-score 0.65, outperforming the baseline by 66.7% and the previous work by 8.3%.

ESE/¬ESE & +/- ESE	P	R	$F1$-score
Baseline	0.32	0.51	0.39
Rule Based System (chapter 5)	0.55	0.65	0.60
DMRNN	0.56	0.78	0.65

Table 19: Results are provided for the combination of the ESE / ¬ESE stage and the +/- ESE stage.

Actual Values

Predicted Values	+ESE	-ESE	¬ESE
+ESE	39	18	37
-ESE	21	150	112
¬ESE	4	11	813

Figure 29: Confusion matrix for the combination of the ESE / ¬ESE stage and the +/- ESE stage.

6.4 Discussion

The results of both the annotation and extraction of ethos show the reliability of both the newly created annotation guidelines and the ethos mining pipeline described in this chapter. In each case improvements have been made above the previous annotation and extraction described in chapter 5, again answering RQ1 and RQ2.

This chapter has focussed on providing an extension and improvement to the previously developed corpus of ethos annotation and a re-developed ethos mining pipeline for automatic extraction which is more easily generalisable. In the former case this has been achieved through a re-annotation step, increasing the size of the data and in turn ensuring that ethos is annotated on the linguistic surface. In the latter case this has been achieved through the application of novel techniques to a text classification task, namely a Deep

Modular Recurrent Neural Network. Both of these improvements, manual annotation and automatic extraction, show the reliability of identifying ethos supports and attacks in parliamentary debate and therefore provide an extensive answer to RQ1 and RQ2. Despite the improvements over the previous rule-based methods in automatic classification, both sets of results show reliability meaning that either can be utilised for ethos mining. Although improvements have been made in annotation and automatic extraction there are areas which can be changed to make each more reliable.

6.4.1 Annotation Improvements

In the case of the annotation task, there was no improvement over the previous chapter when it came to determining if a sentence contained ethos or not. Partly this can be explained due to the contextual information which is still given in each sentence. This means that for an annotator it can be difficult to determine which part of a sentence is ethotic or not especially due to the hostile language in parliament.

One possible solution is a further extension of the corpus, rather than annotating only ethos, to logos. This move would allow the transcripts to be broken down into further segments as part of a larger argument structure. This will benefit ethos identification as it should become more clear which parts of sentences are ethotic. Although the argument annotation will make the task, overall, more difficult, the relations identified could provide pointers to the relevant context in the text. These relations can also rule out some of the possibilities of ethos annotation, and therefore will reduce the set of sentences which can be annotated for ethos. In the case of identifying targets of each ethotic statement, further segmentation of the text should make this task easier. In these cases the segmentation may separate multiple mentions of more than one entity in a sentence in make the target obvious.

Though further segmentation and argument annotation may solve some annotation issues it will make the task more difficult with a choice between annotating an argument relation, an ethotic relation, or no relation. Despite this issue it will mean the corpus created can be more widely used. Not just as a resource for performing ethos mining but also argument mining. To this end improvements can also be made in the automatic classification tasks.

6.4.2 Automatic System Improvements

As eluded to in the previous section the accuracy of the anaphora resolution task can be improved when coupled with other proposed improvements. In this case a further segmentation of the text should make the anaphora resolution task simpler. This segmentation will provide more structure in the text meaning that when a backwards iteration is performed there are a greater number of immediate choices for an entity. Although performance wise this will slow down any decision, it will be a trade off with accuracy. Improvements could also be made by trying to utilise machine learning methods for classification. This would, however, require a further manual annotation step linking targets with entities throughout the dataset.

In the case of the +/-ESE classification a further segmentation may also provide improvements. This should give less text to classify which will therefore contain a smaller number of sentiment holding words. In this case it should improve the overall classification. A further step to improvement can be through the increase in positive words in the sentiment lexicons. The addition of the positive words from the Hansard lexicon did not provide more accurate results, because of the way the corpus was created and the number of false positives. A possible solution would be to use all of the available data manually annotated for ethos, and manually annotate a further set of transcripts as test data. Whilst this may provide positive results the time taken to annotate the further data may outweigh the increase in accuracy.

In the case of the ESE/¬ESE classification a further segmentation and argument annotation may provide a number of opportunities for improvement. The relations gained from this classification could be used as indicators for ethos. For example, a chain of inference relations may be proceeded by one of ethos. Ideally, for any automatic classification the end result of an argument annotation could be used for these indicators. Multi-task learning could also be utilised, performing the argument and ethos mining tasks at the same time.

Other improvements could be made by attempting to encode a domain knowledge for the ethos mining in parliamentary debate task. The DMRNN approach described in this chapter made small steps to this by using a number of word embedding layers. In practical terms, however, this would need to be extended greatly to provide more of a benefit. Training a set of word embeddings on Wikipedia is unlikely to work for this task. The language used and the related words from their most frequent context are very likely to differ to that of parliamentary language. As this thesis has focussed upon one area of parliamentary debate, that of oral answers to questions, word embedding models can be created in other areas, such as , written answers to questions. Despite the fact that the language will differ slightly, the use of each word should remain within a similar context.

As a whole the improved ethos mining pipeline works effectively. There are still questions though as to how these methods can be utilised in a different domain. As a first step, in a new domain, at the very least a small amount of manual annotation should be undertaken. Not only to understand the problem space but also as a set to test the pipeline. In this annotation the entity or entities of interest would have to be identified. As the pipeline in this chapter uses domain specific rules for this purpose, there are two possibilities. Extend or re-create the domain specific rules or use NER if the problem is well enough defined.

In the former case this will require more of a domain knowledge, a larger set of manual annotation, and a larger time frame for creation of automatic methods. In the latter case entity mentions would have to be rather specific although this is dependent on the domain. In all other feature cases the modules should provide enough data for classification although all of the features used in this chapter may not be as relevant in another domain. In the case of the DMRNN, the model would have to be re-defined as it is tailored for the ethos mining task in parliamentary debate. The test or validation set could be utilised for this purpose allowing an extensive set of models to be tested.

This chapter has identified the difficulties which lie in ethos annotation and automatic extraction and demonstrates how improvements can be made. Not only through changes to annotation guidelines and extraction techniques but also in the possibilities of looking to more established areas, like argument mining, for inspiration.

Chapter 7

Ethos Type Mining

This chapter outlines the novel fine grained annotation and automatic classification of ethos in which types of ethotic strategy are identified: ethos supports and attacks on the grounds of practical wisdom, moral virtue and goodwill (Aristotle, 1991, 1378a)[1] in line with RQ3. Elements of ethos (see section chapter 2.1.3) provide a next step in the classification of ethos, using ethos supports and ethos attacks as a base for which annotation, both manual and automatic, can be conducted. As with annotating and classifying ethos, elements of ethos can be annotated using the Aristotelian categories which are less well defined than ethos in general. Due to this fact, some inspiration was taken from textbooks in rhetoric to shape the annotation guidelines. These definitions are coupled with two iterations which work to annotate, evaluate and redefine the guidelines to allow for improvements (see sections 7.1.1 and 7.1.2).

As mentioned in section 2.2 elements of ethos and argument schemes have several similarities, not necessarily in the content of each, but in the application. Both argumentation schemes and elements of ethos can be applied as a final step in any manual or automatic classification to further refine an annotation, thus, methods used for the automatic classification of argument schemes could be used in the automatic classification of elements of ethos. Argumentation schemes apply predominantly to logos rather than ethos, where

[1]The work in this chapter relates to one published work (Duthie and Budzynska, 2018a) where elements of ethos were manually annotated over multiple iterations (in collaboration with Katarzyna Budzynska) and automatically extracted. Some passages are used verbatim from the source.

the steps taken to move from sentences already with a base annotation to a multi-class classification can be emulated as this is a non-trivial step. The step is non-trivial in that the same sentence can be classified into multiple categories depending on the classification method or there may be a lack of balanced data for a purely multi-class step.

Within this thesis, elements of ethos can be considered as an additional step following an ethos mining pipeline. Rather than producing an output of only +/- ESEs, elements of ethos can be identified providing another possible dimension of political analysis. The examples below show three common ethotic strategies: in example 23 Mr. Moore is supporting the experience and knowledge of an entity (Miss Widdecombe in this case); in example 24 Mr. Jenkin is endorsing the Government for having courage; and in example 25 Mr. Moore is referring to Mr. Meyor's good deeds in respect to an audience (his constituents). These strategies correspond to three elements of ethos studied in rhetoric: practical wisdom when the reference is made to having knowledge; moral virtue when there is a mention of truth and courage around knowledge (when the speaker is honest); and goodwill when the knowledge is shared (when the speaker gives the best advice to others).

(23) **Mr. John Moore said,** *I bow to my hon. Friend's distinguished past and detailed knowledge of these matters.*

(24) **Mr. Patrick Jenkin said,** *I believe that the Government were right to have the courage to bring forward the necessary measures to bring public expenditure under control.*

(25) **Mr. John Moore said,** *My hon. Friend is assiduously pursuing his constituents' interests.*

This chapter encompasses the manual and automatic annotation process needed for identifying elements of ethos, describing: the iterative process required to create new annotation guidelines needed for corpus creation in two iterations (see sections 7.1.1 and 7.1.2); the methods which can be utilised for the automatic classification (see section 7.2); an evaluation of both manual iterations (see sections 7.1.1.2 and 7.1.2.3) and automatic methods (see section 7.3); and a discussion of the possible applications of elements of ethos (see section 7.4).

7.1 Annotation

The annotation of elements of ethos, both support and attack, builds upon the annotation of ethos which was undertaken and described in chapter 6. The first step for the classification of elements of ethos is the process of a manual annotation and evaluation of the annotation guidelines. The three ethos elements (wisdom, virtue and goodwill) are annotated using OVA+ with all annotation stored in AIFdb (see chapter 4 for a description of OVA and AIFdb). To annotate ethos types, the base annotation from chapter 6 was used, overall this gave 638 ethotic statements annotated as support and attack. The annotation in OVA involves making an inference or conflict connection between a reconstructed proposition (the proposition is linked to the original text segment), and a general ethos node for an entity. A ”Default Inference” connection is used to show a support of ethos and ”Default Conflict” for an attack on ethos. The ethos element annotation involves changing the ”Default Inference” to Wisdom, Virtue or Goodwill (see figure 30 for this step). The same holds true for ”Default Conflict”, which is also changed in the same way to Wisdom, Virtue or Goodwill.

Given in this section are two annotation iterations the first using basic definitions for wisdom, virtue and goodwill followed by an evaluation and error analysis. The second then implements the suggested improvements from the first annotation and creates a corpus of ethos types which can be used for automatic extraction.

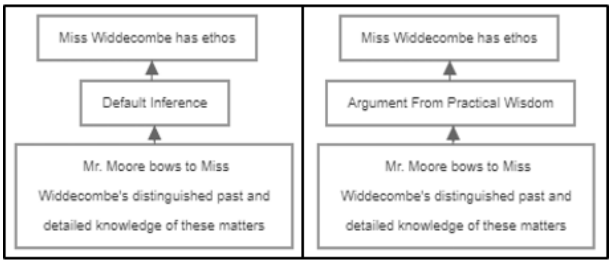

Figure 30: Example (23) annotated in OVA using the AIF format. A reconstructed proposition is connected to an entity ethos node through ”Default Inference”, after applying the Wisdom tag this is changed to ”Argument From Practical Wisdom” (connection to the original text segment via locutions in IAT is not shown for simplicity).

7.1.1 First Annotation Iteration

The manual annotation of ethos types encompasses two iterations taking a set of guidelines on which to annotate the data, an evaluation step where a second annotator completes a subset of the annotation and an annotation guideline editing step. Two iterations are utilised in order to improve the initial annotation of ethos types this is due to the ambiguous definition of these types given by Aristotle. The first annotation iterations allows the basic structure to be defined whilst the second allows this to be refined.

7.1.1.1 Guidelines

In the case of the first annotation iteration, a set of guidelines containing character traits and situations which typically indicate elements of ethos were constructed. While these guidelines are indicative of ethotic support, the opposite of any of the set can be applied for ethotic attack. Due to the ambiguity in what is written by Aristotle (Aristotle, 1991), the source of definitions for each category had to be broadened. The guidelines (Crowley and Hawhee, 2004; Fahnestock and Secor, 2003; Garver, 1994) and use the main tags (Practical Wisdom, Moral Virtue and Goodwill) split into support and attack (Argument and Conflict) resulting in 6 labels according to the following guidelines:

Practical Wisdom should be annotated when the statement refers to an entity having a sufficient knowledge for the purpose at hand or an ability to draw the right conclusions from this knowledge while balancing the moral good and bad and knowing what will benefit man. It may also refer to the practical experience of an entity and an entity's ability to produce the right decision from this practical experience not for one's own benefit. In this case the label "Argument from Practical Wisdom" should be applied replacing "Default Inference". This is for the positive case outlined above. For attacks on ethos the opposite holds true e.g. lack of knowledge. In the case of attacks "Conflict to Practical Wisdom" should be annotated in place of "Default Conflict". This provides four possibilities for which Practical Wisdom can be annotated: (*a*) an entity is said to have sufficient knowledge for the purpose at hand; or (*b*) an entity can draw conclusions from this knowledge; or (*c*) an entity has practical experience; (*d*) an entity can draw conclusions from this experience.

Moral Virtue should be annotated when the statement refers to an entity's positive morality, calmness, justness, selflessness, gracefulness, nobility, positive contributions, liberality, magnanimity or magnificence. It may also refer to an entity's ability to provide the correct information. In this case the label "Argument from Moral Virtue" should be applied replacing "Default Inference". This is for the positive case outlined above. For attacks on ethos the opposite holds true e.g. unjust, selfish etc. In the case of attacks "Conflict to Moral Virtue" should be annotated in place of "Default Conflict". This provides two possibilities for which Moral Virtue can be annotated: (*a*) a statement refers to the character trait of an entity, when the entity shows positive morality, calmness, justness, selflessness, gracefulness, nobility, positive contributions, liberality, magnanimity or magnificence; or (*b*) when an entity provides the correct information.

Goodwill should be annotated when the statement refers to an entity's ability to show goodwill to others with respect to giving sound advice when it is known or caring about who they represent, while ensuring the entity does not deceive and is inclusive avoiding unnecessary repetition of information. It may also take the form of an entity aligning with an audiences values and displaying self sacrifice. In this case the label "Argument from Goodwill" should be applied replacing "Default Inference". This is for the positive case outlined above. For attacks on ethos the opposite holds true e.g. provide bad information or advice that is only partly true etc. In the case of attacks "Conflict to Goodwill" should be annotated in place of “Default Conflict”. This provides three possibilities for which Goodwill can be annotated: (*a*) a statement refers to an entity’s ability to show goodwill to others; or (*b*) an entity gives sound advice when it is know, ensuring the entity does not deceive while being inclusive; or (*c*) an entity aligns with an audiences values, displaying self sacrifice.

Default should be annotated when there is no clear type classification from Practical Wisdom, Moral Virtue and Goodwill. This classification is rare especially within parliamentary discourse where the nature of ethos supports or attacks means that they normally fall into one of the three categories. This can be attributed to the topics of discussion which assume some form of wisdom, virtue and goodwill in each person which can then be supported or attacked. In the case of annotating default there is one possible choice for support: when a statement does not make reference to Moral Virtue, Practical Wisdom or Goodwill and puts the entity in a positive light Default Inference should be applied. For conflict: when a statement does not make reference to Moral Virtue, Practical Wisdom or Goodwill and puts the entity in a negative light Default Conflict should be applied.

Accompanying the annotation guide is an annotation decision tree. Each ethotic statement is identified through a set of attributes, if the statement does not appeal to a character trait (Moral Virtue) or knowledge (Practical Wisdom) then it is considered as being either Goodwill or Default. The decision tree also makes use of a set of key words for each category identified through the literature (see appendix .B for keywords). Figure 31 shows the decision tree with the possible ethos types sectioned on the left side and the possible decisions of the annotator shown in boxes. The intuition behind the decision tree is to aid annotators in making decisions about ethos types with a more concrete set of guidelines. Each type, apart from default, then has two main questions that if positively answered result in the label being chosen.

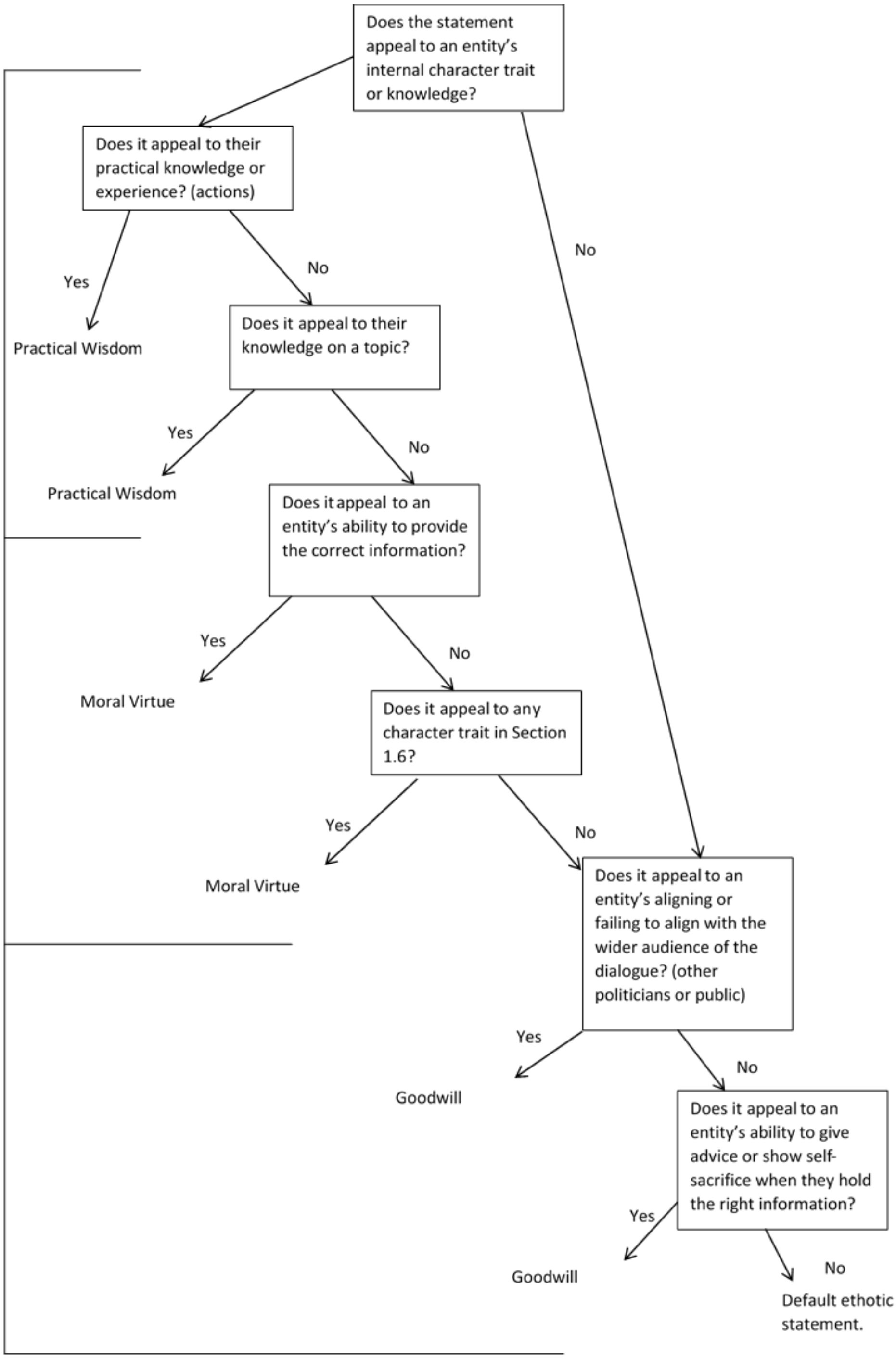

Figure 31: Annotation decision tree for the first annotation iteration.

7.1.1.2 Inter-Annotator Agreement

To determine the reliability of the annotation of ethos types, an evaluation of the first iteration was undertaken. A 12% subset of the data was annotated by a second annotator giving a Cohen's $\kappa = 0.42$ and weighted $\kappa = 0.63$ (see table 20). Weighted kappa is

Pair	κ	**Weighted** κ
Ann1 - Ann2	0.42	0.63
Ann1 - Ann3	0.63	0.79
Ann2 - Ann3	0.51	0.70

Table 20: κ and weighted κ are given for pairwise annotators (Ann1-Ann3).

Pair	**Wisdom / Virtue** κ	**Virtue / Goodwill** κ
Ann1 - Ann2	0.63	0.46
Ann1 - Ann3	0.79	0.72
Ann2 - Ann3	0.64	0.64

Table 21: κ scores for pairwise annotators on polynomial ethos types Wisdom and Virtue against Goodwill and Virtue and Goodwill against Wisdom.

specifically shown to demonstrate the agreement when the classes are more fluid, where near misses are less harshly penalised. The κ score is considered fair with the weighted κ indicating areas of improvement in the guidelines. The weighted κ score gives an indication of the issues in annotation, essentially where the overlapping nature of the ethos types creates errors.

To this extent a third annotator annotated a smaller 8% subset of the data to provide a pairwise agreement score. Comparing all three annotators gave Fleiss $\kappa = 0.51$. A pairwise comparison of annotators one and three gave $\kappa = 0.63$ and weighted $\kappa = 0.79$. Comparing annotators two and three gave $\kappa = 0.51$ and weighted $\kappa = 0.70$. The three coders were experienced in argument annotation, with one a native speaker of English. For the purposes of this task no prior training was given other than a set of examples provided in the annotation guidelines.

As shown through the pairwise evaluation and weighted kappa scores, the annotators had difficulty classifying some sentences due to their closeness to more than label. Combining ethos types into polynomial categories can give an indication of the areas where the guidelines can be improved (see table 21). When combing wisdom and virtue against goodwill annotators one and two had $\kappa = 0.63$, annotators one and three had $\kappa = 0.79$ and annotators two and three had $\kappa = 0.64$. When combing virtue and goodwill against wisdom annotators one and two gave $\kappa = 0.46$, annotators one and three gave $\kappa = 0.72$ and annotators two and three gave $\kappa = 0.64$. Combining wisdom and goodwill did not show any significant increase. These κ values indicate that improving the guidelines would increase the performance of the annotation of ethos types. Specifically, the increase in κ value on the wisdom and virtue class indicate there closeness in the guidelines and the difficulty to tell them apart when annotating. The same conclusion can be drawn from the virtue and goodwill class, where the kappa values increase over the base kappa scores from the annotation indicating the difficulty in classifying between virtue and goodwill.

7.1.2 Second Annotation Iteration

A second manual classification of the data set was undertaken following the evaluation in section 7.1.1.2. Although annotating on the same data with the same tags, the guidelines for each ethos type classification were made more clear through the distinction between knowledge and actions. In turn this should provide an increased agreement between annotators through the further distinctions between each label.

7.1.2.1 Guidelines

The annotation of the ethos types are broken into two categories knowing information (knowledge) or knowing the right actions (actions). Moral Virtue and Goodwill are further distinguished in two ways. Where the two categories apply in general, this is virtue, and when this applies to an audience this is goodwill, wisdom does not have this distinction. The tags are applied in the guidelines (see `http://arg.tech/WVGGuideNew` for full guide) as follows:

Practical Wisdom. Argument From Practical Wisdom should be annotated when an entity: (*a*) knows the right information; or (*b*) knows the right action. Conflict From Practical Wisdom should be annotated when an entity: (*a*) does not know the right information; or (*b*) does not know the right action.

Moral Virtue. Argument From Moral Virtue should be annotated when an entity: (*a*) knows and reveals the right information in general; or (*b*) is honest in general; or (*c*) performs the right action when they know it; or (*d*) does the right action in general. Conflict From Moral Virtue should be annotated when an entity: (*a*) knows information but does not reveal it in general; or (*b*) lies in general; or (*c*) performs an action when they know it is wrong; or (*d*) does the wrong action in general.

Goodwill. Argument From Goodwill should be annotated when an entity: (*a*) knows and shares information with the audience; or (*b*) is honest with the audience; or (*c*) performs the right action for others aligning with their values giving sound advice; or (*d*) does not do wrong to others. Conflict From Goodwill should be annotated when an entity: (*a*) does not share information with the audience; or (*b*) misleads the audience; or (*c*) does not do themselves what they know is right for the audience; or (*d*) does the wrong things for an other or audience.

7.1.2.2 Corpus

Following the re-annotation of the ethos support and attack types, for the second iteration, the publicly available EthosWVG_Hansard corpus (`http://arg.tech/EthosWVG`) was created. As stated in Section 7.1 the corpus builds upon the tags present in the Ethos_Hansard corpus to create, with a total of 638 segments and 18,250 words (see table 22 for further details), the only available corpus containing both support and attack ethos types.

The constructed corpus, (see table 22), shows the imbalance within the dataset. First in regard to the positive and negative aspect of the data, secondly in regard to the word count of the statements in comparison to the full corpora for ethos mining (18,250 against 90,991 in chapter 6), and thirdly with respect to the ethos types practical wisdom, moral virtue and goodwill. Firstly, ethos data is rather sparse within natural language. Taking into account the nature of parliamentary debates and the topics of discussion, it is not surprising that ethotic statements do not make up a higher portion of the data as a whole particularly as the discussion gravitates towards specific legislation. Secondly, when it comes to supporting

Tags	Ethos Supports	Ethos Attacks	Total	Word Count
Wisdom	48	190	238	6,954
Virtue	99	194	293	7,611
Goodwill	20	87	107	3,685
Total	167	471	638	18,250

Table 22: Occurrences of tags with the respective word counts for each segment in EthosWVG_Hansard.

Pair	Wisdom, Virtue, Goodwill		Wisdom / Virtue	Virtue / Goodwill
	κ	Weighted κ	κ	κ
Ann1 - Ann2	0.52	0.70	0.81	0.61

Table 23: κ score and weight κ score for all ethos type annotation and κ scores for polynomial ethos types Wisdom and Virtue against Goodwill and Virtue and Goodwill against Wisdom.

and / or attacking ethos the latter is the more preferred option most likely because within parliamentary discourse there is the idea of opposition. There is then a higher likelihood of cross party ethos attacks rather than cross party ethos supports. Finally, the imbalance over the ethos type labels can be attributed to the nature of each label and the relative ease as to which each label can be called upon. The hypothesis is that it is easier to attack the character trait of a person than it is to attack their knowledge on a topic as a whole unless this person is well known. In the same respect then it is easier to accuse someone of having bad morals than it is to say that this person misleads the general public, as is the case for goodwill. The ease as to which each label can be utilised is reflected by the values in table 22.

7.1.2.3 Inter-Annotator Agreement

To evaluate the redefined guidelines, the annotator pair showing the lowest κ score from Section 7.1.1.2 was compared. The assumption being in this case that the annotators showing the highest agreement values will only improve with added information for annotation. Although this would reflect well on the overall results it does not give a true reflection of the annotation task, therefore, the focus was placed upon the pair with lowest

		Ann 2		
		W	V	G
Ann 1	W	5	16	1
	V	3	39	0
	G	0	8	7

Figure 32: Confusion matrix for inter-annotator agreement between annotators 1 and 2.

agreement. Again a 12% subset was annotated by the second annotator giving $\kappa = 0.52$ and weighted $\kappa = 0.70$ (see Table 23 and figure 32 for a confusion matrix). This result shows a significant improvement over that shown in table 20 (κ of 0.42 against κ of 0.52). Although this is the case there is still room for improvement within the annotation with the weighted κ scores again showing that the distinction between classes could be made more clear.

To test this hypothesis ethos types were again combined into polynomial classes to give an idea of the disagreements between the annotators. When combining wisdom and virtue against goodwill, $\kappa = 0.81$ whilst combining virtue and goodwill against wisdom, $\kappa = 0.61$. This shows that the new guidelines made clear the distinction between virtue and goodwill although the difference between wisdom and virtue could be improved significantly.

Further error analysis showed that the language used within the UK parliament can make the distinction between wisdom and virtue difficult. Example 26 highlights this point where at first glance the entity is referred to as an expert, pertaining to wisdom, however, an expert in doggerel and verse, is considered as being clumsy and irregular, overall an attack on virtue. The same phenomenon is shown in example 27 although the language used is less easy to understand without some wider knowledge of the metaphor and context. In this case there are two possible interpretations as to the overall label. The first being an attack on moral virtue, this is indicated through the use of “never-never land” which could suggest that the person in question is living in a fantasy world and therefore lacks in moral virtue due to the attack on a character trait. Practical wisdom can also be annotated, suggesting that someone who lives in "never-never land" indicates that they have a degree of fantasy in their ideas and therefore would lack common sense knowledge.

(26) *"I understand the right hon. Gentleman to be something of an expert in doggerel and verse."*

(27) *"I will really think that he is living in never-never land."*

7.2 System Architecture

For the automatic classification of types of ethos support and attack, an extension was made to the already existing pipelines for +/- ESE classification where either pipeline in chapter 5 or chapter 6 can be used as a starting point to determine +/- ESE classification. The type classification uses +/- ESEs as its base, meaning that any classification does not need to resolve ¬ESEs as was the case in chapters 5 and 6. For classifying each of the types (wisdom, virtue and goodwill) a pipeline was created (see figure 33). The +/- ESEs were broken into separate component parts as features which create separate components in the pipeline: the entity relations (EXT) from chapter 6 were updated and combined with POS tags to give a new EXT/POS component; the raw ESE text; the polarity of each ESE; and a module which checks for the presence of plural and proper nouns (NNS/NNP). Each of these were passed to a principal component analysis (PCA) module after which a classifier is used to determine the type of ethos support and attack.

7.2.1 Existing Methods

Existing methods used in the pipeline consist of Ethotic Sentiment Expression text, Principal Component Analysis, and classifiers.

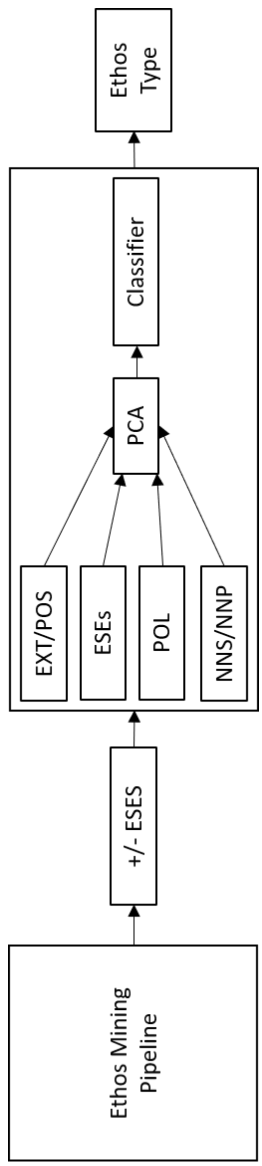

Figure 33: Pipeline for classifying types of ethos support and attacks containing a combination of entity relations and POS tags, ESEs, POL for ESEs and the presence of NNS and NNP tags. These are passed to a PCA module to reduce the dimensionality of the data for classification, which ultimately gives an ethos type.

7.2.1.1 Ethotic Sentiment Expressions (ESE)

This component uses the full ESE text. The text has stop words removed and is transformed into unigrams, bigrams and trigrams as features for classification. Using the standard BOW approach may provide useful features for classification which have otherwise been removed erroneously in the case of the EXT / POS module (see section 7.2.3.1 for a description).

7.2.1.2 Principal Component Analysis (PCA)

This component uses a scikit-learn (Pedregosa et al., 2011) built in PCA to reduce the dimensionality of the data. Fifteen component features were applied as points for the reduction (two components were also specified for reduction to a 2D space for the purposes of visualisation). This step means there is less reliance on more complex classifiers that may take a longer training time period and be more prone to over-fitting.

7.2.1.3 Classifier

Four classifiers were considered for the ethos type classification: Linear Support Vector Machines (SVM), Logistic Regression (LR), Naive Bayes (NB) and a Decision Tree (DT). Scikit-learn implementations were used for the classifiers, all of which performed both classification experiments. More complex methods of classification could be considered in this case, such as deep learning, but the small amount of training data means that in the case of ethos types this was not considered viable.

7.2.2 Adapted Methods

Methods which have been adapted specifically for the pipeline consist of only the polarity component.

7.2.2.1 Polarity (POL)

This component splits all data points into positive and negative due to the varying language that can be used to support and attack. The intuition behind this module is fairly simple, when ethos is attacked a more negative language is used or there is an association with an object that may be deemed to have negative connotations. The same holds true in the case of supports of ethos. For types of ethos the difference between support and attack could be in antonyms.

7.2.3 Novel Methods

Novel methods for the pipeline consist of the entity extraction and part-of-speech component and the plural and proper noun component, which have been developed specifically for this research.

7.2.3.1 Entity Extraction / Part-Of-Speech (EXT / POS)

This component was created to remove any entities which are not relevant to ethos (e.g. removes statements of the kind "Will the hon. Member" which show an entity being addressed to ask a question) leaving only the relevant entities and the words related to them. This is achieved through manual domain rules executed upon UD tags (see chapter 6 for full explanation). POS tags are then matched to these entities to give words plus POS tags and stop words removed. The inclination behind this component is that while the full dependency parse may not provide a robust feature for classification alone, constraining the output to only parts of the dependency tree which are relevant should provide better quality data for classification.

7.2.3.2 Plural Noun / Proper Noun (NNS / NNP)

This component was created to add a binary label where an ESE either contains or does not contain a plural or proper noun. The intuition being that ethos supports or attacks of the type goodwill, will mainly feature an additional entity within the text (e.g. "farmers" or "constintuents") which are not captured using standard NER tools. Although this technique will produce many false positives, the lack of data for the goodwill class means that some form of feature engineering not reliant on BOW must be produced.

7.3 Results and Evaluation

Results are reported for two cases of types of ethos support and attack classification using the data annotated in the second annotation iteration. One vs All classification and pairwise classification have been shown as effective methods for the case of argument scheme classification which is a multi-class problem. Therefore the same procedure as in (Lawrence and Reed, 2016) and (Feng and Hirst, 2011) has been followed.

7.3.1 One vs All classification

In table 24 (also see figure 34 for a confusion matrix), results are reported for the One vs All classification for wisdom, virtue and goodwill with the highest $F1$-scores marked with stars. SVM, LR, NB and DT classifiers are all compared against a baseline which classifies on the class distribution of the dataset. A 10-fold cross validation is used to test each classifier, providing a prediction for every point in the data using only a general set of features for training data[2].

The SVM and LR classifiers perform the most consistently over the three types with all the classifiers above the baseline. Increasing the overall size of the data set would allow for the creation of a hold out test set, where more specific features such as lexical cues can then be extracted for classification. Structural features such as the position of words related to ethos in a sentence or in the text as a whole are not considered, although these features may provide a better classification accuracy, they do not generalise well when the text is changed. The highest $F1$-score of 0.55 was achieved in the virtue classification by the SVM classifier, with a 25% increase over the baseline. The highest macro-averaged $F1$-score of 0.77 was achieved in the goodwill classification by the DT classifier although the relative $F1$-score was only 0.31.

7.3.2 Pairwise Classification

In table 25 (also see figure 35 for a confusion matrix), results are given for the pairwise classification with comparison between a baseline, classifying on the training set distributions, and SVM, LR, NB and DT classifiers. A 10-fold cross validation was used to

[2]This equates to a prediction for 638 tags in total 238 - Wisdom, 293 - Virtue, and 107 - Goodwill.

	Wisdom				Virtue				Goodwill			
	P	**R**	**F1**	**m-F1**	**P**	**R**	**F1**	**m-F1**	**P**	**R**	**F1**	**m-F1**
Baseline	0.34	0.35	0.35	0.51	0.44	0.45	0.44	0.48	0.15	0.16	0.15	0.71
SVM	0.42	0.68	0.52*	0.54	0.58	0.52	0.55*	0.60*	0.27	0.50	0.35	0.72
LR	0.43	0.61	0.50	0.56*	0.55	0.54	0.55*	0.59	0.29	0.55	0.38*	0.73
NB	0.39	0.49	0.43	0.53	0.51	0.57	0.54	0.55	0.25	0.36	0.30	0.73
DT	0.43	0.49	0.43	0.53	0.50	0.55	0.53	0.54	0.31	0.31	0.31	0.77*

Table 24: One vs All classification for Wisdom, Virtue and Goodwill. Precision, recall and $F1$-score and macro-averaged $F1$-score are reported where the $F1$-score relates to the type in question and macro-averaged $F1$-score the combined classification. A baseline classifying on the class distributions is compared against machine learning classifiers using a 10-fold cross validation. * denotes the classifiers with the highest scores.

Actual Values

Predicted Values		W	V	G
	W	106	106	25
	V	84	134	41
	G	48	53	41

Figure 34: Confusion matrix for One Vs All Wisdom, Virtue and Goodwill classification against the manually annotated data.

test all the data points within the dataset. Following the One vs All classification, the pairwise classification makes the distinction between borderline cases, meaning that any classification of the same data point for two types in the One vs All classification would be resolved with a pairwise classification. Again all the classifiers perform significantly above the baselines with the LR classifier performing best overall.

The results of this classification show that a distinction can be made when classifying goodwill against wisdom and virtue (F1 = 0.67), however, the distinction between wisdom and virtue (F1 = 0.57) is less clear. This same issue is apparent from the evaluation in the human annotation task, thus, any improvements made there could be reflected in the automatic classification. New features reliable for human classification could be developed for automatic classification depending on how much contextual or domain based knowledge is required for this classification.

Errors within the classification mainly pertain to the tightly coupled nature of the ethos

	Wisdom / Virtue			Wisdom / Goodwill			Virtue / Goodwill		
	P	R	F1	P	R	F1	P	R	F1
Baseline	0.49	0.48	0.49	0.56	0.56	0.56	0.62	0.60	0.61
SVM	0.54	0.55	0.55	0.65	0.67	0.66*	0.67	0.67	0.67*
LR	0.57	0.58	0.57*	0.67	0.66	0.66*	0.71	0.65	0.67*
NB	0.54	0.54	0.54	0.61	0.62	0.62	0.68	0.64	0.65
DT	0.56	0.56	0.56	0.64	0.65	0.65	0.66	0.67	0.66

Table 25: Pairwise classification results for Wisdom / Virtue, Wisdom / Goodwill and Virtue / Goodwill. Macro-averaged precision, recall and $F1$-score are reported for a 10-fold cross validation. A baseline classifying on the class distributions is compared against machine learning classifiers. * denotes the highest $F1$-scores.

types. In goodwill classification, the language used is very similar to that of virtue with the distinction coming from a mention of entities external to the UK parliament. Although an issue arises in that place names are mistaken for external entities indicating a new solution is needed for identifying particular entities in the UK parliament context. This is illustrated in example 28 where "Scotland" is interpreted as an entity giving a false positive feature. This is eluded to in Section 7.2.3.2 and demonstrates the need for further investigation in future work. "Scottish Office" also provides a false positive as the entity is considered as being part of the UK parliament rather than an external entity which would potential mean a goodwill classification.

Actual Values

Predicted Values		W	V	G
	W	102	112	26
	V	87	128	43
	G	49	53	38

Figure 35: Confusion matrix for pairwise Wisdom, Virtue and Goodwill classification against the manually annotated data.

(28) *"what he has done has brought the Scottish Office into disrepute and indeed the good name of justice in Scotland as a whole?"*

A similar problem arises for wisdom classification (see example 26 in section 7.1.2) where a domain context is needed to realise the difference between wisdom and virtue meaning mistakes are made by the automatic classifier. This is also shown in example 27, classifying “never-never land” is difficult as it does require a general knowledge or context. In some approaches word embeddings trained on wikipedia data can give contextual information, although in these cases this is more commonly used to provide a wider context for commonly occurring words rather than unique words for which a complex meaning is required.

This error analysis takes into account some examples and known classification issues to demonstrate the need for a larger set of annotated ethos types with improved manual annotation guidelines. The results of the classification do demonstrate the same problems which have arisen within the manual classification steps, thus indicating that an overall improvement in manual classification may produce improvements in the automatic classification.

7.4 Discussion

In summary, types of ethos supports and attacks can be reliably annotated using an iterative annotation guideline construction approach. The subsequent automatic classification from the manual annotation then shows room for improvement. Both the annotation and subsequent classification answer RQ3 as to whether fine-grained ethos types can be reliably annotated and extracted. The evaluation and error analysis show the possibilities for improvement yet prove that the classification task is possible and that the results are solid. This line of research is particularly important for progress within the sub-field of ethos mining, following the same path as argument mining will require some degree of specificity when it comes to ethos as it is not enough to simply declare that ethos occurs within a sentence or not. The added fine-grained nature of the ethos types also provides a number of avenues for future application in an analytical sense in politics and as a path to improved ethos extraction.

The results described for the manual classification and automatic classification of ethos types show room for improvement, in particular within the manual annotation of ethos types which could trigger an improvement in the automatic methods. Not only through enhanced training data which will increase performance, but, also through the generation of features. The final results for the manual classification (a Cohen's $kappa$ of 0.52) are promising, they show that with further annotation iterations the overall guidelines can be enhanced which will provide a positive outcome on the reliability of the final corpus. The automatic classification does not match this performance which can be explained due to the nature of ethos types. No specific features have been built to identify any of the classes and the relatively small volume of data means that any automatic solution will struggle to generalise the training data.

7.4.1 Annotation Improvements

The final corpus (EthosWVG_Hansard) does have the scope for extension which will provide two main advantages. The first relates to the empirical approach of the annotation guidelines. With added data and more examples on which to verify the annotation guidelines, they will become more robust. Although this extension step is non-trivial (as it requires extracting transcripts, annotating ethos and then annotating ethos types) the eventual evaluation to be conducted could provide extensions to the list of possible keywords used to identify the wisdom, virtue and goodwill labels. This leads to the second advantage of corpus extension. A larger corpus with more clearly defined keywords for each label will improve or aid the automatic classification process. If a more consistent set of keywords can be identified then simpler classification approaches can be used to identify the ethos types.

7.4.2 Automatic System Improvements

The final results of the automatic classification show room for improvement. The features used for extraction, for example, are fairly simple. This is, however, restricted by the amount of data available for each label within the classification task. A more complex set of features may improve the classification, but, without an extension to the overall corpus these features are likely to overfit to the data provided. On the classification of virtue and wisdom small improvements can easily be made through an increase in the dataset size. Coupled with enhancements to the manual annotation, new features can be identified for the extraction of these labels. In the case of the goodwill label, this is more difficult. Goodwill classification has a reliance on a certain degree of domain and common sense knowledge. Identifying when there are references to entities not within the jurisdiction of the UK parliament is an easy task for manual annotators. In the case of an automated classification, however, the task is more difficult. Providing a list of known entities within the UK parliament, manually defined, will aid in this. Although this is a possibility, it is rather a short term solution which would need to be replaced by an automatic one. As mentioned in section 7.3 external data from Wikipedia or parliamentary sources could be utilised with word embeddings, however, this will not necessarily cover all entities or may provide too many false positive cases to be of use. Ideally through the extension of the corpus, more labels of the type goodwill will be identified this in turn will allow for more elaborate features of extraction to be obtained. The added data would also provide a more comprehensive set of keywords which machine learning techniques can utilise for the classification task.

By improving the automatic classification of ethos types, analytics produced from ethos data are assured to be more reliable. Ethos types do have an advantage over a general ethos support and attack annotation. The more fine-grained nature of the ethos types mean that strategies used by politicians to support and attack one another can be more readily investigated. In particular the ethos type classification could be used to determine if a particular politician or set of politicians use the same techniques to support or attack. If they are of the same political party this could point to a strategy which is being utilised within the parliament to gain an advantage. Otherwise, these strategies could show a rift between members of opposite parties or admiration. In further extensions these types may even show the optimal ways to which politicians can persuade. For example, is an attack on practical wisdom more effective than an attack on virtue? Or is an attack on goodwill the most effective means of persuasion but only to a certain set of people?

When coupled with individual timelines of politicians career, ethos types could reveal particular controversies of the time. If, for example, a politician is accused of lying to the people within the media then it can be expected that this will be reflected with attacks on goodwill within the parliament. On the other hand, this is perhaps a less likely occurrence on positive outcomes as for one they are reported upon less in the media and are not so proficient within the annotated data of this corpus.

Finally, by improving both the manual and automatic classification there is also the possibility to increase the performance of the extraction of ethos. Using the keywords within each ethos type label may create a more specific set of features for which ethos can be extracted, especially when compared to sentences which do not contain ethos. The added dimensions for classification can provide this improvement although making the automatic extraction task more difficult with more steps for evaluation.

Chapter 8

Application of Ethos Mining: Ethos Analytics

This chapter contains the first set of analytics, of its kind, built upon ethos supports and attacks after the automatic extraction of ethos [1]. Thus, data which has largely gone unexplored due to its volume, which in turn makes it difficulty to determine useful insights, can now be explored by developing applications for ethos. In this context ethos mining provides a number of avenues of application - in this thesis - related to political dynamics. By exploring political science publications, newspaper articles and individual political careers, correlations can be made with ethotic statements. This in turn can provide journalists and other political commentators with context in a parliamentary setting that was previously unknown to them. Due to the large volume of Hansard data and the consistency of this data over a large time period within the UK parliament (the format of Hansard has stayed relatively constant since 1915 to present day, although some crisis mean that the data and format of parliament does change, for example, world wars. See chapter 4 for more details.) any ethos mining pipeline constructed on UK parliamentary data can be applied to this large set.

This large set of data means that there are several possible categories of analytics which

[1]The work in this chapter relates to two published works where the scaling up of ethos mining techniques was explored (in collaboration with Katarzyna Budzynska and Chris Reed) and analytics produced for political data (Duthie and Budzynska, 2018b; Duthie et al., 2016a).

can be deployed. In this thesis three types of such analytics have been implemented to answer RQ4: graph visualisations (see section 8.1); qualitative analytics (see section 8.2); and, quantitative analytics (see section 8.3), followed by the combination of all three (see section 8.4) all of which provide an insight into the possibilities of applications without necessarily providing causation analysis. Graph visualisations come from the large number of relationships present within the UK parliament, given the number of politicians and the number of ethotic relations between each politician constructed from ethos mining. Qualitative analytics are created from the political careers of each politician, mapping the supports and attacks on ethos to time points. Spikes in the supports or attacks on a politician can indicate a change in their stature, whether this is moving from a relatively unknown back-bench position (non-ministerial position within the government or shadow government) to a more prominent role or being an active participant in controversial or heated debates. Quantitative analytics can be mapped to political science publications or news articles, where the claims within these pieces are not always backed up by physical evidence. Ethotic supports or attacks can be utilised for this purpose, mapping a political event or observation to them. For example, a common comment may be that a political party is being plagued with infighting (disagreement or volatility between MPs of the same party), the ethotic attacks between members of the same party can then be analysed to confirm or reject this. A combination of all the analytics can then be made showing a pathway through each to the next.

8.1 Graph Visualisations

Graph visualisations, particularly those which depict relationships between politicians, can be constructed on a number of different data points and to show a varying degree of granularity. In the first instance various data time points can be shown to compare the number of politicians involved with ethotic supports and attacks and whether or not the politicians are predominantly attacking each other or supporting each other. The varying time points can also be used to show party political dynamics. In the case of a day-to-day parliamentary session the expectation is for members of the same political party to support one another, rather than attack. Whilst in more controversial debates there is an expectation for there to be attacks between politicians of the same party.

Using a force directed graph can also show the dynamics of the UK political system. By forcing members of opposite parties into different corners of a visualisation, only when there are a large number of relations between them will they be dragged closer together. Thus there is an expectation for Government ministers and the opposite shadow ministers to be dragged closer together within the visualisations. The force directed nature of the graph also shows interesting attack or support dynamics, i.e., does a particular politician always attack a specific other politician?

Extracted ESEs with polarity and source and target person stored in JSON files were used for the purposes of visualisation. To visualise the data, D3.js, a javascript graph visualisation library (available at: `http://d3js.org/`), was used to create force-directed graphs representing positive (coloured as green), negative (coloured as red) and neutral (coloured as blue when the attacks and supports were even) relationships amongst the politicians. Each edge representing a relationship is associated with a set of ESEs depending on the polarity of the ESE. People are visualised as nodes coloured according to their political party (black if information not available) with each relationship between the nodes, showing the attack and support of another speaker's ethos. Speaker nodes were also provided with the frequency of attacks and supports of each speaker to show the overall view of each speakers ethos over the full time frame of transcripts. Nodes in each graph are then clustered by political party using a multi-foci technique which pushes nodes of the same type to the same point in the visualisation: nodes which are pulled closer together (through forces between edges) show that there were either many attacks or supports between them. The relationships between speakers are then coloured as green for a positive relationship showing a support of a speakers ethos and red as a negative relationship showing an attack on a speakers ethos. Nodes can also be defined for organisations with relationships again displayed between the organisations and speakers. In this case the nodes are coloured black unless the political alignment is clear.

Visualisations were produced for several periods during the UK parliamentary debates. Figures 36, 37 and 38 show the output created when visualising ESEs for a period of three months in each of 1978, 1979 and 1989. For all three cases an ethos mining pipeline was used for the automatic extraction of ESEs where the data needed as an input can be extracted by any means due to the generalisable nature of the visualisation.

The three periods shown depict different moments within the UK parliament, 1978 was a difficult period for the Labour party with workers strikes (Taylor, 2013), this concluded with mass infighting between Labour members shown by predominately attack relations between Labour members (depicted as red nodes) in the visualisation. In 1979, there was a general election in the UK where the Labour party were beaten by the Conservative party making Margaret Thatcher the Prime Minister. The visualisation produced reflects this change with positive relations mainly between the Conservative members and negative relations mainly between Labour members. Interestingly, the dynamics normally depicted in the visualisations, of clear lines between the main political parties pulled closer by prominent figures, does not hold as it does in figures 36 and 38. An explanation could be the general election forcing MPs to be more prominent in the parliament.

Finally, in 1997 it was the end of a Conservative government in the time leading to Tony Blair becoming Prime Minister. This was a difficult period for the Conservative party in which it was documented that John Major, the then Prime Minister, was struggling to keep his own party on side (Gov.uk, 2016). Whilst not clearly evident in the abstract visualisation shown, a further analysis providing exact numerical values may unearth these details.

Whilst the abstracted graph visualisations do show a level of detail on the interactions between politicians on relatively short time periods, when this period is extended problems arise with the readability and usability of such a diagram. To demonstrate this problem a larger set of data can be passed through ethos mining pipelines. Due to some inconsistencies

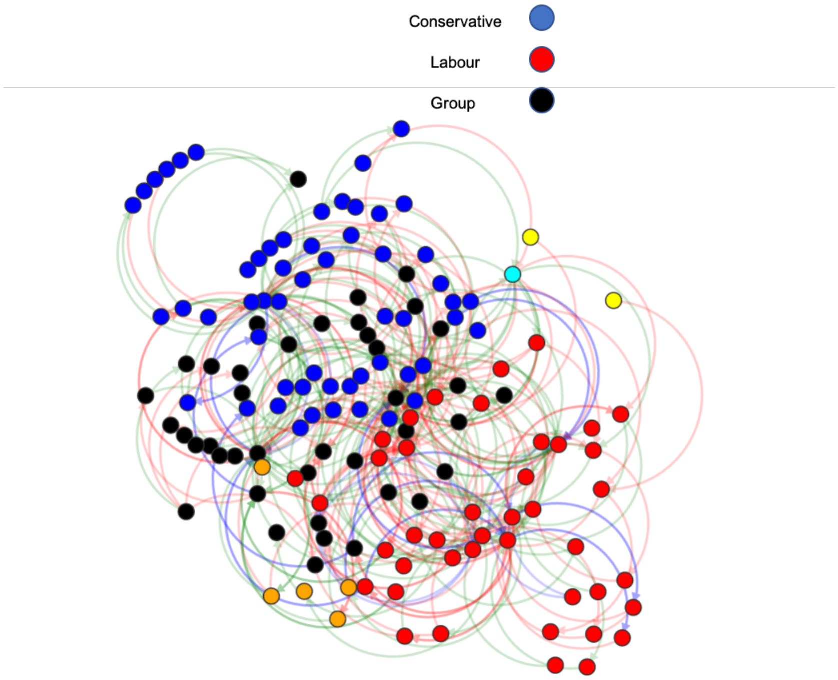

Figure 36: Political diagram for a three month period in 1978. Politicians are shown as nodes, with colours relating to their political party (red - Labour, blue - Conservative, black denotes a group or unknown party). Outgoing edges show attacks and supports of ethos with the colouring relating to the number of attacks and supports per type. See `https://bit.ly/2Q3YXPv` for the interactive graph.

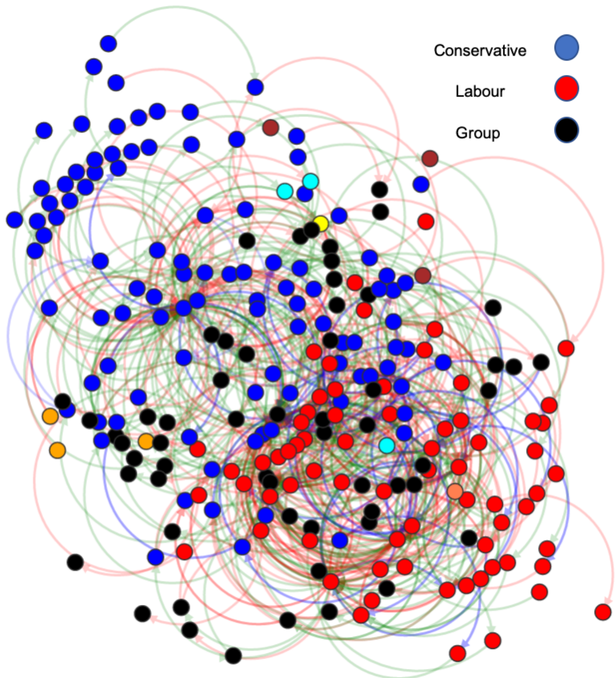

Figure 37: Political diagram for a three month period in 1979. Politicians are shown as nodes, with colours relating to their political party (red - Labour, blue - Conservative, black denotes a group or unknown party). Outgoing edges show attacks and supports of ethos with the colouring relating to the number of attacks and supports per type. See `https://bit.ly/2MetKIn` for the interactive graph.

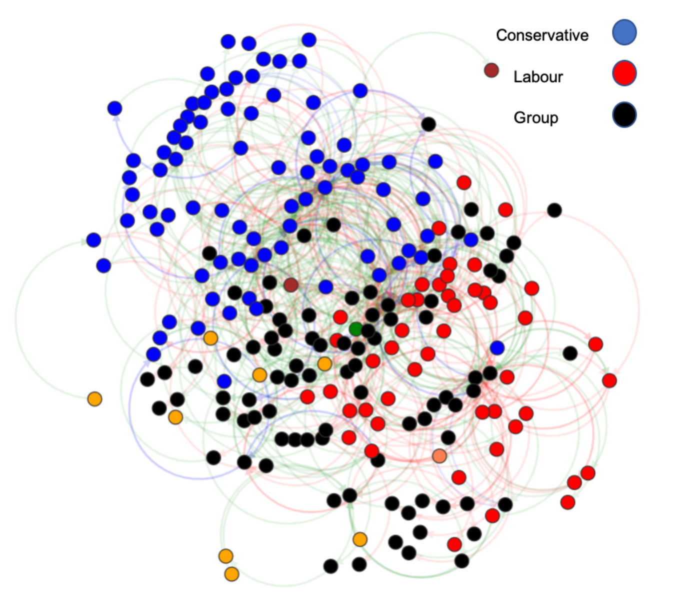

Figure 38: Political diagram for a three month period in 1997. Politicians are shown as nodes, with colours relating to their political party (red - Labour, blue - Conservative, black denotes a group or unknown party). Outgoing edges show attacks and supports of ethos with the colouring relating to the number of attacks and supports per type. See `https://bit.ly/2sOfoYd` for the interactive graph.

in the source data (due to World War 2) the first year available to scrape in a consistent format was 1945. Taking 32 years of data, up to the point at which Margaret Thatcher was appointed Prime Minster (1979) provides a total of 314,465 sentences of which 76112 are considered as ESEs. Due to the large volume of extracted data, using the standard graph visualisations would not be useful. To illustrate this, when only showing unique bi-directional edges and entities as nodes, a total of 23830 unique edges (unique edges refers to a relationship between two entities of which the volume is not taken into account, thus meaning in practice more edges would need to be drawn) between 1371 unique politicians are created. Whilst not a huge volume of data, it is clear that other techniques of analysis must be utilised in order to extract some meaning. One potential avenue is the use of qualitative analytics, essential the volume of supports and attacks on ethos for an individual politician or group and identifying correlations between them and newspaper articles or political science publications.

8.2 Qualitative Analytics

Qualitative analytics can be applied after the extraction of ethos using the attacks and supports upon ethos to explain political events. In particular, individual politicians can be tracked over their period of time within the UK parliament, plotting the supports and attacks on their ethos to their timeline. This process provides a more in depth analysis of the data presented by graph visualisations, although the process for analysis cannot be completely automated. Qualitative analytics do require a level of human knowledge allowing the correlations between ethos supports and attacks and specific news articles or publications to be made. In the case of this thesis, individuals were chosen on the basis of their political career, particularly if this is considered as controversial or integral to UK politics. This step requires some level of human analysis, however, any future work could make use of spikes in ethos supports and attacks as a method for determining if a politician's career should be qualitatively analysed.

For the purposes of analysis +/-ESE data was normalised in two ways: (*1*) the number of sentences where an entity is the target of an attack on or support of ethos, divided by the sentences where the same entity is a target, and (*2*) the number of sentences where an entity supports or attacks the ethos of another entity, divided by the sentences where this entity was the source. Finally a threshold of one standard deviation was used for each individual entity to investigate significant distributions from the norm.

One such interesting political career was that of **Reginald Maudling,** a Conservative party MP who held office from 1950 to 1979. The data relating to his political career is analysed from 1952 (his first mention in the parliament) to 1978, detailing the number of supports and attacks on his ethos (see figure 39). This data provided seven points outside one standard deviation for attacks (1965-1, 1965-2, 1966-1, 1967-1, 1969-2, 1970-1 and 1978-1) and six for supports (1957-2, 1958-1, 1965-1, 1967-2, 1968-2 and 1969-1).

In 1957, Maudling was made a member of the Government cabinet after performing well in is role as paymaster general (Bennett, 2013) this is indicated by a rise of support for

Maudling at this time to 0.18. In the first half of 1958, Maudling put together the pieces for the European Free Trade Association (Beloff, 1963) again indicated in a rise in support for Maudling to 0.20. The next data points occur in the first and second half of 1965 where Maudling unsuccessfully ran in the leadership election for the Conservative party indicated in supports of 0.15 and 0.15 and attacks of 0.54 and 1. In 1966, attacks stayed at a constant as he was made deputy leader.

The attacks rose in the first half of 1967 as Maudling blasted the then Labour Government's approach to Malta (Baston, 2004). In the second half of 1968 and the first half of 1969 supports for Maudling rose again to 0.5, coinciding with his appointment to Shadow Defence Secretary. Ethotic attacks rose again in the second half of 1969 and the first half of 1970 when he was appointed as Home Secretary, a particularly turbulent time in the UK with the rise of "The Troubles" in Northern Ireland and Maudling making a bad impression on his visit to the country (Baston, 2004). Finally, in the first half of 1978 Maudling was attacked in one instance occurring after investigations into his business dealings which had put an end to his career in politics.

A second such interesting career was that of Margaret Thatcher who in 1959 became MP for the constituency Finchley representing the Conservative party. Investigated is the period from her first appearance in the ethos analytics, 1962, to the year she became Prime Minister, 1979. This provides eight data points outside one standard deviation for attacks (1970:12, 1975:1, 1975:11, 1976:6, 1977:1, 1977:6, 1977:11, 1978:5 and 1978:8) and four data points for supports (1962:7, 1964:11, 1973:5 and 1978:5) of Thatcher's ethos (see figure 40). In 1962, when supports increased from 0 to 0.4, she was recently promoted to a junior ministerial position. In 1964, she lost her position, due to the Conservative party losing the general election, but became shadow on pensions. This co-occurs with an increase in supports from 0 to 0.4. In 1970, attacks on Thatcher increased from 0.17 to 0.78 corresponding with her appointment as Education Secretary and the cuts she made to funding (`http://bbc.co.uk/timelines/zqp7tyc`). In 1973, supports of Thatcher's ethos increased from 0 to 0.43. This follows the release of her paper in

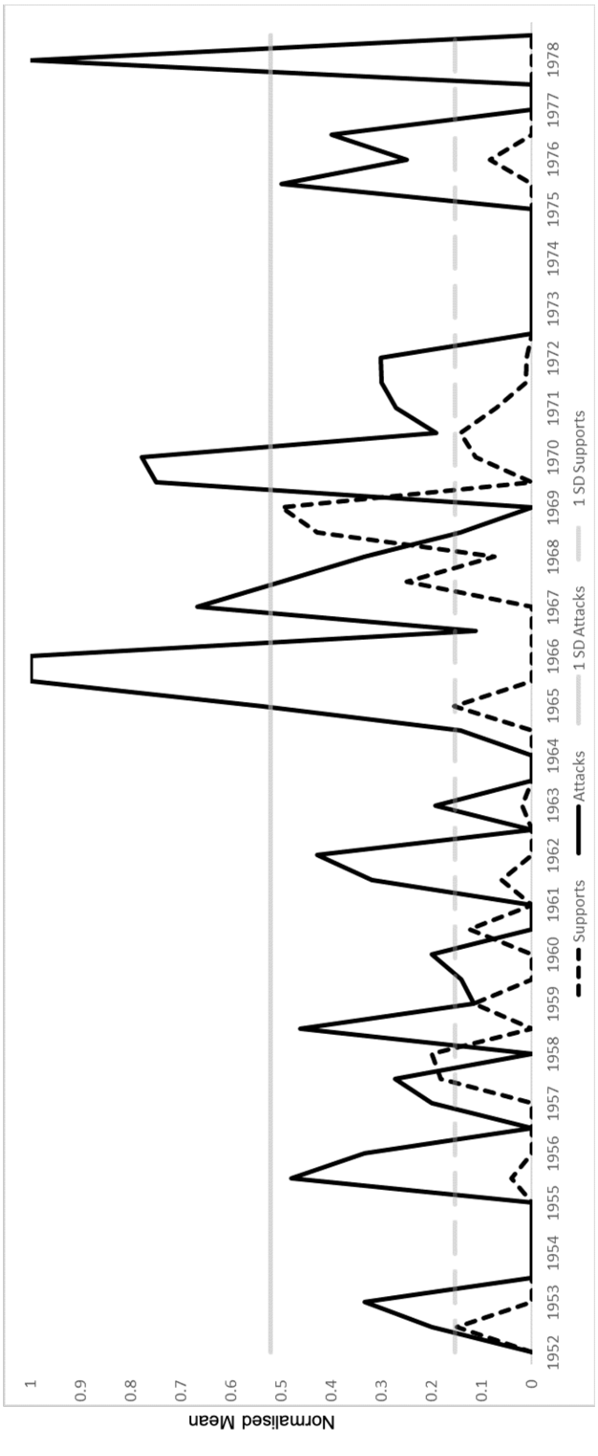

Figure 39: Line chart showing the total number of ethotic attacks and supports on Reginald Maudling a Conservative party MP split by six monthly intervals containing markers for one positive standard deviation (1 SD) from the mean of supports and attacks on ethos.

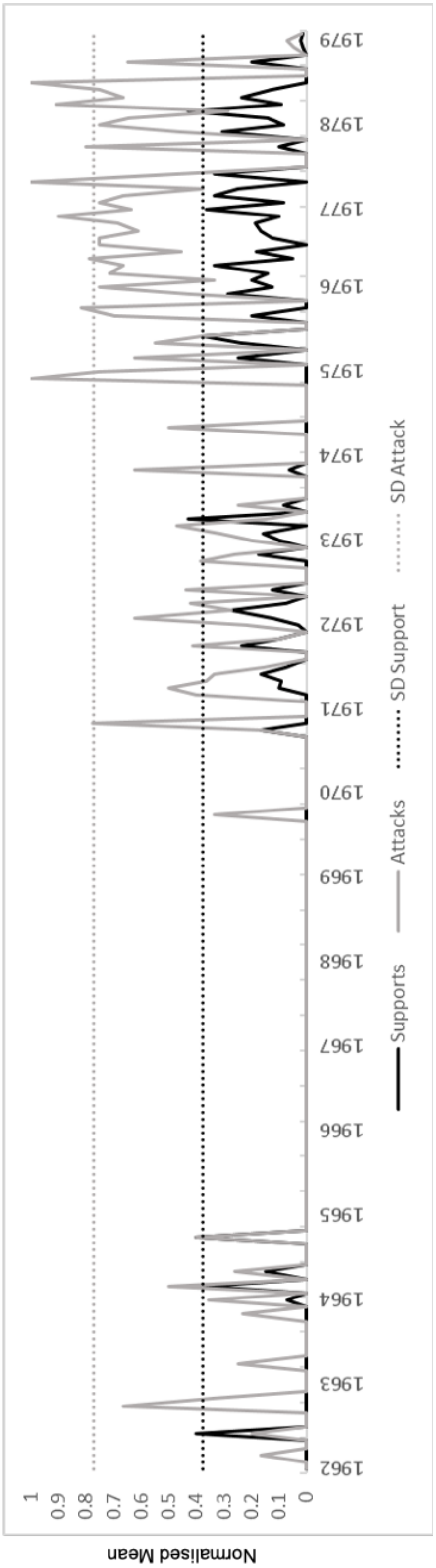

Figure 40: Line chart showing the number of ethotic attacks and supports of Margaret Thatcher split by monthly intervals containing markers for one positive standard deviation (1 SD) from the mean of supports and attacks on ethos.

1972 "Education: A Framework for Expansion" which reformed education. Thatcher became leader of the Conservative party in 1975 which corresponds to an increase of attacks on her ethos before the election (0 to 1) and after (0.7 to 0.8). In 1976, attacks increased on Thatcher (0.67 to 0.79) corresponding to her leadership of the Conservative party and through changing dynamics in the parliament which saw the Government lose their majority. Throughout 1977, attacks increased on Thatcher's ethos (0.69 to 0.9, 0.38 to 1, 0 to 0.8), which correlates with the local council elections in the UK which provided an insight into public attitude. The Conservative party won these resoundingly. Finally, in 1978 attacks increased on Thatcher's ethos (0.29 to 0.91, 0.75 to 1) as did supports (0.14 to 0.43) which co-occurred with an impending general election where she was prominent and her party later won.

8.3 Quantitative Analytics

The use of qualitative analytics related to political discourse, attempts to explain many of the spikes or trends in the ethos supports and attacks data, over time. Another potential application of ethos mining techniques could be in the use of anticipating political events. Much work has gone into the prediction of general elections, particularly from a political science perspective (see chapter 3.2). These works focus upon social media data, assuming that a portion of the general voting population can be identified and their stance towards political figures of parties used to identify the way they will vote in an election. Whilst not necessarily capable of predicting the outcome of elections, ethos supports and attacks within the UK parliament can be used to enhance these predictions looking for correlations in the data. This assumes that any political supports or attacks in the parliament may be in part because of pressure from the general public or may influence the general public if outlined although statistical significance testing would need to be undertaken to determine causation which is beyond the scope of this section which only looks to inform on the possibilities of using ethos in an application.

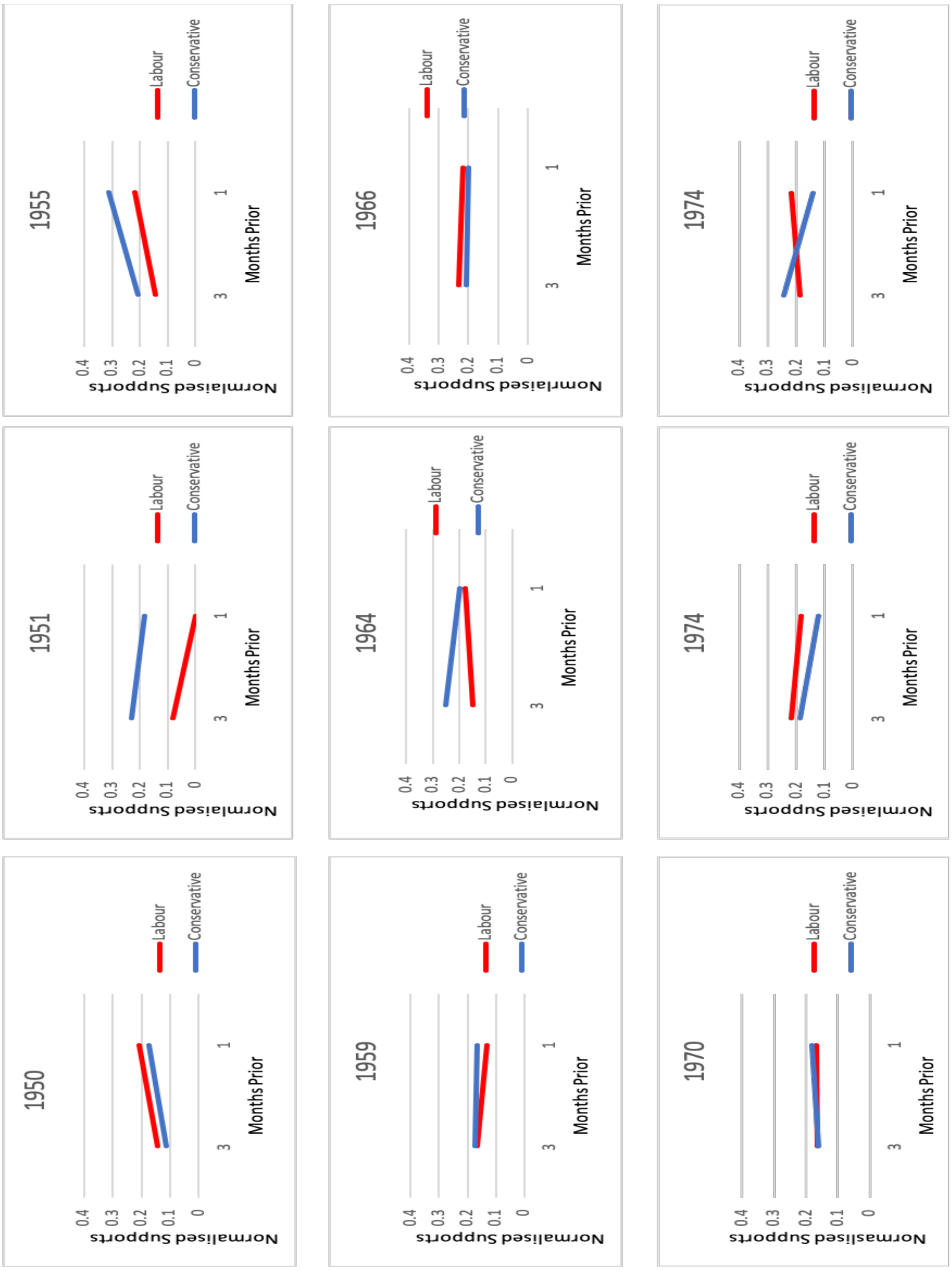

Figure 41: Line charts showing the mean of supports normalised by the volume of total utterances of the Labour and Conservative parties three and one months prior to the general elections between 1950 and 1974.

Year	Labour	Conservative
1950	0.032	0.031
1951	-0.04	0.02
1955	0.038	0.053
1959	-0.017	-0.004
1964	0.015	-0.026
1966	-0.007	-0.003
1970	0.0006	0.0095
1974	-0.017	-0.032
1974	0.016	-0.051

Table 26: Slope values for each election year for both the Labour and Conservative parties.

General elections took place nine times between 1946 and 1979. The two main parties in the UK (Labour and Conservative parties) were explored using two data points: supports in one quarter prior to an election and supports in one month prior to an election, to determine the slope, s = (y_m - y_q) / (x_m - x_q) (where x_m and x_q are the time points of one month and one quarter earlier and y_m and y_q are the ethotic values), and ascertain the election winner. All supports of ethos were taken into account where the political party was tagged except when the Government or the Opposition were mentioned as they appear heavily in oral answers to questions sessions. When comparing the slope values of each party, in all but one case, the party with the most positive slope won the election. The exception occurred in 1966 when the Conservative party had a more positive slope (-0.003), however, the Labour party (-0.007) held a slim majority. Figure 41 shows line charts of three months and one month prior to the general election for both parties and the slope value (see table 26) (on varying scales due to the varying difference in slope values over each general election) for both parties. It should be noted, however, that the difference in slope values are marginal and that the figures provided are an indication to the possible use cases of ethos mining for analytics and cannot directly be used as a means for prediction.

8.4 Combining Applications

By combining the ethos applications - graph, qualitative and quantitative - the more abstract relationship visualisations can be further described using elements of each analytic. In figure 36 the Prime Minister James Callaghan had a total of 3 supports of ethos, where the mean is 1, coming from both Labour and Conservative members and had a total of 11 attacks on his ethos, where the mean across all MPs is 1. Half of these were from Labour party members, reflecting the deep discontent at his leadership. The infighting which followed is also reflected in the graph. Shirley Williams, a Labour member at the time, has for example a total of 8 attacks on her ethos and 17 supports - but of those supports, only two come from other Labour members. In the years following the general election, and after the loss of Williams' seat, she became a founding member of the Social Democrat Party (SDP) (Democrats, 2016), a splinter from the Labour party.

It is expected in figure 37 that there would be many attacks on the Government and Margaret Thatcher due to the Labour party's loss of power and the general election. Many attacks on Margaret Thatcher and the Government can also be attributed to policies which worked to give trade unions less power (Margaret Thatcher promised to "curb the power of the unions" (BBC, 2013)). Thus Labour party members would attack Thatcher due the unions being the primary funding source of the Labour party. This is reflected in the visualisation with 41 attacks of ethos on the Government and 22 attacks on Margaret Thatcher's ethos, compared to a mean of just 2 attacks on a given individual. In this period there were two main trends for the Conservative and Labour parties. The Conservative party (the Government at the time) were prospering which is reflected in the relationship visualisation diagram where 224 of 361 relationships between Conservative members are positive. The opposite situation held for the Labour party with 67 out of 136 relationships as negative. With a difference of one standard deviation between Labour and Conservative relationships showing the difference in positive relationships. In (Hinnfors, 2006), it is documented that up to 1983 there was large-scale infighting in the Labour party.

This is reflected in the visualisation where relations between Labour party members are predominantly attacks with 17 attacks and 15 supports. A final trend is attacks on, and by, ministers of the Government. The data shows that only 17 Labour party members of 47 had no relationship with Government ministers or the Government itself. Government ministers are the proponents of policies, many of which were not well liked by the Labour party and the unions they represented. The attacks from ministers can be seen as a defence of the policies they advocated.

In figure 38, the position of the nodes stays in a consistent layout, unlike figure 37, due to the time period extracted before a general election. Although, as mentioned above, this was a difficult time for the Conservative party with confidence in the leadership, provided by the Prime Minister John Major, lacking. This is evident in the analysis with 8 ethotic attacks coming from his own party where the average number of attacks is 2. Following the loss of the general election to the Labour party a new leader of the Conservatives was elected. Interestingly, in the lead-up to the general election, between these dates, the proposed candidates for the Conservative Leadership election are more prominent in the visualisation, as seen in table 27, where the mean number of attacks for a politician is 2 and for supports is also 2. Many supports and attacks of the potential leaders hint at their impending desire to run for party leadership as a high number of attacks above the average show that the potential leaders are more prominent in debate.

Potential Conservative Leaders	**Supports**	**Attacks**
William Hague	33	30
Ian Lang	17	20
Stephen Dorrell	22	10
Michael Howard	4	4
Peter Lilley	3	0
John Redwood	2	0
Kenneth Clarke	0	0
MEAN AVERAGE	2	2

Table 27: Supports and Attacks on ethos of Conservative Leader proposed candidates.

8.5 Discussion

This chapter has explored three applications of ethos mining and the combination of all three applications used to inform the reporting of political events to answer RQ4. Graph visualisations of the data extracted can be produced showing the relations between political entities. Whilst these visualisations are informative, in the sense that they depict the overall political landscape in an easy to interpret format, they can become cluttered and it can be difficult to grasp the underlying meanings. Qualitative analytics grounded in ethos supports and attacks can be used to explore the graph visualisations and to explain political events using spikes in the data. In particular, individual political timeline analysis can be conducted with political publications used as a backing for the numerical data. A step further then is the use of quantitative analytics choosing a political event and using the ethos supports and attacks to anticipate the outcome of this event overall.

As a last step, all of the applications of ethos mining, leading to ethos supports and attacks, can be combined. Firstly, interesting dynamics can be identified within the graph visualisations. Secondly, the time period can be explored for interesting political events. Thirdly, the events can be grounded in the ethos supports and attacks. This chapter has explored applications of ethos mining, in particular how ethos correlates with political events, but any future work does need to further explore this avenue evaluating this data statistically to also ensure causation. This final step can then allow ethos analytics to be deployed as a suite of tools showing the connection between real world events and parliamentary debates which are often disconnected other than through the subjective analysis of political commentators.

Chapter 9

Conclusion

This final chapter will show the contributions of this thesis as a whole, pinpointing the descriptions given in each chapter and how these relate to the research questions identified. This is followed by a discussion of the possible avenues of future work and then the final closing remarks.

9.1 Contributions

This thesis has contributed the following novel advances in ethos mining: two manually annotated corpora of ethos supports and attacks; the automatic classification of ethos supports and attacks using a domain specific rule set; the creation of deep learning methods (DMRNN) in text classification for extracting ethos; a manually annotated corpus containing ethos types; the development of an NLP pipeline for classifying ethos types; and a set of comprehensive ethos analytics which look to link parliamentary debate and the outside world.

9.1.1 A Corpus and Domain Rule Based Classification for Ethos Mining

Chapter 5 of this thesis outlines the first corpus of ethos supports and attacks and the first automatic classification of ethos. The field of argumentation has focussed mainly upon logos (reasoning structures), whilst ethos, also one of Aristotle's modes of persuasion, has largely been considered as a by-product of these structures. This is particularly the case in argument mining where ethos is only considered as a by-product of argument scheme classification (for example argument from expert opinion). The research in this chapter then outlined a solitary annotation and automatic extraction of ethos.

In the first instance and outlined in chapters 4 and 5 parliamentary debates in the UK were chosen as an exploratory domain precisely because of the hostile environment. Within this domain a historical time period was chosen, 1979 to 1990. This time period in the UK, during Margaret Thatcher's premiership, was particularly hostile with many major political events, such as the troubles in Northern Ireland and the Falklands war, ongoing. From the debate record, 60 transcripts were manually annotated for ethos supports and attacks. Ethos then is defined as a support or attack of an entity which is the participant of communication, using Aristotle's definition of ethos as inspiration.

From this definition ethos supports and attacks were manually annotated for the first time on a sentence level. Inter annotator agreement was calculated for two coders on a 10% subset of the data. Cohen's kappa for recognising whether the sentence is ethotic or not gave the value of $\kappa = 0.67$. For ethotic sentences, $\kappa = 0.95$, when it is a support or an attack. For source-person of an ethotic statement, $\kappa = 1$ and for target-person it was $\kappa = 0.84$.

Following this manual annotation, a pipeline of existing NLP and novel rule-based modules was created to automatically identify ethos supports and attacks. The NLP modules make use of known techniques, part-of-speech tagging and named entity recognition (NER), whilst the rule-based modules identify domain specific entities, conduct anaphora resolution specifically for the UK Parliament and deal with reported speech in sentences. Each of these modules combined give an $F1$-score of 0.69 and 0.70 when NER is removed for determining if a sentence contains ethos or not. The removal of NER is due to the particularly challenging language of the UK parliament where politicians are not referred to by their name but instead through their constituency, as an "hon. Member" or through pronouns. Thus, domain specific rules are needed to ensure accuracy in extraction.

Following the classification of sentences as containing ethos or not, the polarity of each sentence is determined using sentiment classification. An SVM, using a sentiment word lexicon and a novel ethotic word lexicon created for ethos mining, gave a macro-averaged $F1$-score of 0.78. In this case the ethotic word lexicon provides the necessary domain specificity for sentiment classification.

The manual annotation and automatic classification of ethos supports and attacks answers RQ1 and RQ2. The Cohen's kappa score of 0.67 when annotating whether or not a sentence contains ethos shows the reliability of annotating ethos independently of logos. The automatic classification use the rule-based approach also shows the reliability of automatically extracting ethos independently of logos through an $F1$-score of 0.70.

9.1.2 An Extended Corpus and Deep Learning Classification for Ethos Mining

Following the rule-based approach in chapter 5, a more generalisable approach was taken, again to extract ethos supports and attacks. This involved developing new techniques for ethos mining and new techniques for deep learning in text classification. A Deep Modular Recurrent Neural Network (DMRNN) was constructed taking inspiration from methods in image classification.

For the DMRNN to be effective the original corpus created (and described in chapter 5) was re-annotated and extended to alleviate some of the annotation and classification issues observed in this first approach. An altered definition of ethos was used ensuring any reference made to ethos must be expressed on the linguistic surface, rather than relying on implicit, subjective or not widely known general knowledge. From this definition an extra 30 transcripts were annotated and 60 re-annotated, leaving a total of 60 transcripts for training data and 30 for testing.

To evaluate the extended corpus, a second annotator analysed a 10% subset of the data. This gave a Cohen's κ = 0.67 when determining if a sentence contained ethos or not (normalised for word count), κ = 1 for the polarity classification of ethotic sentences (positive and negative) when both annotators have already agreed that the sentence is ethotic, κ = 1 for speaker tags and κ = 0.93 for the target tags.

Subsequently, a natural language processing pipeline made up of standard techniques and those made specifically for ethos mining was used to detect if sentences contained ethos or not. Domain specific rules were utilised for the extraction of entities and words related to those entities from a dependency parse, and were utilised in anaphora resolution. In the case of anaphora resolution this was extended to use external data from Wikipedia to identify politicians. The ordering of the pipeline was also changed from the first approach instead using sentiment as a feature to classify ethos. The intuition here is that the linguistic structure encodes both the target entity of the ethotic statement as well as its polarity.

In the case of sentiment classification an extension was made to the lexicons available. Previously, sentiment classification was undertaken using an ethotic word lexicon and sentiment word lexicon. The domain specific lexicon was then extended using the negative words from a Hansard lexicon (described in chapter 6) created using word embeddings. The addition of this lexicon meant the best classifier performed, 31.3% above the baseline, consisting of a logistic regression classifier and all lexicons (m-$F1$-score 0.84).

All features, raw text, dependency tags, POS tags, entities and related words, and sentiment were used as inputs into the DMRNN. The intuition here was that the overall dataset size could be extended using features for classification, rather than only raw text. Individual features are used for classification are then effectively merged by the DMRNN giving an $F1$-score of 0.74 (21.3% above baseline). Comparing the macro-averaged

$F1$-scores of the previous ethos mining pipeline and the DMRNN there is a 6% increase.

The updated and re-annotated corpus and the automatic classification of ethos supports and attacks through deep learning answers RQ1 and RQ2. The results of the annotation evaluation (Cohen's kappa 0.67) again show the reliability of annotating ethos independently of logos answering RQ1. Although identifying ethos in itself did not improve over previous approaches the annotation task as a whole did. The automatic classification of ethos supports and attacks through a DMRNN again shows the reliability of extracting ethos independently of logos. This novel method also shows improvements over the previous rule-based approach whilst ensuring generalisation. The addition of an updated sentiment lexicon also improves classification as a whole and together answers RQ2.

9.1.3 A Corpus and Multi-class Classification of Ethos Types

Chapter 7 of this thesis investigated the annotation of ethos elements, wisdom, virtue and goodwill. It is clear when investigating the corpus of ethos supports and attacks that the language used to express an attack or support, as expected, is not the same. This point is also carried into both supports and attacks where it is clear the language used to support varies and so does the language used to attack. To explain this difference Aristotle's well known distinction of elements of ethos was used. These ethos types (wisdom, referring to practical experience, virtue, to character traits and goodwill, to aligning with the audience) can then be used to gain an understanding of the strategies used by politicians within debate.

As a first step ethos elements were annotated on top of the already built corpus from chapter 6 using inspiration from several works ((Aristotle, 1991; Crowley and Hawhee, 2004; Fahnestock and Secor, 2003; Garver, 1994)) to define annotation guidelines. Following two annotation iterations, used to refine the annotation guidelines, an evaluation on a 10% subset of the data gave an average Cohen's κ of 0.52. The results show the reliability of annotation and yet the difficulty of the task at hand which can be improved through further annotation iterations.

Using the manual annotation as training and testing data to perform cross validation, a natural language processing pipeline was built to classify these types of ethos. The pipeline used several standard methods to extract features such as sentiment and unigrams, bigrams and trigrams. Previously used modules were updated such as POS tagging on the entities and related words, and a new module identifying proper nouns was created. All features were then passed to a principal component analysis module to reduce the dimensionality of the data. These modules using pairwise classifiers and one versus all classification gave F1-scores averaging 0.62. The results show the reliability of the automatic classification and the difficulty of the task at hand especially when moving from a binary classification for RQ2 to a multi-class classification using less data for RQ3.

Updating the corpus of ethos supports and attacks to add ethos element labels and classifying these ethos elements using an NLP pipeline answers RQ3. Again the Cohen's kappa score shows the reliability of annotation, although with room to be improved by further annotation iterations. The automatic classification results also show the reliability of classification although highlight the difficulty of this task. Together these results answer RQ3 showing the reliability of annotating and extracting more fine-grained ethos strategies.

9.1.4 Ethos Analytics

The ethos support and attack classification pipelines and the extensive data available in UK parliamentary debates left an avenue for developing ethos analytics. Chapter 8 explores the possible applications of ethos supports and attacks in three ways through: graph visualisations; qualitative analytics; and quantitative analytics.

In the first instance the output of ethos supports and attacks, from either the rule-based approach to ethos mining or the deep learning approach to ethos mining, can be used to visualise the interactions between political entities. A force-directed graph was constructed showing an entity as a node and the ethos interactions as edges between them. High interaction between two entities is then shown by closeness on the graph, where nodes are forced in the direction of their respective political party. These graphs can then be analysed to give an outlook of the political landscape at that time period.

Qualitative analytics were developed exploring time series data for supports and attacks on individual politicians allowing for the investigation of correlations with political events such as individual party position appointments. Politicians were investigated over their full time period within the UK parliament. The supports and attacks of their ethos were then counted and compared with the standard deviations of each. Spikes in these values were then investigated for correlations to political science publications, media articles and personal events of the time.

Quantitative analytics were also created to compare supports and attacks on ethos with political events. Here a focus was made upon general elections over a 33 year period in the UK. Supports of ethos in the two main political parties, the Labour and Conservative party, were investigated prior to each general election. Firstly three months before an election and then one month before an election. From this the slope of support in each party was calculated where the party with the most positive slope was anticipated to win the election. This was reflected in all but one election where the outcome resulted in no party holding a majority.

Each of the applications of ethos supports and attacks give a visual, qualitative and quantitative insight into the political landscape of the time. By investigating the ethos interactions, visual representations showing: the closeness of interactions; publications correlating with these interactions; and the change in these interactions leading to political events, can be created. Each of these points individual and together answer RQ4 painting a picture of the political landscape over various time periods.

9.2 Future Work

The research undertaken in this thesis highlights several avenues for future work. These applications range from extensions to the corpora created to advances in the ethos mining technology and ethos applications.

9.2.1 Annotating Argument

One possibility for future research is that of extending the annotated corpus of ethos supports and attacks to also incorporate arguments. As highlighted in the discussion of chapter 6 this extension could mean a number of improvements are made to the extraction of ethos.

In the simplest case this annotation could consider propositional relations only. In that case the argument annotation would require several steps: segmenting all of the text for argument components including those sentences annotated as ethos; resolving proposition text (for example resolving all pronouns and references made to earlier text); connecting up propositions via inference or conflict relations according to an argument annotation format; and re-annotating ethos relations using the new segments.

In a more complex step the argument annotation could take into account the dialogical structure of the text. This approach would require a more complex annotation model, such as the use of Inference Anchoring Theory (IAT) (see chapter 2.3) or combining rhetorical structure theory (RST), Segmented Discourse Representation Theory (SDRT) and argument relations (Peldszus and Stede, 2013). In each of these approaches the motivation is to link the exact utterances made in a dialogue to the argument relations. Both works approach this in a different way. In IAT the utterances are considered as locutions. Locutions are then connected to argument propositions through illocutionary connections, similar to speech acts. This then provides a structure connecting each locution to a proposition. Dialogue moves are also considered between locutions, known as transitions. Each of these transitions then links the segmented locutions. Mirroring the transitions are argument relations of inference and conflict between each of the propositions. The relations of inference and conflict can then be linked to the transitions between the locutions by an illocutionary connection which highlights the intention of the speaker to "argue" or "disagree".

In the case of using RST, SDRT and argument relations, all can be converted into a dependency style annotation. This means that rather than operating within a sentence, these "dependency" relations operate over the full argument component structure. In the case of RST and SDRT only the rhetorical and discourse annotation are considered over the components. This holds true for argument components too, where only the argumentative relations are considered between each component. Once the dependency structures are created correlations between each of the structures can then be investigated.

Either a simple or more complex argument annotation could provide improvements for both the manual annotation of ethos and the automatic extraction. In the first instance the more clearly segmented text with the argumentative context highlighted, could assist the annotators in distinguishing between ethotic and ¬ethotic statements. In the case of automatic identification of ethos, the additional argumentative relations could work as features for ethos classification.

In an analytical sense the extension of the corpus to contain argument relations could also prove advantageous. Instead of only accounting for ethos supports and attacks, these can be compared with argument relations, for individual politicians. This extension would then allow for a strategy comparison, such as does a particular politician tend to give reasons for their assertions or do they tend to attack the opponent? This distinction could then be utilised to determine the optimal strategy when in debate with a particular politician, or to detect the optimal way to persuade in parliament.

A final extension could be the use of Intertextual correspondence (ITC) (Visser et al., 2018a). This method looks to link corpora of different genres but that contain the same topic of argument. Ideally then references made to newspapers within the UK parliament could then be connected to the exact publication. This has two advantages, the first being all context from outwith a particular dialogue can, in theory, be annotated[1] and secondly, the resource for annotation can be increased in volume from a different domain.

In the case of ethos supports and attacks, these can be linked from media articles about a particular person building a picture of their prior ethos to a debate. For example, if a

[1] In some cases the scope of the ITC relations would have to be considered. Unless achieved automatically it would not be make sense to link every mention of a politician in social media, for example, to every time they make an utterance in the UK parliament.

politician has been heavily criticised for their actions in the media this detail can be utilised in a parliamentary setting. Either an argument relation or an ethos relation can then be drawn between the parliamentary transcript and the media report.

One potential issue with this line of improvement is the capability of evaluating the new annotation. In the case of argument annotation an approach had to be taken to ensure that errors created in the segmentation stage were not carried throughout the annotation process when evaluating annotator agreement (Duthie et al., 2016b). In the case of ethos annotation the Combined Argument Similarity Score (CASS) would need to be extended to take into account ethos relations. Ensuring that in cases where the segmentation was not created the same between two annotators, that the relation annotated can still be evaluated.

9.2.2 Improving the Ethos Mining pipeline

In order to improve ethos mining as a whole there are several possible extensions. In the first instance obtaining further classification features may be valuable. Research into new techniques for classification may also prove useful, especially where these techniques have worked for other areas of text classification.

Taking inspiration from argument mining and argumentation schemes (see chapter 3.3.1.4), the ethos element tags can be utilised to extract ethos. By extending the overall size of the ethos corpus, more instances of each tag can be discovered. This in turn will provide different parameters for ethos classification. The intuition modelled here would be that elements of ethos use different language, for both support and attack. This would then provide six potential features for classification, Wisdom, Virtue, Goodwill both support and attack. In this case with enough training data classification using machine learning may become simpler with more clearly defined classes.

In the same area, argument annotation or identification can also be used as features for ethos mining. The intuition here would be that attacks or supports of ethos do not come out of the blue. Instead ethos supports or attacks would be adjacent to arguments and therefore the arguments (or argument relations) can be used as features for extraction.

This line of work can be continued through more advanced machine learning methods. For example, the use of multi-task learning may provide a boost in accuracy for ethos extraction. In a multi-task learning environment a number of tasks can be considered at the same time meaning the features for extraction under one label can be utilised by another. In this case the argument mining task and ethos mining task can be performed at the same time. Also possible is the extraction of ethos and the identification of ethos elements. This is particularly useful if they were to use similar features for identification or in the very least features which may be useful in both cases.

In each of these multi-task learning methods there is a reliance on the other areas of future work, namely argument annotation, to be undertaken. Although in some instances the manual annotation is not completely necessary, it will ensure accuracy. Pilot studies can be undertaken to determine if an argument mining pipeline can provide reliable enough features for ethos classification. The same is true in the case of ethos mining, where features of ethos from an ethos mining pipeline may prove useful for annotating arguments.

Other work has been established in this area instead using rhetorical figures as features of extraction (Lawrence et al., 2017). The hope would be in the case of ethos and argument mining that the tasks are more tightly coupled and therefore would yield more reliable results.

9.2.3 Improving Ethos Analytics

On the development of ethos analytics there are also several possible extensions. The ethos analytics developed in this thesis show the applicability of ethos mining pipelines on a larger. Despite this, the group of analytics created are not necessarily cohesive. From this point of view, the ethos analytics can be further developed to form a suite of analytics rather than individual inclusions.

Inspiration in this case can be taken from argument analytics (Lawrence et al., 2016) where a suite of useful analytics were created to enhance the understanding of a debate. Similarly, a suite of ethos analytics can be developed for the same purpose. That being the inclusion of further analytics can aid in the understanding of debates through ethos. Particularly, combinations between argument analytics and ethos analytics can be achieved. This can then allow for the comparison between operations of logos and ethos which paints an important picture in a debate. A visual representation of a persons willingness or not to give reasons for their assertions is important. Just as important is a representation of the proportion of reasons against the proportion of ethotic manoeuvring that a speaker uses.

A full suite of analytics would allow viewers of a debate to move seamlessly through the various visualisations, in the hope that this will aid in their understanding of the debate. This point is particularly important within the UK parliament. Currently, one must rely upon political commentators or the media in order to digest the large volume of speeches. The goal with an analytics suite would not be to replace these people, rather as a supplement. This means that there is not a reliance upon unbiased media or commentators to paint a full picture. Instead the visualisations can be built and if reasonable then they are digestible for individuals with their own opinion. This is a crucial aspect, as no analytics system should tell a user what to think from a debate, rather they should be able to come to their own conclusions. For example, the use of reasoning versus the use of ethotic attack. In some cases users may prefer one to the other and it is not for the analytics system to sway this.

Further to the analytics suite, particularly the ethos aspect of this, is the potential to add ethos elements. New analytics can be developed showing when a particular speaker, or political party use one of these strategies. This in turn may provide an indicator as to how a political party plans on rebutting a particular bill. It also may provide an indicator to the ideology of the party, perhaps using a people focus (goodwill) versus one of capital gain (perhaps wisdom).

This can allow users to see for themselves whether or not the political party they support uses the strategies they believe in. The same analysis can be made of individuals, especially attempting to determine if a politician follows the party line or whether or not they rebel. In some cases this could even constitute determining if a politicians aligns more with their own political party or another.

Finally, extensions to ethos analytics can also be made using ITC. ITC would allow the prior ethos of a participant in a debate to be accounted for. Visualisations over time can then be created showing how ethos has evolved. This could be particularly interesting in the case of negative media stories and how they affect a debate. Intuition would say that negative stories in the media would be used by political opponents. This fact could then be tracked. Furthering this point could be the use of perception analytics. In this case how an audience perceives a politician could be tracked especially over time and from different sources. This avenue of research would require a large amount of further study. Although through philosophy there is an idea of how effective the use of ethos is, this would have to be extensively studied. Audience based tests would have to be undertaken, analysing a shift or not in perception every time a politician speaks. Whilst interesting, modelling this process would be particularly difficult especially when trying to gauge an opinion which is normally personal.

9.3 Closing Remarks

This thesis has described a set of novel advances in ethos mining, a newly developed sub-field of argument mining. For millennia in rhetoric, ethos has been known to play a crucial role in persuasion and everyday interaction yet despite this fact no attempt has previously been made to create a corpus of such phenomena. From an automatic extraction perspective previous work has focussed upon reasoning structures, also known as logos. From these extraction attempts ethos has been obtained as a by-product not a main feature of identification.

The first step in this research investigated ethos supports and attacks in UK parliamentary debates, precisely because of the volatile yet structured language used. The Aristotelian distinction of ethos was used as an inspiration to define annotation guidelines for the creation of the first corpus of ethos supports and attacks independent of logos. This then allowed for the creation of the first ethos mining pipeline using rule-based methods to extract ethos.

In a second step ethos supports and attacks were further investigated by extending the previous corpus, the largest publicly available. The first ethos mining pipeline was then re-developed using an error analysis as inspiration for the creation of novel modules exploring the connection between an entity and the sentiment shown towards it. New methods for deep learning on the task of text classification were then developed to enhance the reliability of automatic ethos support and attack extraction.

To investigate the intricacies of ethos supports and attacks, Aristotle's well known distinction of elements of ethos wisdom, virtue, and goodwill, were applied for the first time to build a corpus using the ethos supports and attacks as a base annotation. Following two annotation iterations, the first pipeline to classify ethos elements was developed, using a combination of standard and novel modules and machine learning. This then showed the possibility of determining the strategies used by politicians.

In the last step, the output of the ethos mining pipelines (ethos supports and attacks) were used to develop ethos analytics. These analytics investigate and visualise the ethotic interactions between entities. The graph, qualitative and quantitative analytics were constructed to investigate correlations between real political events and the on-goings of the UK parliament. Ultimately the analytics produced show the applicability of ethos to the wider political landscape.

The research in this thesis then contributes a set of novel ethos mining advances: a manual corpus of ethos supports and attacks; the automatic classification of ethos supports and attacks; the creation of deep learning methods (DMRNN) in text classification for extracting ethos; a manually annotated corpus containing ethos types; the development of an NLP pipeline for classifying ethos types; and a set of ethos analytics.

The advances shown in this thesis are widely applicable to various domains not only as a tool to gauge political opinion, but to extract public opinion from various sources of natural language. Although the focus of this thesis has been upon parliamentary debate, the development of general ethos mining pipelines and the future work discussed, highlight this applicability. Social media discussions or public deliberation can be used to build profiles of individuals over large time periods from the natural language and identify individuals or groups at the centre of popular (or unpopular) opinion.

The constant need of fact checking in both social media and debate gives rise to future research possibilities. This is particularly true from an argumentation perspective where automatically extracting the reasoning structures in both domains can aid in this process, determining not only the claims made but the reasons for these claims. Ultimately though, the character of the speaker is just as important, if not more important than the content of what is said in a societal setting. Knowing who the source of a statement is and the strategies that they use to persuade an audience are key tools in the evaluation of speakers which has become an ever present need in modern politics. At the forefront of this research are the fields of argument mining and now ethos mining.

Bibliography

Agarwal, M. and Zhou, B. (2014). Using trust model for detecting malicious activities in twitter. In International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction, pages 207–214. Springer.

Aristotle (1991). On Rhetoric (G. A. Kennedy, Trans.). New York: Oxford University Press.

Artz, D. and Gil, Y. (2007). A survey of trust in computer science and the semantic web. Web Semantics: Science, Services and Agents on the World Wide Web, 5(2):58–71.

Austin, J. L. (1962). How to Do Things with Words. Clarendon Press.

Baston, L. (2004). Reggie : The Life of Reginald Maudling. Sutton.

BBC (2013). Margaret Thatcher: From grocer's daughter to Iron Lady. `http://www.bbc.co.uk/timelines/zqp7tyc\#z2bpv4j` [Last Accessed: 29/01/18].

Beloff, N. (1963). The General Says No:Britain's Exclusion from Europe. Penguin Books.

Bennett, G. (2013). Six Moments of Crisis: Inside British Foreign Policy. Oxford University Press.

Berners-Lee, T. and Hendler, J. (2001). Publishing on the semantic web. Nature, 410(6832):1023.

Braet, A. C. (1992). Ethos, pathos and logos in aristotle's rhetoric: A re-examination. Argumentation, 6(3):307–320.

Brinton, A. (1986). Ethotic argument. History of Philosophy Quarterly, pages 245–258.

Brownstein, O. L. (1965). Plato's phaedrus: Dialectic as the genuine art of speaking. Quarterly Journal of Speech, 51(4):392–398.

Budzynska, K. (2010). Argument analysis: components of interpersonal argumentation. In Proc. of COMMA 2010, pages 135–146.

Budzynska, K. (2012). Circularity in ethotic structures. Synthese, 190:3185–3207.

Budzynska, K. and Reed, C. (2011). Whence inference? University of Dundee Technical Report.

Budzynska, K. and Reed, C. (2012). The structure of ad hominem dialogues. In proc. of COMMA 2012.

Budzynska, K. and Villata, S. (2016). Argument mining. In The IEEE Intelligent Informatics Bulletin, volume 17, pages 1–7.

Burnap, P., Gibson, R., Sloan, L., Southern, R., and Williams, M. (2016). 140 characters to victory?: Using twitter to predict the uk 2015 general election. Electoral Studies, 41:230 – 233.

Burnett, C., Norman, T. J., and Sycara, K. (2011). Trust decision-making in multi-agent systems. In Proceedings of the Twenty-Second Int Joint Conference on Artificial Intelligence, IJCAI, pages 115–120.

Cabrio, E. and Villata, S. (2012). Combining textual entailment and argumentation theory for supporting online debates interactions. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2, pages 208–212. Association for Computational Linguistics.

Carlile, W., Gurrapadi, N., Ke, Z., and Ng, V. (2018). Give me more feedback: Annotating argument persuasiveness and related attributes in student essays. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), volume 1, pages 621–631.

Castelfranchi, C. and Falcone, R. (2000). Trust is much more than subjective probability: Mental components and sources of trust. In System Sciences, 2000. Proceedings of the 33rd Annual Hawaii International Conference on, page 10. IEEE.

Castelfranchi, C. and Falcone, R. (2010). Trust theory: A socio-cognitive and computational model, volume 18. John Wiley & Sons.

Ceolin, D. and Potenza, S. (2017). Social network analysis for trust prediction. In Steghöfer, J.-P. and Esfandiari, B., editors, Trust Management XI, pages 49–56. Springer International Publishing.

Chang, C.-C. and Lin, C.-J. (2011). LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1–27:27. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1):37–46.

Crowley, S. and Hawhee, D. (2004). Ancient Rhetorics for Contemporary Students. Pearson/Longman.

Democrats, L. (2016). Shirley williams. `http://www.libdems.org.uk/shirley_williams` [Last Accessed: 02/02/16].

Duthie, R. and Budzynska, K. (2018a). Classifying types of ethos support and attack. In Proc. of COMMA, pages 161–168.

Duthie, R. and Budzynska, K. (2018b). A Deep Modular RNN Approach for Ethos Mining. In Proceedings of IJCAI, pages 4041–4047.

Duthie, R., Budzynska, K., and Reed, C. (2016a). Mining ethos in political debate. In Proc. of the Sixth Int Conference on Computational Models of Argument (COMMA 2016), pages 299–310. IOS Press, Berlin.

Duthie, R., Lawrence, J., Budzynska, K., and Reed, C. (2016b). The cass technique for evaluating the performance of argument mining. In Proceedings of the Third Workshop on Argument Mining (ArgMining2016), pages 40–49.

Eger, S., Daxenberger, J., and Gurevych, I. (2017). Neural end-to-end learning for computational argumentation mining. arXiv preprint arXiv:1704.06104.

Fahnestock, J. and Secor, M. (2003). A Rhetoric of Argument. McGraw-Hill Higher Education.

Falcone, R. and Castelfranchi, C. (2004). Trust dynamics: How trust is influenced by direct experiences and by trust itself. In Autonomous Agents and Multiagent Systems, 2004. AAMAS 2004. Proceedings of the Third International Joint Conference on, pages 740–747. IEEE.

Feng, V. W. and Hirst, G. (2011). Classifying arguments by scheme. In Proceedings of the The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-2011), pages 987–996.

Finkel, J. R., Grenager, T., and Manning, C. (2005). Incorporating non-local information into information extraction systems by gibbs sampling. Proceedings of the 43nd Annual Meeting of the Association for Computational Linguistics (ACL, 2005), pages 363–370.

Franch, F. (2013). (wisdom of the crowds)2: 2010 uk election prediction with social media. Journal of Information Technology & Politics, 10(1):57–71.

Frogel, S. (2005). The Rhetoric of Philosophy. Controversies Series. John Benjamins Publishing Company.

Gagarin, M. (2001). Did the sophists aim to persuade? Rhetorica: A Journal of the History of Rhetoric, 19(3):275–291.

Garver, E. (1994). Aristotle's Rhetoric: An Art of Character. Philosophy / Classics / Rhetoric. University of Chicago Press.

Golbeck, J. (2008). Weaving a web of trust. Science, 321(5896):1640–1641.

Golbeck, J. et al. (2008). Trust on the world wide web: a survey. Foundations and Trends in Web Science, 1(2):131–197.

Gov.uk (2016). History of sir john major - gov.uk. `https://www.gov.uk/government/history/past-prime-ministers/john-major`[Last Accessed: 02/02/16].

Grandison, T. and Sloman, M. (2000). A survey of trust in internet applications. IEEE Communications Surveys & Tutorials, 3(4):2–16.

Grimmer, J. and Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. Political analysis, 21(3):267–297.

Guha, R., Kumar, R., Raghavan, P., and Tomkins, A. (2004). Propagation of trust and distrust. In Proceedings of the 13th international conference on World Wide Web, pages 403–412. ACM.

Habernal, I., Wachsmuth, H., Gurevych, I., and Stein, B. (2018). Before name-calling: Dynamics and triggers of ad hominem fallacies in web argumentation. In Proceedings of NAACL-HLT, pages 386–396.

Herrick, J. A. (2015). The history and theory of rhetoric: An introduction. Routledge.

Hidey, C., Musi, E., Hwang, A., Muresan, S., and McKeown, K. (2017). Analyzing the semantic types of claims and premises in an online persuasive forum. In Proceedings of the 4th Workshop on Argument Mining, pages 11–21.

Hillard, D., Purpura, S., and Wilkerson, J. (2008). Computer-assisted topic classification for mixed-methods social science research. Journal of Information Technology & Politics, 4(4):31–46.

Hinnfors, J. (2006). Reinterpreting Social Democracy: A History of Stability in the British Labour Party and Swedish Social Democratic Party. Manchester University Press.

Hirst, G., Riabinin, Y., Graham, J., Boizot-Roche, M., and Morris, C. (2014). Text to ideology or text to party status? In From Text to Political Positions: Text analysis across disciplines, pages 61–79. John Benjamins Publishing Company, Amsterdam.

Hochmuth, M. (1952). Kenneth burke and the new rhetoric. Quarterly Journal of Speech, 38(2):133–144.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8):1735–1780.

Hu, M. and Liu, B. (2004). Mining and summarising customer reviews. In Proceedings of the ACM SIGKDD Int Conference on Knowledge Discovery and Data Mining, Seattle, Washington, USA.

Hunter, A. and Booth, R. (2015). Trust-sensitive belief revision. In Proceedings of the 24th International Conference on Artificial Intelligence, IJCAI'15, pages 3062–3068. AAAI Press.

Huynh, T. D., Jennings, N. R., and Shadbolt, N. R. (2006). An integrated trust and reputation model for open multi-agent systems. Autonomous Agents and Multi-Agent Systems, 13(2):119–154.

Janier, M., Lawrence, J., and Reed, C. (2014). OVA+: an argument analysis interface. In Computational Models of Argument (COMMA), pages 463–464.

Johnson, S. (1998). Skills, socrates and the sophists: Learning from history. British Journal of Educational Studies, 46(2):201–213.

Jurafsky, D. (May 2015). Natural language processing. `https://class.coursera.org/nlp/lecture`.

Karan, M., Šnajder, J., Sirinic, D., and Glavaš, G. (2016). Analysis of policy agendas: Lessons learned from automatic topic classification of croatian political texts. In Proceedings of the 10th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities, pages 12–21.

Ke, Z., Carlile, W., Gurrapadi, N., and Ng, V. (2018). Learning to give feedback: Modeling attributes affecting argument persuasiveness in student essays. In International Joint Conference on Artificial Intelligence (IJCAI), pages 4130–4136.

Kennedy, G. A. (2008). The Art of Rhetoric in the Roman World: 300 BC-AD 300, volume 2. Wipf and Stock Publishers.

Kingma, D. and Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Krabbe, E. C. and Walton, D. N. (1993). It's all very well for you to talk! situationally disqualifying ad hominem attacks. Informal Logic, 15:79–91.

Kraut, R. (2017). Plato. In Zalta, E. N., editor, The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University, fall 2017 edition.

Laaksonen, S.-M., Nelimarkka, M., Tuokko, M., Marttila, M., Kekkonen, A., and Villi, M. (2017). Working the fields of big data: Using big-data-augmented online ethnography to study candidatecandidate interaction at election time. Journal of Information Technology & Politics, 14(2):110–131.

Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33 1:159–74.

Larkey, L. S. and Croft, W. B. (1996). Combining classifiers in text categorization. In Proceedings of the 19th Annual Int ACM SIGIR Conference on Research and

Development in Information Retrieval, SIGIR '96, pages 289–297, New York, NY, USA. ACM.

Lawrence, J., Duthie, R., Budzysnka, K., and Reed, C. (2016). Argument analytics. In Baroni, P., Stede, M., and Gordon, T., editors, Proceedings of the Sixth International Conference on Computational Models of Argument (COMMA 2016), Berlin. IOS Press.

Lawrence, J., Janier, M., and Reed, C. (2015). Working with open argument corpora. In European Conference on Argumentation (ECA).

Lawrence, J. and Reed, C. (2015). Combining argument mining techniques. In Proceedings of the Second Workshop on Argumentation Mining, Denver. Association for Computational Linguistics.

Lawrence, J. and Reed, C. (2016). Argument mining using argumentation scheme structures. In Baroni, P., Stede, M., and Gordon, T., editors, Proc. of COMMA, Berlin. IOS Press.

Lawrence, J. and Reed, C. (2017). Mining argumentative structure from natural language text using automatically generated premise-conclusion topic models. In Proceedings of the Fourth Workshop on Argumentation Mining, Copenhagen. Association for Computational Linguistics.

Lawrence, J., Reed, C., Allen, C., McAlister, S., and Ravenscroft, A. (2014). Mining arguments from 19th century philosophical texts using topic based modelling. In Proceedings of the First Workshop on Argumentation Mining, pages 79–87, Baltimore, Maryland. Association for Computational Linguistics.

Lawrence, J., Visser, J., and Reed, C. (2017). Harnessing rhetorical figures for argument mining: A pilot study in relating figures of speech to argument structure. Argument & Computation, 8:1–22.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. Nature, 521(7553):436–444.

Leff, M. (2009). Perelman, ad hominem argument, and rhetorical ethos. Argumentation, 23(3):301–311.

Lippi, M. and Torroni, P. (2016). Argument mining from speech: Detecting claims in political debates. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, AAAI'16, pages 2979–2985.

Liu, B. (2010). Sentiment analysis and subjectivity. Handbook of Natural Language Processing, 2:627–666.

Ma, L., Lu, Z., and Li, H. (2016). Learning to answer questions from image using convolutional neural network. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, pages 3567–3573. AAAI Press.

Mack, P. (2011). A History of Renaissance Rhetoric 1380-1620. Oxford-Warburg Studies. OUP Oxford.

Manning, C. and Klein, D. (2003). Optimization, maxent models, and conditional estimation without magic. In Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: Tutorials-Volume 5, pages 8–8. Association for Computational Linguistics.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., and McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. In Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pages 55–60.

May, J. M. (1988). Trials of character: The eloquence of Ciceronian ethos. UNC Press Books.

May, J. M. (2006). The roman world of cicero's de oratore. Classical World, 99:470.

Maynard, D. and Funk, A. (2011). Automatic detection of political opinions in tweets. In Extended Semantic Web Conference, pages 88–99. Springer.

McCormack, K. C. (2014). Ethos, pathos, and logos: The benefits of aristotelian rhetoric in the courtroom. Wash. U. Jurisprudence Rev., 7:131.

McGregor, S. C., Mouro, R. R., and Molyneux, L. (2017). Twitter as a tool for and object of political and electoral activity: Considering electoral context and variance among actors. Journal of Information Technology & Politics, 14(2):154–167.

McKeon, R. (1942). Rhetoric in the middle ages. Speculum, 17(1):1–32.

Miller, T. (1997). The Formation of College English: Rhetoric and Belles Lettres in the British Cultural Provinces. Composition, Literacy, and Culture Series. University of Pittsburgh Press.

Mitkov, R. (2002). Anaphora resolution. Routledge.

Moens, M.-F. (2013). Argumentation mining: Where are we now, where do we want to be and how do we get there? In FIRE '13 Proceedings of the 5th 2013 Forum on Information Retrieval Evaluation.

Moens, M.-F., Boiy, E., Palau, R., and Reed, C. (2007). Automatic detection of arguments in legal texts. In Proceedings of the Int Conference on AI and Law (ICAIL-2007), pages 225–230.

Mui, L., Mohtashemi, M., and Halberstadt, A. (2002). A computational model of trust and reputation. In System Sciences, 2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference on, pages 2431–2439. IEEE.

Murphy, J. J., Katula, R. A., and Hoppmann, M. (2013). A synoptic history of classical rhetoric. Routledge.

Musi, E., Ghosh, D., and Muresan, S. (2016). Towards feasible guidelines for the annotation of argument schemes. In Proceedings of the third workshop on argument mining, pages 82–93.

Naderi, N. and Hirst, G. (2015). Argumentation mining in parliamentary discourse. Paper presented at 15th Workshop on CMNA, Bertinoro, Italy.

Naderi, N. and Hirst, G. (2017). Recognizing reputation defence strategies in critical political exchanges. In Proceedings of the Int Conference Recent Advances in Natural Language Processing, RANLP 2017, pages 527–535.

Naderi, N. and Hirst, G. (2018). Using context to identify the language of face-saving. In Proceedings of the Fifth Workshop on Argumentation Mining. Association for Computational Linguistics.

Nooy, W. D. and Kleinnijenhuis, J. (2013). Polarization in the media during an election campaign: A dynamic network model predicting support and attack among political actors. Political Communication, 30(1):117–138.

O'Donovan, J., Smyth, B., Evrim, V., and McLeod, D. (2007). Extracting and visualizing trust relationships from online auction feedback comments. In Proceedings of the 20th International Joint Conference on Artifical Intelligence, IJCAI'07, pages 2826–2831. Morgan Kaufmann Publishers Inc.

Onyimadu, O., Nakata, K., Wilson, T., Macken, D., and Liu, K. (2014). Towards sentiment analysis on parliamentary debates in hansard. In Kim, W., Ding, Y. and Kim, H.-G. Semantic Technology. Lecture Notes in Computer Science (8388), pages 48–50. Springer.

Paglieri, F. and Castelfranchi, C. (2014). Trust, relevance, and arguments. Argument & Computation, 5(2-3):216–236.

Palau, R. M. and Moens, M.-F. (2009). Argumentation mining: the detection, classification and structure of arguments in text. In Proceedings of the 12th international conference on artificial intelligence and law, pages 98–107. ACM.

Pang, B., Lee, L., and Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 79–86.

Park, J. and Cardie, C. (2014). Identifying appropriate support for propositions in online user comments. In Proceedings of the First Workshop on Argumentation Mining, pages 29–38.

Parsons, S., Atkinson, K., Li, Z., McBurney, P., Sklar, E., Singh, M., Haigh, K., Levitt, K., and Rowe, J. (2014). Argument schemes for reasoning about trust. Argument & Computation, 5(2-3):160–190.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12:2825–2830.

Peldszus, A. and Stede, M. (2013). From argument diagrams to argumentation mining in texts: A survey. Int Journal of Cognitive Informatics and Natural Intelligence (IJCINI), 7(1):1–31.

Perelman, C. and Olbrechts-Tyteca, L. (1973). The New Rhetoric: A Treatise on Argumentation. University of Notre Dame Press.

Persing, I. and Ng, V. (2016). End-to-end argumentation mining in student essays. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 1384–1394.

Plett, H. (2004). Rhetoric and Renaissance Culture. de Gruyter.

Purcell, W. and Benson, T. (1996). Ars Poetriae: Rhetorical and Grammatical Invention at the Margin of Literacy. Studies in rhetoric/communication. University of South Carolina Press.

Qian, Q., Huang, M., Lei, J., and Zhu, X. (2017). Linguistically regularized lstm for sentiment classification. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1679–1689. Association for Computational Linguistics.

Rahwan, I., Zablith, F., and Reed, C. (2007). Laying the foundations for a world wide argument web. Artificial Intelligence, 171:897–921.

Reed, C. and Rowe, G. (2004). Araucaria: Software for argument analysis, diagramming and representation. International Journal on Artificial Intelligence Tools, 13(04):961–979.

Rheault, L., Beelen, K., Cochrane, C., and Hirst, G. (2016). Measuring emotion in parliamentary debates with automated textual analysis. PLOS ONE, 11:1–18.

Richardson, M., Agrawal, R., and Domingos, P. (2003). Trust management for the semantic web. In International semantic Web conference, pages 351–368. Springer.

Rooney, N., Wang, H., and Browne, F. (2012). Applying kernel methods to argumentation mining. In Twenty-Fifth International FLAIRS Conference.

Rubin, V. (2009). Trust incident account model: Preliminary indicators for trust rhetoric and trust or distrust in blogs. In International AAAI Conference on Web and Social Media, pages 300–303.

Sabater, J. and Sierra, C. (2005). Review on computational trust and reputation models. Artificial intelligence review, 24(1):33–60.

Saint-Dizier, P. (2012). Processing natural language arguments with the textcoop platform. Argument & Computation, 3(1):49–82.

Salah, Z. (2014). Machine learning and sentiment analysis approaches for the analysis of Parliamentary debates. PhD thesis, University of Liverpool.

Sarmento, L., Carvalho, P., Silva, M. J., and de Oliveira, E. (2009). Automatic creation of a reference corpus for political opinion mining in user-generated content. In Proceedings of the 1st International CIKM Workshop on Topic-sentiment Analysis for Mass Opinion, TSA '09, pages 29–36.

Schneider, J. (2014). An informatics perspective on argumentation mining. In ArgNLP.

Schuster, S. and Manning, C. D. (2016). Enhanced english universal dependencies: An improved representation for natural language understanding tasks. In proceedings of the International Conference on Language Resources and Evaluation (LREC), pages 2371–2378.

Searle, J. R. (1969). Speech Acts: An Essay in the Philosophy of Language. Cambridge University Press.

Skalnik, J. (2002). Ramus and Reform: University and Church at the End of the Renaissance. Sixteenth Century Essays & Studies. Truman State University Press.

Stab, C. and Gurevych, I. (2014). Identifying argumentative discourse structures in persuasive essays. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 46–56.

Stab, C. M. E. (2017). Argumentative Writing Support by means of Natural Language Processing. PhD thesis, Technische Universität Darmstadt, Darmstadt.

Taylor, A. (2013). Before thatcher came to power, the uk was literally covered in gigantic piles of garbage. `http://www.businessinsider.com/thatcher-and-the-winter-of-discontent-\\2013-4?IR=T` [Last Accessed: 02/02/16].

Taylor, C. and Lee, M.-K. (2016). The sophists. In Zalta, E. N., editor, The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University, winter 2016 edition.

Teacy, W. L., Luck, M., Rogers, A., and Jennings, N. R. (2012). An efficient and versatile approach to trust and reputation using hierarchical bayesian modelling. Artificial Intelligence, 193:149 – 185.

Thomas, M., Pang, B., and Lee, L. (2006). Get out the vote: Determining support or opposition from Congressional floor-debate transcripts. In Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 327–335.

Toledo-Ronen, O., Bar-Haim, R., and Slonim, N. (2016). Expert stance graphs for computational argumentation. In Proceedings of the Third Workshop on Argument Mining (ArgMining2016), pages 119–123.

Toulmin, S. (1958). The Uses of Argument. Cambridge University Press.

Toutanova, K., Klein, D., Manning, C., and Singer, Y. (2003). Feature-rich part-of-speech tagging with a cyclic dependency network. Proceedings of HLT-NAACL, pages 252–259.

Van Eemeren, F. H. and Grootendorst, R. (1987). Fallacies in pragma-dialectical perspective. Argumentation, 1(3):283–301.

Villalba, M. and Saint-Dizier, P. (2012). Some facets of argument mining for opinion analysis. Proceedings of the Fourth International Conference on Computational Models of Argument, pages 23–24.

Visser, J., Duthie, R., Lawrence, J., and Reed, C. (2018a). Intertextual Correspondence for Integrating Corpora. In Proceedings of International Conference on Language Resources and Evaluation (LREC).

Visser, J., Lawrence, J., Wagemans, J., and Reed, C. (2018b). Revisiting computational models of argument schemes: Classification, annotation, comparison. In Modgil, S., Budzynska, K., and Lawrence, J., editors, Proceedings of the Seventh International Conference on Computational Models of Argument (COMMA 2018), pages 313–324, Warsaw. IOS Press.

Wagemans, J. H. M. (2016). Constructing a periodic table of arguments. In Bondy, P. and Benacquista, L., editors, Argumentation, Objectivity, and Bias: Proceedings OSSA 11, pages 1–12. OSSA.

Walton, D. (1999). Ethotic arguments and fallacies: The credibility function in multi-agent dialogue systems. Pragmatics & Cognition, 7(1):177–203.

Walton, D., Reed, C., and Macagno, F. (2008). Argumentation Schemes. Cambridge University Press.

Walton, D. N. and Koszowy, M. (2014). Two kinds of arguments from authority in the ad verecundiam fallacy. Proceedings of the International Conference on Argumentation (ISSA).

Walton, D. N. and Reed, C. (2003). Diagramming, argumentation schemes and critical questions. In F.H. van Eemeren, J.A. Blair, C. W. and Henkemans, A. S., editors, Anyone Who Has a View: Theoretical Contributions to the Study of Argumentation, pages 195–211. Kluwer, Dordrecht.

Wang, J., Yu, L.-C., Lai, K. R., and Zhang, X. (2016a). Dimensional sentiment analysis using a regional cnn-lstm model. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages 225–230. Association for Computational Linguistics.

Wang, X., Jiang, W., and Luo, Z. (2016b). Combination of convolutional and recurrent neural network for sentiment analysis of short texts. In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pages 2428–2437.

Wyner, A., Mochales-Palau, R., Moens, M.-F., and Milward, D. (2010). Approaches to text mining arguments from legal cases. In Semantic processing of legal texts, pages 60–79. Springer Berlin Heidelberg.

Xiang, Q., Zhang, J., Nevat, I., and Zhang, P. (2017). A trust-based mixture of gaussian processes model for reliable regression in participatory sensing. In Proceedings of the 26th International Joint Conference on Artificial Intelligence, pages 3866–3872. AAAI Press.

Yu, B., Kaufmann, S., and Diermeier, D. (2008). Classifying party affiliation from political speech. Journal of Information Technology & Politics, 5(1):33–48.

Zaremba, W., Sutskever, I., and Vinyals, O. (2014). Recurrent neural network regularization. arXiv preprint arXiv:1409.2329.

Zhang, L., Wang, S., and Liu, B. (2018). Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, page e1253.

Zhou, D. X., Resnick, P., and Mei, Q. (2011). Classifying the political leaning of news articles and users from user votes. In Fifth International AAAI Conference on Weblogs and Social Media, pages 417–424.

Glossary

AH Ad Hominem

AIF Argument Interchange Format

AIFdb Argument Interchange Format Database

CNN Convolutional Neural Network

DMRNN Deep Modular Recurrent Neural Network

DT Decision Tree

DSR Domain Specific Rules

Elements of Ethos see also Ethos Types

ESE Ethotic Sentiment Expression - a sentence which contains ethos referred to through sentiment.

¬ESE ¬Ethotic Sentiment Expression - a sentence deemed to hold no ethos.

+ESE Positive Ethotic Sentiment Expression

-ESE Negative Ethotic Sentiment Expression

Ethos Build or reduce a speakers character and credibility; a property of an identifiable individual or identifiable group of individuals; specifically on the linguistic surface;

Ethos Attack The statement makes explicit mentions of a person, organisation or other entity in a negative frame

Ethos Support The statement makes explicit mentions of a person, organisation or other entity in a positive frame

Ethos Types see also Wisdom, Virtue and Goodwill

EWL Ethotic Word Lexicon

EXT Entity Extraction

F1 F1-score

Goodwill Aligning with the audience

Hansard UK parliamentary debate proceedings

HAN Hansard Specific Lexicon

HAN (neg) Negative Words Hansard Specific Lexicon

IAA Inter-annotator Agreement

IAT Inference Anchoring Theory

Kappa κ Cohen's Kappa

Logos Appealing to Logic, using reasoning to persuade

LSTM Long Short-Term Memory

ME Maximum Entropy

MP Member of Parliament

NB Naive Bayes

NER Named Entity Recognition

NLP Natural Language Processing

Oral Answers Session within the UK parliament where questions are answered by Government Ministers

OVA Online Visualisation of Argument tool

P Precision

Pathos Appealing to the emotions of the audience

POL Polarity

POS Parts Of Speech

PCA Principal Component Analysis

R Recall

RNN Recurrent Neural Network

Source Person Who utters a statement

SVM Support Vector Machine

SWL Sentiment Word Lexicon

Target Person Person or entity described by a statement

Virtue Character traits

Wisdom Practical experience

Appendix

.A Annotation Examples

As a demonstration on annotating ethos in parliamentary debates several examples are given. The examples look to identify the difference between ethotic and non-ethotic statements, annotating logos and ethos and specify the more intricate reasons for annotation which may differ from the annotation conducted in chapter 5.

.A.1 Annotating ESE and non-ESE

This example makes the distinction between ethotic statements and non ethotic statements while applying the tags from the guidelines.

(29) **Speakers**: Mr. Atkins, Secretary of State for Northern Ireland, Mr. Bradford, Member of Parliament for Belfast South.

a. Mr. Bradford: *Will the Secretary of State accept that extradition is consequential on firm evidence being provided to the Eire courts and that the Act to which he has referred does not allow policemen to go in person to the courts to provide that evidence?*

b. Mr. Bradford: *Will he demand that the RUC should be admitted to the Eire courts and that the criminals who are indicted there should be extradited to Northern Ireland?*

c. Mr. Bradford: *If he does not agree to press for those measures, we shall* *have no alternative but to ask the Prime Minister to assume responsibility for security.*

d. Mr. Atkins: *I am not sure that the hon. Gentleman is right in stating that witnesses, whether police officers or anyone else, are not allowed to attend courts in the Republic.*

In the example above (29-a) is annotated as **n-ESE**. Although here a point is asserted through the question it is not an attack on the minister rather a clarification that “the Act to which he has referred does not allow policemen to go in person to the courts to provide that evidence”. (29-b) is also annotated as **n-ESE**. This is a direct question looking for further information from the minister. (29-c) is annotated as **ESE** with the **SOURCE**, Mr. Bradford, the **TARGET**, Mr. Atkins and tagged as **Ethotic Attack**. The source here is attempting to undermine the Ministers position by saying that the Prime Minister should have responsibility for his job. (29-d) is also annotated as **ESE** with the **SOURCE** Mr. Atkins, the **TARGET** Mr. Bradford, and tagged as **Ethotic Attack**. While not a convincing attack on ethos the Minister informs the target in this case that they are not right.

This example, again, makes the distinction between ethotic statements and non ethotic statements while applying the tags from the guidelines.

(30) **Speakers**: Mr. King, Secretary of State for Defence, Sir Peter Emery, Member of Parliament for East Devon.

a. Sir Peter Emery: *‘In view of my right hon. Friend’s answer to the hon. Member for Woolwich (Mr. Cartwright) on question No. 3, and the relationship that Germany must have to that, in the review that is being carried forward will he press the Americans to set a level for the troop requirements in Europe in conjunction with General Galvin and ourselves, as it is essential that we should be able to give a lead to the rest of NATO on what we believe is necessary for the proper defence of Europe even after the conventional force* *reduction treaties?*

b. Mr. King: *I am grateful to my hon. Friend.*

c. Mr. King: *He is exactly right.*

In the example above (30-a) is tagged as **n-ESE**. While containing multiple entity references none of these entities are being attacked or supported, in reality they are used as reference points for multiple points of reasoning. (30-b) is tagged as **n-ESE**. While containing a positive sentiment and an entity, the positive sentiment is from politeness rather than support of character. In essence "I am grateful" could be replaced with "Thank you". (30-c) is tagged as **ESE** with the **SOURCE** Mr. King, the **TARGET** Sir Emery, and tagged as **Ethotic Support**. Here Mr. King states that the prior statement by Sir Emery is right. This is a feature of ethotic language. See ethotic key words.

.A.2 Annotating Ethos against Logos

This example makes the distinction between ethotic statements and reasoning (logos) while applying the tags from the guidelines. When annotating ethos only and not taking into account the logotic context, **logos** is annotated as **n-ESE**.

(31) a. Mr. John: *Since the major obstacle to extradition is the ability of those against whom extradition is sought to raise a political defence, does not the answer lie in the ratification by the Irish Republic Government of the European convention on terrorism?*

b. Mr. Atkins: *The hon. Gentleman is right on the latter point.*

(31-a)

is tagged as **n-ESE**. Here there is reference to entities which are not relevant for the UK parliament. What we also see is logos where the conclusion is that "the answer lies in the ratification by the Irish Republic Government of the European convention on terrorism" which is supported by "the major obstacle to extradition is the ability of those against whom extradition is sought to raise a political defence". What this reasoning allows is the use of supporting or attacking ethos in a particular context. (31-b) is tagged as **ESE** with the **SOURCE** Mr. Atkins, the **TARGET** Mr. John and tagged as **Ethotic Support**. Mr. Atkins again supports Mr. John's ethos by declaring he is right. We see the relation between ethos and logos where Mr. Atkins supports Mr. John's ethos on the conclusion Mr. John has made.

This example, again, makes the distinction between ethotic statements and reasoning (logos) while applying the tags from the guidelines. When annotating ethos only and not taking into account the logotic context, **logos** is annotated as **n-ESE**.

(32) a. Mr. Kaufman: *Is the right hon. Gentleman (Mr. Heseltine) further aware that the director-general of the National Federation of Building Trades Employers said that this was a further body blow to the building industry, which is accelerating into decline and has in prospect the worst recession since the war?*

b. Mr. Kaufman: *The right hon. Gentleman (Mr. Heseltine) is a disaster to housing and is bringing about a disaster to the housing industry and he should resign.*

(32-a)

is tagged as **n-ESE**. Although there are entity mentions again here there are not directly relating to ethos, rather the question is directed towards Mr. Heseltine but in a more assertive fashion. The whole question is rather a premise or using expert opinion to support the final conclusion. (32-b) is tagged as **ESE** with the **SOURCE** Mr. Kaufman, the **TARGET** Mr. Heseltine and tagged as **Ethotic Attack**. Although this can be considered a conclusion in the reasoning. It also plays the roll of an ethotic attack on Mr. Heseltine. Mr. Heseltine is said to be a disaster and is told as a further point he should resign. Despite the clear relation between the logos and ethos here there is also a clear distinction in what each hope to achieve.

.B Ethotic Keywords

Below is a set of ethotic keywords, character traits and situations which typically indicate ethos. While this set is indicative of ethotic support the opposite of any of the set can be applied for ethotic attack. For example, the opposite of calm would be irate. The list is compiled from the following publications: (Aristotle, 1991; Crowley and Hawhee, 2004; Fahnestock and Secor, 2003; Garver, 1994)

- Good moral character
- Know the right information and provide it
- Unselfish
- Graceful
- Calm
- Just
- Courageous (not rash)
- Noble
- Show moral excellence
- Contribute effectively
- Say what they think
- Have an ability for doing good
- Show self control

- Liberality (do good with money)
- Magnanimity (give benefits for others)
- Magnificence (produce something great in expenditure)
- Will always have the right response
- Sound knowledge of the subject
- Have knowledge sufficient for the purpose at hand
- Draw the right conclusions from their knowledge
- Sensible
- Have practical experience
- Have the right decision
- Concerned with doing or action
- Act with regard to human goods
- Able to deliberate well about moral goods not for one's own benefit
- In deliberation they command action
- Balance the moral good and bad
- Know what is good for man
- Use knowledge quickly and reliably
- Treat the audience the way they want to be treated.
- Show goodwill towards others
- Do not deceive

- Inclusive
- Care about who they represent and give good advice
- Consider what needs to be known
- Supply necessary information but do not repeat it
- Say what benefits something will achieve
- Self sacrifice
- Align with the audience
- Give good advice when it is known